syi.hu/cse

Cselekvéselmélet 1.0

Syi

i@syi.hu

February 10, 2014

Contents

Előszó 1

1 Logikai, módszertani alapok **5**
1.1 Metafogalmak 9
1.2 Halmazelmélet 12
1.3 Kijelentéslogika 13
1.4 Predikátumlogika 23
1.5 Relációelmélet 26
1.6 Kategoriális logika 43
1.7 Másodrendű logika 55
1.8 Modális logika 56
1.9 Episztemikus/doxasztikus logika 66
1.10 Dinamikus logika 76
1.11 Preferencialogika 78
1.12 Deontikus logika 116

2 Társadalmi cselekvés **129**
2.1 Cselekvés 129
2.2 Intencionalitás 140
2.3 Késztetés, vágy, érték 148
2.4 Vélekedés, vágy, szándék 155
2.5 Választás 168

3 A társadalmi cselekvés típusai **177**
3.1 Indulati-érzelmi cselekvés 179
3.2 Habituális cselekvés 196
3.3 Célracionális cselekvés 206
3.4 Értékracionális cselekvés 215
3.5 A racionális aracionális és az aracionális racionális 222

4 Társadalmi kapcsolat **233**
4.1 Ágensre irányultság: társas preferencia 234
4.2 Társas értékorientáció 236
4.3 A társadalmi kapcsolat típusai 260
4.4 Interdependens cselekvések 262
4.5 Hatalom 264

5 Kényszermentes társadalmi kapcsolat 269
5.1 Magjátékok . . . 272
5.2 A játékok statikus kiterjesztése . . . 288
5.2.1 Aszimmetrikus játékok . . . 288
5.2.2 Többalternatívás játékok . . . 290
5.2.3 Többszereplős játékok . . . 291
5.2.4 Nem szigorúan rendezett játékok . . . 293
5.3 A játékok magasabbrendű kiterjesztése büntetéssel . . . 294
5.3.1 Aszinkron játékok . . . 297
5.3.2 Alkujátékok . . . 298
5.4 A játékok dinamikus kiterjesztése emlékezettel . . . 302
5.4.1 Reciprocitás . . . 304
5.4.2 Direkt reciprocitás játékok . . . 304
5.5 A játékok közösségi kiterjesztése reputációval . . . 315
5.5.1 Indirekt reciprocitás játékok . . . 315
5.5.2 A direkt és indirekt reciprocitás közös mintázatai . . . 322
5.6 Kooperatív játékok: koalícióelmélet . . . 324

6 Pozíció, norma, uralom 327
6.1 Perszonális deontika . . . 327
6.2 Normatív pozíciók, jogosultságok és kötelezettségek . . . 330
6.2.1 Hohfeld elsőrendű jogviszonyfogalmai . . . 330
6.2.2 Az elsőrendű jogviszonyfogalmak formalizálása . . . 344
6.2.3 Hohfeld másodrendű jogviszonyfogalmai . . . 353
6.2.4 A másodrendű jogviszonyfogalmak formalizálása . . . 358
6.2.5 A jogviszonyfogalmak és szabályfogalmak viszonya . . . 364
6.2.6 A jogviszonyfogalmak teljes logikai rendszere . . . 375
6.2.7 A jogviszonyfogalmak kiterjesztése . . . 376
6.3 Társadalmi szerep . . . 379
6.4 Norma . . . 381
6.5 Normatípusok . . . 389
6.6 Uralom . . . 397
6.7 Uralomtípusok . . . 401

7 Közösségi döntések 411
7.1 Aggregálás mint közösségi döntési technika . . . 411
7.2 Preferenciaaggregálás . . . 425
7.2.1 Manipuláció . . . 426
7.2.2 Algoritmusmanipuláció . . . 427
7.2.3 Condorcet-paradoxon . . . 430
7.2.4 Arrow lehetetlenségi tétele . . . 431
7.2.5 Lehetségességi tételek . . . 439
7.2.6 No-show paradox . . . 444
7.2.7 Reprezentációs paradoxon . . . 446
7.2.8 Szavazatmanipuláció . . . 447
7.2.9 Szavazatcsere . . . 449
7.2.10 Csalásbiztos eljárás . . . 451

7.2.11 Gibbard-Satterthwaite tétel 453
7.2.12 AI-válasz 457
7.2.13 Párosításelmélet 460
7.2.14 Körzethatár-manipuláció 463
7.2.15 Többkörzetes paradoxon 465
7.2.16 Többesetes paradoxon 466
7.2.17 Anscombe-paradoxon 469
7.2.18 Osztrogorszkij paradoxon 471
7.3 Véleményaggregálás 473
7.3.1 Doktrinális paradoxon 473
7.3.2 Diszkurzív paradoxon, List-Pettit lehetetlenségi tétele 475

Előszó

„Dr. Felix Smith, a Stanford Kutatóintézet Molekuláris Fizika Tanszékének vezetője mesélt egy fizikus barátjáról, aki Los Alamos ban dolgozott a II. világháború után. Egy bonyolult probléma megoldása végett elment a híres magyar matematikushoz, Neumann Jánoshoz, aki Los Alamosban volt konzultáns.
A karakterisztikák módszerével egyszerűen meg lehet oldani. — mondta Neumann. — *Félek, nem értem a karakterisztikák módszerét.* — felelte a fizikus. — *Fiatalember*, – válaszolta Neumann – *a matematikában az ember nem megérti a dolgokat, hanem megszokja.*." [597, 208])

A cselekvéselmélet a társadalmi cselekvések leírásával, megértésével, magyarázatával foglalkozik. Ebben a minőségében – definíció szerint – a módszertani individualizmus talaján áll. A cselekvéselmélet fogalmát azonban lehet szűkebb és tágabb értelemben használni. Szűkebb értelemben véve beszélhetünk azokról a modellekről, amelyek azt a kérdést próbálják megválaszolni, hogy mikor tekinthető valami cselekvésnek. A fogalom tágabb értelemben minden olyan szempontot és problémát ide sorolhatunk, amely a társadalmi cselekvések szabályszerűségeinek jobb megértéséhez segíthet minket. Ebben a könyvben a cselekvéselmélet fogalmának kiterjesztett jelentését fogadom el, de szentelek egy külön fejezetet a szűkebb értelmezésnek is.

Jack Hirshleifer egy 1985-ös cikkében leírt egy fontos (és sokakat provokáló) mondatot: "csak egyetlen társadalomtudomány létezik" [237]. Ezt én is így gondolom. Hirshleifer cikke a közgazdaságtan imperializmusával és mindenhatóságával szemben megfogalmazott programként is értelmezhető.[1] Az egyébként közgazdász szerző egyrészt rámutat arra, miért tudott minden társadalomtudományi területre behatolni a közgazdaságtani megközelítés, másrészt elemzi azt is, hogy milyen hiányosságai vannak a közgazdaságtan világképének, hol és miért kell kibővíteni az eszközkészletét, fogalmi modelljeit, harmadrészt felvázolja azt az egységes társadalomtudományi keretrendszert, amelyben a "kigészített" közgazdaságtan és a biológia által teremtett alapokra lehet a további szaktudásterületeket felépíteni. Egy ilyen programmal óvatosan egyetértve egyetlen ponton egészíteném csak ki azt: a társadalmiságunk döntő elemének tartom a társadalmi valóság konstruált jellegét, amit sem a biológia, sem a közgazdaságtan nem képes megragadni, ezért ezt önálló, programalkotó építőelemnek tartom. Könyvemmel e program jegyében szeretnék minél pontosabb képet bemutatni a társadalmi cselekvés világáról.

Több, mint két éven át írtam ezt a könyvet. Talán a legfontosabb célom az volt, hogy az érték, a norma, az érdek, a racionalitás, a racionális cselekvés fogalmait minél pon-

[1] A közgazdasági imperializmusról magyarul lásd: [112].

tosabban definiálni, értelmezni tudjam. Azt gondoltam, hogy ezek azok az alapvető fogalmak, amelyekre támaszkodva a leginkább képesek lehetünk magyarázni a társadalmi cselekvések szabályszerűségeit. Mire a könyvem végére értem sok minden átértékelődött bennem e céllal kapcsolatban. Ma már úgy látom, hogy az embert leginkább az érzelmein és a szokásain keresztül érthetjük meg. Nem könnyű belátni, még nehezebb elfogadni, de én – ma már – úgy gondolom, hogy választásainkat az érzelmeink határozzák meg. Ha meg akarjuk érteni a döntéseinket, cselekvéseinket, akkor bizony az érzelmek kihagyhatatlanok, tehát módszeresen fel kell tárnunk az érzelmeink működését, hatásmechanizmusait. De talán még ennél is fontosabb ténye az életünknek, hogy a cselekvéseink igen nagy részét (amit csak lehet) szokásokká, reflexekké alakítjuk (habituáljuk), és ezek után, illetve ezek alapján sok esetben "egyszerű" cselekvésautomataként működünk. Ingereket, inputokat kapunk, és ezekre reflexszerűen válaszolunk. Sokkoló élmény volt ráébredni arra, amire két híres ember, Neumann János és Alfred North Whitehead mondatai utalnak: még a racionális működésmódunk is – jelentős részben – megszokáson alapul. Az agyunk számára így gazdaságos, tehát metaszinten, evolúciós értelemben racionálisnak mondható, de akkor is nehezen emészthető. Ez az elmozdulás még nem igazán érezhető a könyvemben.[2] Amíg írtam, mindvégig az eredeti célok – ha tetszik, a racionalitás paradigma – mentén haladtam. Mindazt, amit itt megírtam, szükségesnek gondolom a társadalmi cselekvések szabályszerűségeinek magyarázatához. Szükségesnek, de messze nem elégségesnek. Ma már úgy látom, hogy sokkal kisebb – vagy legalábbis más – szerepe van az életünkben a racionalitásnak, a mérlegelésnek, a kalkulációnak, mint ahogy korábban hittük, hittem. De legalább van mit újragondolni, újraértelmezni, feltárni, megérteni.

Amikor szociológushallgatóként először kellett Max Weber munkáit olvasnom, bár éreztem, mennyire fontos szerző, sokáig nem igazán értettem. Sokat küzdöttem azért, hogy befogadjam Weber mondandóját. Volt, hogy oldalakon át szóról szóra leírtam egy papírra a meghatározásait, hogy jobban megérthessem. Szép lassan bigott Weber hívővé váltam. Úgy vélem, hogy a weberi kategóriatan a mai napig a legpontosabb fogalomkészletet nyújtja az emberi cselekvés értelmezéséhez. Van, ahol pontosítani érdemes, van, ahol ki kell egészíteni, de a szerkezet egésze ma is szilárdan áll [519]. Mindez kiderül a könyvemből is, mégcsak figyelmes olvasónak sem kell lenni hozzá. Nem is akarom ezt letagadni. Könyvem egyfajta tisztelgés is az utolérhetetlen Mester munkássága előtt.

Amikor a könyvem témáihoz kerestem tanulmányokat, könyveket az elmúlt években, hol megdöbbenve, hol irigykedve, egyik ámulatból a másikba esve szembesültem azzal, hogy milyen fontos felismeréseket, eredményeket produkáltak a kognitív tudomány, biológia, agykutatás, pszichológia metszetében dolgozó kutatók. Ma már nem mernék belekezdenni a könyvembe anélkül, hogy ne építsem bele a mondandómba a lehető legnagyobb

[2] Egy időközben megírt cikkemben azért már tetten érhető ez a felismerés, lásd [521].

mértékben ezt a tudást. A társadalomtudományi gondolkodást legfontosabb kérdéseit illetően van, amit más megvilágításba helyez, van, amit megerősít és bizonyos értelemben megalapoz.

A könyv (különösen a könyv bizonyos részeinek) megírásakor egy nagyon is gyakorlati cél lebegett a szemem előtt. Kezdetektől fogva úgy gondoltam, hogy a formális imperszonális deontika kidolgozásától talán valamiféle kiindulási pontot, logikai lapot nyújthat egy olyan jogi szemantika (jogi ontológia) kidolgozásához, amely valamikor a normatív, elsősorban persze a jogi szövegek gépi feldolgozását lehetővé teszi. Idővel kiderült, hogy több más fejezetet is meg lehet már úgy írni, hogy a természetes nyelvű szövegfolyamba beillesztem az alapfogalmak formális meghatározásait is, hogy egy valamikori gépi elemzés számára fogodzókat tudjak kínálni. Tudom, hogy ezzel gyakran saját szövegem befogadatóságát nehezítem, de ezt kisebb veszteségnek éreztem a potenciális előnyökhöz képest.

A könyv túlnyomó részét a *Normák, közös tudás, kooperáció* című (OTKA K83887) kutatás keretein belül írtam meg. Köszönöm mindazok támogatását, akik bizalmat szavaztak a kutatásnak, és köszönöm azok segítségét is, akik velem együtt részt vettek ebben a kutatásban.

Chapter 1

Logikai, módszertani alapok

„ ... »a matematika egyik fegfontosabb hozzájárulása a tudományos problémák megoldásához az lehet, hogy az adott problémával kapcsolatban a logikailag létező lehetőségek kimerítő elemzését adja«. Egy ilyen elemzés eredménye a logikailag létező lehetőségek halmazának felosztása egymást kölcsönösen kizáró, de együttesen kimerítő alternatívákra." (idézet Kemeny, Snell és Thompson könyvéből [308, 63])

Ebben a fejezetben bemutatom, értelmezem azokat a módszertani, logikai fogalmakat és elméleteket, amelyeket használni fogok a cselekvéselmélet felépítése során. Nem az a célom ezzel a fejezettel, hogy kimerítő és minden szempontból megalapozott logikai ismereteket adjak át az olvasónak. Csak azt szeretném, hogy egyfelől a könyvem gondolatmenete kellő mértékben megalapozott legyen, másfelől akik birtokában vannak az itt használt logikai apparátusoknak, azok tudják kötni ezeket a fogalmakat az általuk már ismert szerzőkhöz, diszciplinákhoz, akik számára pedig mindezek újdonságot jelentenek, azok tudják, hová kell fordulniuk, ha mélyebb tudásra vágynak.[1]

A logika eszköz ahhoz, hogy a gondolatainkat következetes és egyértelmű módon fejezhessük ki. Az egyértelműség és ezzel a – relatív – konzisztencia lehetősége abból fakad, hogy az elméleteink fogalmait valamilyen formális nyelven, formulák segítségével is reprezentálni tudjuk. Ez azonban nem jelenti azt, hogy csak formulákon keresztül fejtenénk ki mondanivalónkat. A tudományos kommunikációban lényegszerűen két nyelvet használunk. Az emberhez természetes nyelven szólunk, az elméleteinkhez használt formális nyelv(ek) mindig csak kiegészítik a természetes nyelvű kommunikációt. De ha szükséges, az egyértelműsítés kedvéért a formális nyelvekhez fordulhatunk.

További előnye lehet még a logika (matematika) alkalmazásának az a "megnyugtató" érzés, amely akkor keletkezhet bennünk, amikor választ várunk arra a kérdésre, vajon

[1] A logika iránt érdeklődő olvasóknak magyar nyelven Ruzsa Imre könyveit ajánlom (például ezt: [454]), a matematikai logikai kérdésekben Ferenczi Miklós könyvére támaszkodtam [162].

az éppen adott elmélet építése közben lehetne-e más irányban továbbmenni, más fogalmakat használni, más téziseket keresni. A tudományos gondolkodók versenyében mondanak ezt, mondanak azt, és mindig ott lapul a kérdés, nem lehet-e még további fogalmakat, állításokat, elméleteket az addigiak mellé tenni. A fejezet mottájában idézett gondolat arról szól, hogy miután már elköteleztük magunkat bizonyos fogalmak, axiómák, tézisek, elméletek mellett, utána már gyakran kerülhetünk olyan helyzetbe, amikor a logika, matematika összefüggéseire támaszkodva mondhatunk olyat, hogy ennyi és ennyi, ez és ez a lehetőség van elméletileg, megvizsgálhatjuk az összeset, és abban az értelemben megnyugodhatunk, hogy azon a ponton további alternatívákat már nem kell, mert nem is lehet keresnünk.

Az egyértelműség és a konzisztencia kialakításának és fenntartásának lehetőségén túl van még egy érv a logikai apparátus alkalmazása mellett. Egyfajta takarékossági, gazdaságossági szempontból is gondolhatunk a logikára, amikoris az elméletépítéseink során újra és újra alkalmazzuk a logika különböző következtetési szabályait, kalkulusait, szillogizmusait. A következtetési szabályok folyamatos alkalmazása (újrahasznosítása) nagyon fontos hozadéka a logikának.[2]

A gazdaságossági szempontokat azonban nem minden áron érdemes/lehet figyelembe venni. A rokonsági fogalmak (apja, anyja, fia, lánya, testvére, húga, nagyapja, ükapja, sógora, ángya stb.) többszáz tételből álló rendszerét kifejezhetnénk 5-6 alapfogalom segítségével, de nem tesszük ezt[3]. Terminológiai szempontból gazdaságos lenne, hiszen a többszáz helyett csak néhány terminust kellene felvenni a modellünkbe, viszont ennek az lenne az ára, hogy hosszú kifejezéseket kellene képeznünk az alapterminusainkból. 'Ükapád' helyett az 'apád apja apjának apja' alakot (formulát) kellene mondanunk. Ez se kimondani, se befogadni nem lenne jó. Állandóan dolgozni kellene azon, hogy az adott kifejezés mit is jelent. Könnyebb nekünk sok terminust és kevesebb feldolgozási műveletet kezelnünk, mint fordítva. Érdemes hallgatnunk Whitehead fenti – az előszó mottójába foglalt – figyelmeztetésére.

A logika egyfelől egy nagy hagyományokkal bíró, önálló tudományág. Ebben a minőségében behatárolható, közös vizsgálati tárggyal, sajátos elemzési szempontokkal, módszerekkel rendelkezik. Másfelől logikának mondhatunk minden formális nyelven kifejezett elméletet is. A kommunikálni kívánt fogalmakat az adott nyelvnek megfelelő terminusok (nevek és predikátumok) segítségével fejezzük ki, ezek révén kijelentéseket (mondatokat, formulákat) képzünk, melyekhez aztán igazságértékeket rendelhetünk. Mindeközben figyelembe vesszük a nyelvhez kapcsolható szintaktikai szabályok összességét, és mindezt pedig azért csináljuk, hogy a terminusainkkal, mondatainkkal szemantikalag (tehát jelentésüket tekintve) is megfelelő üzeneteket hozhassunk létre. Ebben az értelemben sokféle logikát hozhatunk létre, melyeket aztán különféle szempontok alapján természetesen tipizálhatunk, ezáltal "korlátozhatjuk" a használatba vehető logikák számosságát.

Minden logika számára fontos a **logikai kalkulus** (röviden: kalkulus) fogalma. A kalkulus azon univerzális szabályok (kijelentések, formulák, tételek) összefüggő rend-

[2]Meg kell itt jegyezzem, hogy a következtetési szabályok folyamatos alkalmazása másrészről az elméleti konzisztencia fenntartásához is hozzájárul.

[3]A rokonsági fogalmakról szóló tanulmányomat meghivatkozhatóvá kell tenni.

szere, amelyek mindig igaznak bizonyulnak, bármilyen, az adott logikai nyelvhez tartozó kijelentést (formulát), predikátumot vagy nevet helyettesítünk beléjük. A kalkulus tehát egy axiómarendszer, amely az adott logikai nyelv lehetséges formulái közül részben azokat a logikai igazságokat írja le, amelyek tautológiák (mindig igazak), részben azokat a szabályokat adja meg, amelyek segítségével további formulákat lehet levezetni az adott nyelven. E kettősséggel voltaképpen azt ígéri minden logika, hogy a kalkulusokra támaszkodva minden további állítást, formulát ki lehet fejezni (le lehet vezetni), ami másképp szólva azt jelenti, hogy ami az az adott nyelven "elmondható", kifejezhető, az elmondható, kifejezhető csak a kalkulusba tartozó axiómák használatával – legfeljebb nagyon bonyolult, nagyon hosszú kijelentések, formulák megfogalmazására lehet szükség. Mivel azonban ez a bonyolultság megnehezíti a nyelv – emberek általi – használatát, ezért az adott nyelven nemcsak lehet, de kívánatos is új és új formulák, terminusok bevezetése, ami logikai értelemben redundáns, praktikus szempontból viszont gazdaságos. Említettem már azt a köznapi példát, miszerint hivakozhatnánk valakire úgy is, mint az 'apám apjának az apjának az apja', illetve mint az 'ükapám', mindkétszer ugyanarra a személyre mutathatnánk rá (tekintsünk most el az apai és anyai vonal különbségétől). Ha az első megoldást választom, akkor hosszabban kell beszélnem és nehezebben lehet befogadni az üzenetet, viszont ezen a nyelven nem kell új terminust bevezetnem, megtanulnom és megtanítanom. A második megoldás gyorsabb és könnyebben folytatható kommunikációt kínál, aminek az az ára, hogy új terminust kell a nyelvbe felvenni. Ennek az a hátránya, hogy az ember számára nyilván korlátos a befogadható terminusok száma.[4] Változik a helyzet akkor, amikor számítógépeket is bevonunk a kommunikációba, mert azok számára nem jelentkeznek ilyen feldolgozási problémák. A gépek számára sokszor elhagyható ez a terminológiai redundancia, bár azt mindig mérlegelni kell, hogy a gépek valamilyen ponton már emberekkel is kapcsolatba kerülnek is ilyenkor váltani kell az ember számára kényelmesebb redundáns "beszédmódra".

Minden logika fontos tulajdonsága a logika kifejező- vagy leíróereje, amivel azt fejezhetjük ki, hogy milyen típusú állításokat, fogalmakat tudunk megragadni az adott formális nyelv segítségével. Ez más szempontból azt jelenti, hogy az adott nyelven kifejezett elméletbe milyen ontológiai elköteleződések mentén veszünk fel új és új terminusokat (fogalmakat). Ha olyan logikánk van, amely csak mondatokat (kijelentéseket) tud kezelni, akkor tartalmilag nagyon keveset állíthatunk arról, mi van a világban, és természetesen nagyon keveset (gyakorlatilag semmit sem) tudunk elmondani arról, mit gondolunk a világról. Ha a rokonsági rendszerek világát akarjuk egy elméletben leírni, akkor fel kell vennünk pár alapfogalmat (GYEREKE, FIATALABB, FÉRFI, HÁZASTÁRSA), és ezek segítségével már minden ismert rokonsági fogalmat kifejezhetünk. Ám amíg nem vesszük be a modellünkbe a férfi kategóriáját (vagyis nem köteleződünk el az ember kettős szexuális nembelisége mellett), addig ugyan definiálni tudjuk a szülő, a testvér, a nagyszülő stb. fogalmait, de nem tudunk különbséget tenni az apa és anya, a fia és lánya, a húga és bátyja fogalmak között. Ez a problémakör kapcsolódik a logikai nyelv pontossága, finomsága, granularitása kérdéséhez, ami kijelöli a nyelv helyét az általánosság vs. specifikusság tengelyen.

A következtető rendszer használhatóságát felülről korlátozza a használt ontológia nyelvének kifejezőereje, az a tény, hogy au adott nyelv segítségével mennyire komplex össze-

[4]Ez a probléma nemcsak a formális, de a természetes nyelvek esetében is ugyanúgy (és ugyanazért) létezik.

függések írhatóak le. Azonban minél nagyobb egy nyelv kifejezőereje, annál szűkebb az olyan kérdések köre, melyre (matematikailag) egyáltalán lehetséges a válaszadás. Így találni kell egy egészséges egyensúlyt, melynél már kellően bonyolult összefüggéseket le tudunk írni, ugyanakkor a kérdések megválaszolhatósága még nem reménytelen.

A formális nyelvek bonyolultsága, számíthatósága, eldönthetősége nagyon fontos problémakört alkot, a nyelvek használhatósága szempontjából is kiemelt kérdésnek számít, ám e könyv mondanivalója szempontjából elhagyhatónak tűnik, ezért nem tárgyalom.

Az általam használni kívánt logikák egymásra épülését két dimenzióhoz igazodva lehet kifejezni. Figyelni lehet arra, hogy *(i)* hanyadrendű logikáról van szó, illetve *(ii)* extenzionális vagy intenzionális logikát vizsgálunk-e [454]. Az extenzionális logikákat úgy veszem sorba, hogy kezdem a nulladrendű propozicionális logikával, majd áttérek az elsőrendű predikátumlogikára, valamint az utóbbi alá sorolható kategoriális logika néhány kérdésére, végül röviden kifejtem, hogy hogyan lehet definálni a másodrendű (magasabbrendű) logikákat, és miért fontos figyelembe venni azt, hogy magasabbrendű logikákról (és fogalmakról) van-e szó. Az intenzionális logikák közül csak a modális logikát veszem elő, azt is csak azért és olyan mélységben tárgyalom, hogy érthetővé tudjam tenni azt, miért van szükségünk a deontikus logikára a társadalmi cselekvések vizsgálata során.

A cselekvéselmélet tárgyalásához szükséges fogalmakat – ahol csak tudom – megpróbálom úgy bemutatni, hogy formális nyelven reprezentált terminusokkal fejezem ki a fogalom jelentését. Ezt úgy teszem, hogy megmutatom az új, éppen definiálandó fogalom helyét a már létező, addig már bemutatott fogalmi modell struktúrájában. Ezáltal – remélhetőleg – átláthatóbbá válik, hogy az éppen felvett új fogalomhoz milyen ontológiai elköteleződések szükségesek. Az egyes – jól elkülöníthető – elméleteket önálló logikákkal írom le. A cselekvéselmélet egész belül az alábbi logikákat (és a hozzájuk tartozó jelöléseket) fogom használni.

(AL.xx)	aggregációlogika (HUN)	aggregation logic (ENG)
(BA.xx)	preferenciaalgebra (HUN)	betterness algebra (preferential algebra) (ENG)
(BL.xx)	preferencialogika (HUN)	betterness logic (preference logic) (ENG)
(DL.xx)	deontikus logika (HUN)	deontic logic (ENG)
(EL.xx)	episztemikus logika (HUN)	epistemic logic (ENG)
(FL.xx)	formális logika (HUN)	formal logic (ENG)
(GL.xx)	kategórialogika (HUN)	category logic (ENG)
(HL.xx)	halmazlogika (HUN)	set logic (ENG)
(IL.xx)	interakció logika (HUN)	interaction logic (ENG)
(ML.xx)	modális logika (HUN)	modal logic (ENG)
(PC.xx)	kijelentéskalkulus (HUN)	propositional calculus (ENG)
(PL.xx)	kijelentéslogika (HUN)	propositional logic (ENG)
(QL.xx)	predikátumlogika (HUN)	predicate logic (ENG)
(QC.xx)	predikátumkalkulus (HUN)	predicate calculus (ENG)
(RA.xx)	relációalgebra (HUN)	relation algebra (ENG)
(RL.xx)	relációlogika (HUN)	relation logic (ENG)
(ZL.xx)	intencionalitáslogika (HUN)	intentionality logic (ENG)

1.1 Metafogalmak

A formális nyelveket mindig a természetes nyelvek analógiájára építjük fel abban az értelemben, hogy szavakat (kifejezéseket, terminusokat, fogalmakat) definálunk, amelyekkel a világban létező vagy létezőnek hitt dolgainak valamely halmazára mutatunk rá, illetve ezen szavak egymás mellé illesztésével (relációba állításával) mondatokat (kijelentéseket) állítunk elő azzal az igénnyel, hogy azokkal érvényes állításokat tegyünk a világról. Eközben észrevétlenül használunk olyan fogalmakat, amelyeket magától értetődőnek tekintünk, de nem – vagy csak ritkán – reflektálunk rá. A legfontosabbakat ezek közül bemutatom itt.

A nyelvhasználat során – a dolgokat kissé elegyszerűsítve – kétfajta dolgot teszünk. Egyfelől szószerű nyelvi jeleket képzünk, hogy azokkal létező vagy tételezett dolgok halmazára rámutathassunk, referálhassunk rájuk, a dolgokat reprezentálhassuk a "szavakkal". Ezekkel a szószerű nyelvi jelekkel kevés dolgot fejezhetünk ki, amikor használjuk őket. Talán csak annyit, hogy az adott nyelvi jelet kibocsátva (leírva, kimondva) tételezzük a világ bizonyos dolgainak létezését – legalább annyira, hogy aztán kijelentéseket tehessünk róluk. Az ilyen megnyilatkozások érvényességét csak annyiban lehet vitatni, hogy magát a létezési feltételt támadhatjuk, semmi mást. Akkor tudunk tartalmasabb, mélyebb és szerteágazóbb érvényességvizsgálatba keveredni, ha mondatszerű nyelvi jeleket képzünk. Ezt úgy tehetjük meg, ha a szószerű nyelvi jeleket kapcsolatba állítjuk egymással. Ez a kapcsolatba állítás (valamilyen reláció képzése) az a művelet, amelynek révén szószerű jelekből állításokat keletkeztethetünk, ezért a reláció a mondatszerű megnyilatkozások kulcsa. Mivel a szószerű jelek egymáshoz illesztését rengetegféleképpen tehetjük meg, a mondatszerű megnyilatkozások szintjén megnő a kifejezés gazdagsága, mélysége, szélessége és vitathatósága egyaránt. Amikor egy elméletet építünk fel, akkor ezt a kettősséget kell mindvégig szem előtt tartani.

A logika alapfogalma az állítás. Azt természetesen nem állítom, hogy ez a fogalom ne lenne reflektált a logika világában, de azt fontosnak tartom hangsúlyozni, hogy sokszor zavart okoz az a tény, hogy a logika nyelvek túlnyomó részében nincs önálló jele, reprezentációja ennek a fogalomnak.[5]

(FL.0) PROPOSITION $\equiv A, B, C$

Az **állítás** igazságra vonatkozó érvényességigénnyel fellépő, adott nyelv szabályai szerint érvényes nyelvi megnyilatkozást kibocsátó aktus eredménye. Nagyon sok szinonimája van (jelen írás keretein belül): **kijelentés**, **mondat**, **asszerció**, **propozíció**, **ítélet**, **affirmáció**, **formula**. A tényleges gyakorlatban az állítás fogalmának nincs külön jele. Az állítások logikai értéke igaz vagy hamis lehet, amit szöveges megjelenítésben 'igaz' vagy 'hamis' karakteres formák, táblázatokban a '0' és az '1' értékek feltüntetésével fejezhetünk ki.

Minden nyelvben fontos a tagadás fogalma, amely az állítás "párja". Leggyakrabban olyan műveletként definálják, amely valamely állításhoz képest előállít egy újabb állítást úgy, hogy az eredeti és az új állítás logikai értéke ellentétes egymással.[6]

[5] Frege, a fogalomírás és a modern logika egyik megteremtője még használt külön jelet (ítéletvonalat) az állításra, de a követői mind elhagyták ezt a szokást.

[6] Itt hivatkozni kell a metarelációs fejezet művelet fogalmára.

(FL.1) NEGATION $\equiv \neg \equiv \sim$

A **tagadás** (negáció) az a nyelvi művelet, amely során valamely állítás logikai értékét ellentétére változtatjuk. A tagadás állításra irányul függetlenül attól, hogy egyszerű vagy összetett állításokról van-e szó.

Az állítással, illetve annak tagadásával a mondatszerű nyelvi megnyilatkozásokat reprezentálhatjuk. Ahhoz, hogy a kijelentéseink belső tartalmát, tehát a szószerű jeleket is meg tudjuk ragadni, szükség van a terminus fogalmára. Minden szószerű nyelvi jelre, amelyekkel még nem tudunk állításokat kifejezni, ráhúzhatjuk a terminus kategóriáját. A terminussal a világ dolgaira mutatunk rá, és ezt a kapcsolatot (a terminusok és a világ között) egyfelől a terminus terjedelmébe sorolt dolgokkal, másfelől a terminus jelentésével jellemezhetjük.

(FL.2) TERM $\equiv a, b, c, x, y, z,$ **P,R,Q**

A **terminus** a világ adott dolgaiból álló halmaz létezését reprezentáló nyelvi megnyilatkozás.

A tagadás mellett használjuk gyakran a komplementálás fogalmát is, aminek szoros kapcsolata van ugyan a tagadás fogalmával, sokban hasonlít hozzá, de mégsem azonos vele.

(FL.3) COMPLEMENT $\equiv \neg \equiv \sim$

komplementálás az a logikai művelet, amely során a terminus terjedelme alapján vesszük az univerzális halmazhoz képest komplementer terjedelemmel rendelkező terminust.

A komplementaritás a halmazelmélet gyakran használt fogalma, és mivel valamely halmaznak az univerzumhoz képest vett kiegészítését jelenti, a komplementer halmaz képzése a tagadás műveletével mutat szoros rokonságot. Egy halmaz és annak komplementere úgy viszonyul egymáshoz, mint egy állítás a tagadásához. Előfordul, hogy a tagadás, negálás terminusait is a komplementálás értelmében alkalmazzák, és a **komplementerképzés** terminusa is szinoním a **komplementálás** terminusával, míg a komplementaritás fogalma az egész jelenségegyüttest írja le. A halmazelméletben a komplementer halmaznak (és így áttételesen a komplementálás műveletének) van saját jelölése, ez olykor felbukkan a formális nyelvekben is, de gyakraban nem alkalmazzák a tagadás műveletét jelző szimbólumo(ka) erre a műveletre. A halmazelméleti eredet és analógia rámutat arra, hogy ott lehet alkalmaznunk a komplementálás fogalmát, ahol elemekkel, előfordulásokkal, terjedelemmel van dolgunk, ez pedig a terminusok világa. A komplementálás művelete a terminusok terjedelmébe sorolt előfordulások univerzumhoz képest vett kiegészítése. A tagadás művelet viszont – elsősorban – kijelentésekre vonatkozik, ennyiben szemben áll a komplementálás fogalmával.

Következtetések

Amikor egy nyelvi megnyilatkozásnak ismerjük valamilyen belső szerkezetét, akkor mindig felmerülhet az a kérdés, vajon értelmes-e és ha igen, milyen eredményeket produkál az, ha a belső szerkezetét (vagyis az összetevők egymáshoz való viszonyát) megváltoztatjuk. A későbbi tárgyalandó elméletekben is elő-előkerülnek ilyen-olyan változtatási, transzformációs műveletek, úgyhogy röviden érdemes kitérni erre a kérdéskörre. Már

csak azért is érdekesek ezek a szabályok, mert közülük több is a *közvetlen következtetések* vagyis az egyetlen állításból levonható következtetések körébe tartozik, és mint ilyenek a következtetési szabályok egyik fontos csoportját jelentik.

Egy nyelvi megnyilatkozásnak kétféle értelemben lehet belső szerkezete. Egyfelől ha mondatszerű megnyilatkozásokat vizsgálunk, akkor a "mondatok" belső szerkezetét nem ismerhetjük, viszont elemezhetjük a mondatok összekapcsolását, amikoris az egyes mondatok mint összetevők már a megnyilatkozás mint egész részeiként, belső szerkezeti egységeiként értelmezhetők. Másfelől viszont tekinthetjük a mondatszerű megnyilatkozást (vagy csak annak egy részét) is olyan egységnek, amelynek a belső szerkezete érdekel minket, amikor a szószerű összetevőit és azok egymáshoz, illetve a mondategészhez (vagy csak a mondat egy részéhez) való viszonyát akarjuk megérteni.

Jelölési konvenciók

A tudományos diskurzusban fogalmakat, tételeket, elméleteket gyártunk, és igyekszünk minél pontosabban kifejezni ezeket. Ebebn nagy segítséget jelent minél több formalizmus alkalmazása. De még ha így teszünk, akkor is szavakkal kommunikálunk egymás felé, és ilyenkor mindig könnyen pongyolává válhatunk. Egyszerűen azért, mert amikor az esetleg egyértelműen formalizált fogalmainkat természetes nyelvi szinten fejezzük ki (terminusok, kifejezések, nyelvi konstrukciók alkalmazásával), folyamatosan belebonyolódunk a szinonimitás és homonimitás hálójába. Még a tudományos kommunikációban is tele vagyunk hononím és szinoním szavakkal, kifejezésekkel. Ezen a nehézségen azzal próbálok felülkerekedni, hogy amikor egy-egy új fogalmat vezetek be, akkor mindig megadom azt, hogy hogy hivatkozom rá fogalomként, illetve hogy hivatkoznak rá a különböző tudományterületek képviselői, a különböző szerzők természetes nyelvi konstrukciókkal.

A fogalmakat KISEBB MÉRETŰ, VERZÁL BETŰTÍPUSSAL jelzem és angol nyelvű terminus reprezentálom. A fogalmi síkon megpróbálom teljesen kizárni a homonimitás lehetőségét, és bár ezen a szinten lehetségesnek vélem a szinonimitás jelenségét, de igyekszem ezeket is elkerülni. A természetes nyelvi konstrukciók, terminusok jelölésére – első és/vagy lényeges felbukkanásuk során – **vastagított, talpatlan betűt** használok. A konstrukciók szintjén mind a homonimitás, mind a szinonimitás jelensége gyakran előfordul. A logikai formalizmusokkal támogatott modellépítés egyik fontos célja pont az, hogy ahol csak lehetséges meg lehessen találni és rá lehessen mutatni olyan helyzetekre, amelyekben a tudományos kommunikációban szinoním vagy homoním terminusok használatát tapasztalhatjuk. Amikor a terminusok, konstrukciók szintjén az egyértelműsítés miatt ez szükségesnek látszik, akkor zárójelben megadom, hogy az adott konstrukció melyik fogalomhoz kapcsolódik. Ezt természetesen akkor teszem csak meg, amikor egy-egy új terminust (fogalmat) bevezetek a modell tárgyalásakor, illetve amikor a szövegkörnyezetből nem derül ki egyértelműen, hogy az éppen használt kifejezés melyik jelentéséről van szó. s Minden nyelvhez tartoznak központozást segítő jelek, amelyeket nem tüntetünk fel a nyelv definiálásakor, de mindvégig használjuk őket. Ilyenek a nyitó és csukó zárójelek ('(' és ')'), valamint az elemek felsorolását segítő vessző ',' jele. Ugyancsak "univerzálisan", azaz minden nyelvben használok két ekvivalenciajelet: egyfelől egy új fogalom bevezetésekor a definiálás jelzésére alkalmazom a '$\stackrel{\text{def}}{=}$', illetve a $\stackrel{\text{def}}{\Leftrightarrow}$ **definíciós ekvivalencia** műveletjeleit (attól függően, hogy tárgynyelven vagy a trágynyelv és metanyelv

közötti kapcsolatról van-e szó), másfelől a terminusok szintjén létező szinonimitás kezelésére használom a '≡' **terminológiai ekvivalencia** műveletjelét.

1.2 Halmazelmélet

A halmazoknak, a halmazelméletnek kardinális szerepe van minden formális elméletben. Nincs szükség itt a halmazelmélet tárgyalására, de arra igen, hogy felsoroljam azokat a legfontosabb halmazelméleti fogalmakat (illetve ezek jeleit), amelyeket a könyvben mindvégig használni fogok.

(HL.0) $\mathscr{L}_{SL}(\mathbb{X}, \wp(\mathbb{X}), \varnothing, \mathbb{U}, \mathbb{N}, \mathscr{R}, \{a,b,c\}, \in, =, |\mathbb{X}|, \cup, \cap, \overline{\mathbb{X}}, \setminus, \subset, \subseteq)$, ahol

$\mathbb{X}$	tetszőleges halmaz
$\wp(\mathbb{X})$	$\mathbb{X}$ hatványhalmaz ($\mathbb{X}$ részhalmazainak halmaza)
$\varnothing$	üres halmaz
$\mathbb{U}$	univerzális (teljes) halmaz
$\mathbb{N}(\mathbb{N}^+)$	természetes számok halmaza (nullával együtt vagy nulla nélkül)
$\mathscr{R}$	valós számok halmaza
$\{a,b,c\}$	halmaz elemei
$\in$	eleme reláció
$=$	egyenlőség reláció
$\lvert\mathbb{X}\rvert$	halmaz számossága (kardinalitása) függvény
$\cup$	unióképzés halmazművelet
$\cap$	metszetképzés halmazművelet
$\overline{\mathbb{X}}$	komplementerképzés halmazművelet
$\setminus$	különbségképzés halmazművelet
$\subset$	valódi részhalmaza reláció
$\subseteq$	részhalmaza reláció

A halmazelméleti műveletek tulajdonságait, a köztük levő azonosságokat, a halmazelmélet paradoxonait nem mutatom be, de igazából nem is fogok ezekre támaszkodni a későbbiekben. Egy kivételt teszek csak itt azzal, hogy bevezetem a **partíció** (particionálás) fogalmát, mert erre mindvégig szükségem lesz. A halmazelméletben ezt a fogalmat akkor használjuk, amikor egy halmazt részhalmazokra akarunk bontani, miközben az az elvárásunk, hogy az részek ne legyenek átfedésben, és összeségükben fedjék le teljesen az eredeti teljes halmazt. Ha így osztunk fel egy halmazt, akkor beszélünk partícióról. Ezt a műveletet így lehet definiálni.

(HL.1) PARTITION$(\mathbb{X})$

1) $\forall i \in \mathbb{N}(\mathbb{S}_i \subset \mathbb{X}) \wedge \mathbb{S}_i \in \wp(\mathbb{X}))$
2) $\varnothing \notin \wp(\mathbb{X})$
3) $\bigcup_{\mathbb{S}_i \in \wp} \mathbb{S}_i = \mathbb{X}$
4) $\forall i \forall j((\mathbb{S}_i, \mathbb{S}_j \in \wp(\mathbb{X}) \wedge \mathbb{S}_i \neq \mathbb{S}_j) \rightarrow \mathbb{S}_i \cap \mathbb{S}_j = \varnothing)$

A fenti formulák segítségével biztosíthatjuk, hogy az ($\mathbb{X}$) halmaz olyan részhalmazokra legyen felosztva, amelyek egymástól teljesen elkülönülnek, viszont együttesen kiadják az egész halmazt. Lehetnek, akiknek furcsának tűnhet, mégis megkockáztatom, hogy a megfelelő particionálás a megalapozott fogalmi gondolkodás egyik legfontosabb biztosítéka (↰53.o.). A későbbiekben erre a kijelentésemre még visszatérek.

1.3 Kijelentéslogika

A logikai elméletek közül elsőként a nulladrendű extenzionális logikát veszem elő [454]. A későbbiekben inkább **kijelentéslogikaként** fogok hivatkozni rá, de ebben a könyvben szinonímaként kezelem még az 'állításlogika', illetve a 'propozicionális logika' terminusokat is. A kijelentéslogika a mondatszerű nyelvi megnyilatkozásokra fókuszál azáltal, hogy a mondatokba foglalt elemi állítások egymáshoz fűzésének lehetőségeit, szabályait vizsgálja. Ebből következően a kijelentések belsejével, összetevőivel, vagyis a szószerű nyelvi megnyilatkozásokkal nem foglalkozik. A kijelentéslogika lényegét a kijelentéseket összekapcsoló logikai műveletek adják, ezért először ezeket kell felvenni a fogalmi modellünkbe. A következő fejezetben ezeket veszem sorba. Előtte azonban rögzítenem kell azt, milyen nyelvre van mindehhez szükség.

(PL.0) $\mathscr{L}_{PL}(A, B, C, \neg, \rightarrow)$, ahol

A, B, C	propozíció formula
$\neg$	tagadás logikai művelet
$\rightarrow$	kondicionális logikai művelet

A propozicionális logika kevés ontológiai elkötelezettséget kíván meg, hiszen, láthatjuk, csak a kijelentések és a pár logikai művelet fogalmára van szükség hozzá. A kijelentéslogikát fel lehet építeni a kijelentéskalkulusra támaszkodva, de mielőtt ez utóbbit bemutatnám, egy hosszabb kitérőt teszek a logikai műveletek világába.

Kijelentéslogikai műveletek

Amennyiben veszünk állításokat és ezeket valamilyen művelet segítségével összekapcsol-juk, akkor egy újabb kijelentéshez jutunk, és a kijelentéslogika azt vizsgálja, hogy ezeknek az új állításoknak milyen lesz az igazságértéke. A modellünkbe fel kell vennünk a logikai művelet fogalmát,[7] ami kijelentések, állítások összekapcsolására vonatkozik. Mivel a fogalom nagyon széles körben használt, rengeteg szinonimája termelődött ki.

(PL.1) LOGICAL OPERATION

logikai művelet (HUN) – funktor (HUN) – junktor (HUN) – logikai kapcsoló (HUN) – logikai konnektívum (HUN) – konnektívum (HUN) – mondatoperátor (HUN) – propozicionális kapcsoló (HUN) – mondatkapcsoló (HUN) – propozicionális operátor (HUN) – logical operation (ENG) – logical connective (ENG)

A kijelentéslogikában a műveletek változóit, összetevőit, operandusait az ítéletek alkotják. Ha nem okoz zavart, akkor az ítéletekre mint a műveletek változóira operandusként fogok hivatkozni. Ha a kijelentéslogikában igazából csak azt vizsgálhatjuk, hogy milyen módon lehet összekapcsolni a kijelentéseket, és ezekben az összekapcsolásokban milyen szabályszerűségeket találhatunk, akkor kiemelt kérdés, hogy hányféleképpen és hogyan kapcsolhatjuk össze a kijelentéseket egymással. A válasz részben függ attól, hogy milyen műveletről is beszélünk. Nulláris, azaz nullaváltozós műveletből kettő, unáris, azaz egyváltozós műveletből négy, míg bináris, azaz kétváltozós műveletből tizenhat van.

[7] A művelet fogalmát itt nem definiáljuk, metanyelvi fogalomként használjuk. A tárgynyelvben használt művelet fogalmáról lásd a relációról szóló fejezetet (↬38.o.).

A két nulláris művelet, a tautológia és az ellentmondás jellemzője az, hogy igazságértékük független az általuk összekapcsolt állítások igazságértékeitől. A tautológia mindig igaz, az ellentmondás mindig hamis. Ezért is nevezzük nullaváltozós műveletnek őket, mivel "érzéketlenek" a bennük levő kijelentések mint változók értékeire.[8]

Az unáris műveletek értelmezésekor gyakran mondják azt, hogy a két nulláris művelet, a tautológia és az ellentmondás mellett ide tartozik még az állítás, a tagadás. A tagadásról természetesen lehet azt mondani, hogy az egy unáris művelet, aminek a bemenete egy állítás, és a kimenete egy – az állítással ellentétes igazságértékű – másik állítás. Mindig szem előtt kell azonban tartanunk azt a tényt, hogy a tagadás műveletét tetszőleges – tehát bármilyen összetettségű – állításra lehet vonatkoztatni. Beszélhetünk műveletek által előállított, összetett kijelentések tagadásáról is. A tagadás mindig egyváltozós művelet, de ettől még lehet többargumentumú. Ugyanaz igaz az állításra is természetesen, csak sokkal kevésbé tűnik intuitívnek azt mondani, hogy az állítás maga unáris művelet. Emiatt a fogalmi "kavarodás" miatt én ezen a ponton a 'projekció', illetve ennek tagadásaként a kontraprojekció fogalmát alkalmazom. A projekció olyan kétargumentumú, de egyváltozós művelet, amely a két argumentum közül csak az egyik értékeit veszi figyelembe, tehát csak attól az egy – projektált – változótól függ. A tizenhat bináris logikai művelet közé sorolhatjuk természetesen a négy unáris műveletet is, és ezeken túl van még 12 "szabályos", vagyis két-argumentumú és kétváltozós művelet.

A műveletek jellemzésekor bemutatom azokat a logikai jeleket, amelyeket valamilyen formális logikai nyelvben elterjedt módon használnak, azokat a – magyar – természetes nyelvű szavakat, kifejezéseket, amelyeket a fontosabb elméletekben gyakrabban felbukkantak, megmutatom az adott művelet ellentétét, valamint megadok néhányat a műveletet leíró formulák közül. Végül minden leírás mellett feltüntetek egy ábrát, ami a műveleti igazságtábla geometriai reprezentációjának tekinthető

Minden művelet esetében igaz, hogy a különböző logikai nyelveken olykor más műveleti jeleket alkalmaztak, ezért még itt is, a formális nyelvek világában is létezik a szinonimitás jelensége (sőt, a homonimitás is megfigyelhető, mert elő-előfordul, hogy ugyanazokat a logikai jelekkel más műveleteket reprezentálnak). Még gyakoribb azonban a szinonimitás (és homonimitás) előfordulása a formális elméleteket körbevevő, azokat interpretáló természetes nyelvek területén (főként, ha ide soroljuk a különböző természetes nyelvek által használt terminusrétegeket is). A különböző kontextusokban kitermelődő alternativitás, illetve stilisztikai kényszerek miatt igencsak kiterjedt szinonimitással (homonimitással) kell együtt élnünk.

A műveletekre vonatkozóan sokféle megnevezés verseng egymással, különböző kontextusokban más és más terminusokat használnak ugyanarra a műveletre. A műveletek leírásakor felsorolom majd a leggyakrabban használt neveiket. Amikor döntenem kellett, ezek közül melyiket válasszam magam számára, akkor azt a szempontot vettem figyelembe, hogy a megnevezések önmagukban is minél jobban utaljanak a köztük levő kapcsolatokra. Ezért – ahol csak lehetett – az egymásból származtatható műveletek esetében azokat a neveket választottam, amelyek utaltak két művelet közti konverziós kapcsolatra vagy a köztük levő tagadási viszonyra. Előbbit a 'retro', utóbbit a 'kontra' előtag használata jelzi. Nem lehettem azonban teljesen következetes ezen a téren, mert bizonyos relációk esetében a széles körben használt terminusokat nem lehetett nem figyelembe

[8]Mivel az alacsonyabb változószámú műveletek mindig besorolhatóak a magasabb változószámú műveletek közé, ezért a nullaváltozós műveletek egyben egy- és kétváltozós műveletnek is tekinthetők.

venni. Ilyen esetekben “önkényesen” választottam a versengő nevek közül.

A tizenhat bináris logikai műveletet úgy mutatom be egymás után, hogy megadom a cselekvéselméleti modellben használt nevüket, a gyakran használt megnevezéseiket, valamit adok egy rövid szöveges jellemzést is rájuk. Megmutatom azt is, hogy miként lehet őket definálni más műveletek segítségével, de itt nem törekszem arra, hogy ugyanazt a logikai definíciót adjam meg, amivel a kijelentéskalkulusban meghatározhatjuk őket. A cél itt az, hogy értelmezni lehessen az egyes műveleteket. Akkor hát, vegyük sorba őket!

(PL.2) CONTRADICTION $\equiv \bot \stackrel{\text{def}}{=} A \wedge \neg A$

ellentmondás (HUN) – kontradikció (HUN) – sosem igaz (HUN) – falsum (LAT) – contradiction (ENG)

Az ellentmondás művelete egyfajta konzisztenciát biztosít a kijelentéseinknek abban az értelemben, hogy létével jelzi és kizárja az egymást tagadó kijelentések együttes fennállását. Mondhatjuk azt is, hogy ‘Esik az eső.’ és azt is, hogy ‘Nem esik az eső.’, de ha a kettőt egyszerre állítjuk, akkor ellentmondáshoz jutunk. Az ellentmondás művelete mint két – egymást tagadó – állítás összekapcsolása sosem lehet igaz. Ez a művelet nulláris, konstans.

(PL.3) CONEGATION $\equiv \not\vee \equiv \downarrow \equiv \ \nabla \ \equiv || \stackrel{\text{def}}{=} \neg A \wedge \neg B$

konnegáció (HUN) – sem-sem (HUN) – NOR (LOG) – not OR (LOG) – Webb-művelet (HUN) – egyesített tagadás (HUN) – conegation (ENG) – Webb-operation (ENG)

A konnegáció a diszjunkció tagadása. Ha a hétköznapi nyelvben is használatos ‘sem-sem’ formulát nézzük (‘Sem utódja, sem boldog őse.’), ezen az alakon látszik csak igazán, hogy ez a művelet "közel áll" az ellentmondáshoz, hiszen igazságtáblázataik csak egyetlen esetben térnek el egymástól: amikor mindkét operandus hamis, akkor a konnegáció igaz, míg a kontradikció hamis. Ez egyfajta szimmetricitást kölcsönöz a műveletnek. Az ellentmondáshoz való “közelség” azonban kicsit látszólagos – abban az értelemben legalábbis, hogy a művelet két összetevőjének együttes hamissága a logikai univerzum jóval nagyobb részét fedi le, mint az összes többi opció.

(PL.4) RETROSUBTRACTION $\equiv \nleftarrow \stackrel{\text{def}}{=} \neg A \wedge B$

retroszubtrakció (HUN) – konverz szubtrakció (HUN) – konverz nem-implikáció (HUN) – retrosubtraction (ENG) – converse subtraction (ENG)

Ez a művelet a szubtrakció megfordítása, konverze. A művelet jellemzésére ugyanazok mondhatók el, mint a szubtrakció esetében, a különbség csak a két operandus közti viszony irányultságában van.

(PL.5) CONTRAPROJECTION $\equiv \not\triangleleft \stackrel{\text{def}}{=} \neg A$

kontraprojekció (HUN) – contraprojection (ENG)

Ez a művelet a projekció ellentéte, tartamilag megegyezik az A operandusba foglalt kijelentés tagadásával. Azért nevezem kontraprojekciónak, mert a kontra előtaggal valaminek az ellentétét szokás kifejezni, és ebben az esetben pont erről, a projekció tagadásáról van szó.

(PL.6) SUBTRACTION $\equiv \nrightarrow \stackrel{\text{def}}{=} A \wedge \neg B$

szubtrakció (HUN) – kivonás (HUN) – nem-kondicionális (HUN) – abjunkció (HUN) – subtraction (ENG) – abjunction (ENG) – non-conditional (ENG)

A szubtrakció a kondicionális művelet tagadása, amely csak akkor igaz, amikor A "önmagában" igaz, vagyis amikor B nem igaz, de A igaz. Ez utóbbi összefüggés magyarázza a "kivonás" megnevezést, hiszen a művelet igazságtartományát úgy kapjuk meg, ha A igazságtartományából kivonjuk B igazságtartományát.

(PL.7) CONTRA-RETROPROJECTION $\equiv \not\triangleright \stackrel{\text{def}}{=} \neg B$

kontra-retroprojekció (HUN) – contra-retroprojection (ENG)

Ez a művelet a retroprojekció ellentéte, tartamilag megegyezik a B operandus tagadásával.

(PL.8) BISUBTRACTION $\equiv \not\leftrightarrow \equiv \oplus \equiv \veebar \stackrel{\text{def}}{=} \neg(A \leftrightarrow B) \equiv (A \wedge \neg B) \vee (\neg A \wedge B)$

biszubtrakció (HUN) – exkluzív diszjunkció (HUN) – kizáró diszjunkció (HUN) – erős diszjunkció (HUN) – exkluzív OR (LOG) – XOR (LOG) – vagy-vagy (HUN) – kizáró vagy (HUN) – kontravalencia (HUN) – antivalencia (HUN) – kontradiktórius (HUN) – bisubtraction (ENG) – strong disjunction (ENG) – contravalence (ENG) – antivalence (ENG)

A biszubtrakció művelete a bikondicionális tagadása, de tartalmilag kapcsolatba hozható a diszjunkció műveletével is, amit félreérthetetlenül jelez több névváltozata is. Tételezhető még a szubtrakció és annak konverze, a retroszubtrakció diszjunkciójaként is. A művelet legfontosabb tulajdonsága az, hogy csak akkor igaz, ha a két operandus közül valamelyik igaz, de úgy, hogy egyszerre sosem lehetnek igazak. Ezt fejezi ki a kizáró vagy exkluzív diszjukció megnevezés. A mindennapi beszédben is gyakran használjuk ezt a műveletet. Gondoljunk csak a "csak egy maradhat", a "pénzt vagy életet" vagy a "mindent vagy semmit" szófordulatok felbukkanására. A művelet költői leírása pedig: „Vagy bolondok vagyunk s elveszünk egy szálig, vagy ez a mi hitünk valóságra válik."

(PL.9) EXCLUSION $\equiv \not\wedge \equiv \uparrow \equiv | \equiv || \stackrel{\text{def}}{=} \neg A \vee \neg B$

exklúzió (HUN) – kizárás (HUN) – NAND (LOG) – not AND (LOG) – nem-és (HUN) – Sheffer-művelet (HUN) – vagylagos tagadás (HUN) – alternatív tagadás (HUN) – kontrárius (HUN) – exclusion (ENG) – Sheffer-operation (ENG)

Az exklúzió a konjunkció tagadása. Bizonyos értelemben ez a művelet van legközelebb a tautológiához, hiszen csak a két operandus egyidejű igazsága esetén hamis, minden más esetben igaz. A két összetevő, A és B együttes fennállása erős feltételt jelent, és amikor ez nem teljesül, tehát az esetek nagyobb részében, a művelet igaz. Lehet valami barna, lehet zöld, lehet egyik sem, de egyszerre nem lehet barna is, zöld is.

(PL.10) CONJUNCTION $\equiv \wedge \equiv \& \stackrel{\text{def}}{=} A \wedge B$

konjunkció (HUN) – AND (LOG) – és (HUN) – conjunction (ENG)

A konjunkció abban az értelemben erős állítás (legalábbis a diszjunkcióhoz képest), hogy a művelet igazsága egyszerre függ a két operandus igazságától, vagyis ha fennáll a konjunkció igazsága, akkor tudhatjuk azt is, hogy ez annak "köszönhető", hogy mindkét operandus egyszerre igaz. A művelet által lefedett esetek tartománya kisebb, de az így kifejezett tudásunk "tartalmasabb", hiszen kér dolgot is tudunk (A és B igazságát). Ez az operáció is a széleskörben használt műveletek közé tartozik.

(PL.11) BICONDITIONAL $\equiv \leftrightarrow \equiv \Leftrightarrow \equiv \text{‘} \equiv \text{’} \stackrel{\text{def}}{=} A \leftrightarrow B \equiv (A \wedge \neg B) \vee (\neg A \wedge B)$

bikondicionális (HUN) – ekvivalencia (HUN) – materiális ekvivalencia (HUN) – biimplikáció (HUN) – XNOR (LOG) – akkor és csak akkor (HUN) – csakkor (HUN) – a.cs.a (HUN) – iff (LOG) – biconditional (ENG) – equivalence (ENG) – material equivalence (ENG) – biimplication (ENG)

A bikondicionális művelete a kondicionális és annak konverze, a retrokondicionális egyidejű fennállását jelenti, azaz az implikatív jelleg mindkét irányban érvényes. Fontos művelet a tudományos gondolkodás számára, de a hétköznapi kommunikációban is gyakran előfordul. Példa lehet rá a következő idézet: „A garancia akkor és csak akkor érvényes, amennyiben a terméket az adott termékre vonatkozóan jóváhagyott és a használati útmutatóban feltüntetett környezetben használják."

(PL.12) RETROPROJECTION $\equiv \rhd \stackrel{\text{def}}{=} B$

retroprojekció (HUN) – második projekció (HUN) – jobb projekció (HUN) – right projection (ENG)

A retroprojekció csak annyiban tér el a projekciótól, hogy a két operandus közül nem a bal-, hanem a "jobboldalit" emeli ki, vagyis A és B közül B-t tételezi, míg A "állításától" eltekint. Ebben az értelemben a két műveletet egymás konverzeiként értelmezhetjük. A 'retroprojekció' megnevezés analóg a retrokondicionális név képzésével: a 'retro' és a 'kondicionális' szavak összetételében az előtaggal a megfordítást, a konverzitást fejezzük ki, amit – értelemszerűen – az utótagra vonatkoztatunk. A retroprojekció tartalmilag megegyezik a jobboldali operandus (vagyis B) "egyedüli" állításával.

(PL.13) CONDITIONAL $\equiv \rightarrow \equiv \supset \equiv \Rightarrow \stackrel{\text{def}}{=} \neg A \vee B$

kondicionális (HUN) – materiális kondicionális (HUN) – implikáció (HUN) – materiális implikáció (HUN) – ha, akkor (HUN) – szubaltern (HUN) – conditional (ENG) – implication (ENG) – material implication (ENG) – if, then (ENG)

A kondicionális a hétköznapi kommunikáció talán legfontosabb művelete. Az indoklásainkban, magyarázatainkban gyakran használjuk ezt a műveletet, amely a hipotetikus állítások mintázatát adja. A hétköznapi praxis nem igazán reflektál arra a filozófiai problémára, amely abból fakad, hogy A hamissága esetén a művelet mindig igaz eredményt ad.

(PL.14) PROJECTION $\equiv \lhd \stackrel{\text{def}}{=} A$

projekció (HUN) – első projekció (HUN) – balprojekció (HUN) – projection (ENG) – left projection(ENG)

Ritkán szokták a projekció műveletét ebben a kontextusban használni, helyette valamelyik operandus mint állítás tételezését alkalmazzák – miközben ezt a tételezést unáris műveletként értelmezik. Ez természetesen sok szempontból elfogadható, azonban bináris műveletek teljeskörű bemutatásakor jobbnak tartom azt a megoldást, amikor nem emelünk be unáris műveleteket az elemzésünkbe, csak bináris műveletekkel foglalkozunk. A projekció az a művelet, amely két operandus közül kiemeli az egyiket úgy, hogy a művelet igazságértékei minden ponton megegyeznek a kiemelt állítás igazságértékeivel. Ez természetesen azt is jelenti, hogy a másik operandust ilyenkor nem vesszük figyelembe, úgy tételezzük, mintha nem is létezne, nem is állítanánk.[9] Amennyiben a bal oldali állítást (tehát A-t) projektáljuk, akkor "egyszerű" projekcióról beszélünk. Ez tartalmilag megegyezik a bal oldali propozíció tételezésével.

(PL.15) RETROCONDITIONAL $\equiv \leftarrow \equiv \subset \equiv \Leftarrow \stackrel{\text{def}}{=} A \vee \neg B$

retrokondicionális (HUN) – konverz kondicionális (HUN) – konverz implikáció (HUN) – retroconditional (ENG) –converse conditional (ENG) – converse implication (ENG)

Ez a kondicionális művelet konverze. Tartamilag, formailag mindenben hasonlatos a kondicionálisra, csak éppen minden fordítva, "jobbról-balra van". A megszokás (vélhetőleg a balról-jobbra olvasás dominanciája) miatt a kondicionálist érezzük természetesnek. Enek ellenére a hétköznapi beszédben is előfordul. Példa rá a következő két kifejezés: „Kétszer ad, ki gyorsan ad.", illetve „Akkor szép az erdő, mikor zöld."

(PL.16) DISJUNCTION $\equiv \vee \stackrel{\text{def}}{=} A \vee B$

diszjunkció (HUN) – OR (LOG) – vagy (HUN) – alternáció (HUN) – megengedő vagy (HUN) – gyenge diszjunkció (HUN) – inkluzív diszjunkció (HUN) – disjunction (ENG) – alternation (ENG) – inclusive disjunction (ENG) –

A diszjunkció gyenge művelet, amely két állítást úgy kapcsol össze, hogy a művelet igazsága fennállásához "elegendőnek tartja" az egyik operandus igazságát. Abban az értelemben kell ezt gyengének tartanunk, hogy a művelet eredményeként nem tudjuk, hogy a világról szóló igaz tudásunk mennyiben köszönhető a két operandus által kifejezhető igaz tudásnak. Kicsit másként: ha csak a diszjunkció igazságát ismerjük, nem tudhatjuk, hogy ez A vagy B, esetleg mindkettő egyifejű igazságának köszönhető-e. A diszjunkció a hétköznapi kommunikáció során leggyakrabban használt műveletek közé tartozik.

(PL.17) TAUTOLOGY $\equiv \top \stackrel{\text{def}}{=} A \vee \neg A$

tautológia (HUN) – mindig igaz (HUN) – tautology (ENG)

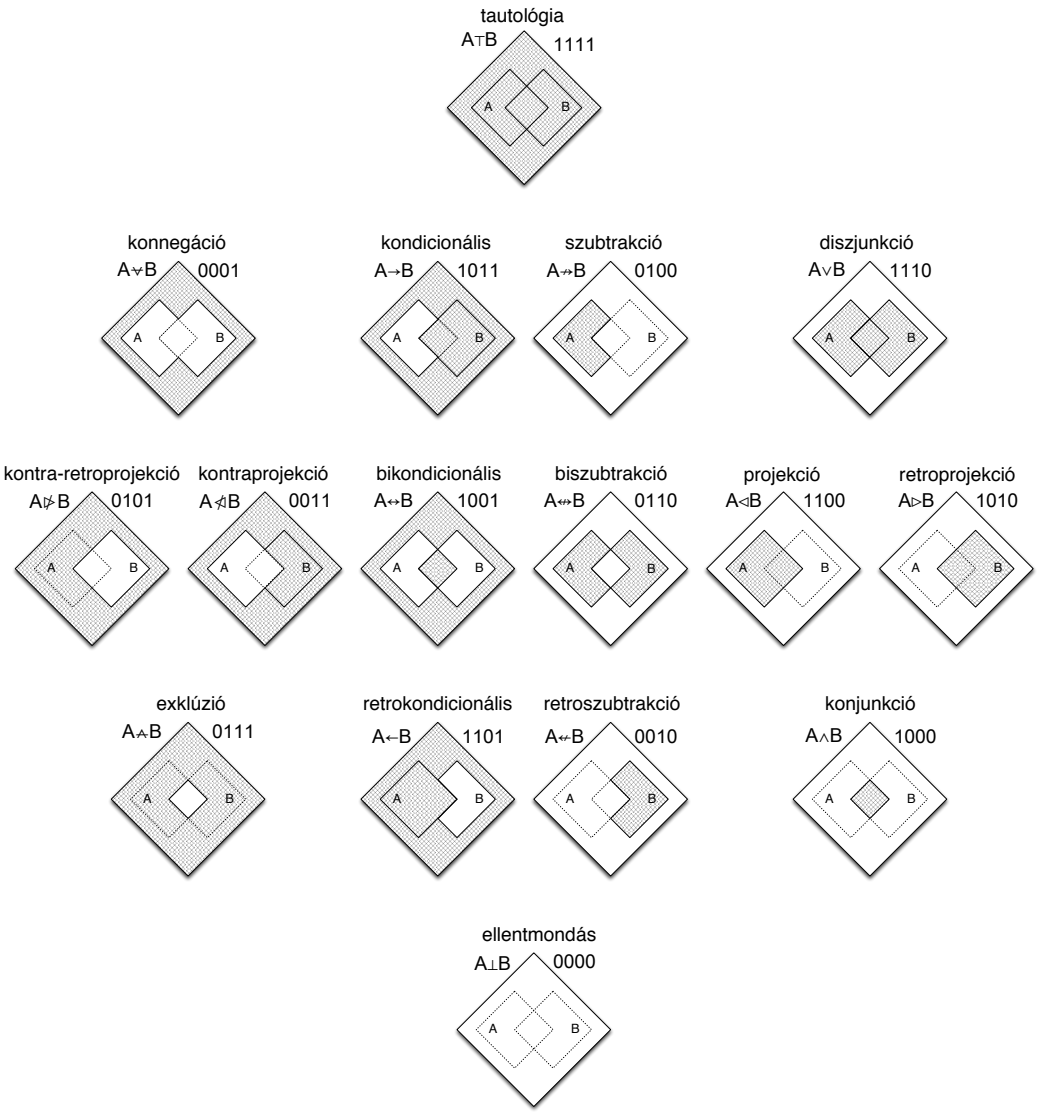
[9] A projekció hasonlít a reifikáció műveletére, hiszen mindkét esetben kiválasztunk egyet két összekapcsolt elemből: logikai műveletek esetén az operandusok, reifikáció esetén a relátumok közül (☞54.o.).

A tautológia az a művelet, amely úgy kapcsol össze két állítást, hogy azok együttese mindig igaz legyen. Ebben a minőségében a kontradikció ellentéte. Ha azt mondjuk, ‘Esik az eső, vagy nem esik az eső.’, akkor ez mindképpen igaz. Persze, kérdés, hogy mi az értelme vagy inkább mi a haszna az ilyen műveletnek. A példamondat alapján nem igazán érezni, pedig igenis nagy jelentősége van a tautológiáknak: a logikai következtetési szabályaink nagy részét ezek adják. A művelet nulláris, konstans.

A tizenhat bináris művelet áttekintése után érdemes őket egy táblázatban egyszerre felsorolni (az ellentéteikkel és igazságtáblázataikkal együtt).

művelet	jele	formula	11	10	01	00	ellentéte
ellentmondás	$\bot$	$A \wedge \neg A$	0	0	0	0	tautológia
konnegáció	$\not\vee, \downarrow, \nabla, \|\|$	$\neg A \wedge \neg B$	0	0	0	1	diszjunkció
retroszubtrakció	$\nleftarrow$	$\neg A \wedge B$	0	0	1	0	retrokondicionális
kontraprojekció	$\not\lhd$	$\neg A$	0	0	1	1	projekció
szubtrakció	$\nrightarrow$	$A \wedge \neg B$	0	1	0	0	kondicionális
kontra-retroprojekció	$\not\rhd$	$\neg B$	0	1	0	1	retroprojekció
biszubtrakció	$\nleftrightarrow, \oplus, \veebar$	$(A \wedge \neg B) \vee (\neg A \wedge B)$	0	1	1	0	bikondicionális
exklúzió	$\not\wedge, \uparrow, \|, \|\|$	$\neg A \vee \neg B$	0	1	1	1	konjunkció
konjunkció	$\wedge$, &	$A \wedge B$	1	0	0	0	exklúzió
bikondicionális	$\leftrightarrow, \Leftrightarrow, \equiv$	$(\neg A \vee B) \wedge (A \vee \neg B)$	1	0	0	1	biszubtrakció
retroprojekció	$\rhd$	B	1	0	1	0	kontra-retroprojekció
kondicionális	$\rightarrow, \supset, \Rightarrow$	$\neg A \vee B$	1	0	1	1	szubtrakció
projekció	$\lhd$	A	1	1	0	0	kontraprojekció
retrokondicionális	$\leftarrow, \subset, \Leftarrow$	$A \vee \neg B$	1	1	0	1	retroszubtrakció
diszjunkció	$\vee$, v	$A \vee B$	1	1	1	0	konnegáció
tautológia	$\top, \vDash$	$A \vee \neg A$	1	1	1	1	ellentmondás

A műveletek igazságtáblázatai alapján észrevehetünk pár érdekességet. Nyolc olyan művelet van, amely a két összetevő igazsága esetén igaz, és nyolc olyan, amely ilyenkor hamis. Nyolc művelet másik nyolcnak a tagadása. Az egymás tagadásaként tekinthető műveletek igazságértékei olyanok, hogy minden ponton egymás ellentétei. Ez szemléletesen megnyilvánul abban a mintázatban, hogy ahol az egyik művelet igaz, ott a másik hamis. Nyolc szimmetrikus és nyolc aszimmetrikus művelet van. A nyolc aszimmetrikus műveletből négy a maradék négy művelet konverzének tekinthető. Nyolc olyan művelet van, amelynél mindkét operandus hamissága esetén a művelet egésze is hamis, míg a másik nyolcművelet esetében a művelet egésze ilyenkor igaz. A logikai műveletek struktúráját, az egyes műveleteken belül az összekapcsolt kijelentések egymáshoz, illetve a tárgyalási univerzumhoz való viszonyát, valamint az egyes logikai műveletek egymáshoz való viszonyát jobban megérthetjük, ha a műveleteket ábrázoló geometriai ábrákat egyszerre, egymás mellé felrajzoljuk.



Kijelentéskalkulus

A kijelentéslogika a propozíciók és a köztük levő logikai műveletek, logikai kapcsolók struktúrájával foglalkozik. E struktúra alapját a **kijelentéskalkulus** (Propositional Calculus, PC) adja. A propozícionális kalkulus axiómakészletének definiálásához nyilván kevesebb elemre van szükség, ezért a PC nyelve szűkebb a kijelentéslogika nyelvénél. A PC logikája a következő nyelvi alapelemekre építhető fel.

(PC.0) $\mathscr{L}_{PC}(A, B, C, \neg, \rightarrow)$, ahol

A, B, C	formula (propozíció)
$\neg$	tagadás logikai művelet
$\rightarrow$	kondicionális logikai művelet

A fenti nyelv segítségével már definiálhatunk egy axiómkészletet, amiből a kijelentéslogika minden további fogalma, tétele levezethető. A PC a következő tételekből áll:

(PC.1) $A \rightarrow (B \rightarrow A)$

(PC.2) $(A \to B) \to ((A \to (B \to C)) \to (A \to C))$

(PC.3) $(\neg A \to \neg B) \to (B \to A)$

Ezek az axiómák a logikai műveletek (kapcsolók) bevezetésére és eliminálására szolgálnak. A logika teljessé tételéhez szükség van következtetési szabályokra is. Ezek olyan formulák, melyek esetében igaz formulákat a változókba írva igaz kijelentésekhez jutunk (ez kevesebb, mint az axiómasémák igazsága, mert utóbbiak esetében bármit írhatunk a változókba, mindig igaz marad a séma formulája. Ezeket az összefüggéseket mindig a logikai következmény relációk segítségével tudjuk kifejezni, amelyeket az alábbi szimbólumokkal jelölünk.

(PC.4) LOGICAL_CONSEQUENCE_RELATION $\equiv \Rightarrow \underline{\vee} \Leftarrow \underline{\vee} \Leftrightarrow$

logikai következmény relációk (HUN) – logical consequence relations (ENG)

A három szimbólumot értelemszerűen az egyik, a másik vagy a mindkét irányból vett következmény fennállásának jelzésére használjuk. A logikai következmény reláció nem az adott nyelvhez tartozik, a nyelvhez képest metanyelvi relációnak kell tartanunk. A (PC) következtetési szabálya a modus ponens (MP), ami ami szűkebb értelemben véve nem tartozik a kalkulushoz, de ettől még nélkülözhetetlen.

(PC.5) $(\mathrm{MP}) \stackrel{\text{def}}{=} \{A, (A \to B)\} \Rightarrow B$

modus ponens következtetési szabály (HUN) – modus ponens inference rule (ENG)

A modus ponens úgy értelmezhetjük, hogy ha egy A kijelentés és a belőle vont bármely B utótagú ($A \to B$) implikáció része igaz az adott nyelven, akkor a kondicionális B utótagja is a igaz. Vannak más következtetési szabályok is, de ezek mind levezethetők a sémaaxiómák és a modus ponens szabály formulái alapján. Ezekre támaszkodva az összes ismert összefüggést, következtetési sémát le lehet vezetni.[10] Csak a szemléltetés kedvéért álljon itt pár további következtetési szabály.

(PC.6) $A \vee B \stackrel{\text{def}}{=} \neg(\neg A \wedge \neg B)$

diszjunkció definiálása (HUN) – definition of conjuction (ENG)

(PC.7) $A \wedge B \Rightarrow A$
$A \wedge B \Rightarrow B$

konjunkció eliminálása (HUN) – elimination of conjuction (ENG)

(PC.8) $A \Rightarrow A \vee B$
$B \Rightarrow A \vee B$

diszjunkció bevezetése (HUN) – introduction of disjunction (ENG)

(PC.9) $\{\neg B, (A \to b)\} \Rightarrow \neg A$

modus tollens következtetési szabály (HUN) – modus tollens inference rule (ENG)

(PC.10) $\neg\neg A \Rightarrow A$

kettős tagadás törvénye (HUN) – double negation law (ENG)

[10] A propozícionális kalkulus kapcsán S.C. Kleene vagy negyven formulát bemutat [281], más felfogás mentén Ruzsa Imre 27 szabályt sorol fel [454, 29-45].

(PC.11) $(A \to B) \Rightarrow (\neg B \to \neg A)$

kondicionális kontrapozíció törvénye (HUN) – law of conditional contraposition (ENG)

(PC.12) $(A \leftrightarrow B) \Rightarrow (A \to B)$
$(A \leftrightarrow B) \Rightarrow (B \to A)$

bikondicionális eliminálása (HUN) – elimination of biconditional (ENG)

(PC.13) $A \leftarrow B \stackrel{\text{def}}{=} B \to A$

retrokondicionális definiálása (HUN) – kondicionális konverzió (HUN) – implikációs konverzió (HUN) – definition of retroconditional (HUN) – conditinal conversion (ENG)

Megállapíthatunk még egy további fontos szabályszerűséget: a nulladrendű kijelentéslogika zárt a logikai operátorokra, ami azt jelenti, hogy minden művelet eredménye propozíció, vagyis a műveletekkel képzett új állítások a logika érvényességi körén belül maradnak. Logikai szempontból még tovább lehetne egyszerűsíteni a kijelentéslogika nyelvét, ha ki akarnánk használni azt a tényt, hogy mind az exklúzió (|), mind a konnegáció (||) azzal az egyedi tulajdonsággal rendelkezik, hogy segítségével az összes többi logikai művelet kifejezhető. A tagadás – például – az exklúzió segítségével a következőképpen írható fel:

(PC.14) $\neg A \Leftrightarrow A \mid A$

Ez azt jelenti, hogy a propozicionális logika minden művelete átalakítható lenne, és minden formulája felírható lenne e két logikai operátor valamelyikével. A következő táblában bemutatom, hogy az exklúzióra támaszkodva hogyan lehet reprezentálni a tizenhat logikai kapcsoló műveletét.

művelet	formula (|)
ellentmondás	$(A \mid (A \mid A)) \mid (A \mid (A \mid A))$
konnegáció	$((A \mid A) \mid (B \mid B)) \mid ((A \mid A) \mid (B \mid B))$
retroszubtrakció	$((A \mid B) \mid B) \mid ((A \mid B) \mid B)$
kontraprojekció	$A \mid A$
szubtrakció	$(A \mid (A \mid B)) \mid (A \mid (A \mid B))$
kontra-retroprojekció	$B \mid B$
biszubtrakció	$((A \mid (A \mid B)) \mid ((B \mid (A \mid B)))$
exklúzió	$A \mid B$
konjunkció	$(A \mid B) \mid (A \mid B)$
bikondicionális	$((A \mid (A \mid B)) \mid ((A \mid B) \mid B)) \mid ((A \mid (A \mid B)) \mid ((A \mid B) \mid B))$
retroprojekció	B
kondicionális	$A \mid (A \mid B)$
projekció	A
retrokondicionális	$(A \mid B) \mid B$
diszjunkció	$(A \mid A) \mid (B \mid B)$
tautológia	$(A \mid A) \mid ((A \mid A) \mid (A \mid A))$

Ez a megoldás egyszerűsítené az adott logikai nyelvet. Ha nem lenne szükség arra, hogy emberek is értelmezzék a nyelv formuláit, akkor gazdaságos lenne ez az átírást megtenni. Amikor számítógépeket lehet használni kijelentéslogikai feladatok megoldására, akkor

ezt meg is teszik, a logikai áramköröket gyakran NAND- vagy NOR-kapuk segítségével építik fel, ami azt jelenti, hogy minden logikai műveletet e két operátort reprezentáló fizikai eszközzel oldanak meg. Az egy-műveletes nyelv gazdaságossága azonban hátránnyá válik akkor, amikor embereknek is értelmezniük kell a nyelv kifejezéseit. Az ember inkább használ több nyelvi terminust, inkább elviseli a redundáns kifejezéseket, mintsem hosszú logikai formulákból bogozza ki állandóan a formula(rész) tényleges jelentését. A fenti táblázatban a leghosszabb formulája a bikondicionálisnak van, ami helyett nekünk könnyebb egy új logikai jelet ($\leftrightarrow$) megtanulni és használni erre a műveletre.

1.4 Predikátumlogika

A nulladrendű logika a kijelentések, a mondatok szintjén keresi a szabályszerűségeket. A mondat, az állítás ilyenkor bonthatatlan egészet alkot, nem tudjuk, nem tudhatjuk, mi van a propozíciók belsejében. Csak az számít, hogy igaz vagy hamis-e az állítás. Ezen a szinten a logika csak 'tudomásul veszi', de érdemben nem tud foglalkozni a logikai igazság/hamisság kérdésével. Az állításlogikában ezért a logikai elemzés fókusza a mondatok közti műveletek vizsgálatára irányul.

Predikátumlogikai alapfogalmak

A logikai apparátus leíróerejét úgy lehet növelni, ha a logikai elemzés hatókörét kiterjesztjük úgy, hogy a kijelentések "belsejét" is vizsgálni tudjuk. Ehhez az kell, hogy "felbontsuk" a mondatot, és megengedjük a kijelentések összetevőkre bontását és azok vizsgálatát. Amikor ezt tesszük, akkor lépünk át az elsőrendű vagy predikátumlogika világába. Ennek leírásához használhatjuk a következő predikátumlogikai nyelvet.

(QL.0) $\mathscr{L}_{QL}(\mathscr{L}_{PL}, m, n, o, t, s, r, \forall, \exists, =)$, ahol

m, n, o	változó
t, s, r	konstans
$\forall$	univerzális kvantor
$\exists$	egzisztenciális kvantor
$=$	egyenlőség

A predikátumlogikában egy mondat egy névre vagy több névre és egy predikátumra bontható fel. A név individuumkonstans: olyan logikai konstans, amely valamely individuumot jelöl. A nevek a predikátumok bemeneteinek, paramétereinek számítanak. A predikátum az a mondatkomponens, amely egy vagy több névre mint bemeneti paraméterekre vonatkozóan fejezi ki az állítás tartalmát (valamilyen cselekvést, tulajdonságot, állapotot, történést, kapcsolatot stb.). Amikor a predikátumlogikát formális nyelvként írjuk le, akkor részben pontosítjuk, részben igazodunk a formális nyelvelmélet terminológiájához, ezért a neveket logikai konstansokkal (individuumok neveivel) fejezzük ki, a predikátumok helyett pedig vagy a relációk vagy a függvények fogalmait használjuk.

A predikátumlogika nyelvébe átvesszük az állításlogika nyelvét (kijelentéseket, formulákat, logikai műveleteket), és ennek kiegészítéseként be kell még vezetnünk további kategóriákat [454]. Az első kettő új komponens az értékét nem változtató név, amely individuumokat azonosító nyelvi megnyilatkozás, valamint a predikátum, ami egy vagy több név mint bemeneti paraméter alapján állítást képező logikai függvény.

(QL.1) CONSTANT $\stackrel{\text{def}}{=} t, s, r$

konstans (HUN) – név (HUN) – constant (ENG) – name (ENG)

(QL.2) PREDICATE $\stackrel{\text{def}}{=} A, B, C, P, R$

predikátum (HUN) – logikai függvény (HUN) – funktor (HUN) – reláció (HUN) – függvény (HUN) – predicate (ENG) – logical function (ENG) – relation (ENG) – function (ENG) – functor (ENG)

Az elsőrendű logika segítségével mondhatunk olyan mondatokat, amelyekben valamely reláció konkrét előfordulására vonatkozóan teszünk egy állítást úgy, mint a következő példában (tekintsünk most el attól, hogy a 'szürke' minősítést pontosan hogyan is kell érteni, kezelni logikailag):

'Ráró szürke.' = 'A Ráró nevű ló szürke.'

Az ilyen mondatok azonban kevéssé érdekesek a logika számára. Ha csak konstansok és relációk lennének ebben a formális nyelvben, akkor ezt a logikát kevéssé lehetne használni, mert így gyenge általánosítóerővel rendelkezne. Azzal léphetünk nagyot az általánosíthatóság, az absztrakció irányában, ha az individuumokat jelző konstansok (konkrét nevek) helyett **változókat** alkalmazunk. Általánosíthatjuk a mondatainkat, ha a relációkba változókat írunk:

'Az x ló szürke.' = 'LÓ(x) szürke'

Ekkor viszont nem tudjuk egyértelműen, mire (itt: melyik lóra) vonatkozik az állítás. Az ilyen kijelentés nem tartalmas, nem egyértelmű. Azt mondjuk rá, hogy nyitott mondat. Az az izgalmas a logika számára, ha olyan általános kijelentéseket fogalmazhat meg, amelyek tartalmasak, zártak. Ilyen mondatok képzésére valók a **kvantorok**, amelyekkel "le lehet zárni" a nyitott logikai mondatokat.

Két kvantort használunk: az univerzális kvantort (jelentése: 'minden x-re igaz, hogy', jele: $\forall$) és az egzisztenciális kvantort (jelentése: 'van olyan x, amire igaz, hogy', jele: $\exists$). A kvantorokkal "leköthetjük" a szabad változókat, ami azt is jelenti egyben, hogy általános (és tartalmas, zárt) állításokat nyerhetünk az alkalmazásukkal. Ha azt mondjuk, hogy:

'minden ló szürke' – $\forall m$ LÓ(x) szürke

'létezik ló, ami szürke' – $\exists x$ LÓ(x) szürke

akkor olyan állításokhoz jutunk, amelyek már logikailag értékelhetőek (és az első kijelentést hamisnak minősíthetjük, hiszen vannak nem szürke lovak is, a másodikat állítást viszont igaznak tarthatjuk, hiszen láttunk szürke lovat). Az elsőrendű logika nyelvét ki kell egészítenünk a fent bemutatott új kategóriákkal. Szükség van *(i)* a változó fogalmára, ami a logikai kifejezésekben valamely halmaz elemeit mint lehetséges értékhordozókat helyettesítő nyelvi jel, illetve szükség van két – a kategóriák terjedelmét jelentő előfordulásokra vonatkozó – új operátorrra is. Az egyik *(ii)* az egzisztenciális kvantor, amely legalább egy előfordulás létezését tételezi a logikai kifejezésben, a másik *(iii)* az univerzális kvantor, amely az összes előfordulás létezését fejezi ki a logikai kifejezésben.

(QL.3) VARIABLE $\stackrel{\text{def}}{=} m, n, o, p$

változó (HUN) – variable (ENG)

(QL.4) EXISTENTIAL_QUANTIFIER $\stackrel{\text{def}}{=} \exists$

egisztenciális kvantor (HUN) – existential quantifier (ENG)

(QL.5) UNIVERSAL_QUANTIFIER $\stackrel{\text{def}}{=} \forall$

univerzális kvantor (HUN) – universal quantifier (ENG)

A két kvantor közül elég csak az egyiket felvenni a nyelvbe alapkategóriaként, a másik már definálható a segítségével. A predikátumkalkulusban így is teszünk majd. A következő fejezetben mutatom meg, hogy hogyan.

Predikátumkalkulus

Természetesen a predikátumlogikának is van kalkulusa, a **predikátumkalkulus** (Predicate Calculus – QC). A QC-kalkulust sem elemzem itt alaposabban, elégnek tartom bemutatni az axiómsémáit. A predikátumkalkulus a predikátumlogika nyelvének azon legszűkebb része, amelyre támaszkodva minden predikátumlogikai formulát ki lehet fejezni valahogyan. A predikátumkalkulusnak egyszerűbb nyelv is elégséges, amit a következőféleképpen lehet definiálni.

(QC.0) $\mathscr{L}_{QC}(A,B,C,m,n,o,t,\forall,\neg,\rightarrow,=)$, ahol

A,B,C	formula
m,n,o	változó
t,s,r	konstans
$\forall$	univerzális kvantor
$\neg$	tagadás logikai művelet
$\rightarrow$	kondicionális logikai művelet
$=$	egyenlőség

Ezen a nyelven már definiálhatjuk a predikátumkalkulus axiómaséma-készletét, amely az alábbi nyolc tételből áll [454, 130]. A negyedik és a nyolcadik sémában a kitevőben szereplő '/' törtjeles kifejezést úgy kell értelmezni, hogy bárhogy is szerepel az A formulában a törtjel utáni változó (o), a séma érvényes marad minden olyan esetben, ha az o helyébe a t terminust (nevet), illetve az m vagy n változót helyettesítjük [454, 81].

(QC.1) $A \rightarrow (B \rightarrow A)$

(QC.2) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

(QC.3) $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$

(QC.4) $\forall o(A \rightarrow A^{t/o})$

(QC.5) $\forall o(A \rightarrow B) \rightarrow (\forall o(A) \rightarrow \forall o(B)))$

(QC.6) $A \rightarrow \forall o(A)$

(QC.7) $o = o$

(QC.8) $(m = n) \rightarrow (A^{m/o} \rightarrow A^{n/o})$

A (QC) első három tétele megegyezik a (PC) első három axiómasémájával. Ebből következik, hogy az öszes logikai műveletet be lehet vezetni a predikátumkalkulusban. A második három séma a kvantifikációra, az utolsó kettő pedig az azonosságra vonatkozik. A

(QC) alapján természetesen további logikai szimbólumokat is be lehet vezetni a propozícionális logika nyelvébe, akár az egzisztenciális kvantort, akár az implikáción túli logikai műveleteket. Az egzisztenciális kvantorra az alábbi definíció adható:

(QC.9) $\text{EXISTENTIAL_QUANTIFIER}(m) \equiv \exists m \stackrel{\text{def}}{=} \neg\forall\neg m$

egzisztenciális kvantor(HUN) – existential quantifier (ENG)

A (QC) levezetési szabálya a (MP) modus ponens, de természetesen minden más állításlogikai következtetési szabályt alkalmazhatunk. A későbbiekben szükségünk lehet arra, hogy amikor az egzisztenciális kvantort alkalmazzuk, akkor mondhassuk azt, hogy abból az individuumból, aminek a létezését állítjuk az egzisztenciális kvantor segítségével, abból csak egyetlen egy van. Ezt a következő formulával fejezhetjük ki.

(QC.10) $\text{UNIQUENESS_QUANTIFIER}(m) \equiv \exists!\, m \equiv \exists!\, m(\mathbf{P}(m)) \stackrel{\text{def}}{=} \exists m \forall n(\mathbf{P}(n) \leftrightarrow m = n)$

unicitás kvantor (HUN) – uniqueness quantifier (ENG) – unique existential quantifier (ENG)

A módosított kvantort (operátort) úgy kell értelmezni, hogy 'létezik egy és csak egy', és mindig mellé kell képzelni egy predikátumot, amire vonatkoztatjuk a kvantort (aminek az egyeleműségét állítjuk). Mivel az operátor segítségével az egyelemű predikátumokat tudjuk kifejezni, és a fenti formulát szokás unicitássémának is nevezni[454, 98], ezért használhatjuk az unicitás kvantor (operátor) megnevezést.

1.5 Relációelmélet

A formális nyelvek által definiált fogalmak valamilyen struktúrát alkotnak, a struktúrák vázát pedig mindig az adott nyelvbe felvett relációk adják. Bármilyen relációs struktúrát építünk is fel, minden esetben szükség van arra, hogy a relációkat magukat is jellemezni tudjuk. A relációelmélet ezt a feladatot végzi el. Három fogalom van az elmélet központjában: a **reláció**, a **függvény** és a **művelet**. Ezek mindegyikét különféle tulajdonságokkal jellemezhetjük, műveletek értelmezhetünk közöttük, és mindegyik levezethető a másikból, tehát bármelyiket megtehetnénk egy formális elmélet alapfogalmául. Én azt a megközelítést fogadom el, amely szerint a reláció fogalmából kell kiindulni, és ebből lehet levezetni előbb a függvény, majd a művelet fogalmát.

Relációelméleti alapfogalmak

A relációelméleti fogalmak rövid bemutatását a relációhoz kapcsolódó fogalomcsokor tárgyalásával kezdem. Első lépésként sorba veszem a relációt leíró tulajdonság- és műveletfogalmakat. Ezután módszertanilag annak kéne következnie, hogy bemutassam a reláció fogalmára építhető kalkulust, de ezt addig nem lehet megtenni, amíg be nem vezettük a művelet fogalmát. Ezért a relációs tulajdonságfogalmak bemutatása után rátérek a függvény és a művelet fogalmainak tárgyalására, és csak miután ezekkel végeztem, azután tudom röviden felvázolni a relációelmélet kalkulusát, a relációalgebrát.

A **reláció** fogalmával kettő vagy több dolog közti kapcsolatot ragadhatjuk meg. Ez azt jelent, hogy mindig meg kell tudnunk adni, hogy milyen dolgok közötti viszonyt írunk le az adott reláció segítségével. Ehhez meg kell adnunk egy (vagy több) halmazt,

amely(ek)nek elemeit összekapcsolhatjuk egymással a reláció keretében. Az összekapcsolt elemeket **relátumoknak** nevezzük. Az összekapcsolt elemeket adjuk meg a reláció argumentumában, és attól függően, hogy hány elemet kapcsolunk össze beszélhetünk a reláció argumentumszámáról. A relációk jellemzéséhez kiindulhatunk az alábbi nyelvből.

(RL.0) $\mathscr{L}_{RL}(\mathscr{L}_{QL}, \mathbf{R}, \mathbf{F}, \mathbf{O}, \langle m,n,o\rangle, \mathbb{X}\times\mathbb{X}, \mathbb{D}\mathrm{om}, \mathbb{I}\mathrm{m})$, ahol

$\langle m,n,o\rangle$	rendezett n-s
R,P,Q	reláció
F	függvény
O	művelet
$\mathbb{X}\times\mathbb{X}$	Descartes szorzat
$\mathbb{D}\mathrm{om}$	értelmezési tartomány
$\mathbb{I}\mathrm{m}$	képtartomány

A relációk minél jobb megragadhatósága érdekében rengeteg tulajdonságot definiáltak, és ezek alapján nevezetes relációkat is elkülönítettek egymástól. Első körben a legfontosabb relációtulajdonságokat, majd a leggyakoribb nevezetes relációtípusokat fogom bemutatni. A relációk egyik első tipizálási lehetőségét az jelent, ha figyelembe vesszük azt a tényt, hogy hány argumentuma van a relációnak. Beszélhetünk unáris, bináris vagy n-áris relációról, ha egy, kettő vagy n elemet kapcsolunk össze a reláció segítségével. A reláció fogalmát halmazelméleti alapon a következőféleképpen határozhatjuk meg.

(RL.1) $\textsc{relation}(m,n) \equiv \langle m,n\rangle \equiv (m,n) \in \mathbf{R} \subset \mathbb{X}\times\mathbb{X} \equiv \mathbf{R}(m,n) \stackrel{\text{def}}{=} \{\{m\},\{m,n\}\}$

reláció

A relációkhoz (függvényekhez) kapcsolódik két fontos osztályfogalom, az értelmezési tartomány és értékkészlet. Ezeket a következő módon emelhetjük be a modellünkbe.

(RL.2) $\textsc{domain} \equiv \mathbb{D}\mathrm{om}(\mathbf{R}(m,n)) \equiv \mathbb{D}\mathrm{om}(\mathbf{R}) \equiv \mathbb{D}\mathrm{om} \equiv \mathbb{D} \stackrel{\text{def}}{=} \{m : \exists n(\mathbf{R}(m,n))\}$

értelmezési tartomány (HUN) – domain (ENG)

(RL.3) $\textsc{image} \equiv \mathbb{I}\mathrm{m}(\mathbf{R}(m,n)) \equiv \mathbb{I}\mathrm{m}(\mathbf{R}) \equiv \mathbb{I}\mathrm{m} \equiv \mathbb{I} \stackrel{\text{def}}{=} \{n : \exists m(\mathbf{R}(m,n))\}$

képtartomány (HUN) – kép (HUN) – értékkészlet (HUN) – image (ENG) – range (ENG) – co-domain (ENG)

A relációk egyik legáltalánosabb tipizálási lehetőségét az adja, ha arra figyeljük, hogy a reláció által összekapcsolt dolgok homogének-e vagy sem, azaz ugyanabból a halmazból vesszük-e őket vagy sem. Ha igen, akkor **homogén** (homogeneous), ha nem, akkor **heterogén** (heterogeneous) relációról beszélünk. Még mindig a legáltalánosabb szinten maradva definiálhatunk négy nevezetes relációt, amelyekre később még szükségünk lesz.

(RL.4) $\textsc{universal_relation}(m,n) \equiv \underset{\perp}{\mathbf{U}}(m,n) \equiv \underset{\perp}{\mathbf{U}} \equiv \underset{\perp}{\mathbf{F}}(m,n) \equiv \underset{\perp}{\mathbf{F}} \stackrel{\text{def}}{=} \forall m \forall n(R(m,n))$

teljes reláció (HUN) – univerzális reláció (HUN) – universal relation (ENG) – full relation (ENG)

(RL.5) $\textsc{identity_relation}(m,n) \equiv \underset{\perp}{\mathbf{I}}(m,n) \equiv \underset{\perp}{\mathbf{I}} \stackrel{\text{def}}{=} \forall m \forall n(R(m,n) \leftrightarrow m = n)$

azonosság (reláció) (HUN) – egység reláció (HUN) – diagonál(is) reláció (HUN) – egyenlőség(i reláció) (HUN) – identitás (reláció) (HUN) – identity (relation) (ENG) – diagonal (relation) (ENG)

(RL.6) $\text{EMPTY_RELATION}(m,n) \equiv \underset{\perp}{\mathbf{E}}(m,n) \equiv \underset{\perp}{\mathbf{E}} \stackrel{\text{def}}{=} \forall m \forall n(\neg R(m,n))$

üres reláció (HUN) – null reláció (HUN) – empty relation (ENG) – null relation (ENG)

(RL.7) $\text{DIVERSITY_RELATION}(m,n) \equiv \underset{\perp}{\mathbf{D}}(m,n) \equiv \underset{\perp}{\mathbf{D}} \stackrel{\text{def}}{=} \forall m \forall n(R(m,n) \leftrightarrow m \neq n)$

különbözőség (reláció) (HUN) – egyenlőtlenség(i reláció) (HUN) – antidiagonális (reláció) (HUN) – differencia (reláció) (HUN) – inkompatibilitás(i reláció) (HUN) – diverzitás(i reláció) (HUN) – difference (ENG) – incompatibility (relation) (ENG) – antidiagonal (relation) (ENG) – diversity (relation) (ENG)

Mivel a fenti négy relációtípus fontos szerepet tölt be a relációelméletben, ezért saját nevet kapott, nem pedig a rá jellemző tulajdonság alapján lehet hivatkozni rájuk. A következő blokkban azokat a relációtulajdonságokat mutatom be, amelyek önmagukban "csak" tulajdonságok, de a relációfogalomhoz kapcsolva őket egy sor nevezetes, gyakran használt relációt azonosíthatunk velük. A különböző korok, diszciplinák, megközelítések, emberek ugyanarra a problémára sokszor nagyon hasonló, de bizonyos pontokon így vagy úgy eltérő tulajdonságokat tartottak fontosnak bevezetni, ezért az egymáshoz közeli fogalmakat csoportokba rendezve mutatom be.

A reflexivitással kapcsolatos tulajdonságok azt a minőséget próbálják megragadni, hogy az elemek között tételezett reláció vonatkoztatható-e az elemekre önmagukra is, vagyis összeköthetők-e az elemek önmagukkal vagy sem. Ezt fejezi ki a reflexivitás tulajdonsága. Hétköznapi példa lehetne a reflexivitásra a SZERETI reláció, hiszen valaki szerethet másokat, de szeretheti magát is, ám ez példa kevés lehet a reflexivitás fennállásához, hiszen nem minden esetben érvényesül. Vannak ugyanis olyan emberek, akik nem szeretik önmagukat. Márpedig a reflexivitás minőségét csak akkor érdemes egy relációhoz kötni, ha az minden elemre vonatkoztatható. Ebben az értelemben elfogadható példa lehet az EGYIDŐS reláció, hiszen egy ember egyidős lehet valaki mással, de mindig egyidős önmagával. Olyan eset is előfordulhat, amikor a reflexivitás sosem érvényesül. Például a SZÜLŐJE vagy a TESTVÉRE relációra hivatkozhatunk, amikoris mondhatjuk, hogy senki sem lehet önmaga szülője vagy testvére.

(RL.8) $\text{REFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m(\mathbf{R}(m,m))$

reflexív (HUN) – reflexive (ENG) – erősen reflexív (HUN) – strongly reflexive (ENG)

(RL.9) $\text{IRREFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m(\neg\mathbf{R}(m,m))$

irreflexív (HUN) – antireflexív (HUN) – irreflexive (ENG) – antireflexive (ENG)

(RL.10) $\text{NON-REFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \exists m(\neg\mathbf{R}(m,m))$

nem-reflexív (HUN) – non-reflexive (ENG)

(RL.11) $\text{SECONDARILY_REFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \rightarrow \mathbf{R}(n,n))$

másodlagosan reflexív (HUN) – szekunder reflexív (HUN) – secondarily reflexive (ENG) – shift reflexive (ENG)

(RL.12) $\text{INVERSE_SECONDARILY_REFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \rightarrow \mathbf{R}(m,m))$

fordítottan másodlagosan reflexív (HUN) – feltételesen reflexív (HUN) – conditional reflexive (ENG) – inverse secondarily reflexive (ENG)

(RL.13) $\text{COREFLEXIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \rightarrow m = n)$

koreflexív (HUN) – üres (HUN) – vacuous (ENG) – coreflexive (ENG)

A relációtulajdonságok másik csoportjába tartoznak a szimmetricitással kapcsolatos tulajdonságok. Ekkor arra kérdezünk rá, hogy ha a reláció két elem között fennáll, akkor vajon fennáll-e a fordított irányban is, vagy másként: ha a reláció igaz, akkor igaz-e a konverze. Szimmetrikus a TESTVÉRE reláció, hiszen ha A testvére B-nek, akkor ez fordítva is igaz, vagyis B testvére A-nak. Léteznek azonban olyan relációk is, amelyekre minden esetben kizárhatjuk a szimmetricitást. Ilyen például az IDŐSEBB reláció: ha A idősebb B-nél, akkor B nem lehet sosem idősebb A-nál.

(RL.14) $\text{SYMMETRIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \rightarrow \mathbf{R}(n,m))$

szimmetrikus (HUN) – symmetric (ENG)

(RL.15) $\text{ASYMMETRIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \rightarrow \neg\mathbf{R}(n,m))$

aszimmetrikus (HUN) – antiszimmetrikus (HUN) – szigorúan antiszimmetrikus (HUN) – asymmetric (ENG) – antisymmetric (ENG)

(RL.16) $\text{ANTISYMMETRIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n((\mathbf{R}(m,n) \wedge \mathbf{R}(n,m)) \rightarrow m = n)$

antiszimmetrikus (HUN) – identitív (HUN) – antisymmetric (ENG) – identitive (ENG)

(RL.17) $\text{NON-SYMMETRIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m))$

nem-szimmetrikus

Fontos csoportot képeznek azok a tulajdonságok, amelyek a teljességgel, összehasonlíthatósággal, összekapcsoltsággal, egyértelműséggel kapcsolatosak. Mindig jó tudnunk egy relációval kapcsolatban, hogy a reláció tartóhalmazának minden elemére kiterjed-e a reláció érvénye vagy sem. Különösen fontos kérdés ez az összehasonlíthatóság miatt. A ISMERI relációt például nem tarthatjuk teljesnek, mert egy csoporton belül nem mindenki ismerheti egymást, ezért a csoport tagjai között nem állíthatunk fel egy teljes ismertségi hálót. Az egyértelműség kérdése pedig azért lehet fontos, mert sokszor akarhatjuk azt, hogy egy reláció fennállásához egyértelmű következményeket rendelhessünk. Ennek "kiderítésére" is érdemes tulajdonságokat definiálni.

(RL.18) $\text{WEAKLY_CONNECTED}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(\mathbf{R}(m,n) \wedge \mathbf{R}(m,o) \rightarrow (n = o \vee \mathbf{R}(n,o) \vee \mathbf{R}(o,n)))$

gyengén összefüggő (HUN) – weakly connected (ENG) – weakly connex (ENG) – összehasonlítható (HUN) – comparable (ENG)

(RL.19) $\text{CONNECTED}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(m \neq n \rightarrow (\mathbf{R}(m,n) \vee \mathbf{R}(n,m)))$

összefüggő (HUN) – konnex (HUN) – connex (ENG) – connected (ENG) – lineáris (HUN) – teljes (HUN) – complete (ENG) – totális (HUN) – total (ENG) – összehasonlítható (HUN) – comparable (ENG)

(RL.20) $\text{STRONGLY_CONNECTED}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \vee \mathbf{R}(n,m))$

erősen összefüggő (HUN) – strongly connex (ENG) – strongly connected (ENG) – totally connected (ENG) – totálisan összefüggő (HUN) – teljes (HUN) – complete (ENG) – totális (HUN) – total (ENG) – összehasonlítható (HUN) – comparable (ENG)

(RL.21) $\text{DICHOTOME}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(m \neq n \rightarrow (\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m)) \vee (\neg\mathbf{R}(m,n) \wedge \mathbf{R}(n,m)))$

dichotóm (HUN) – összehasonlítható (HUN) – comparable (ENG) – totális (HUN) – teljes (HUN) – total (ENG) – dichotomous (ENG) – complete (ENG)

(RL.22) $\text{TRICHOTOME}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n((\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m) \wedge m \neq n) \vee (\neg\mathbf{R}(m,n) \wedge \mathbf{R}(n,m) \wedge m \neq n) \vee m = n)$

trichotóm (HUN) – lineáris (HUN) – trichotomous (ENG)

(RL.23) $\text{COMPARABLE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(\mathbf{R}(m,n) \rightarrow (\mathbf{R}(m,o) \vee \mathbf{R}(o,n)))$

összehasonlítható (HUN) – comparable (ENG)

(BL.24) $\text{INDIFFERENT}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}(m,n) \wedge \mathbf{R}(n,m))$

indifferens (HUN) – erősen indifferens (HUN) – erősen közömbös (HUN) – strongly indifferent (ENG) – indifferent (ENG) – egyenlő értékű (HUN) – equal in value (ENG) – ugyanolyan, mint (HUN) – same as (ENG)

(BL.25) $\text{INCOMPARABLE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\neg\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m))$

összehasonlíthatatlan (HUN) – gyengén közömbös (HUN) – gyengén indifferens (HUN) – incomparable (ENG) – weakly indifferent (ENG) – ugyanolyan értékes, mint (HUN) – same value as (ENG)

(RL.26) $\text{LEFT-TOTAL}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \exists n(\mathbf{R}(m,n))$

balról totális (HUN) – totális (HUN) – sorozatos (HUN) – szeriális (HUN) – serial (ENG) – sorozat (HUN) – left-total (ENG) – extendable (ENG)

(RL.27) $\text{RIGHT-TOTAL}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall n \exists m(\mathbf{R}(m\mathbf{R}(m,n))$

jobbról totális (HUN) – right-total (ENG)

(RL.28) $\text{LEFT-UNIQUE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,o) \wedge \mathbf{R}(n,o)) \rightarrow m = n)$

balról egyértelmű (HUN) – inverz-funkcionális (HUN) – inverse functional (ENG)

(RL.29) $\text{RIGHT-UNIQUE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)) \rightarrow n = o)$

jobbról egyértelmű (HUN) – funkcionális (HUN) – parciális függvény (HUN) – partial function (ENG) – functional relation (ENG) – right-unique (ENG) – univalent (ENG)

(RL.30) $\text{DENSE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall o(\mathbf{R}(m,o) \rightarrow \exists n(\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)))$

sűrű (HUN) – dense (ENG) – gyengén sűrű (HUN) – weakly dense (ENG)

(RL.31) $\text{CONVERGENT}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o \exists p((\mathbf{R}(o,m) \wedge \mathbf{R}(o,n)) \rightarrow (\mathbf{R}(m,p) \wedge \mathbf{R}(n,p)))$

konvergens (HUN) – convergent (ENG)

A későbbiekben talán a legtöbbet azzal a tulajdonságcsoporttal foglalkozunk majd, amely a tranzitivitással, ciklikussággal, aciklikussággal kapcsolatos. A relációk segítségével való következtetéseinket támogathatja az, ha tudjuk, hogy ha az adott reláció fennáll bizonyos elemek között, akkor ebből következtethetünk más minden elem közti fennállás tényére. Példaként hivatkozhatunk a IDŐSEBB relációra: ha A idősebb B-nél, és B idősebb C-nél, akkor már – mivel ez a reláció tranzitív – következtethetünk arra, hogy A idősebb C-nél. A későbbiekben a tranzitivitás tulajdonsága a preferencia fogalmával kapcsolatban fog gyakran előkerülni. Nem könnyíti meg a fogalom értelmezését az a tény, hogy fontossága miatt sokféle módon próbálták meg kifejezni a tranzitivitás minőségét, így sok versengő, egymásra hasonlító definíció született e téren.

(RL.32) $\text{TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow \mathbf{R}(m,o))$

tranzitív (HUN) – transitive relation (ENG)

(RL.33) $\text{INTRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow \neg \mathbf{R}(m,o))$

intranzitív (HUN) – atranzitív (HUN) – antitranzitív (HUN) – intransitive (ENG) – atransitive (ENG) – antitransitive (ENG)

(RL.34) $\text{QUASI-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o) \wedge m \neq o) \rightarrow \mathbf{R}(m,o))$

kvázitranzitív (HUN) – quasi-transitive (ENG)

(RL.35) $\text{ALIO-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow (\mathbf{R}(m,o) \vee m = o))$

aliotranzitív (HUN) – alio-transitive (ENG)

(RL.36) $\text{SEMI-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o \forall p((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow (\mathbf{R}(m,p) \vee \mathbf{R}(p,o)))$

szemitranzitív (HUN) – semi-transitive (ENG)

(RL.37) $\text{FERRERS_PROPERTY}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o \forall p((\mathbf{R}(m,n) \wedge \mathbf{R}(o,p)) \rightarrow (\mathbf{R}(m,p) \vee \mathbf{R}(o,n)))$

Ferrers tulajdonság (HUN) – Ferrers property (ENG)

(RL.38) $\text{NEGATIVELY_TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \rightarrow (\mathbf{R}(m,o)) \vee \mathbf{R}(o,n)))$

negatívan tranzitív (HUN) – negatively transitive relation (ENG)

(RL.39) $\text{NON-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \exists x \exists n \exists z((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \wedge \neg \mathbf{R}(m,o))$

nem-tranzitív (HUN) – atranzitív (HUN) – intranzitív (HUN) – non-transitive (ENG) – intransitive (ENG) – atransitive (ENG)

(RL.40) $\text{CIRCULAR}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow \mathbf{R}(o,m))$

cirkuláris (HUN) – circular (ENG)

(RL.41) $\text{ACIRCULAR}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(n,o)) \rightarrow \neg \mathbf{R}(o,m))$

acirkuláris (HUN) – acircular (ENG)

(RL.42) $\text{EUCLIDEAN}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)) \rightarrow \mathbf{R}(n,o))$

euklideszi (HUN) – euclidean (ENG)

(RL.43) $\text{ANEUCLIDEAN}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)) \rightarrow (\neg \mathbf{R}(n,o) \wedge \neg \mathbf{R}(o,n)))$

aneuklideszi (HUN) – antieuclidean (ENG) – aneuclidean (ENG) – ineuclidean (ENG) – antitransitive (ENG)

(RL.44) $\text{CATENARY}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \exists o(\mathbf{R}(m,n) \rightarrow ((\mathbf{R}(n,o) \vee \mathbf{R}(o,m)))$

katenáris (HUN) – láncolt reláció (HUN) – nem bipoláris (HUN) – catenary (ENG) – non-bipolar (ENG)

(RL.45) $\text{ACATENARY}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(\mathbf{R}(m,n) \rightarrow (\neg \mathbf{R}(n,o) \wedge \neg \mathbf{R}(o,m)))$

akatenáris (HUN) – nem-láncolt (HUN) – bipoláris (HUN) – non-catenary (ENG) – acatenary (ENG) – bipolar (ENG)

(RL.46) $\text{CONFLUENT}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o \exists p((\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)) \rightarrow (\mathbf{R}(n,p) \wedge \mathbf{R}(o,p)))$

konfluens (HUN) – confluent (ENG) – directed (ENG) – incesztuális reláció (HUN) – incestual relation (ENG)

(RL.47) $\text{ACONFLUENT}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o \exists p(\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)) \rightarrow \neg(\mathbf{R}(n,p) \wedge \mathbf{R}(o,p)))$

akonfluens (HUN) – aconfluent (ENG)

Ahhoz, hogy további fontos relációtulajdonságokat tudjunk definiálni, szükség van arra, hogy bevezessünk olyan új fogalmakat, amelyek elkülönítik a reláció valamely részét adott tulajdonság mentén, és csak az így lehatárolt részre vonatkoznak.

(RL.48) $\text{SYMMETRIC_PART_OF_}\mathbf{R}(m,n) \equiv \mathbf{I}(\mathbf{R}) \stackrel{\text{def}}{=} \{(m,n) \in \mathbb{X} \times \mathbb{X} | (m,n) \in \mathbf{R} \wedge (n,m) \in \mathbf{R}\}$

R szimmetrikus része (HUN) – symmetric part of **R** (ENG) – **R** szimmetrikus faktora (HUN) – symmetric factor of **R** (ENG)

(RL.49) $\text{ASYMMETRIC_PART_OF_}\mathbf{R}(m,n) \equiv \mathbf{P}(\mathbf{R}) \stackrel{\text{def}}{=} \{(m,n) \in \mathbb{X} \times \mathbb{X} | (m,n) \in \mathbf{R} \wedge (n,m) \notin \mathbf{R}\}$

R aszimmetrikus része (HUN) – asymmetric part of **R** (ENG) – **R** aszimmetrikus faktora (HUN) – asymmetric factor of **R** (ENG)

A későbbiekben még definiálni fogunk további származtatott relációkat (↪41.o.). A fenti relációrész-fogalmak segítségével bevezethetünk itt olyan tulajdonságokat, amelyekre támaszkodnunk majd a későbbiekben a racionalitás problémájának tárgyalásakor. Adottnak veszünk egy relációt, és annak egy részére írunk elő követelményeket.

(RL.50) $\text{TRIPLE_ACYCLIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((m,n) \in \mathbf{P}(\mathbf{R}) \wedge (n,o) \in \mathbf{P}(\mathbf{R}) \rightarrow (o,m) \notin \mathbf{P}(\mathbf{R}))$

hármas aciklusos (HUN) – triple acyclic relation

(RL.51) $\text{ACYCLIC}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall t \in \mathbb{N}^+, \forall m^1, m^2, \ldots, m^t \in \mathbb{X}(\forall \tau \in \{1,2,\ldots,t-1\}(m^t, m^{t+1}) \in \mathbf{P}(\mathbf{R})) \rightarrow (m^t, m^1) \notin \mathbf{P}(\mathbf{R})$

aciklusos (HUN) – acyclic relation (ENG)

(RL.52) $\text{CONSISTENT}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall t \in \mathbb{N}^+, \forall m^1, m^2, \ldots, m^t \in \mathbb{X}((m^1, m^2) \in \mathbf{P}(\mathbf{R}) \wedge \forall \tau \in \{2,3,\ldots,t-1\}((m^\tau, m^{\tau+1}) \in \mathbf{R})) \rightarrow (m^t, m^1) \in \mathbf{R}$

(Suzumura) konzisztens (HUN) – (Suzumura) consistent relation (ENG)

A konzisztencia fogalmára később bemutatok másfajta definíciókat is (↪41.o.).
Mivel a tranzitivitás tulajdonsága sok cselekvéselméleti fogalomhoz kapcsolódik, a gazdasági cselekvések területén pedig a legfontosabb fogalmak közé tartozik, a tranzitivitás tulajdonságára támaszkodva további fogalmakat határozhatunk meg. Ha van egy **R**(m,n) tranzitív relációnk, akkor definiálhatjuk a következő tulajdonságokat [476].

(BL.53) $\text{PP-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(((\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m)) \wedge (\mathbf{R}(n,o) \wedge \neg\mathbf{R}(o,n))) \rightarrow (\mathbf{R}(m,o) \wedge \neg\mathbf{R}(o,m)))$

PP-tranzitív (HUN) – PP-transitive (ENG) – kvázitranzitív (HUN) – quasi-transitive (ENG)

(BL.54) $\text{IP-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(((\mathbf{R}(m,n) \wedge \mathbf{R}(n,m)) \wedge (\mathbf{R}(n,o) \wedge \neg\mathbf{R}(o,n))) \rightarrow (\mathbf{R}(m,o) \wedge \neg\mathbf{R}(m,o))$

IP-tranzitív (HUN) – IP-transitive (ENG)

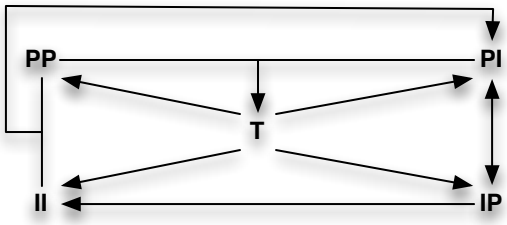
(BL.55) $\text{PI-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(((\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m)) \wedge (\mathbf{R}(n,o) \wedge \mathbf{R}(o,n))) \rightarrow \mathbf{R}(m,o) \wedge \neg\mathbf{R}(o,m)))$

PI-tranzitív (HUN) – PI-transitive (ENG)

(BL.56) $\text{II-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(((\mathbf{R}(m,n) \wedge \mathbf{R}(n,m)) \wedge (\mathbf{R}(n,o) \wedge \mathbf{R}(o,n))) \rightarrow (\mathbf{R}(m,o) \wedge \mathbf{R}(o,m)))$

II-tranzitív (HUN) – II-transitive (ENG)

Az így definiált tulajdonságok között fontos összefüggések állapíthatók meg, de ezeket nem mutatom be formális módon, csak megemlítek párat közülük [476], [219, 324]. Igaz például az, hogy ha egy reláció tranzitív (jelöljük ezt itt T-vel), akkor fennáll mind a négy belőle levezethető PP-, II-, IP- és PI-tranzititás tulajdonság is, de fontos tézis az is, miszerint a PP- és PI-, illetve a PP- és II-tulajdonságok együttes fennállásából következik a T tranzitivitás létezése. Ezeket a kapcsolatokat szemléltethetjük egy ábra segítségével.



A preferencia fogalmának tárgyalásakor még visszatérünk ezekre az összefüggésekre, és ott további relációk megnevezésével egyszerűsítünk az itt bevezetett tranzitivitási tulajdonságok formuláin (↰106.o.). Akkor lehet majd rámutatni arra, hogy honnan származnak a most definiált tulajdonságok megnevezései. Az eddig definiált relációtulajdonságok között természetesen nem teljesen függetlenek egymástól, rengeteg összefüggés van közöttük. Itt csak azokat mutatom be, amelyekre a későbbiekben szükség lesz [514].[11]

(RL.57) ASYMMETRIC → IRREFLEXIVE

(RL.58) STRONGLY_CONNECTED → IRREFLEXIVE

(RL.59) TRANSITIVE → CONSISTENT

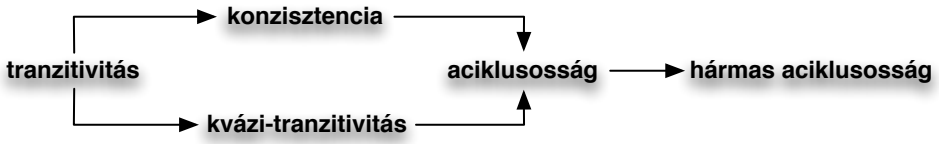
(RL.60) TRANSITIVE → QUASY_TRANSITIVE

(RL.61) CONSISTENT → ACYCLIC

(RL.62) QUASY_TRANSITIVE → ACYCLIC

(RL.63) ACYCLIC → TRIPLE_ACYCLIC

A fenti tételek közül a tranzitivitással kapcsolatos összefüggéseket grafikusan is ábrázolva szemléletesebbé tehetők a fogalmak közti összefüggések. Az ábráról könnyen leolvasható, hogy a legerősebb tulajdonság a tranzitivitás, míg a "leggyengébb" hármas aciklusosság, és köztük vannak a többiek.



[11] A kvázitranzitivitás terminusa korábban két fogalomhoz kapcsolódva is felbukkant. Az itteni kontextusban a PP-tranzitivitás értelemben szerepel a terminus. De ugyanezt a tartalmat fejezhetjük ki azzal, ha azt mondjuk, hogy a kvázitranzitív reláció ekvivalens a reláció tanzitív aszimmetrikus részével [65].

Nevezetes relációk

A legelemibb rendezésfogalmakat a különféle relációtulajdonságok segítségével lehet elkülöníteni egymástól. Sajnos, a rendezési fogalmak világában sincs terminológiai tisztaság, gyakran találkozhatunk mind a homonimitás, mind a szinonimitás jelenségével. Ebben a könyven leginkább a rendezési relációkat fogom használni, de fontosnak tartom megemlíteni itt a többi nevezetes relációt is.

(RL.64) $\text{TOLERANCE} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{SYMMETRIC}$

tolerancia reláció (HUN) – kompatibilitási reláció (HUN) – hasonlósági reláció (HUN) – compatibility relation (ENG) – similarity relation (ENG) – tolerance relation (ENG)

(RL.65) $\text{PARTIAL_EQUIVALENCE} \stackrel{\text{def}}{=} \text{TRANSITIVE} \wedge \text{SYMMETRIC}$

részleges ekvivalencia reláció (HUN) – partial equivalence relation (ENG) – per (ENG)

(RL.66) $\text{EQUIVALENCE} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{SYMMETRIC} \wedge \text{TRANSITIVE}$

ekvivalenciareláció (HUN) – ekvivalencia (HUN) – equivalence (ENG)

(RL.67) $\text{PARTIAL_ORDER} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{ANTISYMMETRIC} \wedge \text{TRANSITIVE}$

részben rendezés (HUN) – parciális rendezés (HUN) – részleges rendezés (HUN) – gyenge rendezés (HUN) – gyenge részben rendezés (HUN) – nem szigorú rendezés (HUN) – weak order (ENG) – weak partial order (ENG) – partial order (ENG)

(RL.68) $\text{WEAK_ORDER} \stackrel{\text{def}}{=} \text{COMPLETE} \wedge \text{TRANSITIVE}$

gyenge rendezés (HUN) – weak order (ENG)

(RL.69) $\text{LINEAR_ORDER} \stackrel{\text{def}}{=} \text{COMPLETE} \wedge \text{ANTISYMMETRIC} \wedge \text{TRANSITIVE}$

lineáris rendezés (HUN) – linear order (ENG)

(RL.70) $\text{STRICT_ORDER} \stackrel{\text{def}}{=} \text{IRREFLEXIVE} \wedge \text{TRANSITIVE}$

szigorú rendezés (HUN) – szigorú rendezési reláció (HUN) – erős részben rendezés (HUN) – irreflexív rendezés (HUN) – strict order (ENG) – strict weak order (ENG) – strong order (ENG) – strong weak order (ENG)

(RL.71) $\text{LINEAR_PARTIAL_ORDER} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{ANTISYMMETRIC} \wedge \text{TRANSITIVE} \wedge \text{DICHOTOMOUS}$

nem szigorú elrendezés reláció (HUN) – teljes parciális rendezés (HUN) – lineáris parciális rendezés (HUN) – totális parciális rendezés (HUN) – total partial order (ENG) – linear partial order (ENG)

(RL.72) $\text{LINEAR_STRICT_ORDER} \stackrel{\text{def}}{=} \text{IRREFLEXIVE} \wedge \text{TRANSITIVE} \wedge \text{DICHOTOMOUS}$

szigorú elrendezés (HUN) – teljes szigorú rendezés (HUN) – lineáris szigorú rendezés (HUN) – totális szigorú rendezés (HUN) – total strict order (ENG) – linear strict order (ENG) – total strong order (ENG) – linear strong order (ENG)

(RL.73) $\text{PREORDER} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{TRANSITIVE}$

előrendezés (HUN) – kvázirendezés (HUN) – preorder (ENG) – quasi order (ENG)

(RL.74) $\text{WEAK_ORDER} \stackrel{\text{def}}{=} \text{REFLEXIVE} \wedge \text{TRANSITIVE} \wedge \text{DICHOTOMOUS}$

gyenge rendezés (HUN) – lineáris előrendezés (HUN) – weak order (ENG) – linear preorder (ENG) – total preorder (ENG)

(RL.75) $\text{WEAK_PARTIAL_ORDER} \stackrel{\text{def}}{=} \text{ANTISYMMETRIC} \wedge \text{TRANSITIVE}$

gyenge részben rendezés (HUN) – partial order (ENG) – weak partial order (ENG)

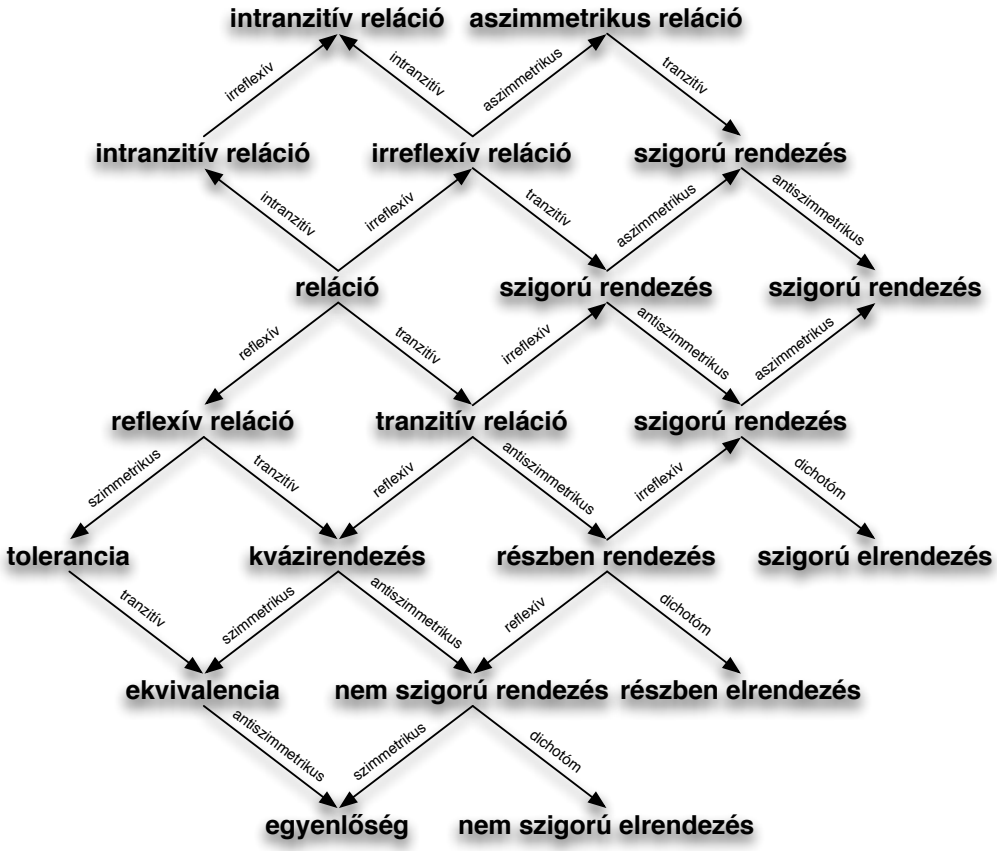
(RL.76) $\text{LINEAR_WEAK_PARTIAL_ORDER} \stackrel{\text{def}}{=} \text{ANTISYMMETRIC} \wedge \text{TRANSITIVE} \wedge \text{DICHOTOMOUS}$

részben elrendezés (HUN) – teljes gyenge részben rendezés (HUN) – total weak partial order (ENG) – linear weak partial order (ENG)

(RL.77) $\text{TOURNAMENT} \stackrel{\text{def}}{=} \text{TRANSITIVE} \wedge \text{DICHOTOMOUS}$

turnament (HUN) – tournament (ENG)

A relációról szóló könyvében a rendezési fogalmakról Jurij A. Srejder készített egy érdekes áttekintő ábrát (jelölve a relációk egymáshoz való viszonyát is) [508], amit korábban további relációkkal egészítettem ki [526], és most még teljesebbé téve mutatok be.



A relációkat mindig a hozzájuk tartozó (tartó)halmazon definájuk. Ha veszünk egy halmazt és definiálunk rajta egy vagy több relációt, akkor már reláció **struktúráról** beszélhetünk. A struktúra fogalmának tehát nagyon egyszerű (mert nagyon általános) meghatározása van: egy halmazból és egy (vagy több) reláció együttese.

(RL.78) $\text{STRUCTURE} \equiv \mathfrak{S} \equiv \mathfrak{F} \equiv \mathfrak{T} \stackrel{\text{def}}{=} (\mathbb{T}, \{\mathbf{R}_i\}_{i \in \mathbb{N}^+})$

struktúra (ENG) – structure (ENG) – relációs struktúra (ENG) – relation structure (ENG) – modell (HUN) – model (ENG) – algebra (HUN) – algebra (ENG)

Az eddig tárgyalt rendezési fogalmak is felfoghatók valamilyen struktúrának, de persze nemcsak rendezési fogalmak léteznek. Definiálhatunk egy nagyon fontos relációs struktúrát (ami még mindig rendezési fogalom). A neve **fa** vagy fastruktúra [61, 6], de nevezhetjük hierarchiának is [525]. Ezt a struktúrát mind a tudományban, mind a hétköznapi életben rengeteget használjuk, a szervezési, szerveződési technikák egyik legfontosabb típusa.

(RL.79) $\text{TREE}(m,n) \stackrel{\text{def}}{=} \forall n \exists! r \exists! m (\mathbf{R}^*(r,n) \wedge (m \neq r \rightarrow \mathbf{R}(m,n)) \wedge \neg \mathbf{R}^+(n,n))$

fa (HUN) – tree (ENG) – monohierarchia – (HUN) – mono-hierarchy (ENG)

Az első részformula azt határozza meg, hogy a fának egy és csak egy **csúcspontja** (root) van, a formula második összetevője pedig azt definálja, hogy a csúcselemen kívül minden más elemhez egy és csak egy **közvetlenül megelőző** elem (predecessor) tartozik.[12] A harmadik formula az aciklikusságot mondja ki a reláció $\mathbf{R}^+$ tranzitív lezártja segítségével – más formában, mint az aciklikusságra adott korábbi definíció (↪32.o.).

Be kell itt vezetnem két olyan további rendezésfogalmat, amelyekre a későbbiekben még hivatkozni szeretnék. Ha van két részben rendezett halmazunk (A és B), a két halmazon értelmezett $A \times B$ Descartes-szorzatán kétféle rendezést definiálhatunk: a sorrendezést és a lexikografikus rendezést. Ezeket a következőképpen határozhatjuk meg.

(RL.80) $\text{ARCHIMEDEAN_ORDER}((a_1,b_1),(a_2,b_2)) \equiv (a_1,b_1) \leq (a_2,b_2) \stackrel{\text{def}}{=} a_1 \leq^* a_2 \wedge b_1 \leq' b_2$

sorrendezés (HUN) – product order (ENG) – arkhimédeszi rendezés (HUN) – archimedean order (ENG)

(RL.81) $\text{LEXICOGRAPHIC_ORDER}((a_1,b_1),(a_2,b_2)) \equiv (a_1,b_1) \leq (a_2,b_2) \stackrel{\text{def}}{=} a_1 <^* a_2 \vee (a_1 = a_2 \wedge b_1 \leq' b_2)$

lexikografikus rendezés (HUN) – alfabetikus rendezés (HUN) – ábécé szerinti rendezés (HUN) – lexicographic order (ENG) – dictionary order (ENG)

Mindkét rendezés nagyon fontos. A sorrendezés példaként hivatkozhatunk arra, ahogyan sorba rakjuk a természetes számokat, míg a lexikografikus rendezés legismertebb példája a természetes nyelv szavainak szótári (ábécé szerinti) rendezése.

Függvény

A reláció egyik típusának fogható fel a függvény fogalma, de – ahogy azt már említettem – a függvényt is lehetne alapkategóriának tekinteni és le lehetne vezetni belőle a reláció fogalmát. A reláció kategóriatanát azzal folytatom, hogy előbb veszem a függvény mint speciális reláció fogalmát, majd bemutatom, hogy milyen fontosabb függvénytulajdonságok vannak.

[12] Az első részformulában a hierarchia csúcspontjára, illetve a második összetevőben a közvetlenül megelőző elemre megkövetelt egyediséget az unicitás kvantorral biztosítottam (↪26.o.).

(RL.82) $\text{FUNCTION}(m,n) \equiv \mathbf{F}(m,n) \equiv \mathbf{F}(m) = n \equiv \mathbf{F} \equiv \mathbf{F} : \mathbb{D}\mathrm{om} \mapsto \mathbb{C}\mathrm{od} \stackrel{\text{def}}{=} \text{LEFT-TOTAL}(\mathbf{F}) \wedge \text{RIGHT-UNIQUE}(\mathbf{F})$

függvény (HUN) – leképezés (HUN) – function (ENG)

(RL.83) $\text{INVERSE_FUNCTION}(m,n) \equiv \mathbf{F}^{-1}(m,n) \stackrel{\text{def}}{=} \text{RIGHT-TOTAL}(\mathbf{F}) \wedge \text{LEFT-UNIQUE}(\mathbf{F})$

inverz függvény (HUN) – inverz leképezés (HUN) – inverse function (ENG)

(RL.84) $\text{BIUNIQUE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{LEFT-UNIQUE}(\mathbf{F}) \wedge \text{RIGHT-UNIQUE}(\mathbf{F})$

egy-egyértelmű (HUN) – egy-egyértelmű (HUN) – biunique (ENG) – unique (ENG)

(RL.85) $\text{BITOTAL}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{LEFT-TOTAL}(\mathbf{F}) \wedge \text{RIGHT-TOTAL}(\mathbf{F})$

bitotális (HUN) – 1:1-es reláció (HUN) – bitotal (ENG) – bitotal relation (ENG) – 1-1 relation (ENG)

(RL.86) $\text{INJECTIVE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{FUNCTION}(\mathbf{F}) \wedge \text{LEFT-UNIQUE}(\mathbf{F})$

injektív (HUN) – egyrétű (HUN) – injective (ENG)

(RL.87) $\text{SURJECTIVE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{FUNCTION}(\mathbf{F}) \wedge \text{RIGHT-TOTAL}(\mathbf{F})$

szürjektív (HUN) – ráképzés (HUN) – szuperjektív (HUN) – lefedő (HUN) – surjective (ENG)

(RL.88) $\text{BIJECTIVE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{INJECTIVE}(\mathbf{F}) \wedge \text{SURJECTIVE}(\mathbf{F})$

bijektív (HUN) – egyrétűen lefedő (HUN) – bijective (ENG)

(RL.89) $\text{SERIES}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \text{FUNCTION}(\mathbf{F}) \wedge \mathbb{D} = \mathbb{N}$

sorozat (HUN) – series (ENG)

Ha speciális relációstruktúraként értelmezünk olyan struktúrát, amelyben ugyanazon a halmazon értelmezünk egy rendezési relációt és egy függvényt, akkor további fontos tulajdonságokat definiálhatunk. Többször lehet fontos az a kérdés, vajon egy rendezés által kialakított sorrend megőrződik-e akkor, ha definiálunk egy függvényt ugyanazon a halmazon. Ezt írhatjuk le a monotonicitás körébe tartozó tulajdonságokkal.

(RL.90) $\text{MONOTONE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (m \leq n \rightarrow \mathbf{F}(m) \leq \mathbf{F}(n))$

monoton (HUN) – monoton növekvő (HUN) – monotone (ENG) – monotonic (ENG) – monotonically increasing (ENG) – isotone (ENG) – order-preserving (ENG)

(RL.91) $\text{ANTITONE}(\mathbf{F}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (m \leq n \rightarrow \mathbf{F}(n) \leq \mathbf{F}(m))$

antiton (HUN) – komonoton (HUN) – monoton csökkenő (HUN) – co-monotone (ENG) – monotonically decreasing (ENG) – antitone (ENG) – anti-monotone (ENG) – order-reserving (ENG)

Művelet

A **műveletek** (operációk) speciális függvényként értelmezhető, de természetesene fordítva is igaz, és definiálható lenne a függvény a művelet fogalmára támaszkodva. A műveletre gyakran gasználják még az operáció kifejezést is, ez magyarázza azt a tényt, hogy

a műveletek argumentumaira, változóira operandusként hivatkozunk. A műveletek esetében szinte kizárólagos az ún. infix reprezentációs mód, amikor a művelettel összekapcsolt változókat (operandusokat) a műveleti jel két oldalára írjuk. A következő blokkban bemutatom a művelet és a legfontosabb művelettulajdonságok definícióit.

(RL.92) $\text{OPERATION}(m,n) \equiv \mathbf{O}(m,n) \equiv m\mathbf{O}n \stackrel{\text{def}}{=} \text{FUNCTION}(m,n) \wedge \mathbb{D}\mathrm{om} = \mathbb{C}\mathrm{od}$

művelet (HUN) – operáció (HUN) – belső összekapcsolás (HUN) – operation (ENG)

(RL.93) $\text{ASSOCIATIVE}(\mathbf{O}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o (m\mathbf{O}(n\mathbf{O}o) = (m\mathbf{O}n)\mathbf{O}o)$

asszociatív (HUN) – társítható (HUN) – associative (ENG)

(RL.94) $\text{COMMUTATIVE}(\mathbf{O}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (m\mathbf{O}n = n\mathbf{O}m)$

kommutatív (HUN) – felcserélhető (HUN) – commutative (ENG)

(RL.95) $\text{DISTRIBUTIVE}(\mathbf{O}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o ((m\mathbf{O}(n\mathbf{Q}o) = (m\mathbf{O}n)\mathbf{Q}(m\mathbf{O}o))$

O disztributív művelet Q-ra HUN) – szétosztható (HUN) – széttagolható (HUN) – distributive (ENG)
$\text{CONCEPT:RELATION:DISTRIBUTIVE}(\mathbf{Q}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o ((n\mathbf{Q}o)\mathbf{O}m = (n\mathbf{O}m)\mathbf{Q}(o\mathbf{O}m))$

Q disztributív O-ra (HUN) – szétosztható (HUN) – széttagolható (HUN) – distributive (ENG)

(RL.96) $\text{IDEMPOTENT}(\mathbf{O}(m,n)) \stackrel{\text{def}}{=} \forall m (m\mathbf{O}m = m)$

idempotens (HUN) – önhatványozó (HUN) – azonos hatványú (HUN) – idempotent (ENG)

(RL.97) $\text{ABSORPTIVE}(\mathbf{O}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (m\mathbf{Q}(m\mathbf{O}n) = m)$

abszorbtív Q-ra (HUN) – elnyelő (HUN) – adjunktív (HUN) – absorptive (ENG)
$\text{ABSORPTIVE}(\mathbf{Q}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (n\mathbf{O}(m\mathbf{Q}n) = m)$

abszorbtív O-ra (HUN) – elnyelő (HUN) – adjunktív (HUN) – absorptive (ENG)

Relációműveletek

Miután már ismerjük a műveletek tulajdonságait, típusait, visszatérhetünk annak a kérdésnek a tisztázásához, hogy milyen műveletek definiálhatók a relációk között. A relációműveletek azért fontosak, mert ha relációkat kapcsolunk össze relációkkal (műveletek mint speciális relációk segítségével), akkor valamilyen halmazon (a relációk halmazán) értelmezett reláció(k)ról, művelet(ek)ről van szó, vagyis ilyen módon is valamilyen struktúrát állítunk elő. A műveletek között unáris és bináris műveletet definiálunk.

(RL.98) $\text{R-CONVERSION} \equiv \mathbf{R}^{-1}(m,n) \equiv \mathbf{R}^{-c}(m,n) \equiv \mathbf{R}^{-1} \equiv \mathbf{R}^{-c} \stackrel{\text{def}}{=} \forall n \forall n (\mathbf{R}^{-1}(m,n) \leftrightarrow \mathbf{R}(n,m))$

relációkonvertálás (HUN) – konverzképzés (HUN) – inverzképzés (HUN) – megfordítás (HUN) – relation conversion operation (ENG) – relation converse operation (ENG) – relation inverse operation (ENG)

Ha van egy $\mathbf{R}(m,n)$ relációnk, akkor annak a konverze (vagyis a konverzreláció) úgy állítható elő, hogy a relációba kapcsolt elemek, a relátumok sorrendjét felcseréljük egymással. A relációkonvertálást a köznyelvben sokszor új terminus bevezetésével oldjuk meg.

Például a GYEREKE(*Béla,Pál*) reláció konverze a GYEREKE^{-1}(*Béla,Pál*), de nem ezt alkalmazzuk, hanem inkább egy új kifejezést alkotunk és a SZÜLŐJE(*Pál,Béla*) relációt használjuk. A definícióból adódik, hogy ha a konverziót kétszer egymás után alkalmazzuk, akkor az eredeti relációhoz jutunk vissza. A konverzreláció a megfordított reláció, a konverzió maga a megfordítási művelet. A konverzió műveletét mind a mondatszerű, mind a szószerű megnyilatkozások szintjén értelmezhetjük. Amikor az összekapcsolt mondatok belső struktúráját vizsgáljuk, akkor **kondicionális konverzióról** beszélünk (↪22.o.), amikor a mondatokon belüli összetevők szerkezetét, akkor **kategoriális konverzióról** (↪51.o.). Az első esetben a művelet értelmezési tartománya a propozíciókból, a második esetben a kategoriális propozíciók alkotórészeiből, predikátumokból, kategóriákból áll.[13]

Timothy Williamson egy tanulmányában kifejtette, hogy a konverzreláció nem tekinthető önálló relációnak, sőt, voltaképpen teljesen megegyezik az alaprelációval [583], vagyis nem tekinthető másnak, mint a szokásostól eltérő, alternatív jelölési konvenciónak. Ahogy a formális nyelvekben ugyanazt a relációt is többféleképpen jelölhetjük (például $\mathbf{R}mn$, $m\mathbf{R}n$ vagy $\mathbf{R}(m,n)$), úgy mutatunk ugyanarra a tartalomra az alapreláció $\mathbf{R}(m,n)$, illetve a konverzreláció $\mathbf{R}(n,m)$ vagy $\mathbf{R}^{-1}(m,n)$ jeleivel, csak éppen más sorrendben rögzítjük a relátumokat – ami csak megállapodás kérdése, semmi más. Ebben az értelemben a konverzrelációt jelző terminusok mind szinonímái az alapreláció terminusainak, tehát a GYEREKE relációnak szinonímája a SZÜLŐJE reláció, mivel tartalmilag teljesen ugyanazt jelentik. Ez természetesen csak logikai szempontból igaz. A hétköznapi nyelvhasználatban ezt az összefüggést nem is érzékeljük, nem is érvényesítjük.

Sok helyen, sokan szinonimaként használják az **inverz** és **konverz** terminusokat, és vannak, akik megkülönböztetik őket egymástól. Én is így teszem, de kiderül majd az is, hogy vannak helyzetek, amikor valóban megegyezik a két művelet eredménye (tehát bizonyos feltételek esetén valóban kezelhetjük őket szinonimaként).

(RL.99) $\text{R-INVERSION} \equiv \mathbf{R}^{-i}(m,n) \equiv \mathbf{R}^{-i} \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{R}^{-i}(m,n) \leftrightarrow \mathbf{R}(\neg m, \neg n))$

relációinvertálás (HUN) – inverzképzés (HUN) – relation inverse operation (ENG)

Ha van egy $\mathbf{R}(m,n)$ relációnk, akkor annak az inverze (vagyis a inverzreláció) úgy állítható elő, hogy a relációba kapcsolt elemek, a relátumok komlementereit (tagadásait) kapcsoljuk össze. A definíció alapján könnyű belátni, hogy két egymás utáni inverzió visszaadja magát az eredeti relációt.

Transzformációs művelet még az **obverzió** (átalakítás) is, de ezt csak később tudom bemutatni és definiálni (↪52.o.), mivel tágyalásához szükség van még további fogalmak ismeretére. Itt elemzem viszont az általános kontrapozíció műveletét, hogy aztán a későbbi fejezetekben kitérjek majd a további típusaira is.

(RL.100) $\text{R-CONTRAPOSITION} \equiv \mathbf{R}^{-p}(m,n) \equiv \mathbf{R}^{-p} \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{R}^{-p}(m,n) \leftrightarrow \mathbf{R}(\neg n, \neg m))$

kontrapozíció (HUN) – transzpozíció (HUN) – contraposition (ENG) – transposition (ENG)

Ha van egy $\mathbf{R}(m,n)$ relációnk, akkor annak a kontrapozíciója (vagyis a reláció kontraponáltja) úgy állítható elő, hogy a relációba kapcsolt elemek, a relátumok komplementereit fordított sorrendben kapcsoljuk össze. A kontrapozíciót is értelmezhetjük mind

[13] Lásd a 'Cambridge Dictionary of Philosophy' szótár "converse" lexikai egységét [26].

a mondatszerű, mind a szószerű megnyilatkozások szintjén. Összekapcsolt mondatok szerkezetének átalakításakor kondicionális kontrapozícióról (↫22.o.), mondatokon belüli összetevők átstrukturálásakor kategoriális kontrapozícióról (↫52.o.) beszélünk. A kontrapozíció definíciójából az is látszik, hogy a kontrapozíció nem más, mint a konverzió inverze, illetve az inverzió konverze.[14]

(RL.101) $\textsc{r-contraposition} \leftrightarrow \textsc{r-conversion}(\textsc{r-inversion}) \leftrightarrow \textsc{r-inversion}(\textsc{r-conversion})$

Két további fontos unáris relációműveletet a reláció halmazelméleti kötődése alapján határozhatunk meg. Ha a relációt speciális halmazként fogjuk fel, akkor alkalmazható rá a halmazokon értelmezett komplementálás művelete (↫**??**.o.), ami az univerzális halmazra vonatkoztatva a reláció tagadásaként is felfogható.

(RL.102) $\textsc{r-complementation} \equiv \mathbf{R}^{-t}(m,n) \equiv \mathbf{R}^{-t} \equiv \overline{\mathbf{R}}(m,n)) \equiv \overline{\mathbf{R}} \stackrel{\text{def}}{=} \forall m \forall n(\overline{\mathbf{R}}(m,n) \leftrightarrow \neg \mathbf{R}(m,n)$

relációkomplementálás (HUN) – relation complementation (ENG) – relation transzpozíció (HUN) – relation transposition (ENG)

Ha pedig egymás után alkalmazzuk a reláció konvertálásának és komplementálásának (tagadásának) műveletét, akkor a reláció duális párjához jutunk.[15]

(RL.103) $\textsc{r-dual} \equiv \mathbf{R}^{-d}(m,n)) \equiv \mathbf{R}^{-d} \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}^{-d}(m,n) \leftrightarrow \neg \mathbf{R}(n,m) \equiv \overline{\mathbf{R}^{-c}}(m,n)$

duálreláció (HUN) – duális (reláció) (HUN) – dual relation (ENG)

A relációk között bináris műveleteket is értelmezhetünk. A két halmazelméleti művelet (az únió- és metszetképzés), valamint a két aritmetikai művelet (a szorzás és összeadás) mintájára bevezethetjük a következőket.

(RL.104) $\textsc{r-intersection} \equiv \times(\mathbf{R}(m,n), \mathbf{Q}(m,n)) \equiv \times(\mathbf{R}, \mathbf{Q}) \equiv \mathbf{R} \times \mathbf{Q} \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R} \times \mathbf{Q}(m,n) \leftrightarrow (\mathbf{R}(m,n) \wedge \mathbf{Q}(m,n))$

relációmetszés (HUN) – relation intersection (ENG)

(RL.105) $\textsc{r-union} \equiv +(\mathbf{R}(m,n), \mathbf{Q}(m,n)) \equiv +(\mathbf{R}, \mathbf{Q}) \equiv \mathbf{R} + \mathbf{Q} \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R} + \mathbf{Q}(m,n) \leftrightarrow (\mathbf{R}(m,n) \vee \mathbf{Q}(m,n))$

relációegyesítés (HUN) – relációúnió (HUN) – relációegyesítés (HUN) – relation union (ENG)

(RL.106) $\textsc{r-composition} \equiv \ast(\mathbf{R}(m,n), \mathbf{Q}(m,n)) \equiv \ast(\mathbf{R}, \mathbf{Q}) \equiv \mathbf{R} \ast \mathbf{Q} \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R} \ast \mathbf{Q}(m,n)) \leftrightarrow (\mathbf{R}(m,o) \wedge \mathbf{Q}(o,n))$

relációszorzás (HUN) – relatív szorzat (HUN) – kompozíció (HUN) – multiplikáció (HUN) – szuperpozíció (HUN) – relációkompozíció (HUN) – relation product (ENG) – relative multiplication (ENG) – relative product (ENG) – superposition (ENG) – composition (ENG)

(RL.107) $\textsc{r-addition} \equiv \dagger(\mathbf{R}(m,n), \mathbf{Q}(m,n)) \equiv \dagger(\mathbf{R}, \mathbf{Q}) \equiv \mathbf{R} \dagger \mathbf{Q} \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R} \dagger \mathbf{Q}(m,n)) \leftrightarrow (\mathbf{R}(m,o) \vee \mathbf{Q}(o,n))$

relációösszeadás (HUN) – összeadás (HUN) – relatív összeadás (HUN) – relative addition (ENG) – relative sum (ENG)

[14] Ezen a ponton elméletileg hiányzik még a relációk egymás utáni alkalmazásának vagyis a relációkompozíció műveletének fogalma, de ezt a hiányt még ebben a fejezetben pótlom.

[15] Itt is szükség van a relációkompozíció műveletére.

Bár a fenti bináris relációműveletek mindegyike szerepet kap a relációalgebra struktúrájában, közülük is "kiemelkedik" a relációszorzat vagy másik gyakori nevén a relációkompozíció művelete. A mindennapi fogalmi készletünkben is előfordulnak olyan fogalmak, amelyek ezt a műveletet használják. Egyszerű példát lehet hozni erre a rokonsági fogalmak köréből. Vegyük a GYEREKE relációt, amivel összeköthetjük Pétert és Katalint (GYEREKE(*Péter,Katalin*)). Ezután vegyük a TESTVÉRE relációt, amivel összekapcsolhatjuk Máriát és Pétert (TESTVÉRE(*Mária,Péter*)). A két relációt összekomponálva a következő eredményt kapjuk: TESTVÉRE(*Mária,Péter*)$*$GYEREKE(*Péter,Katalin*), és hogy ne kelljen ilyen hosszan kifejeznünk ezt a kapcsolatot Mária és Katalin között, használjuk inkább a NAGYNÉNJE relációt.

A relációműveletek segítsévégel további származtatott relációkat definiálhatunk, amelyek fontosak lesznek a későbbiekben. Az itt definiált tranzitív lezártja reláció voltaképp ugyanúgy relációművelet, mint a fentiek, csak épp itt végtelen sorozat segítségével tudjuk a pontos meghatározást megadni.

(RL.108) $\text{TRANSITIVE_CLOSURE}(\mathbf{R}) \equiv \mathbf{R}^+ \stackrel{\text{def}}{=} (\mathbf{R}^1 = \mathbf{R} \wedge (i > 1 \rightarrow \mathbf{R}^{i+1} = \mathbf{R} * \mathbf{R}^i)) \rightarrow \bigcup_{i \in \mathbb{N}^+} \mathbf{R}^i$

tranzitív lezártja (HUN) – transitive closure (ENG)

(RL.109) $\text{REFLEXIVE_TRANSITIVE_CLOSURE}(\mathbf{R}) \equiv \mathbf{R}^* \stackrel{\text{def}}{=} a$

reflexív tranzitív lezártja (HUN) – reflexive transitive closure (ENG)

Ha egy reláció megegyezik a tranzitív lezártjával, akkor a reláció egyben tranzitív, de a legtöbbször nem ez a helyzet, és a reláció, illetve a tranzitív lezártja különbözik egymástól. A reláció tranzitív lezártja fogalom segítségével a korábbinál jóval egyszerűbben is definiálhatjuk a – racionalitás leírásában oly fontos – konzisztencia fogalmát (↳32.o.), ehhez csak a reláció tranzitívját, illetve aszimmetrikus részét kell összekapcsolni egymással [64].

(RL.110) $\text{CONCISTENT_RELATION} \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{R}^+(m,n) \rightarrow \neg \mathbf{P}(\mathbf{R})(n,m))$

konzisztens reláció (HUN) – consistent relation (ENG)

A formula egyszerűen értelmezhető: egy reláció konzisztens, ha tetszőleges két elemet, m-t és n-t a reláció tranzitív lezártjával összekötünk, akkor a két elem inverz viszonyban (tehát n és m) nem tartozik bele a reláció aszimmetrikus részébe. Ennek alapján Suzumura társaival javasolta bevezetni a reláció konzisztens lezártja reláció fogalmát [64].

(RL.111) $\text{CONCISTENT_CLOSURE} \equiv \mathbf{R}^\times(m,n) \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{R}(m,n) \vee (\mathbf{R}^+(m,n) \wedge \mathbf{R}(n,m)))$

reláció konzisztens lezártja (HUN) – consistent closure relation (ENG)

Az $\mathbf{R}^\times(m,n)$ relációra támaszkodva már kimondható és bizonyítható a tétel, miszerint egy reláció akkor konzisztens, ha megegyezik a konzisztens lezártjával.

Relációalgebra

Minden formális elmélet számára kiemelten fontosak a relációs fogalmak, hiszen a relációk mint predikátumok adják a kijelentéseink általánosítható ismétlődéseit, ezek biztosítják kijelentéseink univerzális jellegét, ezek az állításaink "igéi". A struktúra, a strukturáltság fogalma és a relációra épül rá. Egy ilyen központi fogalom esetében is keresni

lehet további absztrakciós lehetőségeket. Felmerül a kérdés: lehet-e a relációk és műveleteik között valamilyen ismert matematikai struktúrát találni? A válasz pedig igenlő, van egy ilyen absztrakt stuktúra, amit **relációalgebrának** neveznek. A reláció és a köztük végezhető műveletek fogalmaira támaszkodva definiálták ezt a struktúrát. Ez azért fontos, mert az **algebra** egy alaposan leírt és nagyon erős tulajdonságokkal rendelkező matematikai struktúra. A matematika nagyon sok mindent tud egy struktúrára nézve mondani, ha kiderül, hogy az algebra. Vigyázni kell azonban, hogy ne tévesszük össze a relációs algebrával, ami a relációs adatbázis-kezelés matematikai alapjait biztosító elmélet. A relációalgebra felépítéséhez a következő relációs fogalmakból álló nyelvre van szükségünk.

(RA.0) $\mathscr{L}_{RA}(\mathbf{R},\mathbf{P},\mathbf{Q},\mathbf{U},\mathbf{E},\mathbf{I},\mathbf{D},\neg,\wedge,\rightarrow,=,\mathbf{R}^{-1},\overline{\mathbf{R}},\times,+,\divideontimes,\dagger)$, ahol

$\mathbf{R},\mathbf{P},\mathbf{Q}$	relációk
$\underset{\perp}{\mathbf{F}}$	teljes reláció
$\underset{\perp}{\mathbf{E}}$	üres reláció
$\underset{\perp}{\mathbf{I}}$	azonosság reláció
$\underset{\perp}{\mathbf{D}}$	különbözőség reláció
$\neg$	tagadás logikai művelet
$\wedge$	konjunkció logikai művelet
$\rightarrow$	kondicionális logikai művelet
$=$	egyenlőség reláció
$\mathbf{R}^{-1}$	relációkonvertálás művelet
$\overline{\mathbf{R}}$	relációkomplementálás művelet
$\times$	relációmetszés művelet
$+$	relációegyesítés művelet
$\divideontimes$	relációszorzás művelet
$\dagger$	relációösszeadás művelet

A fenti nyelvre támaszkodva már definiálhatjuk a relációalgebrát. Ehhez a következő axiómákat kell felvennünk.

(RA.1) $(\mathbf{R}=\mathbf{P}\wedge\mathbf{R}=\mathbf{Q})\rightarrow\mathbf{P}=\mathbf{Q}$

(RA.2) $\mathbf{R}=\mathbf{P}\rightarrow(\mathbf{R}+\mathbf{Q}=\mathbf{P}+\mathbf{Q}\wedge\mathbf{R}\times\mathbf{Q}=\mathbf{P}\times\mathbf{Q})$

(RA.3) $\mathbf{R}+\mathbf{Q}=\mathbf{Q}+\mathbf{R}\wedge\mathbf{R}\times\mathbf{Q}=\mathbf{Q}\times\mathbf{R}$

(RA.4) $(\mathbf{R}+\mathbf{P})\times\mathbf{Q}=\mathbf{R}\times\mathbf{Q}+\mathbf{P}\times\mathbf{Q}\wedge(\mathbf{R}\times\mathbf{P})+\mathbf{Q}=(\mathbf{R}+\mathbf{Q})\times(\mathbf{P}+\mathbf{Q})$

(RA.5) $\mathbf{R}+\underset{\perp}{\mathbf{E}}=\mathbf{R}\wedge\mathbf{R}\times\underset{\perp}{\mathbf{F}}=\mathbf{R}$

(RA.6) $(\mathbf{R}+\overline{\mathbf{R}}=\underset{\perp}{\mathbf{F}})\wedge(\mathbf{R}\times\overline{\mathbf{R}}=\underset{\perp}{\mathbf{E}})$

(RA.7) $\neg\underset{\perp}{\mathbf{F}}=\underset{\perp}{\mathbf{E}}\wedge\underset{\perp}{\mathbf{F}}=\neg\underset{\perp}{\mathbf{E}}$

(RA.8) $(\mathbf{R}^{-1})^{-1}=\mathbf{R}$

(RA.9) $(\mathbf{R}\divideontimes\mathbf{Q})^{-1}=\mathbf{R}^{-1}\divideontimes\mathbf{Q}^{-1}$

(RA.10) $\mathbf{R}\divideontimes(\mathbf{P}\divideontimes\mathbf{Q})=(\mathbf{R}\divideontimes\mathbf{P})\divideontimes\mathbf{Q}$

(RA.11) $\mathbf{R}\divideontimes\underset{\perp}{\mathbf{I}}=\mathbf{R}$

(RA.12) $\mathbf{R}\divideontimes\underset{\perp}{\mathbf{F}}=\underset{\perp}{\mathbf{F}}\wedge\underset{\perp}{\mathbf{F}}\divideontimes\overline{\mathbf{R}}=\underset{\perp}{\mathbf{F}}$

(RA.13) $((\mathbf{R}\divideontimes\mathbf{P})\times\mathbf{Q}^{-1}=\underset{\perp}{\mathbf{E}})\rightarrow((\mathbf{P}\divideontimes\mathbf{Q})\times\mathbf{R}^{-1}=\underset{\perp}{\mathbf{E}})$

(RA.14) $\underset{\perp}{\mathbf{D}} = \overline{\underset{\perp}{\mathbf{I}}}$

(RA.15) $\mathbf{R}\dagger\mathbf{Q} = \overline{\overline{\mathbf{R}} * \overline{\mathbf{Q}}}$

A relációkalkulust meg lehet fogalmazni kevesebb tételből is, nekem itt nem fontos, hogy a legaktuálisabb változatot mutassam be. Értelme, mint minden kalkulusnak, az, hogy segítségével definiálni lehet minden relációfogalmat (relációtulajdonságokat, nevezetes relációkat), és természetesen le lehet vezetni belőlük relációlogika minden további tézisét. A relációkalkulust Alfred Tarski dolgozta ki elsőként [538], a relációalgebra és -kalkulus fejlődéséről alapos áttekintést nyújt Roger D. Maddux [316], míg a relációalgebra területét behatóbban tárgyalja [Andréka Hajnal, Németi István és Sain Ildikó kézikönyve [13].

1.6 Kategoriális logika

A kategoriális logika részben a predikátumlogika egyfajta előzményének minősíthetjük, részben a predikátumlogikával párhuzamosan fejlődő, azzal részleges átfedésben levő logikai diszciplinának tekinthetjük. A kategoriális logika a kategorikus ítéletek logikai összefüggéseivel foglalkozik. Arisztotelész dolgozta ki az alapjait. A logika két fő kategóriája a szubjektum és a predikátum. Mivel ezek a terminusok más kontextusokban – legalább részben – más jelentéssel bírnak, ezért itt minősítő jelzőket kapcsolok hozzájuk, és a kategoriális szubjektum és a kategoriális predikátum fogalmát fogom definiálni. A meghatározások előtt azonban tisztázni kell azt a kérdést, mi a kategoriális logika viszonya a predikátumlogikához. A predikátumlogika az általánosabb, mivel bármilyen predikátum használatát megengedi, ezért a kijelentések belső szerkezete is tetszőleges lehet. Vehetünk egy három-argumentumú predikátumot, amit három individuumnév behelyettesítésével lehet megítélhető állítássá formálni.

elad(x,y,v,w) + ‘Laci’, ‘Zoli’, ‘BMW’ – ‘Laci eladta Zolinak a BMW-jét.’

A kategoriális logika speciálisabb, mert nem enged meg bármilyen mondatstruktúrát. Ahogy már említettem, kétféle kategóriát enged meg, a szubjektumot és a predikátumot, és ezeket csak speciális állításszerkezetekbe engedi ‘beletenni”. Például:

Minden veréb [van] madár.

Ez egy lehetséges kategoriális állítás, és az egyik legfontosabb minősége az, ami a magyar nyelvben “láthatatlan”: két kategória közti kopulával (a magyar nyelvben elhagyott ‘van’-nal) azt fejezzük ki, hogy a két kategória (illetve azok terjedelmébe eső előfordulások) között valamilyen ontológiai kapcsolat van. A kategoriális szubjektumról állítjuk azt, hogy annak valamilyen köze van a kategoriális predikátumhoz. A kategoriális logikában két kategóriát kapcsolatára mondunk valamit úgy, hogy valamilyen módon (mennyiségileg vagy minőségileg) módosítani lehet a kategóriákat, de a két kategóriát mindenképpen a kopulával kapcsoljuk össze.

(GL.0) CATEGORICAL_SUBJECT ≡ LOGICAL_CONSTANT ≡ INDIVIDUAL_NAME ≡ **S**

kategoriális szubjektum: amire a kategoriális kijelentés szerkezetén belüli másik kategória, a kategoriális predikátum vonatkozik.

(GL.1) CATEGORICAL_PREDICATE ≡ RELATION ≡ FUNCTION ≡ **P**

kategoriális predikátum: amivel a kategoriális kijelentés szerkezetén belüli másik kategóriára, a kategoriális szubjektumra vonatkozóan állítunk valamit.

Az nyilvánvaló, hogy a kategoriális predikátum a predikátumlogikai értelemben felfogott predikátum fogalma alá sorolható, a kérdés az, hogy mit mondhatunk a kategoriális szubjektumra vonatkozóan. A válaszhoz tudnunk kell, hogy ilyen szerkezeti feltételek tehetünk szinguláris állításokat, amikor a szubjektum egy individuum neve, illetve általános állításokat, amikor a szubjektum valamilyen predikátumlogikai értelemben vett predikátum. Rögzítenünk kell tehát, hogy mind a kategoriális szubjektum, mind a kategoriális predikátum lehet "hagyományos" predikátum.

A kategoriális logika kidolgozásakor és továbbépítésekor azonban nem ez a kérdés foglalkoztatta a logikusokat. Kezdetekben két szempont szerint osztották fel a kategoriális kijelentéseket, és ez alapján tettek meg fontos – évezredeken át fenntartott – állításokat. Egyfelől figyeltek arra, hogy partikuláris vagy univerzális állításról van-e szó, a másik szempont szerint pedig azt nézték, hogy az állításba az adott kategóriákat állító vagy tagadó módban vették-e föl. Az univerzális kijelentések a kategória terjedelmébe sorolt minden előfordulásra vonatkoztak, míg a partikuláris kijelentések hatóköre csak néhány ilyen elemre irányult. Az állító, pozitív vagy megerősítő kijelentésekben a kopula szerepelt a kategóriák között, míg a tagadó, negatív kijelentésekben a negált, tagadott kopula. A négyféle kijelentést az alábbi táblázatba rendezhetjük:

kijelentéstípus	jele	latin mondat	magyar mondat
egyetemes állító	**(A)**	Omne S est P.	Minden S [van] P.
egyetemes tagadó	**(E)**	Nullum S est P.	Egyetlen S sem P.
partikuláris állító	**(I)**	Quoddam S est P.	Néhány S [van] P.
partikuláris tagadó	**(O)**	Quoddam S non est P.	Néhány S nem P.

Az évezredeken át formálódó közös tudás formális szempontból inkonzisztenciát tartott fent, mert az egyetemes tagadás nem a konzisztens alakot tartotta számon (ami a 'Minden S [van] nem P' lett volna), de tartalmilag mindvégig helyesen kezelték ezeket az állításokat.[16] Arisztotelésznél még nem így volt, de a követői a négyféle állítást már egyetlen rendszerbe szervezték szemléletes ábrával reprezentálva az állítások közti összefüggéseket, és innentől kezdve logikai négyszögként hivatkoztak rájuk. Ezt a következő fejezetben bontom ki alaposabban. Előtte azonban felveszem az új fogalmakat az ontológiánkba.

(GL.2) UNIVERSAL_AFFIRMATIVE_PROPOSITION ≡ **S(A)P** ≡ **(A)** $\stackrel{\text{def}}{=} \forall m(\mathbf{S}(m) \rightarrow \mathbf{P}(m))$

egyetemes állító kijelentés (HUN) – universal affirmative proposition (ENG)

[16] kell-e ez, ide kell-e ez?
Azt is meg kell itt jegyezzem, hogy akkor is előfordult inkonzisztencia, amikor a kopula tagadása helyett a predikátum (tehát egy kategória) tagadásáról beszéltek (ez nyilvánul meg a modernkori logikai formulákban is egyébként). Abban az értelemben tartom ezt inkonzisztenciának, hogy elméletileg nem szabad összemosni a tagadás műveletét a komplementálással. Előbbi ugyanis a kijelentésekre vonatkozik, utóbbi viszont terminusokra (predikátumokra, kategóriákra). A negált terminus, negált kategória értelmes fogalom, de szigorú értelemben komplementált vagy komplementer terminusnak kellene neveznünk. Amennyiben a kontextusok alapján ez egyértelműen kiderül, akkor természetesen megengedhető ez a szinonimitás.

(GL.3) UNIVERSAL_NEGATIVE_PROPOSITION $\equiv \mathbf{S(E)P} \equiv \mathbf{(E)} \stackrel{\text{def}}{=} \forall m(\mathbf{S}(m) \rightarrow \neg\mathbf{P}(m))$

egyetemes tagadó kijelentés (HUN) – universal negative proposition (ENG)

(GL.4) PARTICULAR_AFFIRMATIVE_PROPOSITION $\equiv \mathbf{S(I)P} \equiv \mathbf{(I)} \stackrel{\text{def}}{=} \exists m(\mathbf{S}(m) \rightarrow \mathbf{P}(m))$

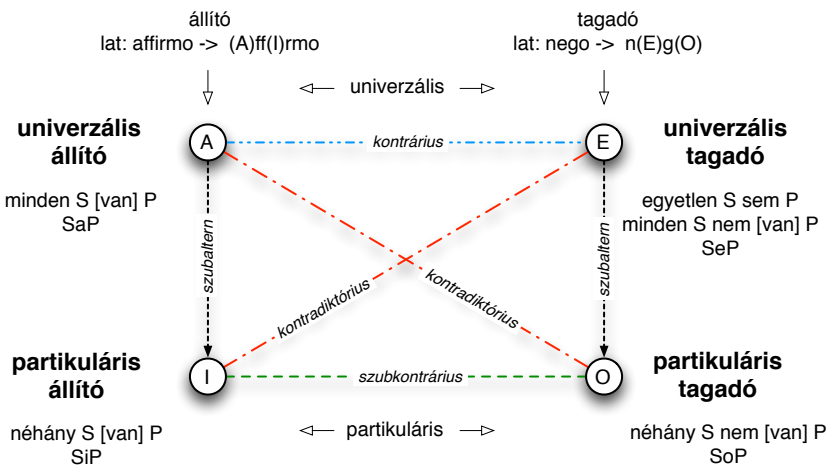
partikuláris állító kijelentés (HUN) – particular affirmative proposition (ENG)

(GL.5) PARTICULAR_NEGATIVE_PROPOSITION $\equiv \mathbf{S(O)P} \equiv \mathbf{(O)} \stackrel{\text{def}}{=} \exists m(\mathbf{S}(m) \rightarrow \neg\mathbf{P}(m))$

partikuláris tagadó kijelentés (HUN) – particular negative proposition (ENG)

Logikai négyszög

A kategoriális logika alapjait Arisztotelész fektette le, de ő még nem rajzolt ábrákat. Pár évszázad múlva Apuleius, majd Boethius volt az, aki a négyféle kijelentés közti kapcsolatokat ábrában kezdte mutatni. Az ilyen az ábrákat szokták Apuleius négyszögének vagy Boetius-féle négyzetnek is nevezni, de ezeken túl gyakran használják még a **logikai négyszög**, a logikai négyzet, valamint az ellentétek négyszöge vagy az oppozíciós négyszög kifejezéseket is [52][366].[17] Az ábra így néz ki:



A négyszög négy sarokpontját elnevezték ((**A**)-, (**E**)-, (**I**)-, (**O**)-pontoknak), és így hivatkoznak rájuk a mai napig. Az ábrán már feltüntettem, de az eddigiekből még nem derült ki, miért is nevezték ellentétek négyszögének ezt az alakzatot. Ugyan Arisztotelész nem használta még ezt a fenti vizuális megjelenítést, azonban ő volt az, aki alaposan leírta a kijelentések közti viszonyrendszert, és kiderült, a kijelentéstípusok közti kapcsolatok majd mindegyike valamifajta ellentétet fejez ki. A logikai négyzet elemei közt évszázadokon keresztül három ellentétes viszonyt tételeztek: *(i)* a **kontradiktórius** (ellentmondó), *(ii)* a **kontrárius** (ellentétes), valamint *(iii)* a **szubkontrárius** (alárendelt ellentétes) kapcsolatot. Ezeken felül rögzítettek még egy alárendeltségi viszonyt kifejező **szubaltern** (alárendelt), **szuperaltern** (fölérendelt) kapcsolatpárt is. Ezen kapcsolatokat, illetve modernkori megfelelőiket az alábbi táblázatban mutatom meg:

[17] Mivel a négyszög továbbfejlesztésekor tartalmilag sokféle irányban elmozdultak, az eredeti négyszöget ma már **kvantifikációs négyszögnek** is nevezik.

ellentét	formulája	logikai művelet neve	jele
szubkontrárius	$\neg(\neg A \wedge \neg B)$	diszjunkció	$A \vee B$
kontradiktórius	$\neg(A \wedge B) \wedge \neg(\neg A \wedge \neg B)$	biszubtrakció	$A \not\leftrightarrow B$
kontrárius	$\neg(A \wedge B)$	exklúzió	$A \not\wedge B$
szubaltern	$A \rightarrow B$	implikáció	$A \rightarrow B$

A különféle oppozíciók meghatározását természetesen nem a táblázatban látható modern képletekkel fejezték ki évszázadokon keresztül, hanem szöveges értelmezést adtak. Kontradiktórikus viszony esetén a két elem nem lehet egyszerre sem hamis, sem igaz, vagyis csak az lehetséges, hogy vagy az egyik igaz és a másik hamis vagy fordítva. A kontrárius ellentét azt jelenti, hogy a két összetevő nem lehet egyszerre igaz, de lehet egyszerre hamis (és az is lehetséges, hogy az egyik igaz és a másik hamis vagy fordítva). A szubkontrárius ellentét azt az esetet fejezi ki, amikor a két elem nem lehet egyszerre hamis, de lehet egyszerre is és külön-külön is igaz. Végül a szubaltern kapcsolat azt fejezi ki, hogy ha valami igaz, akkor annak a szubalternje is igaz (a szuperaltern pedig ennek a konverze).

A logikai négyzethez hozzátartozott még pár átalakítási szabály is (pl. az (**E**)- és (**I**)-sarokban a konverzió művelete is igaz állítást eredményez), de ezeket itt nem részletezem. A logikai négyzet, a maga ellentmondás-fogalmaival nagyon népszerű volt évszázadokon keresztül. Idővel kiterjesztették használatát más – a kvantifikációs kapcsolatok kezelésén túlmutató – célokra is. Erről még bővebben is írni fogok a későbbiekben. Előtte azonban meg kell említenem, hogy a logikai négyszög elemei közti kapcsolatokat a modern logika revízió alá vette, és – legalább részben – felülbírálták az ókor óta hangoztatott téziseket. Kiderült ugyanis, hogy az üres, illetve az univerzális terminusokkal (vagy osztályokkal), valamint az egzisztenciális import jelenségével problémák vannak – olyannyira, hogy e felfedezések óta sokan már csak a kontradiktórius kapcsolatot tartják érvényesnek [406]. A kritikák egyik vonulata úgy szólt [458], hogy ha az (**A**)-pontba egy univerzális osztályt, tehát egy mindig igaz logikai igazságot, azaz tautológiát teszünk ('Minden négyzet derékszögű.'), akkor arról tudjuk, hogy sosem lehet hamis. Egy tautológiával szemben azonban semmi sem állhat kontrárius viszonyban, hiszen a tautológia sosem hamis, viszont a kontrárius kapcsolat létezéséhez ezt kell feltételeznünk, hogy a két összetevő lehet egyszerre hamis. Ugyancsak ellentmondáshoz jutunk, ha az (**I**)-sarokba egy mindig hamis állítást, tehát egy kontradikciót teszünk ('Néhány négyzet köralakú'). Egy ellentmondás sosem lehet igaz, így nem kerülhet szubkontrárius kapcsolatba egy másik kijelentéssel, hiszen utóbbi tételezése azt kívánná meg, hogy a két kijelentés egyszerre legyen igaz. Gondok adódnak abban az esetben is, ha egy üres osztályt, üres terminust teszünk be az (**A**)-kijelentés szubjektumába [282]. Mivel egy üres terminusról (üres osztályról) minden állítható[18], így az (**A**)-mondat igaz. Az üres terminus esetén viszont nem lehet igaz az (**I**)-sarok, hiszen ha nincs az S terjedelme alatt semmi, akkor nem lehet igaz a 'Néhány S [van] P' kijelentés. Ekkor viszont az (**I**)-pont kontradiktórius párja az (**E**)-mondatnak igaznak kell lenni, amiből viszont az következne, hogy a vele szubaltern viszonyban álló (**O**)-pontnak is igazságot kellene tartalmaznia. Az (**O**)-sarokban viszont (**A**) igazságból következően – és a köztük levő kontradiktórius viszony miatt – hamis állításnak kellene szerepelni, így ellentmondáshoz jutunk. Az üres osztályok megjelenése es-

[18] Az egyetemes állítás mindig átalakítható a két kategóriát kondicionális művelettel összakapcsoló kijelentéssé (a 'Minden S [van] P' propozíció ekvivalens azzal, hogy 'Ha S, akkor P'), és ha egy kondicionális előtagja hamis (üres), akkor abból következik az utótag igazsága.

etén tehát felborul a logikai négyzet rendje. A gondok ellenére – ha a jelzett problémákra vonatkozó kizáró feltételekkel eliminájuk a logikai ellentmondásáok kelektkezését – a logikai négyszög a mai napig fennmaradt, és sok szempontból hasznosnak bizonyul(t).

A logikai négyzet eredeti formájában kvantifikációs állításokat tartalmazott, ezért is nevezték el idővel kvantifikációs négyzetnek. Azért kellett új nevet kitalálni, mert az évszázadok során másféle kijelentésekkel, kategóriákkal töltötték fel a logikai négyzet sarokpontjait, amelyek esetében ugyanúgy működni látszott a séma, az új négyszögek elemei és a köztük levő kapcsolatok ugyanazt a mintázatot mutatták, mint az eredeti négyzet. A modális logika kifejlődésével párhuzamosan az egyes modális logikák kulcskategóriáit is négyszögekbe rendezték.

	(A)	**(E)**	**(I)**	**(O)**
alethikus logika	szükségszerű	lehetetlen	lehetséges	nem-szükségszerű
deontikus logika	kötelező	tilos	megengedett	nem-kötelező

Sok mindenre használták még a logikai négyzetet, például alkalmazták a cselekvéselmélethez egészen közel áll fogalmakra is. Canterburyi Szent Anzelm olyan cselekvéstipológiát dolgozott ki, ami a logikai négyzetre épült. A cselekvéselmélet szempontjából vett fontossága miatt az anzelmiánus elméletet részletesebben bemutatom majd, de nem itt, hanem a cselekvsé fogalmáról szóló fejezetben (↪132.o.)

Logikai hatszög

A logikai négyzettel kapcsolatban felmerültek másfajta problémák is, ezek azonban sokkal inkább a továbbfejlesztési lehetőségek keresésére ösztönözték a kutatókat, mint a négyszög-gondolat feladására. A négyszög hatszögesítését Augustin Sesmat és Robert Blanché egymástól függetlenül, közel egyidőben végezte el az 1950-es években, de közel ötven éven keresztül eredményeiknek nem volt jelentős hatása a tudományos gondolkozásra. Csak a 2000-es években változott ez a helyzet [52]. Kiderült, hogy a logikai négyzet négy sarokpontja egyszerre volt kevés is és sok is a tényleges összefüggések pontos megragadására. Zavaró volt, hogy a hétköznapi és a tudományos-logikai nyelvhasználat több ponton eltért egymástól. Az **(I)**-sarokba tartozó kategóriákat sokszor az **(A)**-sarokba tartozó kategóriákkal ellentétesnek érezzük, szemben a logikai négyzet "állításával", mely szerint a két pont állításai között szubaltern viszony létezik, ami egyáltalán nem számít ellentétes kapcsolatnak (a megengedett vagy a lehetséges fogalmát sokszor sokkal inkább érezzük ellentétesnek, mint alárendeltnek a kötelező vagy a szükségszerű fogalmához képest). Magyarázatra szorult az a furcsa tény is, hogy az **(O)**-pontba került kategóriáknak szinte sosem volt megfelelő hétköznapi terminusa. A modális logikában például a nem-szükségszerű, nem-kötelező "mesterséges" terminusokat kellett használni a hiányzó természetes szavak hiányában. Az ilyen és ehhez hasonló problémák feloldhatónak látszottak azáltal, hogy két új pólust veszünk fel a négyzethez, és ezáltal olyan logikai hatszöghöz jutunk, amely két háromszögből áll. Első lépésként az "eredeti" logikai négyszög helyett érdemesebb egy olyan logikai háromszöget képezni, amelyben megmarad az **(A)**- és az **(E)**-sarok, és ezek mellé kell egy új pólust felvenni. Ezt az új pólust – Blanché után – **'Y'** karakterrel kezdték el jelölni. Az az új logikai háromszög nagyon hasznosnak ígérkezett. Egyre több olyan fogalomhármast kezdtek felfedezni a gondolkodás, a kategóriaképzés legkülönfélébb területein, amelyek mind egyidejű igazságukat

kölcsönösen kizáró, tehát kontrárius háromszöget alkottak egymással. Íme néhány példa:

	(A)	**(E)**	**(Y)**
kvantifikáció	minden	semmi	valami
tér	mindenhol	sehol	valahol
idő	mindig	soha	valamikor (néha)
alethikus modaitás	szükségszerű	lehetetlen	kontingens
deontikus modalitás	kötelező	tiltott	megengedett

A kontrárius háromszögek mellett természetesen képezhetők voltak a szubkontrárius háromszögek is. Ehhez új elemként az (**U**)-sarokpontot vették fel, ami az "eredeti négyzet" (**I**) és (**O**) pólusait kötötte össze. A két háromszög egymáshoz illesztésével pedig már felrajzolhatóvá vált az új minőség, a logikai hatszög, amely egyben tartalmazta az eredeti logikai négyszöget is, de benne volt még a két "új" logikai háromszög is. Kiderült az is, hogy az új pólusok felvételével a hatszög végpontjai mind kifejezhetőek a szomszédos sarokpontok segítségével.

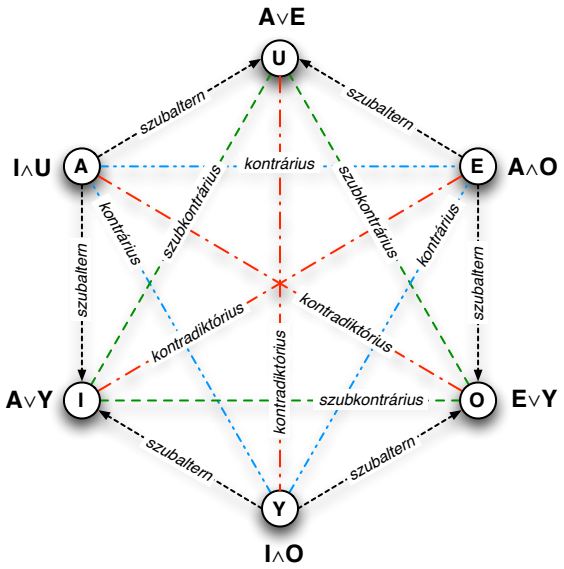
(GL.6) UNIVERSAL_PROPOSITION $\equiv$ **S(U)P** $\equiv$ **(U)** $\stackrel{\text{def}}{=}$ **(A)** $\vee$ **(E)**

egyetemes kijelentés (HUN) – universal proposition (ENG)

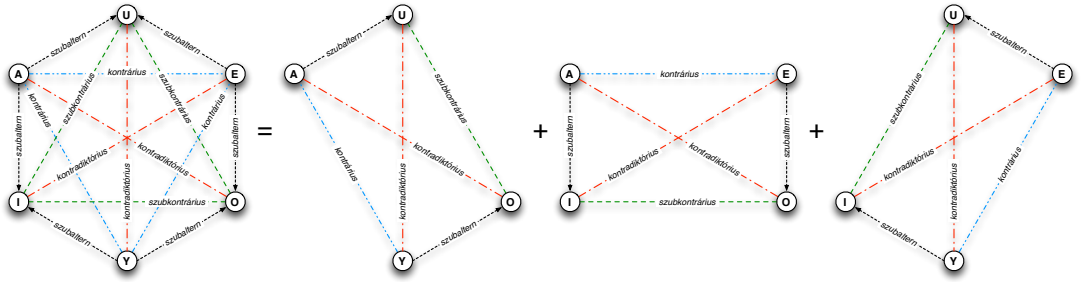
(GL.7) PARTICULAR_PROPOSITION $\equiv$ **S(Y)P** $\equiv$ **(Y)** $\stackrel{\text{def}}{=}$ **(I)** $\wedge$ **(O)**

partikuláris kijelentés (HUN) – particular proposition (ENG)

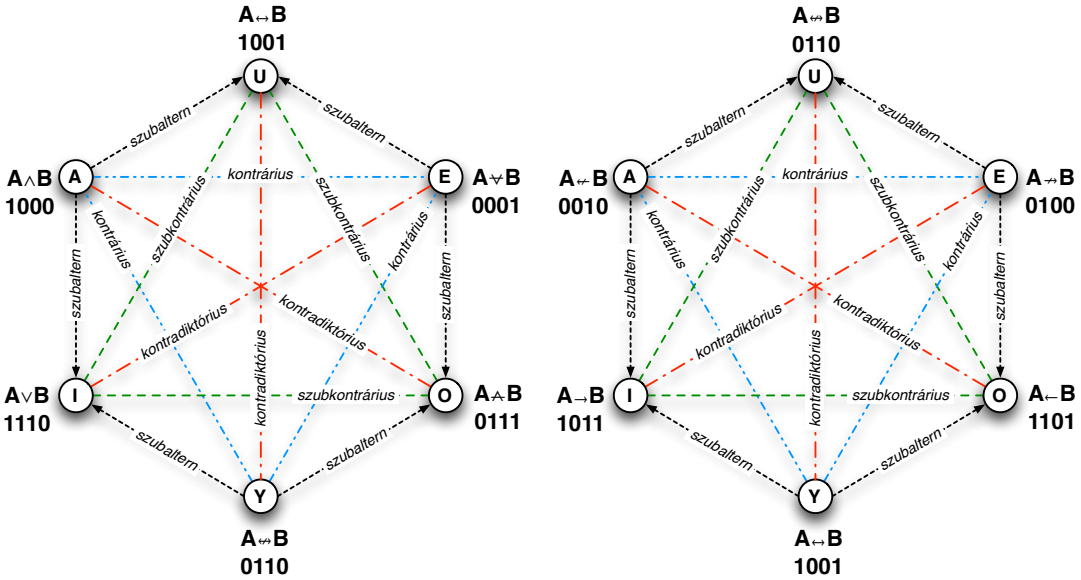
Az új elemek felvétele után felvetődik a kérdés, vajon milyen összefüggések léteznek az új (és régi) kategóriák között. Kiderült, hogy a már jelzett kapcsolatokon túl a szubaltern relációk még több sarokpont között behúzhatóak. A teljes kapcsolatrendszert mutatja a következő ábra.



A hatszög kapcsolatrendszerét felvázolva észrevehetünk még egy további rejtett összefüggést is. A hatszögbe nemcsak egy, hanem összesen három négyszög van “elrejtve”, és a két “új” négyszög ugyanolyan belső szerkezettel rendelkezik, mint az “eredeti” logikai négyszög. Az ellentétes pólusaikban kontradiktórius, a szomszédos sarokpontok közt kontrárius vagy szubkontrárius, valamint szubaltern viszonyokat találhatunk. A három négyszöget kiemeltem a hatszögből és eredeti helyzetüket megtartva mutatom be őket.



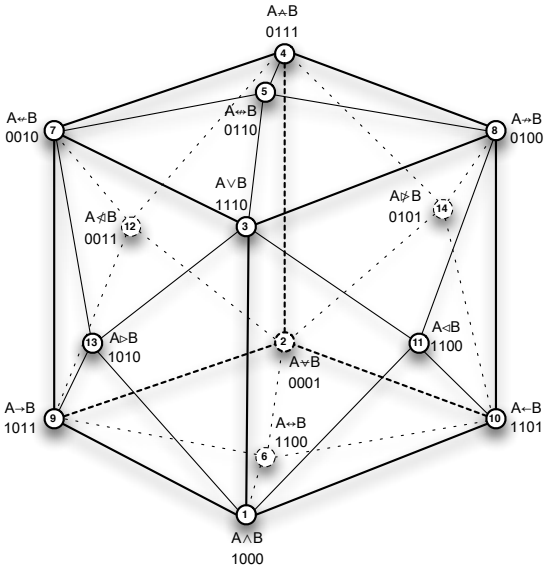
Idővel kiderült, hogy a logikai hatszög sok területen használható. A logikai operátorokat vizsgálva Blanché rájött arra, hogy a tíz legfontosabb operátor ábrázolható két logikai hatszög segítségével. A nullad- és az elsőrendű operátorok (a tautológia és ellentmondás, valamint a négyféle projekciós művelet) nélkül marad tíz műveletünk, ezek a “tisztán” másodrendű logikai operátorok, és ezeket rendezte Blanché egy összefüggő szerkezetbe.

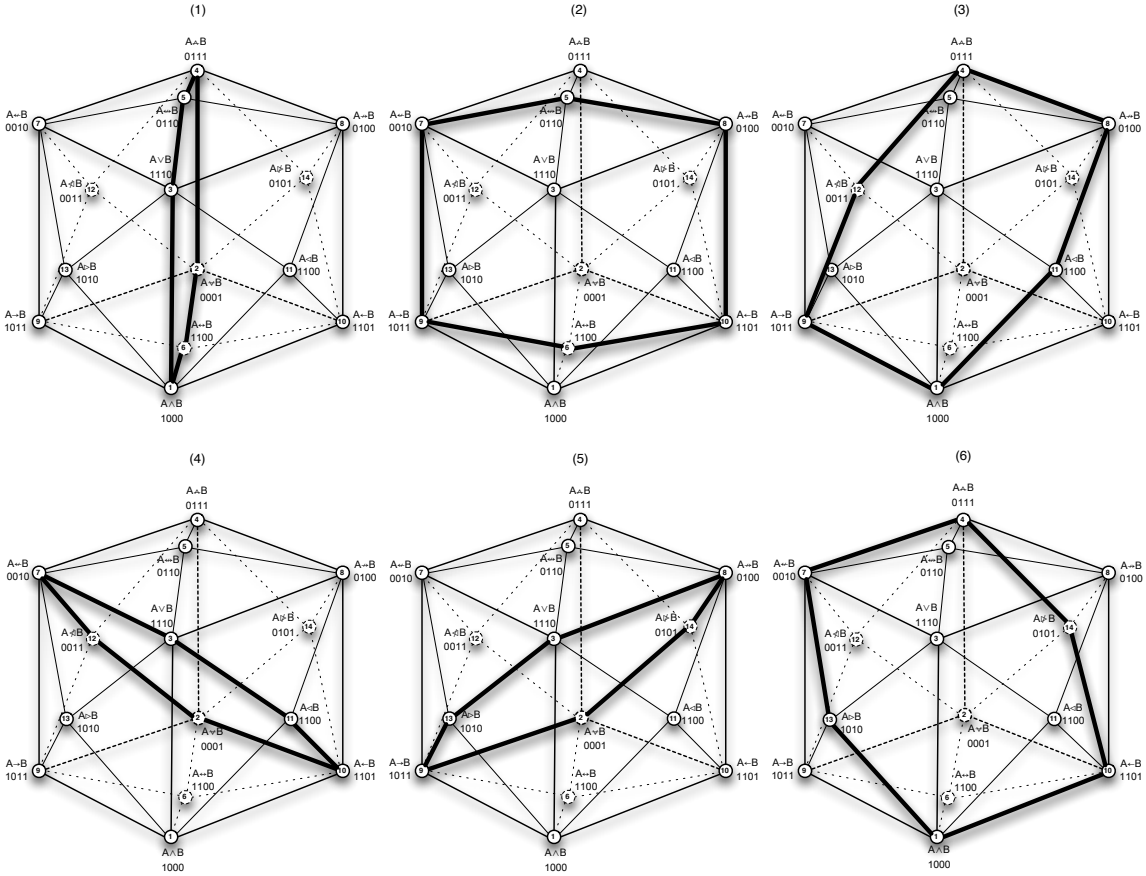


A geometriai szemlélet érvényesítését, további hasznosítását azonban itt már nem lehetett leállítani.[19] Amíg a logikai operátorok egymáshoz való viszonyát Blanché még úgy fejezte ki, hogy az operátorok közül tizet két hatszögbe foglalt bele, addig Sauriol már mind a tizenhat operátort egyetlen alakzatban tudta “megmutatni” egy tetrahexahedron segítségével. A tetrahexahedron egy kockából és hat darab piramisból álló geometriai forma, aminek a drótvázára hat darab hatszögből is összeállítható. Ezt meg lehet jeleníteni

[19] A logikai relációk háromdimenziós megjelenítésről lásd: [497].

háromdimenziós formában, és hogy könnyebben felismerhetők legyenek a benne “rejtőző” hatszögek, egyenként kiemelve őket megmutatom a helyzetüket is a tetrahexahedronon belül.





A logikai hatszögek, majd a további geometriai formák alkalmazhatósága sikere bizonyos értelemben "kijelölte" a továbblépési lehetőségeket a geometriai általánosítás irányába, és megjelentek azok a cikkek, amelyek logikai n-szögekről és a problémakör geometrizálásáról szóltak [363], [365]. Ezekkel azonban itt nem foglalkozom.

Közvetlen következtetések, átalakítási műveletek kategoriális kijelentések esetén

A logika négyszög kapcsán is sokat foglalkoztak a kategoriális kijelentések átalakítási lehetőségeivel. Azért volt mindez fontos, mert ezek olykor közvetlen következtetési szabályként is működtek [26, 895]. Előre fontos hangsúlyozni, hogy ezek az átalakítási szabályok nem mindig eredményeznek igaz állításból igaz állítást, tehát nem minden esetben működnek következtetési szabályként. Ahhoz, hogy mind a négy kategorikus állítástípusra egyszerre megfogalmazhassuk majd az átalakítási szabályokat, illetve az érvényes következtetési formulákat, érdemes bevezetni a **(C)** általános kategorikus propozíció fogalmát. Ebből persze egyből kettőt kell definiálnunk, hogy – szükség esetén – jelezni tudjuk, melyik esetről teszünk állítást, négy vagy hat típusról, a logikai négyszögről vagy a logikai hatszög elemeiről beszélünk.

(GL.8) $\text{GENERAL_C-PROPOSITION} \equiv \mathbf{S(C)}_4\mathbf{P} \equiv \mathbf{(C)}_4 \stackrel{\text{def}}{=} \mathbf{(A)} \underline{\vee} \mathbf{(E)} \underline{\vee} \mathbf{(O)} \underline{\vee} \mathbf{(I)}$

(GL.9) $\text{GENERAL_C-PROPOSITION} \equiv \mathbf{S(C)}_6\mathbf{P} \equiv \mathbf{(C)}_6 \stackrel{\text{def}}{=} \mathbf{(A)} \underline{\vee} \mathbf{(E)} \underline{\vee} \mathbf{(O)} \underline{\vee} \mathbf{(I)} \underline{\vee} \mathbf{(U)} \underline{\vee} \mathbf{(Y)}$

Az új fogalom bevezetésének csak annyi értelme van, hogy segítségével egyszerűsíthessünk a későbbi definíciókon, a fogalomnak önálló szemantikája nincs, csak egy belső mintázattal rendelkező változóként érdemes felfogni. A korábban már bevezetett átalakítási műveleteket itt – igaz, új értelmezési tartományon – újraértelmezhetjük. A konverzió meghatározásában a korábbiakhoz képest (↪38.o.) csak annyi változásra van itt szükség, hogy az értelmezési tartományt le kell szűkíteni a kategoriális kijelentések halmazára. A műveletet az általános kategoriális propozícióváltozó segítségével definiálhatjuk:

(GL.10) $\text{C-CONVERSION} \equiv \mathbf{S(C)}_4^{-1}\mathbf{P} \equiv \mathbf{(C)}_4^{-1} \stackrel{\text{def}}{=} \forall m\big(\mathbf{S(C)}_4^{-1}\mathbf{P} \leftrightarrow \mathbf{P(C)}_4\mathbf{S}\big)$

kategoriális konverzió (HUN) – category conversion (ENG)

A kategoriális konverzió olyan konverzió, amelynek értelmezési tartománya a kategoriális kijelentések halmaza. A kategoriális konverzió praktikusan annyit tesz, hogy a kijelentésben felcseréljük a szubjektumot és a predikátumot. Ha a tradicionális oppozíciós négyzet **(E)**-propozíciója ('Egyetlen S sem P.') igaz, akkor annak megfordítása ('Egyetlen P sem S.') is igaz, tehát ez a konverzió közvetlen következtetési szabályként is alkalmazható. Ugyanúgy elmondható ez az **(I)**-propozícióra is ('Néhány S [van] P.'), mert annak a megfordítása ('Néhány P [van] S.') megintcsak igaz. A négyzet másik két állítása is megfordítható, de ezekben az esetekben nem áll fenn a konverz kijelentések igazsága.

Az átalakítási műveletek korábbi tárgyalásakor nem mutattam be az **obverzió** műveletét, mert ez igazából csak a kategoriális kijelentések esetében értelmezhető. Az új művelet bevezetéséhez szükség van a kategóriák (relációk) komplementumaira is, és a kategoriális predikátum tagadására.

(GL.11) $\text{OBVERSION} \equiv \mathbf{S(C)}_4^{-o}\mathbf{P} \equiv \mathbf{(C)}_4^{-o} \stackrel{\text{def}}{=} \forall m\big(\mathbf{S(C)}_4^{-o}\mathbf{P} \leftrightarrow \overline{\mathbf{S}}\mathbf{(C)}_4\neg\overline{\mathbf{P}}\big)$

obverzió (HUN) – átalakítás[20] (HUN) – category obversion (ENG)

Az obverzió olyan átalakítási művelet, ami megtartja a szubjektum és predikátum sorrendjét, de kicseréli őket a komplementumaira, valamint a predikátum komplementumát tagadva állítja. A propozicionális logikában az obverzió műveletét nem használják. Vélhetőleg azért nem, mert az állításlogikában hiányzik a pozitív és negatív (állító és tagadó) kijelentések megkülönböztetése, ami viszont a kategoriális logikában adott.

A kategoriális kijelentések színterén is értelemezhető a kontrapozíció művelete, ami itt kifejezhető a konverzió és obverzió segítségével is. Ezt azonban akkor lehet jól szemléltetni, ha használjuk a részleges kontrapozíció műveletét.

(GL.12) $\text{PARTIAL_C-CONTRAPOSITION} \equiv \mathbf{S(C)}_4^{-r}\mathbf{P} \equiv \mathbf{(C)}_4^{-r} \stackrel{\text{def}}{=} \forall m\big(\mathbf{S(C)}_4^{-r}\mathbf{P} \leftrightarrow \neg\overline{\mathbf{P}}\mathbf{(C)}_4\overline{\mathbf{S}}\big)$

részleges kategoriális kontrapozíció (HUN) – partial category contraposition (ENG)

A részleges kategoriális kontrapozíció az a művelet, amely egyszer-egyszer alkalmazza az obverzió és a konverzió műveletét. A művelet "képlete" tehát egyszerű: részleges kontrapozíció = átalakítás + megfordítás.

Ide valami értékelés kell még. 'Minden bogár rovar.'

(GL.13) $\text{C-CONTRAPOSITION} \equiv \mathbf{S(C)}_4^{-p}\mathbf{P} \equiv \mathbf{(C)}_4^{-p} \stackrel{\text{def}}{=} \forall m\big(\mathbf{S(C)}_4^{-p}\mathbf{P} \leftrightarrow \neg\mathbf{P}\mathbf{(C)}_4\neg\mathbf{S}\big)$

teljes kategoriális kontrapozíció (HUN) – kategoriális kontrapozíció (HUN) – categoric contraposition (ENG)

A teljes kategoriális kontrapozíció olyan kontrapozíció, amelynek az értelmezési tartománya a kategoriális kijelentések halmaza. A kategoriális kontrapozíció praktikusan annyit tesz, hogy megcseréli a kijelentés szubjektumának és predikátumának komplementumait. A művelet felírható a következő alakban is: kategoriális kontrapozíció = átalakítás + megfordítás + átalakítás.

(GL.14) $\big(\mathbf{(C)}_4 = \mathbf{(A)} \veebar \mathbf{(C)}_4 = \mathbf{(O)}\big) \rightarrow \big(\text{C-CONTRAPOSITION} \leftrightarrow \text{OBVERSION}\big(\text{C-CONVERSION}\big(\text{OBVERSION}\big)\big)\big)$

A kondicionális alak azt jelzi, hogy a fenti formula nem minden esetben igaz: a négyből csak két esetben. Ha az (**A**)-propozíció igaz, akkor annak kategoriális kontraponáltja is igaz, tehát ez az átalakítás használható közvetlen következtetési szabályként. Ugyanez áll az (**O**)-propozícióra is, hiszen a kontraponáltja igaz.

A relációműveletek között szerepelt az inverzió is. Ezt is definiálhatjuk itt, bár a jelentősége, hasznossága kisebb a többi átalakítási szabályhoz képest. A definíciót így adhatjuk meg:

(GL.15) $\text{C-INVERSION} \equiv \mathbf{S(C)}_4^{-i}\mathbf{P} \equiv \mathbf{(C)}_4^{-i} \stackrel{\text{def}}{=} \forall m\big(\mathbf{S(C)}_4^{-i}\mathbf{P} \leftrightarrow \overline{\mathbf{S}}\mathbf{(C)}_4\overline{\mathbf{P}}\big)$

kategoriális inverzió (HUN) – category inversion (ENG)

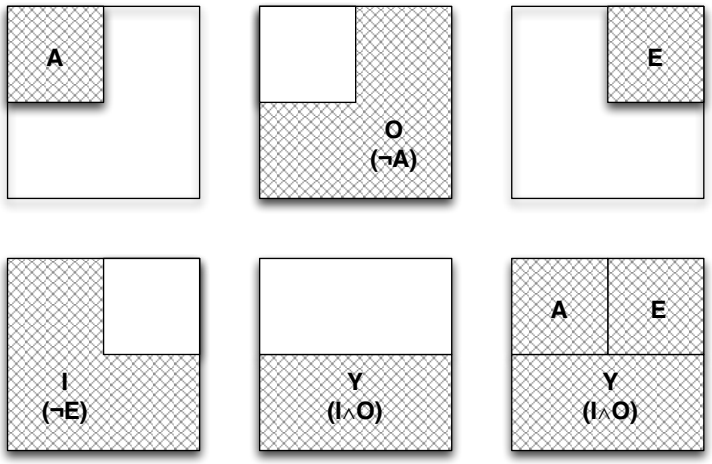
[20] Az átalakítás terminus használata zavaró lehet, hiszen az átalakítást mint obverzió az átalakítási műveletek egyik típusaként értelmezve részleges, de mindenképpen inkonzisztens homonímához jutunk. Ha a kontextusból egyértelműsíthető, hogy az átalakítás terminusát az obverzió szinonimájaként kell érteni, akkor persze használni lehet ezt a terminust is.

A kategoriális inverzió az a művelet, amelynek az értelmezési tartománya a kategoriális kijelentések halmaza. A kategoriális inverzió kicseréli a kijelentés két kategóriáját azok komplementjeire, de megtartja a sorrendjüket. Az inverziónak nincs direkt következtetési potenciálja.

JEPD-elv

Amikor elméleti modellt építünk, akkor a generikus alárendeltje reláció segítségével képzünk új fogalmakat úgy, hogy egy már definiált fogalom alá egy tulajdonságnyaláb segítségével szűkebb terjedelmű, viszont bővebb jelentésű alosztály-fogalmakat rendelünk. Az alosztály (altípus, faj) fogalmak képzésekor sokdszor érdemes törekednünk arra, hogy igazodjunk a JEPD-elv elvárásaihoz [59]. Ez az elv azt rögzíti, hogy egy fogalom alá rendelt fogalmak **együttesen kimerítőek** legyenek és **kölcsönösen kizárják egymást** (JEPD – Jointly Exhaustive and Pairwise Disjoint).[21] A generikus struktúra konzisztenciájához nélkülözhetetlen vagy legalábbis fontos, hogy az azonos fogalmi szinten egymás mellett levő, egy közös fölérendelt kategóriához tartozó fogalmak tartalma kölcsönösen kizáró és együttesen kimerítő legyen. Ezáltal elégíthetjük ki azt az elvárást, hogy egy adott fogalom alá kidolgozott fogalmak együttesen a fogalom terjedelmébe tartozó minden előfordulást egyértelműen lefedjenek.

A matematika világából a JEPD-elvnek megfeleltethető egy osztály alatt létrehozott alosztályokból történő **partíció** képzése. Ebben az esetben is az az elvárás, hogy az alosztályok egyetlen példányt tekintve se fedjék át egymást (vagyis az alosztályok partíciók legyenek), illetve az alosztályok terjedelmükkel teljes mértékben lefedjék a fölöttes osztályfogalom (szülőosztály) terjedelmét (vagyis a partíció teljes legyen). Ez ugyanazt jelenti, mint amit a halmazelméletben a **partíció** (partition) fogalma a halmaz és részhalmazai közti viszonyban (↰12.o.). Ha egy fogalom alá olyan generikusan alárendelt fogalmakat hozunk létre, amelyek kielégítik a JEPD-elv elvárásait, akkor azt mondhatjuk, hogy a fogalmat particináltuk, **partíciókra** bontottuk fel. A JPED-elv érvényesül például a logikai hatszög azon kategóriái között, amelyek kontrárius viszonyban állnak egymással, vagyis az (**A**), az (**E**) és az (**Y**) kategóriák között. Nem teljesül viszont a logikai négyszög elemeire.



[21] Angolul szokták még a ‘mutually incompatible and jointly exhaustive’ kifejezést is használni.

A logikai hatszög bevezetése ezért is fontos, mert ezáltal olyan kategóriafelosztáshoz juthatunk, amely partíciónak minősül, vagyis érvényes rá a JEPD-elv. A későbbiek során, amikor csak lehet, alkalmazni fogom a JEPD-elvet. Az elméleti építkezés során mindig fontos annak tisztázása, vajon a fogalmi modell bővítésekor mindent figyelembe vettünk-e. Ha ilyenkor az új fogalmakra vonatkozóan állítani lehet a JEPD-elv teljesülését, akkor az éppen adott szempont szerint teljesnek mondható az elméleti keret bővítése.

Reifikáció

Az egyértelműséggel és konzisztenciával kapcsolatos elvárások miatt említést kell tennem még egy nagyon fontos jelenségről. A relációk a világ dolgai közti kapcsolatokat fejezik ki. Ez az jelenti, hogy – bináris relációk esetén – egy relációt jelző terminussal mindig két dolog, két relátum közti viszonyt ragadunk meg. A relációfogalmak bevezetése után azonban a fogalom (terminus) használata során előfordulhat, hogy ezt a viszonyfogalmat "tárgyiasítjuk" és valamelyik relátummal megragadott "dologra" vonatkoztatjuk a fogalmat (terminust). Egy hétköznapi példával szemléltetem ezt a jelenséget. Vegyük az 'apja' relációt (terminust), ami egy közvetlen biológiai leszármazási viszonyt ír le két ember között, amikor az "örökítő" pozícióban levő egyed hímnenű. Az 'apja' tehát viszonyt fejez ki. A kommunikáció során azonban nagyon gyakran úgy használjuk ezt a fogalmat, hogy sokkal inkább arra a dologra, arra az emberre gondolunk, aki a kapcsolat egyik felét testesíti meg. Magyarul ekkor az 'apa' terminust használjuk. Amikor így teszünk, akkor reifikáljuk a relációfogalmat, a relációból kiemeljük az egyik relátumot és csak erre hivatkozunk.[22] Természetesen nincs semmi probléma ezzel a jelenséggel, csak arra kell figyelni, hogy mindig tudjuk, éppen a relációfogalmat magát vagy épp annak származtatott változatát, a reifikátumát használjuk. Azt következményt viszont fontos figyelembe venni, hogy amikor reifikálunk egy relációt, akkor információt vesztünk, hiszen a reifikátum már nem mond semmit a reláció másik relátumáról. A reláció másik relátumát általában a reláció inverze segítségével tudjuk könnyen reifikálni.

állítás	reláció	reifikátum	inverz reláció	inverz reifikátum
Laci apja Katinak.	$apja(Laci, Kati)$	$apa(Laci)$	$gyereke(Kati, Laci)$	$gyerek(Kati)$
Laci apja Lilinek.	$apja(Laci, Lili)$	$apa(Laci)$	$gyereke(Lili, Laci)$	$gyerek(Lili)$
Zoli apja Petinek.	$apja(Zoli, Peti)$	$apa(Zoli)$	$gyereke(Peti, Zoli)$	$gyerek(Zoli)$

A bemutatott példák alapján már meghatározhatjuk a reifikáció fogalmát. Mivel egy – bináris – relációt két irányból is tárgyiasíthatunk, ezért a reifikáció műveletét kétféleképpen is definiálhatjuk, így beszélhetünk bal- és jobboldali reifikációról.

(GL.16) $\text{LEFT-REIFICATION} \equiv \text{REIFICATION} \equiv \text{`}\mathbf{R}(m,n) \equiv \text{`}\mathbf{R} \equiv \mathbf{R}' \stackrel{\text{def}}{=} \forall m \forall n(\text{`}\mathbf{R}(m,n) \leftrightarrow \mathbf{R}(m))$

bal-reifikáció (HUN) – reifikáció (HUN) – reification (ENG)

(GL.17) $\text{RIGHT-REIFICATION} \equiv \text{REIFICATION} \equiv \mathbf{R}'(m,n) \equiv \mathbf{R}' \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{R}'(m,n) \leftrightarrow \mathbf{R}(n))$

jobb-reifikáció (HUN) – reifikáció (HUN) – reification (ENG)

A reifikáció művelete – a dolgokat absztraktan szemlélve – hasonlít a propozicionális logika projekció műveleteihez, hiszen itt is, ott is arról van szó, hogy két valamilyen

[22] Ez a művelet olyan, mint a logikai műveletek között a projekció (↳18.o.).

módon összekapcsolt elem közül a művelet végrehajtása során (eredményeként) csak az egyiket vesszük figyelembe. A reifikáció műveletét még tovább absztrahálhatjuk, és tetszőlegse absztrakt individuumokat létrehozhatunk, ami a tudásreprezentáció számára nagyon praktikus előnyökkel szolgálhat [72, 41]. Donald Davidson az események logikai reprezentálhatóságát keresve tett egy javaslatot arra, hogy miként lehetne az eseményeket megfelelő módon logikai formulákkal leírni [119]. A davidsoni megoldásnak van egy reifikációs vonzata is, ami az események, illetve az arra építhető cselekvések logikai leírásában nagyon jól hasznosítható. Jerry Hobbs 'Ontológiai promiszkuitás' címen írott cikkében olyan reifikációs leíró nyelvre tett javaslatot, ami Davidson ötletét vitte tovább, de annál nagyobb rugalmasságot, jobb kezelhetőséget biztosított a jelenségek logikai leírása számára [240]. A davidsoni-hobbsi ajánlásokat én is megpróbálom figyelembe venni az elemzett jelenségek formális leírása során. A reifikáció itteni értelmezése nagyon eltér az olyan társadalomtudományi interpretációktól, mint a Berger-Luckman-féle reifikációfelfogás vagy a Marx-féle tárgyiasulás/tárgyiasítás fogalomhasználat.

1.7 Másodrendű logika

Abban a pillanatban, hogy a logikai nyelvünkbe felvesszük a változókat és kvantorokat, lehetővé válik, hogy egy új szempont szerint különíthessünk el különböző logikákat, logikai szinteket. Nem véletlenül nevezik a kijelentéslogikát nulladrendű extenzionális logikának, a predikátumlogikát pedig elsőrendű logikának [454]. Mivel a kvantálás, a kvantorok alkalmazásának lehetősége a predikátumlogikában jelenik meg először, ezért ez a logikát (és ezt a logikai szintet) nevezzük elsőrendűnek. Elsőrendű logikákban a kvantorokat az individuumok valamely csoportját jelölő változókra lehet alkalmazni, semmi másra. Ha egyáltalán nem használhatunk kvantorokat (mert az adott logikában ez nincs is implementálva), akkor nulladrendű nyelvről beszélhetünk. Az ítéletlogikában nincsenek kvantorok, ezért az a logika nulladrendű.

A másodrendű logikák felé úgy lehet elmozdulni, hogy kiterjesztjük a változók, illetve a kvantorok hatókörét. Ezt úgy tehetünk meg, hogy egyfelől megengedjük, hogy változókat ne csak individuumokra, hanem predikátumokra (relációkra, függvényekre) vonatkozóan is lehessen képezni, másfelől megengedjük azt is, hogy a kvantorokat az individuumváltozókon túl a predikátumokra irányuló változókra is lehessen alkalmazni [162, 238], [454, 197]. A formulák szintjén ez úgy ismerhető fel, hogy az individuumváltozók mellett a predikátumváltozókra is érvényesítjük a formulában használt kvantorokat, valahogy így:

$$\forall m \forall P \forall Q(P(x) \rightarrow Q(x))$$

Ez a formula a **generikus alárendeltje** (GENERIC SUBORDINATE) relációt definiálja, amely a predikátumok, fogalmak közti alá-fölérendeltségi viszonyt (vagy másként a faja-neme, ISA relációt) írja le.

Ha a másodrendű logikában megengedett, hogy a predikátumokra, kategóriákra is alkalmazzuk a változókat és kvantorokat, akkor ez azt is jelenti egyben, hogy a fogalmi szinten is képezhetünk másodrendű fogalmakat. Ezek formálisan úgy jeleníthetőek meg, hogy az adott predikátum (fogalom) változóhelyeire, a predikátum paraméterei közé olyan predikátumot (fogalmat) írhatunk be, amely szerepel az elsőrendű predikátumok (fogalmak) között. Ez a beágyazás egyértelműen jelzi a predikátum (fogalom) másodrendű

minőségét. Ha vesszük a 'tudja, hogy' predikátumot, aminek két változóhelyet tartunk fent, amikor jelölni akarjuk, hogy 'ki az, aki tud valamit', illetve 'mit tud az a valaki'. Ezt a predikátumot így definiálhatjuk:

$\mathsf{TUD}(x,p)$ – *x tudja, hogy p*

Előfordulhat azonban az is, hogy a p paraméterre újra a 'tudja hogy' predikátumot alkalmazzuk, amikor például azt mondjuk, hogy:

$p = \mathsf{TUD}(y,q)$ – *y tudja, hogy q*

Ha a p tudásra vonatkozó második mondatunkat behelyettesítjük az első mondatba, akkor ilyen formulát kapunk:

$\mathsf{TUD}(x,p) = \mathsf{TUD}(x,\mathsf{TUD}(y,q))$ – *x tudja, hogy y tudja, hogy q*

Ezen a ponton szépen látszik a predikátum önmagába való beágyazottsága. Amikor ilyet tapasztalunk, akkor mindig másodrendű fogalomról van szó, aminek leírásához már másodrendű logikára van szükség. Ilyen esetekben ugyanis ha általános kijelentéseket akarunk tenni (az univerzális kvantor használatával), akkor meg kell engednünk, hogy a változókat és kvantorokat predikátumra is lehessen alkalmazni. A hétköznapi kommunikációban és a társadalomtudományi diskurzusban egyaránt rengeteg másodrendű fogalmat használunk, igaz, reflektálatlanul. A későbbiekben mindig jelezni fogom, amikor másodrendű fogalmakhoz jutunk valamely téma tárgyalása során. Teszem ezt egyfelől a fogalmi egzaktság érdekében, de másfelől teszem ezt azért is, mert a másodrendű fogalmak használata könnyen paradoxonhoz vezet.

Ha már beszélünk másodrendű logikáról, akkor természetesen beszélhetünk magasabbrendű (harmadrendű, negyedrendű stb.) logikáról is. Ezeket a másodrendű logikához hasonlóan lehet képezni. Minél magasabbrendű fogalmat képzünk, annál inkább egyre mélyebben egymásba ágyazott fogalmakhoz jutunk. A cselekvéselméleten belül gyakran kell másodrendű fogalmakat használnunk, jóval ritkábban, de azért felbukkannak harmadrendű fogalmak, viszont harmadrendnél magasabb szintre már nem nagyon kell felmennünk. A végtelen egymásba ágyazás, a tetszőlegesen magasabb rendű fogalmak elvi lehetőségét nem használjuk ki a tényleges fogalomhasználati gyakorlatunkban.

1.8 Modális logika

A társadalmi cselekvés modelljének felépítése során több alkalommal is szükségünk lesz olyan elméletekre, amelyek a deontikus logika területéhez sorolhatók (már a cselekvés fogalmának értelmezéséhez is szükségünk van erre a diszciplinára). A deontikus logikát a modális logika egyik ágaként kezeljük. A modális logika nem elsősorban a kijelentések igazságának kérdéseivel, hanem inkább a kijelentések igazságának modalitásával (minőségével, "erősségével", megalapozottságával) foglalkozik. Szükségszerű-e egy kijelentés igazsága; mindig fent kell-e tartanunk a kijelentés igazságát vagy azt csak bizonyos feltételek, adott világ, világállapot létezéséhez tudjuk kötni; van-e olyan kijelentés, ami sosem lehet igaz, tehát lehetetlen; melyek azok a kijelentések, amelyek lehetségesen igazak, de nem szükségszerűek; ha valami szűkségszerűen igaz, akkor igaz is-e egyben – ilyen és ezekhez hasonló kérdésekre próbál meg válaszokat találni a modális

logika.[23] Mivel sokféleképpen módosíthatjuk az állításainkat, sokféle modalitást kezelhetünk. Egy A kijelentés kapcsán mondhatjuk azt, hogy ‘ tudom, hogy A’, ‘azt hiszem, hogy A’, ‘gyakran A’, ‘biztos, hogy A’, ‘tilos A’, ‘szükségszerű, hogy A’ stb. Ilyen esetekben az állításunkat (A-t) valamilyen modalitásértékkel “kiegészítjük”, módosítjuk, amit úgy is kifejezhetünk, hogy az állításhoz egy modális operátort kapcsolunk hozzá. Amikor az állításaink szükségszerűségét vizsgáljuk, akkor alethikus modalitásról van szó. A normativitás, a konstruált kötelezőség világát a deontikus, a kijelentések időbeliségét a temporális, a hitek logikáját a doxasztikus, a tudásokét az episztemikus modalitással próbáljuk megragadni.

A modális logika nyelvébe fel kell vennünk a modális operátorokat és egy olyan relációt, amely a lehetséges világok közti elérhetőségi kapcsolatot képezi le. A különböző modális logikákat azon az alapon különíthetjük meg egymástól, hogy feltételezzük, sokféle világ létezhet, melyek közül némelyikben igazak bizonyos állítások, másokban viszont hamisak, megint más állítások minden lehetséges világban fennállnak, míg vannak olyanok is, amelyek sehol sem igazak. Az elérhetőségi reláció azt a tényt fejezi ki, hogy adott világból szemlélve mely más világok érhetők el (aminek inverze írja le azt, hogy melyek nem). Ennek az elérhetőségi relációnak fontos szerepe van a modális logika tipizálásában.

A lehetséges világok közti elérhetőségi relációra vonatkozóan különböző relációtulajdonságokat lehet megállapítani, és ezek a tulajdonságok egyben jellemzik a különböző modális logikákat is. Egy-egy ilyen relációtulajdonsággal egy-egy kalkulust definiálhatunk, amely az adott modális logika ‘szükségszerű’ (‘erős modálisan igaz’), illetve ‘lehetséges’ (‘gyenge modálisan igaz’) fogalmára vonatkozik. A reflexivitás tulajdonságával például az ún. (T) kalkulust határozhatjuk meg, amelyre a következő formula írható fel:

$$\Box A \to A \quad \text{(T)}$$

A (T) formulát úgy interpretálhatjuk, hogy ‘ami szükségszerűen igaz, az igaz’. Első “nekifutásra” ez kijelentés magától értetődőnek tűnhet, pedig könnyen lehet olyan értelmezést találni, amikor ez az állítás nem igaz. Jobban meg lehet érteni mindezt, ha az előző mondatot átfogalmazzuk a következő alakra: ‘ez így van, mert így kell lennie’. Ha a ‘kell’ fogalmát egy természettörvényre hivatkozva értelmezzük, akkor a “kellés” (szükségszerűség) valóban maga után vonja az állítás igazságát is: ha feldobunk egy követ, akkor annak – a tömegvonzás törvénye miatt – le kell esnie, és, valóban, a földobott kő le is esik a földre. Ha azonban a ‘kellés’ fogalma alatt a kötelezőséget értjük, akkor közel sem annyira magától értetődő, hogy ami kell (kötelező), annak meg is kell valósulnia, hiszen rengeteg esetet ismerünk, amikor valamilyen kötelező parancsba foglalt állítás a valóságban nem teljesült.

Egy közlekedési példával szemléltetve ezt a különbséget tegyük fel, hogy éppen azt meséljük egy barátunknak, hogy egy közös ismerősünket az úttesten elütötte egy jobbról érkező autó, amire a barátunk ennyit mond:

> Jobbról nem is jöhetett autó azon a helyen.

[23] A modális logika két gyakran hivatkozott forrása Brian Farrell Chellas [92] és G.E. Hughes és M.J. Cresswell műve [245], én Patrick Blackburn, Maarten de Rijke és Yde Venema könyvére [61], valamint Robert A. Bull és Krister Segerberg tanulmányára [86] támaszkodom leginkább.

Ez a mondat kétféleképpen is értelmezhető. Nem jöhetett jobbról autó, mert *(i)* a gázolás egy éles kanyarban történt, és a baleset színhelyétől jobbra házfalak meredeznek, vagy *(ii)* egy egyenes, ám – balról jobbra – egyirányú utcában jobbról nem érkezhetett semmi a kötelező haladási iránnyal szemben. Kicsit másként: nem jöhetett jobbról az autó, mert az fizikailag lehetetlen lett volna, illetve nem jöhetett jobbról az autó, mert a tiltó jogszabály miatt nem lett volna szabad. A fenti megnyilatkozást átfogalmazhatjuk a következő formába:

Nem lehetséges, hogy az adott helyen jobbról jöjjön az autó.

Ebben az esetben ugyanúgy fennáll a kettős jelentés lehetősége, csak ekkor a 'nem lehetséges' kifejezésnek adhatunk kétféle értelmezést: *(i)* 'nem lehetséges, mert fizikailag lehetetlen', illetve *(ii)* 'nem lehetséges, mert nem megengedett'. A 'lehetséges' kétféle interpretációja teljesen eltérő magyarázósémákat és különböző modális logikákat eredményez. Mielőtt bemutatnám, hogy miként is határozhatjuk meg a legfontosabb modális kalkulusokat (modális logikákat), át kell tekintenem a modális logika nyelvnek elemeit.

A modális logikai nyelv elemei

Szükség van tehát egy modális logikai nyelvre, ami a következőből áll [61].

(ML.0) $\mathscr{L}_{ML}(p, q, r, A, B, C, \bot, \neg, \vee, \Diamond)$, ahol

$p, q, r \in \Phi$	propozíció
A, B, C	formula
$\bot$	sosem igaz (ellentmondás)
$\neg$	tagadás logikai művelet
$\vee$	diszjunkció logikai művelet
$\Diamond$	lehetséges modális logikai operátor

A modális nyelv jól-formált formulái a következő induktív felsorolással adhatók meg (ahol az egyes opciókat a '|' jel választja el).

(ML.1) $A \stackrel{\text{def}}{=} p \mid \bot \mid \neg B \mid B \vee C \mid \Diamond B$

jól-formált formula (HUN) – well-formed formula (ENG)

A felsorolást úgy kell értelmezni, hogy *(i)* jól formált formula bármely propozíció vagy az *(ii)* ellentmondás; *(iii)* ha van egy jól-formált formulánk, akkor jól-formált marad, ha tagadjuk; *(iv)* ha van két jól-formált formulánk, akkor ezek diszjunkciója is jól-formált; *(v)* ha egy jól-formált formulát összekapcsoljuk a lehetséges operátorral, akkor jól-formált formulához jutunk. Ezeket formulákat természetesen tetszőleges formában és mértékben egymásba ágyazhatjuk, így a többi ismert logikai műveletet is bevezethetjük a nyelvbe.

A modális logika nyelvében két modális operátort használunk, hogy kifejezhessük velük a szükségszerűség és lehetségesség fogalmait. A két új operátor között természetesen szigorú összefüggés létezik, egyik a másikból levezethető. A különböző iskolák, szerzők különböznek abban, hogy melyik kategóriát veszik fel definiálatlan alapfogalomként (és ebből következően melyiket vezetik be a másik segítségével). Én a 'lehetséges' kategóriáját vettem fel alapként, a szükségszerű fogalmát a fent ismertetett módszer segítségével már definiálni lehet.

(ML.2) $\text{POSSIBLE}(A) \equiv \Diamond(A) \equiv \Diamond A$

lehetséges (HUN) – gyenge modálisan igaz (HUN) – gyémánt (HUN) – possible (ENG) – diamond (ENG)

(ML.3) $\text{NECESSARY}(A) \equiv \Box(A) \equiv \Box A$

szükségszerű (HUN) – erős modálisan igaz (HUN) – necessary (ENG) – box (ENG)

(ML.4) $(\text{DUAL}) \stackrel{\text{def}}{=} \Box A \leftrightarrow \neg \Diamond \neg A$

(Dual)-séma (HUN) – (Dual)-schema (ENG)

A modális logika következő fontos fogalma a relációs struktúra (↪36.o.), amit azonban itt másként, **keretnek**, neveznek. A keret segítségével léphetünk tovább következő fontos modális kategóriára, a **modell** fogalmára. Ez felfogható egy másodrendű relációstruktúrának, amiben egy keret és egy reláció van definiálva. A modellhez szükséges reláció egy $\mathbf{V}$ értékelési (valuation) függvény, amely a $A \in \Phi$ propozíciókat rendeli hozzá a $\mathbb{W}$ halmaz elemeihez. Ezt a három új fogalmat fel kell vennünk az elméletünkbe.

(ML.5) $\text{FRAME} \equiv \mathfrak{F} \stackrel{\text{def}}{=} (\mathbb{W}, \mathbf{R})$

keret (HUN) – frame (ENG)

(ML.6) $\text{VALUATION} \stackrel{\text{def}}{=} \mathbf{V} : \Phi \mapsto \mathbf{V}(A) \subseteq \mathbb{W}, p \in \Phi$

értékelés(i függvény) (HUN) – valuation (function) ((ENG))

(ML.7) $\text{MODEL} \equiv \mathfrak{M} \stackrel{\text{def}}{=} (\mathfrak{F}, \mathbf{V})$

modell (HUN) – model ((ENG))

A modell kategóriája azért (is) fontos, mert vele lehet az **igazság** vagy kielégíthetőség fogalmait értelmezni, a keret pedig az **érvényesség** meghatározásához szükséges. Nézzük meg előbb, hogy lehet az igazság fogalmát meghatározni. Feltételezzük, hogy ϕ, ψ tetszőleges formulák, $A \in \Phi$ adott propozíció, $w, u \in \mathbb{W}$ adott állapotok ($\mathbb{W}$ elemei). Ekkor az ϕ, ψ formulák igazságát induktív felsorolással lehet meghatározni az alábbi módon.[24]

(ML.8) $\text{SATISFIED} \equiv \models \stackrel{\text{def}}{=}$

1) $\mathfrak{M}, w \models A \stackrel{\text{def}}{\Leftrightarrow} w \in \mathbf{V}(A), p \in \Phi$
2) $\mathfrak{M}, w \stackrel{\text{def}}{\neg\models} \bot$
3) $\mathfrak{M}, w \models \neg\phi \stackrel{\text{def}}{\Leftrightarrow} \mathfrak{M}, w \stackrel{\text{def}}{\neg\models} \phi$
4) $\mathfrak{M}, w \models \phi \vee \psi \stackrel{\text{def}}{\Leftrightarrow} \mathfrak{M}, w \models \phi \stackrel{\text{def}}{\vee} \mathfrak{M}, w \models \psi$
5) $\mathfrak{M}, w \models \Diamond\phi \stackrel{\text{def}}{\Leftrightarrow} \stackrel{\text{def}}{\exists} v \in \mathbb{W}(\mathbf{R}(w, v)) \stackrel{\text{def}}{\wedge} \mathfrak{M}, v \models \phi$

kielégítő (HUN) – satisfied (ENG) – igaz (HUN) – true (ENG) – szemantikai következmény (HUN) – semantic consequence (ENG)

A feltételeket meglepően egyszerű értelmezni. Mind az öt feltételt az adott $\mathfrak{M}$ modellre és az adott w állapotra kell értelmezni. Az első feltétel azt mondja ki, hogy az A propozícióra vonatkozóan a modell akkor (és csak akkor) elégíti ki az igazságfeltételt az adott w állapotban, ha az állapot benne van a $\mathbf{V}$ értékelési függvény értékkészletében. A második feltétel szerint az ellentmondás nem igaz az adott modell szerint az adott w állapotban. A harmadik formula azt mondja ki, hogy ha egy ϕ formula igaz (kielégíti az adott modellt, az adott w állapotban), akkor (és csak akkor) a tagadása nem igaz. A negyedik feltétel két

[24] A definiáláshoz használni kell olyan logikai műveleteket, amelyeket nem lehet a tárgynyelvben alkalmazott szimbólumokkl jelölni, még akkor sem, ha az értelmük ugyanaz. Az ilyen metanyelvi műveleteket a műveletek fölé írt ‘def’ jelzéssel látom el.

formula diszjunkciójára mondja ki, hogy annak igazsága a komponensformulák vagylagos igazságától függ az adott w állapotban. Végül az ötödik formula rögzíti azt, hogy az adott modell szerint egy ϕ formula akkor LEHETSÉGES a w állapotban, ha van egy másik v állapot, amely $\mathbf{R}$-relációban van a w állapottal, és amelyben a formula igaz. Ezekkel a kielégíthetőségi feltételekkel már tetszőleges formuláról eldönthetjük, hogy igaz-e az adott modell szerint a w állapotban. Azért pedig, hogy kifejezhessük azt is, ha valami nem kielégíthető, vagyis hamis, érdemes bevezetni a szemantikai következmény reláció tagadását.

(ML.9) FALSIFIED $\equiv \not\models \overset{\text{def}}{=} \neg \models$

nem kielégítő (HUN) – hamis (HUN) – false (ENG) – falsified (ENG) – refuted (ENG)

Ha ismerjük már a lehetséges operátor használatát, akkor nézzük meg, hogyan lehet a szükségszerű operátort alkalmazni. Az eddigiekből következik, hogy a definícióban a korábbi egzisztenciális kvantifikációt ki kell cserélni univerzálisra.

(ML.10) LOCALLY_NECESSARY $\overset{\text{def}}{=} \mathfrak{M}, w \models \Box A \overset{\text{def}}{\Leftrightarrow} \overset{\text{def}}{\forall} v \in \mathbb{W}(\mathbf{R}(w,v)) \overset{\text{def}}{\wedge} \mathfrak{M}, v \models A$

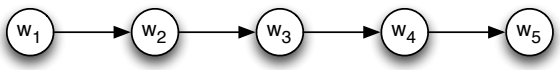
lokálisan szükségszerű (HUN) – locally necessary (ENG)

Ezek lokálisan érvényes kijelentések. Globálisan vagy univerzálisan igaz formula definíciója.

(ML.11) UNIVERSALLY_TRUE $\overset{\text{def}}{=}$

univerzálisan igaz (HUN) – globálisan igaz (HUN) – universally true (ENG) – globally true (ENG)

Az értelmezést megkönnyítendő bemutatok egy példát [61]. Vegyük a következő öt-elemű $\mathbb{W}$-halmazt, amelynek elemei közti relációt a nyilak jelzik. Az elérhetőségi reláció legyen: $\mathbf{R}(w_i, w_j) \leftrightarrow j = i+1$. Ekkor a keretünk így néz ki: $\mathfrak{F} = (\{w_1, w_2, w_3, w_4, w_5\}, \mathbf{R})$. Az állapotok közti kapcsolatot az alábbi grafikonnal lehet szemléltetni.

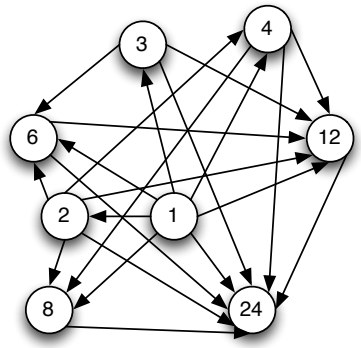


Az elérhetőségi reláció itt azt jelenti, hogy w_1-ből csak w_2, w_2-ből csak w_3, w_3-ból csak w_4 és w_4-ből csak w_5 állapot érhető el, utóbbiból pedig semmi sem. Ezen a kereten vegyük a következő értékelési függvényt: $\mathbf{V}(A) = \{w_2, w_3\}$, $\mathbf{V}(B) = \{w_1, w_2, w_3, w_4, w_5\}$, illetve $\mathbf{V}(C) = \varnothing$. Ekkor a $\mathfrak{M} = (\mathfrak{F}, \mathbf{R})$ modellen a következő formulák kielégítőek (igazak).

1) $\mathfrak{M}, w_1 \models \Diamond\Box A$
2) $\mathfrak{M}, w_1 \not\models \Diamond\Box A \to A$
3) $\mathfrak{M}, w_2 \models \Diamond(A \wedge \neg C)$
4) $\mathfrak{M}, w_1 \models B \wedge \Diamond(B \wedge \Diamond(B \wedge \Diamond(B \wedge \Diamond B)))$
5) $\mathfrak{M} \models \Box B$

Az első formula megértéséhez elsőként meg kell állapítanunk, hogy a $\Box A$ formula igaz a w_1 és w_2 állapotban, mivel mindkettőre igaz, hogy belőlük elérhető a w_2, illetve w_3 állapot, és ezekben az A propozíció igaz. Mivel w_1-ból és w_2-ből más állapotok nem érhető el, ezért igaz az az állítás, hogy a belőlük elérhető összes állapotban fennáll az A propozíció, ez pedig w_1-r és w_2-re is $\Box A$ igazságát jelenti. Ezek után nézzük meg a $\Diamond\Box A$ formula kielégíthetőségét w_1-ben! Jelöljük a formulánkat így: $\Diamond(D)$. Ez akkor igaz, ha van olyan w_1-ből elérhető állapot, ahol a D igaz. Ilyen pedig van: mivel w_1-ből csak a w_2 állapot érhető el, és ott a $D(=\Box A)$ igaz (ahogy az előbb megállapítottuk). Az első formula tehát igaz. A második sor azt fejezi ki, hogy a $\Diamond\Box A \to A$ formula nem kielégíthető az adott modellen, a w_1 pontban. Ez a formula azért nem igaz, mert a kondicionális előtagja igaz (ezt az előbb néztük át), viszont a kondicionális utótagja (maga az A propozíció) nem igaz, hiszen a $\mathbf{V}$ értékelési függvény az A propozíciót csak a w_2 és W_3 állapotokban adja meg igaznak. Emiatt a vizsgált formulánk (kondicionálisunk) nem lehet igaz. A harmadik formula azt fejezi ki, hogy a w_2 állapotból elérhető egy olyan állapot (ez a w_3), amelyben igaz az A propozíció és nem igaz a C. Ez így is van, hiszen w_3-ban a $\mathbf{V}$(A) igaz, a $\mathbf{V}$(C) nem igaz (mivel utóbbi sehol sem igaz). A negyedi formula értelmezéséhez azt kell figyelembe vennünk, hogy az értékelési függvény szerint a B propozíció minden állapotban igaz. A vizsgált – kielégíthető – formula azt fejezi ki, hogy a w_1 állapotban a B propozíció igaz, és van olyan belőle elérhető állapot, amelyben a B propozíció igaz (ez a w_2 állapot), és van olyan állapot, amelyben a w_2-ből elérhető és ott a B igaz és így tovább addig, amíg az utolsó állapotig el nem érünk. Az ötödik formula tartogat egy "meglepetést". A formula ugyanis azt állítja, hogy a B propozíció univerzálisan igaz, vagyis minden állapotban van egy olyan elérhető másik állapot, ahol a propozíció fennáll. Ez triviális az első négy állapotra, hiszen a belőlük elérhető szomszédos állapotokban igaz a B. Az utolsó w_5 állapotból azonban egyetlen más állapot sem érhető el. Erre viszont azt mondhatjuk, hogy a lokális kieléígíthetőséghez csak azt a feltételt kell kielégíteni, hogy az adott állapotból elérhető minden más állapotban igaznak kell lenni a propozíciónak. Mivel egy állapot sem érhető el a w_5-ből, ezért e feltétel üresen (vacuously) teljesül, és emiatt a vizsgált teljes formula igazsága is fennáll.

Érdemes megnéznünk egy másik példát is [61], hogy még közelebb kerülhessünk az eddigi fogalmainhoz. Legyen a halmazunk $\mathbb{W} = \{1,2,3,4,6,8,12,24\}$ és az $\mathbf{R}(m,n)$ legyen az 'n osztható m-mel' reláció ($m \neq n$). Ekkor az 1.-ből elérhető minden más állapot, tehát $\{2,3,4,6,8,12,24\}$, a 2.-ből elérhető a $\{4,6,8,12,24\}$ állapot, a 3.-ból a $\{6,12,24\}$ állapot, a 4.-ből a $\{8,12,24\}$ állapot, a 6.-ból a $\{12,24\}$ állapot, a 8.-ból a $\{24\}$ állapot, a 12.-ből a $\{24\}$ állapot, és a 24.-ből nem érhető el semmi. Ezt a következő ábrával szemléltethetjük.

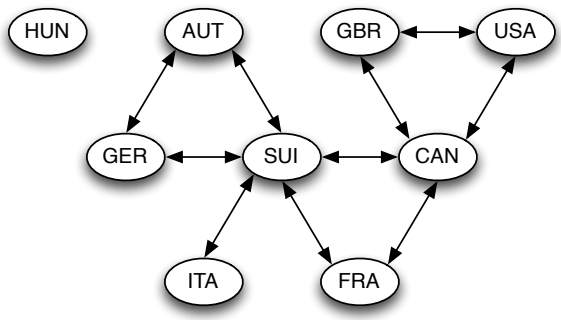


Legyen az értékelési függvényünk a következő: $\mathbf{V}(A) = \{4,8,12,24\}$ és $\mathbf{V}(B) = \{6\}$ (triviális értelmezésként mondhatjuk, hogy A a 4-gyel, B a 6-tal való oszthatóságot jelenti). Ekkor az adott modellünkön az alábbi formulák elégíthetők ki.

1) $\mathfrak{M},4 \vDash \Box A$
2) $\mathfrak{M},6 \vDash \Box A$
3) $\mathfrak{M},2 \nvDash \Box A$
4) $\mathfrak{M},2 \vDash \Diamond(B \wedge \Box A) \wedge \Diamond(\neg B \wedge \Box A)$

Nézzük meg, miért. Az 1) formula azt fejezi ki, hogy a 4. állapotból elérhető a 8., a 12. és a 24. állapot, és ezekben az értékelési függvény szerint az A formula igaz, vagyis minden olyan világállapotban, amely a 4.-ből elérhető, az A formula kielégítő. Ez pedig pont megfelel annak, hogy a $\Box A$ formula kielégíti a modellünket. A 2) sor szerint a 6. állapotból látszik a 12. és 24. állapot, és ezek mindegyikében igaz az A formula, amiből következően a 6. állapotban igaz lesz a $\Box A$ formula is. A 3) formula értelmezése: a 2. állapotból elérhető $\{4,6,8,12,24\}$, amelyek közül a 6. esetében nem teljesül az A propozíció igazsága, ezért a 2. állapotra nézve nem áll fent a $\Box A$ formula igazsága. Az utolsó formula pedig azt fejezi ki, hogy egyszerre igaz két állítás. Egyfelől a 2. állapotból elérhető legalább egy olyan világ, amelyben B is és $\Box A$ is igaz (ilyen a 6. állapot, hiszen ott a B igaz az értékelési függvény alapján, a $\Box A$ formula pedig igaz a fenti 2. formula alapján). Másfelől van legalább egy olyan állapot is, amelyben $\neg B$ és $\Box A$ egyszerre igaz (ilyen a 4. állapot, hiszen ott B nem igaz az értékelési függvényalapjá, viszont $\Box A$ fennáll a fenti 1. formula szerint).

Végül nézzünk át egy olyan példát, amit talán jobban köthetünk a mindennapi életünkhöz, mint az előző kettőt. Vegyünk pár országot: Németország (GER), Ausztria (AUT), Svájc (SUI), Olaszország (ITA), Franciaország (FRA), Magyarország (HUN), Kanada (CAN) és Egyesült Államok (USA). Nézzük meg, hogy ezekben az országokban milyen hivatalos nyelvek vannak: német (N), olasz (O), francia (F), magyar (M), angol (A). Ha az elérhetőségi reláción azt értjük, hogy két ország közt beszélnek-e közös nyelven az emberek, akkor a következő kapcsolatokat kapjuk: $\mathbf{R}(HUN, \varnothing)$, $\mathbf{R}(AUT, \{GER, SUI\})$, $\mathbf{R}(GBR, \{USA, CAN\})$, $\mathbf{R}(USA, \{CAN, GBR\})$, $\mathbf{R}(GER, \{AUT, SUI\})$, $\mathbf{R}(CAN, \{FRA, SUI\})$, $\mathbf{R}(ITA, \{SUI\})$, $\mathbf{R}(SUI, \{GER, AUT, ITA, FRA, CAN\})$, $\mathbf{R}(FRA, \{CAN, SUI\})$. Tételezzük fel azt, hogy az értékelési függvényünkkel azt írjuk le, hogy a különböző országokban (világokban) hivatalos-e az adott nyelv vagy sem. Vagyis: $\mathbf{V}(N) = \{GER, SUI, AUT\}$, $\mathbf{V}(O) = \{SUI, ITA\}$, $\mathbf{V}(F) = \{SUI, FRA, CAN\}$, $\mathbf{V}(M) = \{HUN\}$. Az elérhetőségi reláció az alábbi módon kapcsolja össze a világokat.



Ebben a modellben – sok egyéb mellett – az alábbi formulák kielégíthetőek.

1) $\mathfrak{M}, HUN \vDash \Box M \wedge \Box A \wedge \Box N \wedge \Box O \wedge \Box F$
2) $\mathfrak{M}, SUI \vDash \Box(O \vee F \vee N)$
3) $\mathfrak{M}, USA \vDash \Box A$
4) $\mathfrak{M}, CAN \nvDash \Box A$
5) $\mathfrak{M}, ITA, FRA, GER, GBR, SUI, AUT, CAN, USA \nvDash \Diamond M$

Az 1) formula azt írja le, hogy a minden HUN-világból elérhető más világban szükségszerűen beszélik az összes nyelvet. Ez a logikai igazság azonban üres abban az értelemben. hogy csak azért igaz, mert nincs egyetlen olyan másik világ, amely elérhető lenne ebből az állapotból. Tartalmas jelentése van a 2) formulának, ez ugyanis azt takarja, hogy a Svájból elérhető világokban szükségszerűen beszélnek a német, a francia vagy az olasz nyelv valamelyikén. A 3) sorban az szerepel, hogy az Egyesült Államokból elérhető országokban szükségszerűen (mindenhol) beszélnek angolul. A 4) formula tanulsága szerint azonban ez nem igaz Kanada esetén (mivel van olyan ország, ahol az angol nem hivatalos nyelv, például Svájc). Végül az 5) sor azt a tényt reprezentálja, hogy nincs olyan ország, amelyből elérhető lenne egy olyan világ, ahol a magyart beszélik (azaz a magyar nyelv szigetként működik).

A bemutatott példákból kiderül az is, hogy az **R** elérhetőségi relációnak tetszőleges jelentést tulajdoníthatunk. Sokféle kontextus, sokféle feltételrendszer képzelhető el, amelyeken belül más és más értelmezéseket adhatunk az elérhetőségi relációnak, és ezáltal persze másként lehet az ezekre építhető modelleket is interpretálni.

Modális sémák, kalkulusok

Ahogy korábban már láttuk, a logikai rendszerekhez axiómákat, sémákat definiálhatunk, amelyek alapján a logikai rendszer fogalmai, tételei mind levezethetőek. Ezt megtehetjük a modális logikában is, vagyis modális sémákat definiálhatunk, amelyek alapján modális kalkulusok építhethetünk fel. Egymás után definálhatunk újabb és újabb sémaaxiómákat, amelyek a korábbi axiómákkal együtt egy-egy modális kalkulust alkotnak majd – miközben mind átveszik a kijelentéskalkulus három axiómáját (↪20.o.). Az axiómasémákat pedig jellemezhetjük az elérhetőségi relációra vonatkozóan különböző tulajdonságokkal is. A következőkben sorban bemutatom a (K), a (D), a (T), a (B), az (S4) és az (S5) modális axiómasémák definícióit.

(ML.12) $(\mathrm{K}) \stackrel{\text{def}}{=} \Box(A \to B) \to (\Box A \to \Box B)$

(K) (LOG) – disztribúciós axióma (HUN) – distribution axiom (ENG)

A (K)-séma jelentése annyi, hogy ha egy $A \to B$ kondicionális szükségszerűen igaz, akkor a kondicionális két tagjának szükségszerűségét is összekapcsolhatjuk a kondicionálissal. Ez tétel a modális operátorra vonatkozó disztributivitás (↪38.o.) tulajdonságát is kimondja egyben (a kondicionális műveletre nézve).

(ML.13) $(\mathrm{D}) \stackrel{\text{def}}{=} \Box A \to \Diamond A$

(D) (LOG) – szeriális axióma (HUN) – seriality axiom (ENG)

A (D)-séma azt jelenti, hogy ha valami szükségszerű(en igaz), akkor az lehetséges(en igaz). Első pillantásra egy kicsit intuícióellenesnek bizonyulhat sokak számára ez a tétel,

hiszen ‘miért ne lenne igaz, ami szüksészerű’. Pedig ez a séma biztosítja a deontikus séma és a deontikus logika speciális minőségét. A (D)-sémában az a lényeg, ami csak implicit módon van benne: azokban a világokban, amelyek erre sémára támaszkodnak, ha valami szükségszerű (kötelező), az nem feltétlenül igaz egyben. Vagyis, ami szükségszerű (kötelező), az nem szükségszeűen teljesül, csak a lehetősége van meg ennek. Ez a séma összakapcsolható az elérhetőségi relációra vonatkozó szerialitás tulajdonságával (↬30.o.).

(ML.14) $(\mathrm{T}) \stackrel{\text{def}}{=} \Box A \to A$

(T) (LOG) – (M) (LOG) – reflexivitási axióma (HUN) – reflexivity axiom (ENG)

Ami a (D)-sémában nem teljesül, azt mondja ki a (K)-axiómaséma: ha valami szükségszerű(en igaz), akkor az igaz. Azokra a világokra tehát, amelyeket ez a séma jellemez, teljesül az a – mint láthattuk, nem teljesen megalapozott – intuíciónk, hogy valami akkor lehet szükszerűen igaz, ha igaz is egyben. Ezt a tulajdonságot az elérhetőségi reláció reflexivitásával hozhatjuk párhuzamba (↬28.o.). Ez annyit tesz, hogy a (K)-szükségszerűség fennállása esetén nemcsak az aktuális világból elérhető más világokra vonatkozó elvárásokat az aktuális világra is ki kell terjeszteni.

(ML.15) $(\mathrm{B}) \stackrel{\text{def}}{=} A \to \Box\Diamond A$

(B) (LOG) – szimmetricitási axióma (HUN) – symmetricity axiom (ENG)

A (B)-séma jelentése az, hogy ha valami igaz, akkor szükségszerű(en igaz), hogy lehetséges(en igaz). Ebben az esetben az elérhetőségi reláció szimmetrikus (↬29.o.). Első értelmezésre elfogadhatónak látszik a tétel, hiszen ha egyszer valami már fennáll, akkor ennek a lehetőségét szükségesnek érezzük, de a helyzet mégsem ilyen egyszerű. Tekintsük ugyanis a következő levezetést:

1) $A \to \Box\Diamond A$
2) $\neg\Box\Diamond A \to \neg A$
3) $\neg(\neg\Diamond\neg)\Diamond A \to \neg A$
4) $\Diamond\neg\Diamond A \to \neg A$
5) $\Diamond\neg\Diamond(\neg\neg A) \to (\neg A)$
6) $\Diamond\Box(\neg A) \to (\neg A)$
7) $\Diamond\Box B \to B$

A levezetés tanulsága az, hogy a kondicionális kontrapozíció, a kettős tagadás, a zárójelezési, a behelyttesítési és a (Dual) szabályok segítségével a (B)-sémából levezethető a hetedik sorban látható tétel, amit már közel sem érzünk triviálisnak, sőt inkább intuícióellenesnek tartunk. A tétel ugyanis azt állítja, hogy ha valami lehet, hogy szükségszerű, akkor az igaz. Ennek az igazsága egyáltalán nem nyilvánvaló. A (B) séma elfogadása ezért mégsem annyira magától értetődő, mint az az első látásra feltételezhető lenne. Sokkal inkább elfogadható a következő axiómaséma, bár bizonyos modalitásértelmezések esetén ennek fennállását is meg lehet kérdőjelezni.

(ML.16) $(\mathrm{S4}) \stackrel{\text{def}}{=} \Box A \to \Box\Box A$

(S4) (LOG) – (4) (LOG) – tranzititási axióma (HUN) – transitivity axiom (ENG)

A séma azt rögzíti, hogy ha valami szükségszerű(en igaz), akkor szükségszerű(en igaz), hogy szükségszerű(en igaz). Ez az elérhetőségi reláció tranzitivitását is jelenti egyben (↬30.o.), ami abban nyilvánul meg, hogy ha valami szükségszerű egy aktuális világban, akkor nemcsak belőle elérhető szomszédos világokban, de az azokból elérhető "másodszomszédos" világokban is szükségszerű lesz. Az erre a sémára épülő S4-kalkulusban interpretáljuk úgy – Gödel nyomán – a $\Box A$ formulát, hogy annak az 'A bizonyított állítás' jelentést tulajdonítjuk. Ezzel az értelmezéssel az S4 kalkulus a bizonyítottság[25] (vagy bizonyíthatóság) logikájának tekinthető [454, 263]. Már csak ezért is, de persze más okok miatt is, az (S4)-sémára szükség lesz a cselekvéselmélet tárgyalása során.

(ML.17) $(\mathrm{S5}) \stackrel{\text{def}}{=} \Diamond A \to \Box \Diamond A$

(S5) (LOG) – (E) (LOG) – euklideszi axióma (HUN) – euclidean axiom (ENG)

A legerősebb S5-kalkulus alapsémája az (S5), amelynek jelentése: ha valami lehetséges(en igaz), akkor szükségszerű(en igaz), hogy lehetséges(en igaz). Ez az értelmezés a sémát az elérhetőségi reláció euklideszi tulajdonságával kapcsolja össze (↬31.o.). Ebben a kalkulusban adhatjuk azt az értelemzést a $\Box A$ formulának, hogy 'analitikusan igaz, hogy A'. Ha így teszünk, akkor a kalkulussal az analicitás, vagyis az analitikus állítások logikáját kapjuk [454, 265]. Vigyázni kell azonban ezzel az értelmezéssel, mert nem jelent minden esetben egyértelmű segítséget akkor, amikor el akarjuk dönteni egy állításról, hogy analitikus-e vagy sem.[26]

Fontosnak tartom újra hangsúlyozni, hogy a fenti definíciók csak egy-egy axiómasémát határoznak meg, és csak úgy jutunk kalkulusokhoz a segítségükkel, ha az "alattuk levő" többi axiómát is beemeljük az adott kalkulusba (vagy ha nem is mindet, de legalábbis a legtöbbet). És az axiómák mellé szükség van még levezetési szabályokra is. Egyfelől minden modális kalkulus átveszi a modus ponens szabályt (↬21.o.), illetve a behelyettesíthetőségre is kimondanak egy szabályt, de ezzel itt nem foglalkozom. A modális kalkulusokhoz azonban kell még egy további levezetési szabály is, amit többféleképpen is neveznek, de persze a definíció minden esetben ugyanaz.

(ML.18) $(\mathrm{NEC}) \stackrel{\text{def}}{=} \vdash A \Rightarrow \vdash \Box A$

szükségességi szabály (HUN) – modális generalizálás szabálya (HUN) – Necessitation Rule (ENG) – Modal Generalization rule (ENG) – (Nec) – (LOG) – (Gen) – (LOG)

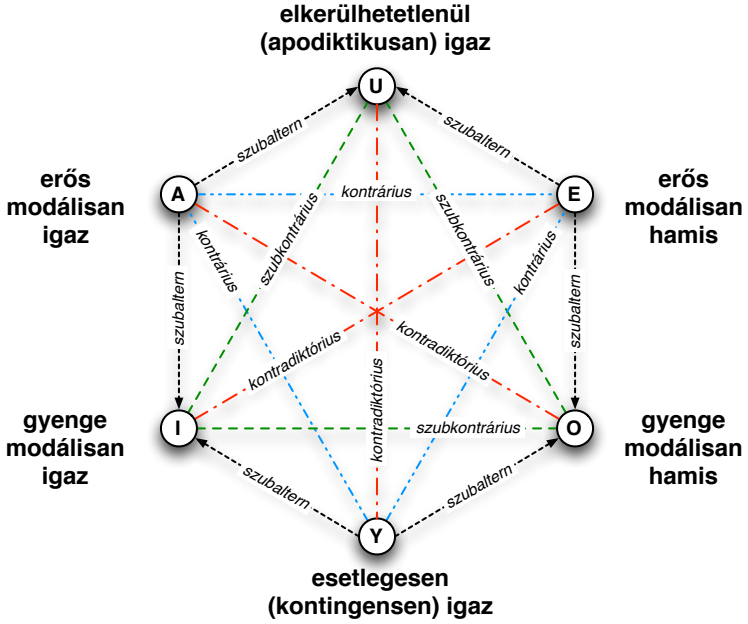
Ez a levezetési szabály tehát olyan, mint a modus ponens, ami szükséges az axiómákból és tételekből való következtetések levonásához. Az értelmezése egyszerű: ha egy A formula(osztály) eleme egy logikai nyelvnek, akkor a $\Box A$ formula is eleme lesz.

A fent bemutatott sémát (és kalkulusok) közül a továbbiakban a (D) séma és a hozzá tartozó modális logika a deontikus logikában (↬116.o.) és a doxasztikus logikában (↬71.o.) lesz különösen fontos, a (K), (T), (B), (S4) és (S5) sémákat pedig az episztemikus (↬67.o.) és a preferencialogikában (↬110.o.) fogjuk majd alkalmazni.

[25] Ruzsa Imre megjegyzi, hogy csak akkor tartható ez az interpretáció, ha a bizonyítottságon nem formalizált elméleten belüli bizonyíthatóságot, levezethetőséget értünk. Ebben az értelemben az S4-es kalkulus inkább a tapasztalati (kísérleti) tudáson alapuló bizonyítottság fogalmához köthető.

[26] „Ennek oka az, hogy az emberiség tapasztalati ismereteinek törzsanyaga fölszívódott a szavak jelentésébe, és ezért nehéz megvonni a határt a tapasztalati ismeretek és a nyelvhasználaton, a szavak jelentésén alapuló ismeretek között." [454, 265].

A modális logikai fejezet lezárásaként ki kell mégtérnem a modális logika és a logikai négyszög kapcsolatára. Utóbbit ugyanis alkalmazhatjuk a modális logikai operátorokra is. Sőt, a modális operátorokra nemcsak a négyszögbe, de a hatszögbe rendezés is értelmes. Ha a legáltalánosabb modális logikai hatszöget akarjuk felrajzolni, akkor a szükségszerűség és lehetségesség terminusai helyett az erős és gyenge modális operátorokat használva az alábbi ábrát kaphatjuk:



Hasonló ábrát lehetne felrajzolni az alethikus logikára, ami a szükségszerűség és a hozzá kapcsolódó fogalmak összefüggéseit mutatná, és a többi modális logika esetében is ugyanúgy érvényesnek bizonyulna a logikai hatszög az operátorok közti összefüggések szemléltetésére [51][52][366].

1.9 Episztemikus/doxasztikus logika

Az episztemikus logika, illetve a doxasztikus logika a modális logikának az az ága, amely a vélekedés, tudás, ismeret, hit logikai leírásának lehetőségeivel foglalkozik. Azonnal képet kaphatunk, mire lehet jó ez a diszciplina, ha felidézünk egy arab közmondást. Ez az ősrégi bölcselet olyasvalamit állít az ember és tudás viszonyáról, ami – időben sokkal, sokkal később – bekerült az episztemikus logikai diskurzusba is.

> „Négyféle ember van.
> Aki nem tudja, és nem tudja, hogy nem tudja: a balga – kerüld őt;
> Aki nem tudja, de tudja, hogy nem tudja: az egyszerű – tanítsd őt;
> Aki tudja, de nem tudja, hogy tudja: alszik még – ébreszd fel őt;
> Aki tudja, és tudja, hogy tudja: ő a bölcs – őt kövesd.” [150]

Az idézet egyetlen fogalom, a tudás segítségével különít el négy embertípust egymástól. A balga és bölcs alakja közé kihúzott ívben a balga az, aki “nem tudva nem tud”, és a bölcs az, aki “tudva tud”. A hétköznapi beszédben nem különböztetjük meg ennyire élesen ezeket a terminusokat, s bár a **tudás**, ismeret (knowledge), illetve **vélekedés**, hit (be-

lief) kettőssége mentén érezni valamilyen elkülönítési szándékot, de a terminológiában inkább zavar van e kérdésben. Az biztos, hogy amiről ennek kapcsán szó van (a tudás és a vélekedés kettőssége), az általában valamilyen mentális állapottal hozható kapcsolatba, olyan mentális állapottal, amelyek tartalma egy propozíció. A tudáshoz képes a vélekedést gyengébbnek, megalapozatlanabbnak érezzük. Egy vélekedés lehet megalapozatlan (hamis) és megalapozott (igaz), a tudásra ez nem mondható. Fontos feladat, hogy megtaláljuk, hogyan tehetünk különbséget e fogalmak között.

Angol nyelvben a 'knowledge' és 'belief' szavak vannak használatban, ezek közül a 'belief' fordítása nem egységes. Az angol 'belief' terminust lehetne 'hit'-ként vagy 'hiedelem'-ként fordítani, de a magyarban mindkét terminusnak olyan konnotációja van, ami itt nem kívánatos értelmezést sugallna, amit szeretnék elkerülni. A magyar nyelvhasználatban ugyanis mind a hit, mind a hiedelem terminusához egyfelől valamiféle megalapozatlanságot, másfelől nem tudományos jellegű világmagyarázó intézményeket (vallást, ezoteriát) társítunk. Angolban a 'belief' terminus ebben az értelemben semleges, amit magyarul jobban kifejezhetünk a 'vélekedés' terminusával. Ennek ellenére használni fogom a 'hit' terminust is, ha az adott nyelvi kontextusban nem sugall transzcendentális kapcsolatot.

Episztemikus logika

Az episztemikus logikát úgy építhetjük fel, hog először átértelmezzük a modális logika (K)-kalkulusát, majd azt további axiómákkal bővítve egyre erősebb nyelvhez jutunk [150]. Induljunk ki tehát a (K)-kalkulus nyelvéből.

(EL.0) $\mathscr{L}_{EL}(\mathscr{L}_{(K)C}, \boldsymbol{K})$, ahol

$\mathscr{L}_{(K)C}$	(K)-kalkulus
$\boldsymbol{K}_i$	tudja

Az episztemikus logikában a $\Box$ szükségszerű modális operátort úgy értelmezzük át, hogy annak jelentése az 'i ágens tudja, hogy a ϕ formula igaz', a jelölése pedig a $\boldsymbol{K}_i(\phi)$ szimbólum legyen. A $\boldsymbol{K}$ tudásoperátorhoz kapcsolt alsó index jelzi, hogy ennek a logikának a szemantikai újraértelmezés mellett van még egy másik újdonsága. Ez egy multimodális logika, ami azt jelenti, hogy több modális operátort is fel lehet venni a kalkulusba. Erre azért van szükség, hogy kezelni tudjuk azt a fontos társadalmi tényt, hogy a tudást, a hitet sok esetben nem általában, hanem az egyes ágensekhez rendelve értelmezzük. Természetesen előfordulhat az, hogy úgy akarunk, úgy tudunk mondani valamit a tudásra vonatkozóan, hogy el lehet tekinteni a – tudáshordozó – ágensektől, de nagyon sok más esetben pont arra van szükség, hogy jelezni, kezelni tudjuk azt, hogy melyik ágens tudásáról, hitéről van szó egy társadalmi kapcsolaton belül. Ilyen megfontolások miatt kell ellátni a tudásoperátort – alapértelmezés szerint – az ágensekre utaló indexekkel (az ágensekre itt is a természetes számok segítségével hivatkozunk).

(EL.1) $\text{TUDJA} \equiv \boldsymbol{K}_i(\phi) \equiv \boldsymbol{K}_i\phi \stackrel{\text{def}}{=} \forall i(\Box_i\phi), i \in \mathbb{N}^+$

tudja (HUN) – tudja, hogy (HUN) – knows (ENG) – knows that (ENG)

A modális logika (K)-kalkulusát (↰63.o.) itt úgy vehetjük át, hogy a szükségszerűség operátort a fentiek szerint átértelmezzük. A különböző szerzők többféle alakban is kifejezik

az axiómát, amelyek természetesen teljesen ekvivalensek egymással. Kettőt bemutatok közülük: elsőként a modális (K)-kalkulus alakhű átértelmezését, majd utána egy átalakított formulát [150]. Az episztemikus disztribúciós szabályt tehát – kétféle formában – a következőképpen rögzíthetjük.

(EL.2) (K') $\stackrel{\text{def}}{=} \boldsymbol{K}_i(\phi \to \psi) \to (\boldsymbol{K}_i\psi \wedge \boldsymbol{K}_i\psi)$

(EL.3) (K") $\stackrel{\text{def}}{=} (\boldsymbol{K}_i\phi \wedge \boldsymbol{K}_i(\phi \to \psi)) \to \boldsymbol{K}_i\psi$

episztemikus disztribúciós axióma (HUN) – epistemic distribution axiom (ENG)

A disztibúciós axiómát úgy is interpretálhatjuk, hogy a tudás zárt a logikai következményre (implikációra), azaz ha valaki tuda, hogy ψ állítás a ϕ-nek következménye, és még azt is tudja, hogy ϕ fennáll, akkor tudnia kell azt is, hogy a következmény, vagyis ψ-nak is fenn kell állnia.

Természetesen az episztemikus kalkulushoz is következtetési szabály(oka)t kell rendelnünk. Itt is érvényesnek tarthatjuk a modus ponens szabályt (↬21.o.), a modális generalizálás szabályát (↬68.o.) pedig át kell értelmezzük a $\boldsymbol{K}$ tudásoperátorra vonatkozóan.

(EL.4) (MP) $\stackrel{\text{def}}{=} \{A, (A \to B)\} \Rightarrow B$

modus ponens következtetési szabály (HUN) – modus ponens inference rule (ENG)

(EL.5) (NEC) $\stackrel{\text{def}}{=} \vdash \phi \Rightarrow \vdash \boldsymbol{K}_i\phi$

tudásgeneralizáció szabálya (HUN) – (Nec) (LOG) – (Gen) (LOG) – Knowledge Generalization rule (ENG)

Növelni lehet az episztemikus logika erősségét, ha további modális sémákat veszünk fel a kalkulusba. Első lépésként felvehetjük a modális (T)-sémát (↬64.o.) az axiómák közé (természetesen episztemikus módon interpretálva). Ennek az lesz az eredménye, hogy a tudás fogalma összekapcsolódik az igazságéval. További hozadéka lesz még a (T)-séma alkalmazásának az a tény, hogy ennek segítségével meg tudjuk különböztetni a **tudás** és a **vélekedés** fogalmát egymástól, amikor pont a (T)-axiómához való eltérő viszonyulás alapján definiáljuk a vélekedés fogalmát, s ezzel magát a doxasztikus logikát (↬71.o.). A modális (T)-séma episztemikus átértelmezése eredményeként az alábbi formulához jutunk.

(EL.6) (T) $\stackrel{\text{def}}{=} \boldsymbol{K}_i\phi \to \phi$

tudás axióma (HUN) – knowledge axiom (ENG) – igazság axióma (HUN) – truth axiom (ENG) – episztemikus (T)-séma (HUN) – epistemic (T)-schema (ENG)

Nagyon fontos minőséget jelent ez tétel a tudás fogalmának értelmezésekor. Ha azt mondjuk, hogy tudunk valamit, akkor annak igaznak kell lennie. Nagyon erős állítás ez – különösen a vélekedés, hit definíciójával szembeállítva.

Tovább bővítjük az episztemikus kalkulust, ha felvesszük az axiómák közé a $\boldsymbol{K}$ tudásoperátorra vonatkozó (S4) modális sémát (↬64.o.). Ennek a lépésnek az lehet az értelme, hogy ezáltal az ágens saját tudásának tudásáról, vagyis introspekcióról, egyfajta reflektáltságról beszélhetünk. A formulában természetesen ebben az esetben is ki kell cserélnünk a modális operátorok szimbólumait.

(EL.7) (S4-PIK) $\stackrel{\text{def}}{=} \boldsymbol{K}_i\phi \to \boldsymbol{K}_i\boldsymbol{K}_i\phi$

pozitív introspekciós axióma (HUN) – positive introspection axiom (ENG) – (S4) (LOG) – (KK) (LOG) – reflektált pozitív tudás (HUN) – reflected positive knowledge (ENG)

Az axiómát úgy interpretálhatjuk, hogy ha az i ágens tud valamit, akkor tudja azt is, hogy tudja azt a valamit. Ez abban az értelemben erős tudás, hogy reflektált. A fejezetet elején mottóként közölt versrészlet szereplői közül az – utolsóként megénekelt – bölcsre illik ez a jellemzés. Az ősi versben dramaturgiai szerepe van, ahogy követik egymást az egyre inkább reflektált tudással bíró szereplők, így lehet eljutni a teljesen (reflektálatlanul) tudatlan embertől a reflektáltan tudó bölcsig. Az episztemikus logika másként "működik". Ha az episztemikus kalkulusba felvesszük az (S5) modális sémát is (↰65.o.), akkor olyan tudásoperátorhoz (tudástípushoz) jutunk, amit a köznapi gondolkodás még nem sokra tart. Ennek az új episztemikus sémának a jelentését és értelmét csak akkor érthetjük meg igazán, ha átalakítjuk az (S5)-sémát az $A = \neg\phi$ egyenlőség, illetve a (Dual)-séma (↰59.o.) alapján.

1) $\Diamond A \to \Box\Diamond A$
2) $\neg\Box\neg A \to \Box\neg\Box\neg A$
3) $\neg\Box\phi \to \Box\neg\Box\phi$

Az átalakítás első lépésben a (Dual)-szabályt alkalmazva kicseréljük a lehetséges modális operátort a szükségszerűre. A második lépésben behelyettesítjük az új változót a negált A-formulákra vonatkozó egyenlőség alapján, a következő lépésként már csak a $\boldsymbol{K}$ tudásoperátor szimbólumát kell átvezetni, és mindezek után az (S5)-axiómát a következő alakban ábrázolhatjuk.

(EL.8) (S5-NIK) $\stackrel{\text{def}}{=} \neg\boldsymbol{K}_i\phi \to \boldsymbol{K}_i\neg\boldsymbol{K}_i\phi$

negatív introspekciós axióma (HUN) – negative introspection axiom (ENG) – (S5) (LOG) – reflektált negatív tudás (HUN) – reflected negative knowledge (ENG)

Az axióma értelmezése az lehet, hogy ha az i ágens nem tud valamit, akkor tudja, hogy nem tudja. Logikai szempontból ez az – átalakított (S5)-sémával kiegészített – kalkulus a legerősebb. Eszerint az igazi bölcs az, aki egyaránt tudja azt, hogy mit tud, és azt is, hogy mit nem.

Definiálhatunk még további levezetési szabályokat is, például a **monotonitás** vagy a **kongruencia** szabályait. Ezeket azért mondhatjuk levezetési szabályoknak, mert a nyelv zártságát definiálják a kongruencia és monotonitás tulajdonságaira nézve.

(EL.9) (MON) $\stackrel{\text{def}}{=} \vdash (\phi \to \psi) \Rightarrow \vdash (\boldsymbol{K}_i\phi \to \boldsymbol{K}_i\psi)$

monotonitási szabály (HUN) – monotony rule (ENG) – (Mon) (LOG)

(EL.10) (CGR) $\stackrel{\text{def}}{=} \vdash (\phi \leftrightarrow \psi) \Rightarrow \vdash (\boldsymbol{K}_i\phi \leftrightarrow \boldsymbol{K}_i\psi)$

kongruencia szabály (HUN) – congruence rule (ENG) – (Cgr) (LOG)

Ezeket a szabályokat úgy is értelmezhetjük, hogy ezek azt fejezik ki, hogy a tudással rendelkező ágensek képesek tudássá formálni a logikai állításokat, vagyis ha egy adott episztemikus modellben érvényes egy kondicionális vagy egy bikondicionális művelet,

akkor az összetevőkre vonatkozó tudások köztött is fennáll ugyanaz a művelet. Lehetne még további szabályokat definálni, amelyek mind az ágensek logikai következtető képességét fejeznék ki más és más logikai műveletekre fókuszálva, de ezt a képességet kifejezhetjük egy még általánosabb formában is. Rögzíthetünk ugyanis egy nagyon erős, és ezért sokak által vitatott következtetési szabályt is, ami azt írja le, hogy ha a logika szabályai szerint egy adott formális elméletből (mondhatjuk így is: formulák valamilyen konjunkciójából) levezethető egy tétel, akkor az elmélet minden formulája (vagyis a konjunkciók elemi komponensei és a konjunkcióból következő "új" formula egyaránt) ismert az adott ágens számára. Formálisan kifejezve ezt a következő képletet kapjuk.

(EL.11) $(\mathrm{RK}) \stackrel{\text{def}}{=} \vdash (\phi_1 \wedge \phi_2 \wedge \ldots \wedge \phi_n \rightarrow \psi) \Rightarrow \vdash (\boldsymbol{K}_i\phi_1 \wedge \boldsymbol{K}_i\phi_2 \wedge \ldots \wedge \boldsymbol{K}_i\phi_n \rightarrow \boldsymbol{K}_i\psi), n \in \mathbb{N}$

logikai mindentudás szabálya (HUN) – knowledge logical omniscience rule (ENG) – (RK) (LOG)

Akik ezt a szabályt "bevezették", érvényesnek tartották, azt abból a megfontolásból tették, hogy ennek segítségével lehet az episztemikus képességekkel rendelkező ágensek viselkedését minél jobban lemodelezni, számítógépes szimulációját minél jobban megvalósítani. Tették ezt abban a – jogos – meggyőződésben, hogy az ágenseik mindig megbízhatóan képesek a logikai szabályok figyelembe vételével helyes következtetéseket levonni. Ez a számítógépes modellezés során teljesen elfogadható. Akik viszont ezt a szabályt támadják, azok arra hivatkoznak, hogy az emberek esetében sokszor nem teljesen megalapozott ez az elvárás, az emberek nem képesek minden és mindenfajta logikai feladatot megoldani. A fenti szabályra valóban rá lehet húzni a "logikai mindentudás" kifejezést, de az ember esetében ennek teljes körű érvényeségét nem szabad feltételeznünk. Ez nyilván nem logikai és – bizonyos értelemben – nem elméleti probléma. Egy cselekvéselmélet építése során azonban mégiscsak figyelembe kell venni az ember korlátait, ezért úgy gondolom, hogy ezt a szabályt fenntartásokkal kell kezelnünk. Nem eldobni kell, csak mindvégig figyelni kell az emberek valóságos logikai következtető képességeinek tényleges korlátaira. Ezt azonban nem ebben a fejezetben fejtem ki, hanem majd később.[27]

> Laci tudja, hogy Kati nem tudja, hogy Béla nem hiszi, hogy Klára tudja, hogy Péter azt hiszi, hogy a németek mind tudják, hogy az amerikaiak hisznek abban, hogy *A*.

Egy ilyen állítást tényleges tartalmát már nem igazán vagyunk képesek befogadni, noha az egymába ágyazott elemi állítások mindegyike könnyen értelmezhető. A hétköznapi gyakorlatban azonban sosem fordulnak elő ilyen hosszúságú (és összetettségű) mondatok, ezért nem tudjuk értelmezni. Egy számítógép számára viszont – persze csak akkor, ha az általa olvasható formában áll rendelkezésre – ugyanolyan könnyedén feldolgozható ez a mondat is, mint azok az egyszerűek, amelyeket még mi is kezelni tudunk. A fenti példamondat a közösségi helyzeteket jellemzi, amikor nem az összetett mondat bonyolultsága jelenti a nehézséget, hanem az egy csoportba tartozó, tudással rendelkező ágensek egyéni tudásainak egymásra vonatkozása. A másikfajta feldolgozási nehézséget az összetett mondatok logikai szerkezetének bonyolultsága okozhatja. A jogi nyelvben könnyedén találhatunk olyan sokszorosan egymásba ágyazott, összetett mondatokat, amelyek értelmezése a laikus embereknek mindenképpen, de akár még a jogászok számára

[27] amikor már meglesz, ugró hivatkozást betenni ide

gondot okoz. Az episztemikus logika egyik komoly feltételezése, hogy az ágensek teljesen racionálisak abban az értelemben, hogy tökéletes következtetési képességgel rendelkeznek – mind minőségi, mind mennyiségi értelemben. Ez a felételezés azonban túlságosan erős, ha a tényleges társadalmi gyakorlat tényleges ágenseire akarjuk mindezt vonatkoztatni. Ez a feltételezés nyilván komoly korlátja az elméletnek.

Doxasztikus logika

A (T)-séma azért lehet alkalmas a tudás és vélekedés fogalmainak elválasztására, mert amíg a tudás fogalmával szemben fenntarthattuk a séma érvényességét (☞68.o.), addig a vélekedésre vonatkozóan pont az ellenkezőjét kell mondanunk: ebben az esetben érvénytelennek kell tartanunk az igazság axiómát. Azt ugyanis mondhatjuk, hogy valaki igaznak vél valamit, ami hamis, csak ilyenkor hozzá kell tennünk, hogy tévesen véli, tévesen hiszi azt, ami nem igaz. Azt viszont nem mondhatjuk, hogy valaki tud valamit, ami hamis. Ha valami hamis, akkor az nem tudás, csak – hamis – vélekedés.[28] Ezen a ponton lehet elágaztatni az episztemikus és doxasztikus logikát egymástól azon az alapon, hogy amíg az előbbi a tudás, addig az utóbbi a vélekedés fogalmának formalizálásával foglalkozik. Mielőtt belekezdenék a doxasztikus logika vagyis a vélekedés (hit) logikájának bemutatásába, jeleznem kell, hogy ezt a logikát annyira közelállónak tekintem az episztemikus logikához, hogy nem definiálok számára új nyelvet, a tételeket ugyanúgy az episztemikus logikára utaló 'EL' szimbólummal azonosítom, a tételek számozását folytatom. Annak ellenére tehát, hogy a doxasztikus logikát külön fejezetben tárgyalom, sok szempontból .

A doxasztikus logika központi kategóriája a **vélekedés**, amit a tudás fogalmával szemben kell értelmeznünk. Ha a tudásra érvényesnek, a vélekedésre érvénytelennek tartjuk a (T) tudás axiómát, akkor kérdés, hogy milyen tulajdonsággal jellemezhetjük a vélekedés fogalmát. Ehhez a modális logika (D)-sémáját kell elővenni, amit persze újfajta módon kell értelmeznünk. Ahhoz, hogy megértsük, mi lehet az megfelelő új interpretáció, át kell alakítani kicsit a modális logika (D)-sémáját.

1) $(\mathrm{D}) \equiv \Box\phi \rightarrow \Diamond\phi$
2) $\Box\phi \rightarrow \neg\Box\neg\phi$
3) $\boldsymbol{B}_i\phi \rightarrow \neg\boldsymbol{B}_i\neg\phi$

Az 1) és 2) között kicseréltük a $\Diamond$ lehetséges operátort a (Dual)-szabály alapján a $\Box$ szükségszerű operátorra (külső és belső tagadással együtt), majd a 2) és 3) között a modális szükségszerű operátort a $\boldsymbol{B}_i$ vélekedésoperátorral helyettesítettük. Ezek alapján a következő formulával határozhatjuk meg a vélekedés egy nagyon fontos tulajdonságát.

(EL.12) $(\mathrm{D\text{-}CB}) \stackrel{\mathrm{def}}{=} \boldsymbol{B}_i\phi \rightarrow \neg\boldsymbol{B}_i\neg\phi$

konzisztencia axiómája (HUN) – Consistency of Belief axiom (ENG) – doxasztikus (D)-séma (HUN) – doxastic (D)-schema (ENG)

Ez a formula a vélekedés (hit) konzisztenciájáról szól, ugyanis azt rögzíti, hogy ha valaki hisz valamiben (ϕ-ben), akkor nem hihet annak az ellenkezőjében ($\neg\phi$-ben). Ez a tudásra

[28] A köznyelv természetesen nem határolja el egymástól ezt a két fogalmat ennyire pontosan és egyértelműen.

nem lenne igaz, csak a vélekedésre. Ezért ez a (T)-axiómánál gyengéb episztemikus (D)-séma a vélekedést, a hitet jellemző doxasztikus logika specifikumát adja. A vélekedéshez a $\boldsymbol{B}$ szimólumot és a következő axiómát rendelhetjük.

(EL.13) $\text{BELIEF} \stackrel{\text{def}}{=} \neg \boldsymbol{B}_i \bot$

vélekedés (HUN) – belief (ENG)

A formulát úgy értelmezhetük, hogy valaki (az i ágens) nem hiheti azt, hogy egy ellentmondás igaz lehet (nagyon fontos nem "elfelejteni", hogy ebben az esetben az episztemikus (T)-sémát nem tartjuk érvényesnek). A (T)-séma érvénytelensége miatt a doxasztikus logika nem olyan erős, mint az episztemikus logika (T)-sémával megerősített változata.

(EL.14) $(\text{S4-PIB}) \stackrel{\text{def}}{=} \boldsymbol{B}_i\phi \rightarrow \boldsymbol{K}_i\boldsymbol{B}_i\phi$

a hit pozitív introspekciós axiómája (HUN) – positive introspection of belief (ENG) – doxastic (S4) (LOG) – (BB) (LOG)

A doxasztikus logikában is értelmezhetjük az (S4) és (S5) axiómákat tartalmilag a tudás kapcsán bemutatott interpretációhoz hasonlóan.

(EL.15) $(\text{S5-NIB}) \stackrel{\text{def}}{=} \neg\boldsymbol{B}_i\phi \rightarrow \boldsymbol{K}_i\neg\boldsymbol{B}_i\phi$

a hit negatív introspekciós axiómája (HUN) – negative introspection of belief (ENG) – (S5) (LOG)

A tudás és vélekedés viszonyára vonatkozóan kimondható egy fontos axióma (amiről azért tudni érdemes, hogy vannak, akik vitatják az érvényességét).

(EL.16) $\text{KB} \stackrel{\text{def}}{=} \boldsymbol{K}_i\phi \rightarrow \boldsymbol{B}_i\phi$

a hit tudás alapú axióma (HUN) – knowledge imply belief axiom (ENG)

A doxasztikus logikában is felvehető a (RK)-szabály "megfelelője", annyit kell csak tennünk, hogy a tudásoperátort kicseréljük a vélekedésoperátorra.

(EL.17) $(\text{RB}) \stackrel{\text{def}}{=} \vdash (\phi_1 \wedge \phi_2 \wedge \ldots \wedge \phi_n \rightarrow \psi) \Rightarrow \vdash (\boldsymbol{B}_i\phi_1 \wedge \boldsymbol{B}_i\phi_2 \wedge \ldots \wedge \boldsymbol{B}_i\phi_n \rightarrow \boldsymbol{B}_i\psi), n \in \mathbb{N}$

a logikai mindenhit szabálya (HUN) – belief logical omniscience rule (ENG) – (RB) (LOG)

Kollektív tudás

Sok szó esett már eddig a tudás, a hit fogalmáról, és bár mindvégig nyilvánvaló volt, hogy az ezeket reprezentáló operátorokat ($\boldsymbol{K}_i$, illetve $\boldsymbol{B}_i$) össze kell kapcsolnunk azokkal az ágensekkel, akit "hordozzák" a tudás- és hittartalmakat, viszont eddig csak jeleztük ezt a kapcsolatot, de nem igazán használtuk ki ezt az információt. A társadalmi cselekvések és társadalmi kapcsolatok értelmezése során rendkívüli jelentősége van annak a ténynek, hogy a tudás, a hit lényege szerint kollektív jelenség, az ágensek a saját egyéni tudásaikat kicserélhetik, megoszthatják egymás között, az egyéni tudásaikat mások tudásaira "vonatkoztathatják", és ezáltal az egyéni tudások mellett másfajta, kollektív tudások is

megjelennek a társadalmi térben. A kollektív tudás jelenségének fontosságát önmagában jelzi az a tény, milyen sokféle terminust használunk ebben a fogalomkörben. A későbbiekben a **közös tudás**, a **közismert tudás**, illetve a **elosztott tudás** fogalmait fogom majd a társadalmi cselekvés modelljébe beemelni, de ezekkel együtt fordulnak elő a terepen az olyan kategóriák, mint az együttes tudás, kölcsönös tudás, köztudott tudás, magától értetődő tudás, nyilvános tudás, kumulatív tudás, kollektív bölcsesség, szétszórt tudás, csoporttudás. Mindezeket pedig együttesen lefedi a kollektív tudás felettes fogalma.

A három kiemelten kezelni kívánt fogalom közül igazából csak a közös tudás és az elosztott tudás fogalma a fontos, a közismert tudás kategóriájára csak azért van szükségünk, hogy a közös tudás jelenségét értelmezni tudjuk a segítségével. A következőkben ezért előbb a közös tudás fogalmát mutatom be, majd ezt követi az elosztott tudás fogalmának rövid ismertetése. A közös tudás jelenségét először David Lewis vizsgálta monografikus igényességgel a 'Convention' című könyvében [305], majd még sokan követték őt a jelenség értelmezésében. A jelenség megragadására John McCarthy, a mesterséges intelligencia egyik úttörő kutatója javasolta használatba venni a "minden bolond tudja" ("any fool knows") kifejezést [150], de ez – számomra – inkább a közismert tudás jelentését sugallja. Nehéz ezekkel a kifejezésekkel pontosan operálni. nem is tehetünk mást, mint hogy egy-egy kifejezést kiragadunk a szokásos diskurzusból, és pontos, rögzített jelentéssel látjuk el őket. Ezáltal persze elszakadunk attól a lehetőségtől is, hogy a hétköznapi gyakorlatból eredő intuícióinkra hivatkozzunk.

Ahhoz, hogy a kollektív tudás és más hasonló fogalmak elemzésébe kezdjünk, először be kell vezetnünk a társadalmi csoport fogalmát. A csoportba tartozó ágensek tudásait, hiteit vizsgálva, a csoport egyes tagjait indexekkel jelölve, a csoport fogalmára hivatkozva definiálhatjuk majd a kollektív tudás általános kategóriája alá sorolható további tudásfogalmakat. A tudást hordozó ágens, az ágenseket összefogó csoport fogalmával itt még nagyon kell foglalkoznunk, ezekre a fogalmakra később még úgyis kitérek még. Most elég csak rögzíteni, hogy vannak ágensek, akik csoportokba rendeződ(het)nek. A csoportot G-vel jelöljük, az ágenseket, ha ez nem zavaró, akkor az i természetes számokkal azonosítjuk, ha szükséges, akkor az a_i szimbólumot alkalmazzuk rájuk. A közösség ekkor a tagjainak uniója.[29]

(EL.18) $$G \stackrel{\text{def}}{=} \bigcup_{i=1}^{n} a_i$$

---

csoport (HUN) – group (ENG) – közösség (HUN)

Az i. ágenshez kapcsolható $\boldsymbol{K}_i$ egyéni tudás operátorát már sokat használtuk korábban, most kell kidolgoznunk azokat a tudásfogalmakat, amelyek segítségével már meg tudjuk ragadni az egyéni tudások kölcsönösségét, egymásra vonatkozatottságát, illetve le tudjuk írni a csoport egészéhez köthető tudásokat is. Ehhez azonban először meg kell tudnunk adni azt, hogy az ágensek egyénileg milyen tudásokkal rendelkeznek. Ehhez nem kell mást tennünk, mint összegezni az egyes egyénekre az elemi tudásaikat. Ekkor is használhatjuk a már bevezetett $\boldsymbol{K}_i$ tudásoperátort, és ha csak magában használjuk, akkor az i. egyén össze tudását érttjük alatta, ha pedig a tudásoperátor után teszünk még az adott tudáselemre utaló jelet, akkor értelemszerűen az arra a konkrét tudáselemre vonatkozó tudás tényét állítjuk. Most tehát az egyén teljes tudását kell definiálnunk.

[29] Amíg nem zavaró, a közösség és a csoport terminusait felváltva, itt még azonos jelentés mentén alkalmazom.

(EL.19) $\boldsymbol{K}_i \stackrel{\text{def}}{=} \bigcup_{j=1}^{m} \boldsymbol{K}_i \phi_j$

---

egyén teljes tudása (HUN) – agent's full knowledge (ENG)

Az egyén teljes tudása (tudattartalma) az egyéni tudáselemek uniója.[30] Ez a fogalom segíteni fog minket a kollektív tudás körébe tartozó fogalmak értelmezésében. Lássuk, hogyan!

A közös tudás fogalmának értelmezéséhez először a **közismert tudás** kategóriáját kell bevezetnünk, ami nem egy bonyolult fogalom. Azt a jelenséget ragadhatjuk meg vele, amikor egy csoporton belül egyénileg mindenki rendelkezik ugyanazzal a ϕ tudással, vagyis amit "minden bolond tud". Ha a G csoport bármelyik i tagja tudja ϕ-t, akkor azt a következő formulával írhatjuk le.

(EL.20) $\boldsymbol{K}_G^E \phi \stackrel{\text{def}}{=} \bigcap_{i \in G} \boldsymbol{K}_i \phi$

---

közismert tudás (HUN) – köztudott tudás (HUN) – együttes tudás (HUN) – magától értetődő tudás (HUN) – közös tudás (HUN) – mutual knowledge (ENG)

Nehéz széles körben elfogadott, a fenti formulával definiált, pontos jelentés mentén értelmezett kifejezést rendelni ehhez a fogalomhoz. A definíció alatt felsorolt nyelvi alakzatok mindegyike használható lenne, csak a választási kényszer miatt emeltem ki közülük egyet (a 'közismert tudás' kifejezést), de ha nem lesz zavaró, a későbbiekben is megengedhetőnek tartom, hogy szinonímaként használjuk az alternatív kifejezéseket a fogalomra vonatkozóan. [31] A közismert tudás definíciója a G csoport egészére, de egyetlen ϕ tudáselemre vonatkozik. Továbbléphetünk egyet, ha erre a fogalomra támaszkodva definiáljuk a csoport teljes közismert tudását.

(EL.21) $\boldsymbol{K}_G^E \stackrel{\text{def}}{=} \bigcap_{i=1}^{n} \bigcup_{j=1}^{m} \boldsymbol{K}_i \phi_j$

(csoport teljes) közismert tudás(a) (HUN) – group's full mutual knowledge (ENG)

A $\boldsymbol{K}_G^E$ közismert tudáskészlet minden $\boldsymbol{K}_i$ egyéni tudáskészletben benne van, vagyis a $\boldsymbol{K}_i^E$ egyéni tudatok össztartalma mindig szélesebb vagy azonos terjedelmű, mint a $\boldsymbol{K}_G^E$ közismert tudat. Lehetséges (bár nem valószínű), hogy a $\boldsymbol{K}_G^E$ közismert tudat megegyezik valamelyik i személy $\boldsymbol{K}_i$ egyéni tudattartalmával. Amennyiben nem ez a helyzet áll fent (és ez a valószínűbb kimenet), akkor azt mondhatjuk, hogy $\boldsymbol{K}_G^E$ közismert tudat különbözik minden egyes $\boldsymbol{K}_i$ egyéni tudattól. Ilyen esetben hiába van jelen a $\boldsymbol{K}_G^E$ kollektív tudat minden eleme az összes $\boldsymbol{K}_i$ egyéni tudatban, mégis különbözik mindegyiktől, vagyis a

[30] Lehetne logikai és halmazelméleti műveletekkel definiálni a kollektív tudás alá tartozó fogalmakat. Előbbi esetben predikátumokat kellene összefűzni a definíciókban (az 'x tudja' és 'y tudja' állításokat kellene logikai operátorokkal összekapcsolni), utóbbi esetben a tudások összesítését kellene elvégezni (x tudását és y tudását kellene halmazműveletekkel összekapcsolni). Én a halmazelméleti jelölést választottam itt.

[31] Csak a kölcsönös tudás terminust zárnám ki (még akkor is, ha az angolban az ennek megfelelő 'mutual knowledge' alakot használják a leggyakrabban). A kölcsönösség számomra sokkal inkább sugallja a szimmetricitást és reflektáltságot, ami majd a közös tudás fogalmához lesz elengedhetetlen. A $\boldsymbol{K}_G^E$ terminusban az E paraméter a közismert tudásra utal – az angol 'everybody' kezdőbetűjével.

kollektív tudat mindegyik egyéni tudathoz képest más. Ez a fontos meglátás az alapja Emilé Durkheim társadalomelméletének is.

A tudást közismertnek mondhatjuk abban az értelemben, hogy mindenki részesül belőle. Azonban ez nem jelenti azt, hogy annak is tudatában lennének a csoport tagjai, hogy ezt a valamennyiük által egyénileg birtokolt tudást valamennyien ismerik. Sőt, elvileg az is elképzelhető, hogy mindenki azt feltételezi, hogy csak ő rendelkezik az adott tudással, senki más. Andersen meséje a meztelen királyról szép példája annak, hogy ami közismert tudás egy közösségen belül, az még nem feltétlen jelent közös tudást is egyben. Mindenki látta, tudta, hogy a királyon nincs ruha, de amíg valaki el nem kiáltotta magát, hogy "a király meztelen", addig senki sem tudhatta, hogy amit egyénileg mind tudnak, azt mindenki más is tudja, vagyis a tudás közös. Az Andersen-mesének persze létezik más olvasata is, de ez a tény nem teszi érvénytelenné a rá való hivatkozást ezen a ponton.

Ha azt akarjuk megragadni, ami az Andersen-mese tanulsága, akkor a **közös tudás** (vagy kölcsönös tudás) fogalmát kell definiálnunk. Ehhez új elemként fel kell vennünk a nyelvünkbe a közös tudást reprezentáló $\boldsymbol{K}_G^C$ modális operátort, amit a következőképpen definiálhatunk. Ehhez a $\boldsymbol{K}_i$ egyéni tudások alapján képezhető közismert tudásoknak az egymásra vonatkozó – iterációval kifejezhető – közös metszetét kell előállítanunk, amihez egy segédformulát is igénybe kell vennünk. Utóbbival azt a közismert tudást kell kifejeznünk, hogy valamelyik (j) csoporttag tudja, hogy a másik csoporttag ($j-1$) tudja, hogy ϕ. A közös tudás fogalmát ez alapján úgy ragadhatjuk meg, hogy egy "végtelen" iterációba állítjuk ezt a segédfogalmunkat.

(EL.22) $$\boldsymbol{K}_G^{E,j}\phi \stackrel{\text{def}}{=} \boldsymbol{K}_G^E \boldsymbol{K}_G^{E,j-1}\phi$$

(j. csoporttag) iterált tudása (HUN) – iterated knowledge of jth group member (ENG)

(EL.23) $$\boldsymbol{K}_G^C \stackrel{\text{def}}{=} \bigcap_{j=1}^{\infty} \boldsymbol{K}_G^{E,j}\phi$$

(csoport teljes) közös tudás(a) (HUN) – kölcsönös tudás (HUN) – common knowledge (ENG) – mutual knowledge (ENG)

A fogalom értelmezése: a G csoport minden tagja tudja, hogy a másik tudja, és tudja, hogy a másik tudja, hogy ő tudja, és így tovább. Elméletileg egy végtelen processzussal tudjuk kifejezni a közös tudás fogalmát (nem is kevesen vannak, akik épp ezért nem fogadják el sem a fenti definíciót, sem az effajta értelmezési irányt).

Fentebb említettem, hogy Andersen 'A császár új ruhája' című meséjéhez többféle értelmezést is hozzárendelhetünk. Legalább három lehetőség adódik. Azért interpretálhatunk többféleképpen is, mert van egy nagy adag bizonytalanság abban, hogy miként lehet előállítani azt a helyzetet, ami a közös tudáshoz eléréséhez szükséges "tudom, hogy tudja" állapotot eredményezi. Az első lehetőség az, hogy valóban nem tudják a csoporttagok, hogy a többiek tudják, ami maguk egyénileg mind tudnak, és a nyilvános bejelentés mindenkit meglep. Ilyen esetekben az a bejelentés, aminek eredményeként a közös tudás létrejön, csodálkozást, meghökkentést, meglepetést vált ki a csoport tagjaiban. A második értelmezés szerint a csoporttagok mind sejthetik, de nem tudhatják biztosan, hogy a többiek is a közös tudás birtokában vannak. Ekkor a mindenki által tudott tudás közzététele a korábbi bizonytalanságból adódó feszültség feloldódásához, egyfajta megkönnyebbüléshez, megnyugváshoz vezet. A harmadik értelmezési lehetőséget nem

lehet teljes mértékben elválasztani a második interpretációtól, de érdemes elkülöníteni tőle. Előfordulhat ugyanis, hogy az egyetemesen birtokolt egyéni tudások azért nem válhatnak sokáig közös tudássá, mert a közzététel útjában társadalmi akadályok vannak, valamiért nem lehet kimondani a közösség előtt a mindenki által tudott tudást. Ez is feszültséget okozhat a csoport tagjaiban, ám az ilyen helyzetekhez másfajta érzelmek kapcsolódnak, és ha mégis megszületik a nyilvános bejelentés, akkor azt felszabadultság követi (az Andersen-mesében ez utóbbival állunk szemben, amit a tömeg reakciója bizonyít: a királyt leleplező felkiáltás után a tömeg nevetésben robban ki).

A közös tudás fogalmának tárgyalásakor az volt a fontos, hogy mi az a tudáselem, amit a közösség tagjai mind közösen birtokolnak, vagyis hogy mi az, ami ott van a csoport mindegyik tagjának a fejében. Az **elosztott tudás** fogalmával nem az ebben az értelemben vett közös részt akarjuk megragadni, hanem valami mást. A kollektív bölcsesség elve, a 'sok kicsi sokra megy' mondása és még sok minden más ősrégi megfigyelés fejezte és fejezi ki azt a társadalmi tényt, hogy a közösség olykor több lehet, mint a részei, vagyis a tagjai összessége. Azt a tudásösszességet, ami a közösség egésze rendelkezésére áll nevezhetjük elosztott tudásnak. A korábban bevezetett egyéni teljes tudás fogalmára támaszkodva már könnyen definiálhatjuk ezt a fogalmat.

(EL.24) $$\boldsymbol{K}_G^D \stackrel{\text{def}}{=} \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} \boldsymbol{K}_i \phi_j$$

csoport teljes elosztott tudása (HUN) – group's full distributed knowledge (ENG)

Ez a tudásfogalom nem más, mint az egyéni tudattartalmak (tudáselemek) uniója, vagyis a csoport tagjainak fejében levő tudás valahogyan képzett összessége.

Francis Galton 1907-ben publikált egy cikket a Nature-ben egy Plymout városában tartott állatvásár tapasztalatairól [182]. A látogatókat versenyre invitálták: meg kellett becsülni egy levágandó bika vágás utáni súlyát. A versenynek jelentkezési díja volt, ami miatt gyakorlatilag ki lehetett zárni a "vicces becsléseket", viszont komolynak mondható díjat ajánlottak fel a nyertes számára, ami miatt sokan beszálltak a játékba. A verseny után a közel nyolcszáz játékos által leadott papírokat Galton megkapta, és elemezte a rajtuk található adatokat. Meglepő eredmény született: az 1198 font súlyú állatra leadott becslések mediánja 1207 font lett, ami az egy százalékos hibahatáron belül maradt. Galton a "kísérlet" eredményét a demokratikus szavazás, a Vox Populi elvének igazolásának is szánta, bár ebben nem volt teljesen igaza (mivel itt tényre vonatkozó átlagolásról, a demokratikus szavazás során értékre vonatkozó átlagolásról van szó).

1.10 Dinamikus logika

A **dinamikus logikát** (dynamic logic) eredetileg a számítógépes programok leírására hozták létre, hogy legyen olyan eszköz, amelynek segítségével meg lehet vitatni a programok helyességét, de a nyelv idővel egészen másfajta alkalmazási területeken is hasznosnak bizonyult. A dinamikus logika a modális logika egyik típusa. Központi kategóriája a **program(darab)** fogalma, ami értelmezhető úgy is, hogy valamely cselekvést (action) kell végrehajtania. A program(darab)ot a-val jelöljük. Két deontikus operátort használunk a **szükségszerű** és **lehetséges** operátorok helyett:

$$\Box p \quad \text{helyett:} \quad [a]p$$
$$\Diamond p \quad \text{helyett:} \quad \langle a\rangle p$$

ahol az a program(darab) valamely p állapotra (azt leíró propozícióra) vonatkozik. Az első operátor azt fejezi ki, hogy ha végrehajtjuk az a programot, akkor p szükségszerűen igaz kell legyen, vagyis a szükségszerűen létrehozza p-t, míg a második operátor ennek csak a lehetőségét teremti meg. A programokkal (cselekvésekkel) különböző műveleteket lehet végezni (ezeket értelmezhetjük úgy, hogy 'futtasd le' vagy másként 'tedd meg'). A legfontosabb műveleteket a következőféleképpen jelölhetjük:

(ML.0) $\mathscr{L}_{DL}([\,],\langle\rangle,1,0,\cup,\cap,;,*,?)$, ahol

$[a]p$		ha megteszed a-t, kötelező, hogy p
$\langle a\rangle p$		ha megteszed a t, lehetséges, hogy p
1	SKIP/NOP	ne csinálj semmit, állj le a programmal (terminálj)
0	BLOCK, konstans	ne csinálj semmit
$a\cup b$	egyesítés	tedd meg a-t vagy b-t
$a;b$	komponálás	tedd meg a-t, utána b-t
$a*$	iterálás	ismételten tedd meg (véges sokszor) a-t
$a\cap b$	egyszerre futtatás	tedd meg a-t és b-t
$a?$	tesztelés	ha a igaz, akkor tedd meg, ha hamis, akkor állj le

A műveletek logikai definíciójához szükség van néhány modális logikai axiómára (egyre a modális operátorok közti transzformáció, kettőre a következtetési szabályok miatt):

$[a]p \equiv \neg\langle a\rangle\neg p$	operátorok közti transzformáció
$(\vdash p \wedge \vdash p \to q) \to \vdash q$	modus ponens
$\vdash p \to \vdash [a]p$	szükségességi szabály (necessitation rule)

A fent bemutatott dinamikus logikai műveleteket a következő axiómákkal határozhatjuk meg (ahol a és b programot, p tetszőleges kifejezést, x - változót, e konstansokból és változókból álló kifejezést, $\Phi(x)$?????-t jelent):

(DL.1) $[0]p$

(DL.2) $[1]p \equiv p$

(DL.3) $[a\cup b]p \equiv [a]p \wedge [b]p$

(DL.4) $[a;b]p \equiv [a][b]p$

(DL.5) $[a*]p \equiv p \wedge [a][a*]p$

(DL.6) $p \wedge [a*](p \to [a]p) \to [a*]p$

(DL.7) $[x := e]\Phi(x) = \Phi(e)$

(DL.8) $[p?]q \equiv p \to q$

Az általános axiómák és a műveletekre vonatkozó axiómák alapján levezethetőek azok az egyenlőségek is, amelyek a másik dinamikus logikai operátorra állapítanak meg összefüggéseket:

(DL.9) $\neg\langle 0\rangle p$

(DL.10) $\langle 1\rangle p \equiv p$

(DL.11) $\langle a \cup b\rangle p \equiv \langle a\rangle p \vee \langle b\rangle p$

(DL.12) $\langle a; b\rangle p \equiv \langle a\rangle\langle b\rangle p$

(DL.13) $\langle a*\rangle p \equiv p \vee \langle a\rangle\langle a*\rangle p$

(DL.14) $\langle a*\rangle p \rightarrow p \vee \langle a*\rangle(\neg p \wedge \langle a\ \rangle p)$

(DL.15) $\langle p?\rangle q\ \equiv p \wedge q$

Nem tartozik a könyv témájához, de ha már itt vagyunk, érdemes megjegyezni, hogy a programozás két fontos konstrukciója, az 'if-then elágazás' és a 'while-ciklus' is egyszerűen definiálható a dinamikus logika operátoraival.

(DL.16) IF p THEN a ELSE $b \stackrel{\text{def}}{=} (p?; a) \cup (\neg p?; b)$

(DL.17) WHILE p DO $a \stackrel{\text{def}}{=} (p?; a)*; \neg p?$

Logikából nagyon sokféle van. A mérnökök, informatikusok ezerfélét kidolgoztak a legkülönfélébb célokra, a legkülönbfélébb szintaxissal, a legkülönfélébb képességekkel. A dinamikus logika csak egy a sok közül, ami csak azért kell számunkra, mert a társadalmi cselekvés szempontjából fontos fogalmak közül többet is dinamikus logikai alapon formalizáltak.

1.11 Preferencialogika

Ha azt mondjuk, hogy A-t preferáljuk B-vel szemben, akkor nem teszünk mást, mint valamiféle sorrendet alakítunk ki a két dolog között. A sorrendállítás pedig rendezést jelent. Rendezésből persze többféle van, erről bővebben írtam a könyv relációelméleti részében (↪34.o.). A matematikai rendezésfogalmak azonban túl általánosak, ezért sok jelenségre "ráhúzhatók". Ha speciálisabb rendezési fogalmakat szeretnénk definiálni, szükségünk lehet további tulajdonságokat kikötni. A reflexív, antiszimmetrikus és tranzitív tulajdonságok egyidejű feltételezésével többféle rendezési relációt is jellemezhetünk. A **része** (partitív), a **tartalmazási**, a **nem öregebb** vagy a **kisebb vagy egyenlő** reláció esetén egyaránt fennállnak a jelzett tulajdonságok, és érezzük, hogy szükség lenne további minőségeket hozzájuk rendelni, hogy jól elkülöníthessük őket egymástól. A partitív reláció esetében megtette ezt Peter Simon vagy Archello Varzi. Előbbi könyvet írt a rész fogalmáról [493], utóbbi axiómarendszert dolgozott ki a partitív reláció fogalmára [559], amelyben a három alaptulajdonságon túl még 4-5 további tulajdonságot definiált.[32]

(a_1) $\forall x(\mathbf{P}(x,x))$

(a_2) $\forall x \forall y(\mathbf{P}(x,y) \wedge \mathbf{P}(y,x) \rightarrow x = y)$

(a_3) $\forall x \forall y \forall z(\mathbf{P}(x,y) \wedge \mathbf{P}(y,z) \rightarrow \mathbf{P}(x,z))$

Az első három axióma rögzíti, hogy a $\mathbf{P}$ partitív reláció reflexív (a_1), antiszimmetrikus (a_2) és tranzitív (a_3). A következő lépésben a $\mathbf{P}$ relációval néhány új mereológiai relációt definiálunk, majd az újakkal még újabbakat hozunk létre:

[32] A partitív reláció Varzi-féle axiomatizálását bemutattam magyarul [527].

(1) $\mathbf{PP}(x,y) \stackrel{\text{def}}{=} \forall x \forall y(\mathbf{P}(x,y) \wedge \neg \mathbf{P}(y,x))$

(2) $\mathbf{O}(x,y) \stackrel{\text{def}}{=} \forall x \forall y \exists z(\mathbf{P}(z,x) \wedge \mathbf{P}(z,y))$

(3) $\mathbf{U}(x,y) \stackrel{\text{def}}{=} \forall x \forall y \exists z(\mathbf{P}(x,z) \wedge \mathbf{P}(y,z))$

(4) $\mathbf{OX}(x,y) \stackrel{\text{def}}{=} \forall x \forall y(\mathbf{O}(x,y) \wedge \neg \mathbf{P}(x,y))$

(5) $\mathbf{UX}(x,y) \stackrel{\text{def}}{=} \forall x \forall y(\mathbf{U}(x,y) \wedge \neg \mathbf{P}(x,y))$

(6) $\mathbf{PO}(x,y) \stackrel{\text{def}}{=} \forall x \forall y(\mathbf{OX}(x,y) \wedge \mathbf{OX}(y,x))$

(7) $\mathbf{PU}(x,y) \stackrel{\text{def}}{=} \forall x \forall y(\mathbf{UX}(x,y) \wedge \mathbf{UX}(y,x))$

Sorba definiáljuk a következő relációkat: valódi része (1), átfedése (2), kívüllevősége (3), részhatáros átfedése (4), részhatáros kívüllevősége (5), valódi lefedése (6), valódi kívüllevősége (7). Ezekután a **P** része, **O** átfedése és **U** kívüllevősége relációk segítségével először meghatározhatjuk a kiterjeszthetőség (erős) elvét (a_4), majd az (a_5)-(a_7) mereológiai axiómák elfogadásával definiálhatjuk a zárt (extenzionális) mereológia elméletét.

(a_4) $\forall x \forall y(\neg \mathbf{P}(x,y) \exists z(\mathbf{P}(z,x) \wedge \neg \mathbf{O}(z,y)))$

(a_5) $\forall x \forall y(\mathbf{U}(x,y) \rightarrow \exists z \forall w(\mathbf{O}(w,z) \leftrightarrow (\mathbf{O}(w,x) \vee O(w,y))))$

(a_6) $\forall x \forall y(\mathbf{O}(x,y) \rightarrow \exists z \forall w(\mathbf{P}(w,z) \leftrightarrow \mathbf{P}(w,x) \wedge P(w,y))))$

(a_7) $\forall x \forall y \exists z((\mathbf{P}(z,x) \wedge \neg \mathbf{O}(z,y)) \rightarrow \exists z \forall w(\mathbf{P}(w,z) \leftrightarrow (\mathbf{P}(w,x) \wedge \neg \mathbf{O}(w,y))))$

További axiómák felvételével egyrészt definiálhatjuk az általános (extenzionális) mereológia elméletét (a_8) másrészt meghatározhatjuk az atomos (a_9), illetve atom nélküli mereológiákat (a_{10}).

(a_8) $\exists x \phi \rightarrow \exists z \forall y(\mathbf{O}(y,z) \leftrightarrow (\exists x(\phi \wedge \mathbf{O}(y,x))$

(a_9) $\forall x \exists y(\mathbf{P}(y,x) \wedge \neg \exists z \mathbf{PP}(z,y))$

(a_{10}) $\forall x \exists y(\mathbf{PP}(y,x))$

A mereológia elméletének bemutatásához, megérttetéséhez nyilván nem elégséges, hogy formulákkal definiáljuk az alaprelációkat, meg az axiómákat, és kész. Én mégis ezt tettem meg itt. Azért csak ennyit foglalkoztam a kérdéssel, mert csak azt akartam megmutatni, hogy a RÉSZE reláció meghatározásához milyen logikai apparátust kell megmozgatni. Nem elég rögzíteni az első három relációtulajdonság meglétét, mert azzal csak egy "szimpla" rendezési relációt határozhatunk meg, ami túl általános, még nagyon sokféle dolgot képes lefedni. Láthattuk, hogy további 5-6 axióma kell ahhoz, hogy a partitív reláció specifikumát megtaláljuk.

Ez az egész pedig csak azért érdekes e könyv mondanivalója szempontjából, mert a preferenciarelációra is ki lehet jelenteni, hogy az aszimmetrikus és tranzitív egyben, ám ezzel még biztos nem tudjuk megragadni a preferenciareláció különlegességét, hiszen több más ismert relációt jellemezhetnek ezek a minőségek, ezért keresni kell még további tulajdonságokat hozzá. Ezt végzik el azok a szerzők, akik a preferencialogika különböző elméleteit dolgozzák ki. Georg Henrik Von Wright is, a követői is szokták a preferencia terminusa helyett a – magyarra lefordíthatatlan – **betterness** terminust használni. A

kategória tartalmát talán a meglehetősen kicsavart **jobbnak tartottság** kifejezéssel adhatnánk vissza, de e jelzésen túl, hogy létezik, ezt a terminust a természetes nyelvi szövegben inkább nem használom.[33] Hansson szerint amikor másokról beszélünk, akkor inkább a "preferálja" terminust használjuk, és amikor magunkra utalunk vagy imperszonális módon nyilatkozunk, akkor inkább a "jobb" terminust alkalmazzuk [219].

A preferenciarendezés valamilyen halmaz elemeinek kívánatosság szerinti sorbarendezését jelenti. A preferenciarendezés reláció. Mint minden relációnak, úgy a preferenciarendezésnek is vannak argumentumai, relátumai. A preferencia jelenségéről beszélve könnyen átválthatunk, átváltunk reifikált tárgyalásmódba. Amikor a 'preferencia' terminust használjuk (például feltesszük a kérdést: "Neked mi a preferenciád?"), akkor észrevétlenül a preferenciarendezés eredményére irányul a figyelmünk. Ilyenkor a preferenciarelációt tárgyiasítjuk. A preferenciarendezés fogalma az összehasonlítandó alternatívák viszonyát vonatkozik, nem az összehasonlítás eredményére. Ez a beállítódás arra hasonlít, amikor a "vágy" terminusát használjuk. A vágyunk tárgya egy preferenciareláció reifikátuma. Amikor a vágyunkra gondolunk, már nem a valamivel való összehasonlításra, hanem csak a rendezés eredménye jár a fejünkben. Az angol nyelvben ritkán használt, de létező terminus a 'preferent', magyarban nincs megfelelője, ezért a **preferált alternatíva** kifejezést használjuk a preferenciarendezés előnyben részesített relátumára vonatkozóan.

A preferenciarendezést el kell választani a **preferenciális választástól** (preferential choice) is. A választásainkban valamilyen módon mindig megnyilvánulnak a preferenciáink, de nem minden preferenciánk jelent választást is egyben. Von Wright két példát is ad erre vonatkozóan. Az egyik példa szerint preferálhatjuk ugyan a különféle időjárás jelenségek közül (napos idő, viharos idő, kánikula stb.) egy egyiket a másikkal szemben, de nem választhatunk belőlük. De ugyanez igaz az egészségi állapotunkra is. Hiába tarjuk jobbnak az egészséget a betegségnél, nem áll módunkban választani ezen állapotok közül (legfeljebb csak áttételesen, az életmódválasztásokon keresztül).

A preferencia fogalmilag feltételezi a **diszkrimináció** létezését: ha m alternatívát preferáljuk n-vel szemben, akkor ezennel n-t diszkrimináljuk is egyben. A diszkrimináció (vagy a diszkriminálás) a preferencia (vagy a preferálás) konverze. Természetesen a tárgyiasítás a másik irányban is működhet, vagyis a reifikálás a rendezési reláció másik relátumára a **diszkriminált alternatívára** is vonatkozhat. Ezzel azonban kevesebbet foglalkoznak a szakirodalomban. Amikor a hasznosság (utility) fogalmát elemzik a közgazdászok, akkor olykor felbukkan az a nézet, hogy a hasznosság dipólusos fogalom, tehát a szokásos, pozitív értelméhez képest létezik a másik véglete is: ez a negatív hasznosság, kellemetlenség, károsság (disutility), de ez ritkán történik meg. Pedig a döntéseink során figyelünk mindkét pólusra: a munkába álláskor a munkás hasznosságának számít a munkabére, a kellemetlensége pedig a munkavégzés okozta fáradság, kimerülés.

A preferenciarelációt, a szemantikailag ide kapcsolható további relációkat, valamint az ezek közti összefüggéseket a preferencialogika (BL) segítségével lehet definálni. Mielőtt bemutatnám, hogyan tehetjük meg ezt, előtte még el kell választani a preferenciák két típusát. Ez a megkülönböztetés azért nagyon fontos, mert ez alapján lehet majd a

[33] Praktikus megfontolások miatt azonban a preferenciarelációt leíró formális logika azonosítására, az egyes tételek jelzésére a 'BL' szimbólumot fogom használni, ami a 'Betterness Logic' kifejezés rövidítése. Ezt a megoldást csak azért választottam, mert a logikák azonosítójaként használt jelek között a BL még szabad volt, míg a PL jelzést más logika már korábban "lefoglalta".

döntéseinkben, választásainkban, mérlegeléseinkben tetten érhető kétféle értékválasztási, preferenciarendezési mód, helyzet között különbséget tenni. Elsőként az intrinzikus preferencia fogalmát elemzem, és majd másodjára térek rá az extrinzikus preferenciák tárgyalására.

Intrinzikus preferencia

Von Wright saját esszéjében nem használta ki, de a társadalmi cselekvés modellezése során komoly jelentősége lesz annak a megkülönböztetésnek, amelynek segítségével von Wright definiálta a **intrinzikus** (intrinsic), illetve a **extrinzikus** (extrinsic) preferencia fogalmát.

> „...egy személy például preferálhatja a bordói vörös bort a rajnai fehérrel szemben azért, mert az orvosa azt ajánlotta neki, vagy azért, mert sajátmaga kitapasztalta, hogy az első bor jobbat tesz a gyomrának vagy általában az egészségének. Ebben az esetben az 'ez jobb, mint az' ítélet alapjául vagy indokául szolgál a preferenciának. Azt a preferenciát nevezem extrinzikusnak.
> Előfordulhat azonban az is, hogy ez a személy nem azért preferálja a bordói bort a rajnaival szemben, mert azt gondola, hogy az jobbat tesz neki, hanem pusztán csak azért, mert az elsőt jobban szereti. Ilyenkor az egyik bor szeretete nem indoka, értelme a preferenciájának. ...” [565, 14]

Ez utóbbit nevezi von Wright intrinzikus preferencának. Mondhatnám, hogy ez a megkülönböztetés azért fontos, mert von Wright a preferenciákról szóló esszéjében csak az intrinzikus preferenciákkal foglalkozik. De nyilván nem véletlenül. Von Wright azt akarja megfogni, hogy mi a lényege azoknak a kedveléseknek, preferenciakinyilatkozásoknak, amelyek a legbensőnkből jönnek, amelyek a legkevésbé és a leglassabban változnak, amelyek kinyilvánításakor nem mérlegelünk, “nem gondolkozunk”, amelyek az értékelköteleződéseink legmélyebb rétegét jelentik. Egy későbbi tanulmányában így ír erről.

> „Az x és y közötti preferencia extrinzikus, ha létezik egy olyan (nem cirkuláris) érv, amely megadja, x-t *miért* preferáltuk y-nal szemben. Minden más esetben a preferencia intrinzikus. Azt, hogy x intrinzikusan preferált y-nal szemben, úgy is kifejezhetjük, hogy azt mondjuk, x 'önmagában' vagy 'önmagáért' preferált y-nal szemben. Az intrinzikus preferencia kinyilvánítása, a legtöbbb esetben, kifejezi azt, amit *szeretünk*.” [566]

Ezzel a megkülönböztetéssel persze sok esetben nagyon nehéz mit kezdeni. Az, hogy mikor számít valami extrinzikus vagy intrinzikus preferenciának, mikor jön valami belülről, mikor támaszkodunk külső észérvekre, sok esetben nagyon nehezen dönthető el, ha eldönthető egyáltalán. De egyelőre nem a nehézségekkel kell foglalkozni, hanem a jelenségek jobb megértésével, amihez nyilván hasznos eszköz az ideáltípusokra támaszkodó megközelítés alkalmazása.

Igazából minden preferenciát négy-argumentumú relációként kellene kifejeznünk. Az összehasonlítandó alternatívák mellett ugyanis mindig meg kellene adnunk, hogy az adott preferencia kinek az értékítéletét fejezi és mikor. Hiszen, tudjuk jól, ugyanazon dolgokra vonatkozó preferenciarendezés más lehet attól függően, hogy kiknek a preferenciájáról van szó, illetve egy személy preferenciája is változhat az időben. Ennek ellenére a preferenciával foglalkozó elméletek nem veszik figyelembe az ágencia és az idő dimenzióit. Leginkább azért nem, mert a preferencia jelenségének megértéséhez érdemben nem (vagy

csak keveset) ad hozzá e két szempont figyelembe vétele. Ezért a továbbiakban személytelenített és időtlenített (imperszonális és atemporális) módon fogom én is elemezni a preferencia fogalmát. A teljeskörű leírást biztosító mondat a következő lenne:

> A preferálja m-t n-nel szemben t időpontban

Ebből az állításból azonban – elhagyva a személyi és időbeli vonatkozásokat – csak annyit hagyunk meg a továbbiakban, hogy:

> valaki preferálja m-t n-nel szemben

Ez a megoldás egyszerűbbé teszi a jelenség tárgyalását. Von Wright is erre hivatkozva fogadta el ezt a szűkítést. Ezen túl azonban tett még egy olyan további leegyszerűsítést is, amit a preferenciával foglalkozó későbbi kutatók már nem fogadtak el. Von Wright ugyanis nemcsak az extrinzikus preferenciákra zárta ki a tárgyalásból, de lemondott az intrinzikus preferenciák változásainak vizsgálatáról is:

> „Csak azokat a preferenciákat tanulmányozzuk, amelyek egy személy intrinzikus preferenciái egy adott pillanatban. Ezzel a tárgyalásból kizárjuk mind a preferenciákat támogató érveket, mind a preferenciák változásának lehetőségét.” [565, 23]

A fenti szűkítések miatt von Wright preferencialogikája statikus maradt, mivel nem tudta leírni a preferenciák változásait. Sem a mérlegelés nélküli, “belülről jövő” preferenciák változásait, sem a preferenciák mérlegelés, befolyásolás hatására bekövetkező, “külsődleges” változásait. A von wright-i preferencialogikát ezért ki kell egészíteni az extrinzikus preferenciákat leíró logikával, illetve bővítenünk kell olyan módon, hogy dinamikus preferencialogikához jussunk. Ezt azonban csak később végezzük el. Előtte definálnunk kell egy nyelvet, ami az intrinzikus preferenciák leírására alkalmas lehet.

(BL.0) $\mathscr{L}_{BL}(\mathscr{L}_{QL}, \textsc{pref}, \mathbf{P})$, ahol

$\mathscr{L}_{QL}$	predikátumlogika
$\mathbf{P}$	preferenciareláció

Bár a preferencia fogalmának első formális elméletét Sören Halldén dolgozta ki még 1957-ben [205], a preferencia fogalmának tárgyalásakor nem az ő könyve lett a legtöbbet hivatkozott forrás. A mai napig sokkal inkább Georg Henrik von Wright ‘The Logic of Preference’ című esszéje számít hivatkozási alapnak e tárgyban [565]. Halldén is, von Wright is szintaktikai alapon definálta a preferencia fogalmát, ahogy a kortársak közül még sokan mások, aztán a hetvenes években már megjelentek azok az elméletek, amelyek – döntő módon – a lehetséges világ szemantika keretein belül igyekeztek modellt adni ennek a fogalomnak [216]. A következőkben előbb bemutatom, hogy milyen tulajdonságokkal próbálták meg definiálni a preferenciarelációt a legfontosabb szerzők, majd egyenként elemzem, hogy az egyes tulajdonságok elfogadásának milyen következményei lehetnek, végül áttérek arra, hogy a modális logika segítségével hogyan lehet teljesebbé tenni a preferencia fogalmának leírását.

Az intrinzikus preferencia tulajdonságai

Von Wright abból indult ki, hogy bár a preferenciarelációt sok minden “dolog” között alkalmazhatjuk, elég csak a világállapotok közti preferenciarelációt vizsgálni, mert minden

más eset visszavezethető erre.[34] Ezután megadott öt axiómát, amivel definiálni lehet a preferencia fogalmát.

Von Wright első lépésben két relációalgebrai tulajdonság (az aszimetrikusság és a tranzitivitás) tételezésével a preferenciarelációt a SZIGORÚ RENDEZÉSI RELÁCIÓ egyik típusának minősítette. Ezután pedig megállapított három további ekvivalenciát, amelyek segítségével transzformációs (konjunkciós, disztribúciós, amplifikációs) műveletek végezhetők el. Az öt kritérium formulája a következő:

(BL.1) $\text{STRICT_PREFERENCE}(m,n) \equiv \mathbf{P}(m,n) \equiv \mathbf{P} \equiv \mathbf{R}^{>} \equiv m > n \equiv >$

szigorú preferenciarendezés (HUN) – erős preferenciarendezés (HUN) – preferenciarendezés (HUN) – preferencia (HUN) – strict preference (ENG) – strong preference (ENG)

(BL.2) $\text{ASYMMETRIC}(\mathbf{P}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n(\mathbf{P}(m,n) \rightarrow \neg\mathbf{P}(n,m))$

P reláció **aszimmetrikus**

(BL.3) $\text{TRANSITIVE}(\mathbf{P}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(\mathbf{P}(m,n) \wedge \mathbf{P}(n,o) \rightarrow \mathbf{P}(m,o))$

P reláció **tranzitív**

(BL.4) $\forall m \forall n(\mathbf{P}(m,n) \leftrightarrow \mathbf{P}(m \wedge \neg n, \neg m \wedge n))$

P relációra érvényes a **konjunktív expanzió** elve

(BL.5) $\forall m \forall n \forall o \forall p(\mathbf{P}(m \vee n, o \vee p) \leftrightarrow (\mathbf{P}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{P}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p)$
$\wedge\ \mathbf{P}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{P}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p)))$

P relációra érvényes a **disztribúció** elve

(BL.6) $\forall m \forall n \forall o(\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge o, n \wedge o) \wedge \mathbf{P}(m \wedge \neg o, n \wedge \neg o)))$

P relációra érvényes az **amplifikáció** elve

A preferenciareláció fenti tulajdonságaira támaszkodva von Wright további két fontos diadikus preferenciarelációt határozott meg, ami azonban – tágabb kontextusba helyezve a dolgokat – három relációval hozható kapcsolatba. Az egyik az **indifferencia** reláció, amit von Wright azzal a minőséggel definál, hogy két állapot egyike sem preferált a másikkal szemben [565, 52]. Ezen a ponton vigyázni kell, mert könnyen zavart okozhat az a tény, hogy ugyanezt a relációt más szerzők egészen más szemantikával értelmezik, és hol összehasonlíthatatlansági relációnak [14], hol 'ugyanolyan, mint' relációnak nevezik [98]. Von Wright is tisztában van azzal, hogy kétfajta relációról van szó, ezért a most tárgyalt relációnak kicsit később a **gyenge indifferencia** nevet adja [565]. A későbbi félreértések elkerülése végett von Wright névadási javaslatát csak a jelzős szerkezetben viszem tovább. Mindez persze csak terminológiai kérdés, jelentését tekintve az új relációt egyféleképpen vezethetjük be a formális modellünkbe.

(BL.7) $\text{INCOMPARABILITY}(m,n) \equiv \mathbf{R}^{\#}(m,n) \equiv \mathbf{R}^{\#} \equiv \mathbf{J}(m,n) \equiv \mathbf{J} \stackrel{\text{def}}{=} \forall m \forall n(\neg\mathbf{P}(m,n) \wedge \neg\mathbf{P}(n,m))$

[34] Von Wright a preferenciák lehetséges irányultságaiként az állapotok mellett a dolgokat, eszközöket, illetve a cselekvéseket említi még. Abban nem nagyon van vita a preferenciával foglalkozó szerzők között, hogy elég az állapotokkal foglalkozni, mert minden másra irányuló preferencia kifejezhető ezen az alapon. Ennek a döntésnek komoly előnye, hogy az állapotokra vonatkozóan könnyű kijelentéseket tenni, és ezekre azonnal alkalmazható a propozícionális logika teljes eszköztára.

összehasonlíthatatlanság (HUN) – gyenge közömbösség (HUN) – gyenge indifferencia (HUN) – incomparability (ENG) – weak indifference (ENG) – ugyanolyan értékes, mint (HUN) – same value as (ENG)

A közömbösségi reláció terminusra vonatkozóan pedig, amit von Wright az erős indifferenciának nevezett el, meghagyom a sokkal inkább elterjedt – jelző nélküli – indifferencia vagy közömbösség megnevezést. Ebben az esetben tagadás nélkül állítjuk az egyidejűleg két irányú preferálás tényét.

(BL.8) $\text{INDIFFERENCE}(m,n) \equiv \mathbf{I}(m,n) \equiv \mathbf{I} \equiv \mathbf{R}^{\equiv} \equiv \mathrel{\text{III}}(m,n) \equiv m \mathrel{\text{III}} n \equiv \mathrel{\text{III}} \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{P}(m,n) \wedge \mathbf{P}(n,m))$

indifferencia (HUN) – erős indifferencia (HUN) – erős közömbösség (HUN) – strong indifference (ENG) – indifference (ENG) – értékegyenlőség (HUN) – value equality (ENG) – ugyanolyan, mint (HUN) – ugyanolyan jó, mint (HUN) – equally good as (ENG) – same as (ENG)

Utóbbi relációra vonatkozóan von Wright bevezetett még egy újabb megnevezést is, és **értékegyenlőség** relációnak nevezte el, de ezt a terminust nem fogom használni, csak a történeti érdekesség miatt említem meg. Von Wright a közömbösségi relációt nem a fent megadott képlet alapján definiálta, hanem ahelyett ugyanolyan axiómákat rögzített, mint amilyeneket a preferenciareláció bevezetésekor kötött ki.

(BL.9) $\text{SYMMETRIC}(\mathbf{I}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{I}(m,n) \rightarrow \mathbf{I}(n,m))$

I reláció **szimmetrikus**

(BL.10) $\text{TRANSITIVE}(\mathbf{I}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o (\mathbf{I}(m,n) \wedge \mathbf{I}(n,o) \rightarrow \mathbf{I}(m,o))$

I reláció **tranzitív**

(BL.11) $\forall m \forall n (\mathbf{I}(m,n) \leftrightarrow \mathbf{I}(m \wedge \neg n, \neg m \wedge n))$

I relációra érvényes a **konjunktív expanzió** elve

(BL.12) $\forall m \forall n \forall o \forall p (\mathbf{I}(m \vee n, o \vee p) \leftrightarrow \mathbf{I}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{I}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p) \wedge \mathbf{I}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{I}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p))$

I relációra érvényes a **disztribúció** elve

(BL.13) $\forall m \forall n \forall o (\mathbf{I}(m,n) \leftrightarrow \mathbf{I}(m \wedge o, n \wedge o) \wedge \mathbf{I}(m \wedge \neg o, n \wedge \neg o))$

I relációra érvényes az **amplifikáció** elve

A fenti formulák együttes igazsága – von Wright intuíciója szerint – szükséges és elégséges feltétel az értékegyenlőségi reláció meghatározásához. A preferenciareláció axiómáival összevetve látható, hogy az ötből tétel közül csak az elsőben különbözik a két reláció: a preferencia aszimmetrikus, az értékegyenlőség szimmetrikus. Ez azt is jelenti egyben, hogy a von Wright értelmezésben vett preferenciareláció és értékegyenlőség reláció segítségével előállíthatunk egy olyan preferenciarendezési fogalmat is, amelyben a preferencia fogalma nem a 'szigorú értelemben vett jobb', hanem a 'legalább olyan jó' jelentés mellett értelmezhető.

(BL.14) $\text{WEAK_PREFERENCE}(m,n) \equiv \mathbf{R}(m,n) \equiv \mathbf{R} \equiv \mathbf{R}^{\geq}(m,n) \equiv \mathbf{R}^{\geq} \equiv m \geq n \equiv \geq \stackrel{\text{def}}{=} \forall m \forall n (\mathbf{P}(m,n) \vee \mathbf{I}(m,n))$

gyenge preferencia (HUN) – preferencia (HUN) – legalább olyan jó, mint (HUN) – jobb vagy ugyanolyan (HUN) – weak preference (ENG) – preference (ENG)

A legalább olyan jó értelmezésű **R** gyenge preferenciarelációra ugyanazokat a tulajdonságokat rá lehet vetíteni, mint amiket a szigorú preferenciarendezés, illetve közömbösségi relációval szemben von Wright kikötött. A három preferenciarelációt, a szigorú és gyenge preferenciarendezést, valamint a közömbösségi relációt (és olykor még az összehasonlíthatatlansági relációt is) együttesen (**P**, **I**, **J**) **preferenciastruktúrának** is szokták nevezni, és bizonyos tételeket a struktúra egészére vonatkozóan szoktak kimondani [451].

A preferenciareláció eddigi axiómái alapján von Wright már meghatározhatónak gondolta két fontos monadikus (vagy unáris) preferenciareláció, a **jó** (GOOD) és a **rossz** (BAD) fogalmát is. Mivel von Wright a preferenciát a világállapotokra vonatkoztatva definiálta, ezért ebben az elméletben a jó és rossz kategóriája is (világ)állapotokra vonatkozóan érvényes. Szerinte a jó (állapot) így tehát az az állapot, amely feltétel nélkül preferált saját ellentétével szemben, míg a rossz (állapot) az, amelynek az ellentéte feltétel nélkül preferált magával szemben.[35] A meghatározások megfelelő formulái [565].

(BL.15) $\text{GOOD}(m) \equiv \mathbf{G}'(m) \equiv \mathbf{G}' \stackrel{\text{def}}{=} \mathbf{P}(m, \neg m)$

(BL.16) $\text{BAD}(m) \equiv \mathbf{B}'(m) \equiv \mathbf{B}' \stackrel{\text{def}}{=} \mathbf{P}(\neg m, m)$

Voltak, akik elfogadták ezt a meghatározást, és voltak, akik elutasították (erre később még visszatérek). Ez persze általában is igaz. Halldén és Von Wright preferencialogikájára sokan reagáltak, bizonyos pontokon támogató, máshol kritikus módon. Voltak szerzők nem osztották Von Wright intuícióit, és részben vagy teljesen más axiómákat tartottak meghatározónak. Roderick M. Chisholm és Ernest Sosa egy egészen eltérő formalizmust dolgozott ki von Wright elméletéhez képest [98]. Az aszimmetrikusság tulajdonságait fenntartották ők is, de ezután von Wright megoldásához képest teljesen más axiómákat javasoltak. Másodiknak a tranzitivitás duális fogalmát, a negatív vették fel az axiómák közé.[36]

(BL.17) $\text{CONTRA-TRANSITIVE}(\mathbf{P}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\neg\mathbf{P}(m,n) \wedge \neg\mathbf{P}(n,o)) \rightarrow \neg\mathbf{P}(m,o))$

negatívan tranzitív (HUN) – negatively transitive (ENG) – contra-transitive (ENG)

Ezután definiáltak még további három axiómát, de mielőtt bemutatnám ezeket, előtte érdemesebbnek tűnik felsorolni azt, hogy milyen további monadikus és diadikus preferenciarelációkat vezettek még be Chisholmék az első két axiómájukra támaszkodva. Azért érdemes előbb megismerni ezeket, mert az ismeretükben könnyebb lesz majd értelmezni a további axiómákat. Lássuk tehát az öt új Chisholm-Sosa-féle preferenciarelációt.[37]

(BL.18) $\text{SAME_VALUE_AS}(\mathbf{S}(m,n)) \equiv \mathbf{S}(m,n) \stackrel{\text{def}}{=} \forall m \forall n(\neg\mathbf{P}(m,n) \wedge \neg\mathbf{P}(n,m))$

ugyanolyan (HUN) – ugyanolyan értékes, mint (HUN) – same value as (ENG)

(BL.19) $\text{INDIFFERENT}(\mathbf{I}''(m)) \equiv \mathbf{I}''(m) \stackrel{\text{def}}{=} \forall m(\neg\mathbf{P}(m,\neg m) \wedge \neg\mathbf{P}(\neg m, m))$

indifferens (HUN) – közömbös (HUN) – indifferent (ENG)

[35] Von Wright egy másik tanulmányában megjegyezte, hogy ezt a meghatározást először A.P. Brogan dolgozta ki még 1919-ben [566].

[36] Chisholmék nem használtak semmilyen megnevezést a szóbanforgó tulajdonságra. A formulájukat azonban ekvivalens alakra lehet hozni a **negatívan tranzitív** relációtulajdonság képletével (↬31.o.), ahogy azt M. Roubens és Ph. Vincke megmutatta [451, 3]. Ezért használom itt ezt a megnevezést.

[37] Chisholm és Sosa az öt preferenciareláció megnevezésébe mindig beleírta, beleértette az 'intrinzikusan' jelzőt, amit én itt – kényelmi okok miatt – elhagyok.

(BL.20) $\text{NEUTRAL}(\mathbf{N}''(m)) \equiv \mathbf{N}''(m) \stackrel{\text{def}}{=} \forall m \exists n(\mathbf{I}(m) \wedge \mathbf{S}(m,n))$

neutrális (HUN) – semleges (HUN) – neutral (ENG)

(BL.21) $\text{GOOD}(\mathbf{G}''(m)) \equiv \mathbf{G}''(m) \stackrel{\text{def}}{=} \forall m \exists n(\mathbf{I}(m) \wedge \mathbf{P}(m,n))$

jó (HUN) – good (ENG)

(BL.22) $\text{BAD}(\mathbf{B}''(m)) \equiv \mathbf{B}''(m) \stackrel{\text{def}}{=} \forall m \exists n(\mathbf{I}(m) \wedge \mathbf{P}(n,m))$

rossz (HUN) – bad (ENG)

A továbblépés előtt érdemes itt pár megjegyzést tenni. Az ugyanolyan, mint reláció megegyezik a korábban bemutatott von wrighti gyenge indifferencia relációval (↫84.o.). A fenti öt relációból csak az első diadikus, a többi monadikus, vagyis csak mindegyiknek egy argumentuma van. Arra is figyelni kell nagyon, hogy amíg Chisholm és Sosa itt mindegyikhez egy-egy pontos és egyértelmű jelentés rendel, addig a köznyelv ezeket a terminusokat sokféle értelemben használja.

Az $\mathbf{S}(m,n)$ ugyanolyan és az $\mathbf{I}(m)$ indifferens reláció csak abban különbözik egymástól hogy az első két-, a második egyargumentumú. Két állapot (alternatíva) ugyanolyan, ha egyik sem jobb, mint a másik, egy állapot pedig önmagában indifferens, ha nem jobb a tagadásánál, és a tagadása sem jobb az állapotnál. Az $\mathbf{S}(m,n)$ és az $\mathbf{I}(m)$ segítségével lehet definiálni a neutrális állapotot. E fogalom alatt Chisholm és Sosa egy olyan m indifferens állapotot értenek ($\mathbf{I}(m)$), amely legalább egy másik n állapothoz képest ugyanolyan értékű ($\mathbf{S}(m,n)$). Hasonló logika mentén a $\mathbf{G}(m)$ jó az az $\mathbf{I}(m)$ indifferens állapot, amely legalább egy másik állapothoz képest jobb ($\mathbf{P}(m,n)$), a $\mathbf{B}(m)$ rossz állapot pedig az az $\mathbf{I}(m)$ indifferens állapot, amelyhez képest legalább egy másik állapot jobb ($\mathbf{P}(n,m)$).[38] Ezek alapján könnyebben meg lehet majd érteni, milyen minőségeket akartak Chisholmék az alábbi három axiómájukkal megragadni.

(BL.23) $\forall m \forall n((\neg\mathbf{P}(m,\neg m) \wedge \neg\mathbf{P}(\neg m,m) \wedge \neg\mathbf{P}(n,\neg n) \wedge \neg\mathbf{P}(\neg n,n)) \rightarrow (\neg\mathbf{P}(m,n) \wedge \neg\mathbf{P}(n,m)))$

(BL.24) $\forall m \forall n((\neg\mathbf{P}(n,\neg n) \wedge \neg\mathbf{P}(\neg n,n) \rightarrow \mathbf{P}(m,n)) \rightarrow \mathbf{P}(m,\neg m))$

(BL.25) $\forall m \forall n((\neg\mathbf{P}(n,\neg n) \wedge \neg\mathbf{P}(\neg n,n)) \rightarrow \mathbf{P}(n,\neg m) \rightarrow \mathbf{P}(m,\neg m))$

Ha fenti három formula közül az elsőbe behelyettesítjük az $\mathbf{I}(m)$ és $\mathbf{S}(m,n)$ relációk képleteit, akkor egy jóval egyszerűbb, áttekinthetőbb és így könnyebben értelemezhető formulához jutunk.

(BL.26) $\forall m \forall n((\mathbf{I}(m) \wedge \mathbf{I}(n)) \rightarrow \mathbf{S}(m,n))$

A képletet úgy értelmezhetjük, hogy ha két állapot indifferens önmagában, akkor egymáshoz képest ugyanolyan értékűek lesznek. A második és harmadik axióma képletébe is behelyettesítve az $\mathbf{I}(m)$ reláció formuláját, újra csak egyszerűbb eredményt kapunk.[39]

[38] Újra hangsúlyozom, hogy a köznyelvben ezeket a fogalmakat sokkal pongyolább módon használjuk, nem teszünk ilyen éles és pontos különbséget e terminusok között.

[39] Egy másik tanulmányban Chisholm és Sosa ezt a két axiómát egy közös diszjunktív formulába összevonva fejezte ki [97].

(BL.27) $\forall m \forall n((\mathbf{I}(n) \rightarrow \mathbf{P}(m,n)) \rightarrow \mathbf{P}(m,\neg m))$

(BL.28) $\forall m \forall n((\mathbf{I}(m) \rightarrow \mathbf{P}(n,\neg m)) \rightarrow \mathbf{P}(m,\neg m))$

Az első formulát úgy interpretálhatjuk, hogy ha van egy n indifferens állapot, és ebből az következik, hogy egy másik m állapot preferált n-hez képest, akkor az m állapot preferált a saját $\neg m$ tagadásához képest. A második formula szerint ha az n indifferens állapot létezéséből következik, hogy ez az n állapot preferált egy másik m állapot tagadásához, tehát $\neg m$-hez képest, akkor az m állapot preferált a saját $\neg m$ tagadásához képest.

Láthattuk tehát, hogy Chisholm és Sosa öt axiómája – egy kivételével – teljesen eltér von Wright elméletétől, de hogy ennek mi az oka, a vitában kinek milyen érvei voltak, erre a kérdésre nem egyszerre és nem is itt válaszolok, hanem az elméletek által kezelt egyes problémákra különö fejezeteket szentelve fogom összehasonlítani a különböző nézőpontokat.

Itt érdemes röviden áttekinteni, hogy Bengt Hansson hogyan kezelte a preferenciareláció fogalmát [213]. Hansson először megmutatta, hogy mind a von Wright, mind a Chisholm-Sosa szerzőpár által lefektetett elmélet túl erős, és a preferenciareláció mindennapi használatával kapcsolatos intuíciónknak több ponton ellentmondó következményei vannak. Ezért Bengt Hansson gyengébb axiómákat vezetett be, bár tanulmánya végén ezekre vonatkozóan is érvényesnek tartotta azt az elvárását, hogy a preferenciák logikájának kidolgozásához talán erősebb logikai nyelvre lehet szükség. Bengt Hansson a preferenciarelációt a 'jobb vagy ugyanolyan jó' értelmeben fogta fel, tehát az $\mathbf{R}(m,n)$ gyenge preferenciarendezést elemezte, és erre a relációra kikötött két rendezési, illetve két disztribúciós axiómát. Az első bengt hanssoni tétel a tranzitivitás, a második a **totalitás** tulajdonságát rögzítette. A tranzitivitási feltételt már felvettük a listába, a totalitás formulája így néz ki:

(BL.29) $\forall m \forall n(\mathbf{R}(m,n) \vee \mathbf{R}(n,m))$

R reláció **erősen összefüggő** (totális)

Hansson totalitási kritériuma megfelel a relációk tárgyalása során bemutatott erősen összefüggő tulajdonságnak (↪29.o.), a tranzitivitás és totalitás alapján pedig azt mondhatjuk, hogy a bengt hanssoni **R** reláció totális (összefüggő) előrendezés (↪34.o.). A következő lépésben Hansson először megmutatta, hogy milyen továbbbi tételek vezethetők le a két axiómájából, majd – miután több, túl erős, elfogadhatatlan következményekkel járó változatot kipróbált és elvetett – javasolt két gyengített disztribúciós axiómát.

(BL.30) $\forall m \forall n(\mathbf{R}(m,n) \wedge \mathbf{R}(m,o)] \rightarrow \mathbf{R}(m,n \vee o))$

(BL.31) $\forall m \forall n(\mathbf{R}(m,o) \wedge \mathbf{R}(n,o)] \rightarrow \mathbf{R}(m \vee n,o))$

A két tétel egy balról és egy jobbról érvényes disztribúciós kritériumot fogalmaz meg. Az első formula azt írja le, hogy ha egy m állapot preferáltabb mint két másik (n vagy o),

akkor az m állapot a két másik (n és o) állapot diszjunkciójánál is preferáltabb. A második tétel ennek a fordítottját követeli meg, vagyis azt, hogy ha egy o állapotnál két másik (m vagy n) is preferáltabb, akkor azok konjunkciója is preferáltabb az o állapothoz képest. Beng Hansson hosszan érvelt amellett, hogy végül miért ezeket a tételeket fogadta el, erre én is visszatérek a disztribúciós elv tárgyalásakor.

A preferencialogika történetéhez hozzátartozik még, hogy von Wright egy évtizeddel később visszatért a kérdés tárgyalására. Az 1963-as könyvével kapcsolatban megjelent kritikák, alternatív formalizálási kísérletek után újra elővette a preferencia problémakörét, és 1972-ben részben újraformálta elméletét [566]. Saját értékelése szerint nem is annyira a kritikákat fogadta meg, mint saját maga törekedett egyfelől a kérdések egyértelműsítésére, másfelől a válaszok egyszerűsítésére. Tanulmányában von Wright először felsorolta azokat a kérdéseket, amelyeket minél pontosabban tisztázni kell egy preferencialogika kidolgozásakor, majd megállapította, hogy nem igazán lát esélyt arra, hogy ki lehessen dolgozni egy széles körben elfogadott elméletet. Inkább csak azt az állapotot tartotta elérhetőnek, hogy a különböző intuíciók mentén bár szabatos, de párhuzamos értelmezések alakulhassanak ki. Von Wright újra megpróbálkozott a preferencia fogalmának formalizálásával. Von Wright az új elmélet központjába is a **P** szigorú preferenciarendezést tette. Első axiómaként felvette az aszimetrikusságot, másodikként pedig egy új elvet rögzített, amit az (érték)összehasonlíthatóság elvének nevezett el. Ez az összehasonlítható relációtulajdonsággal egyezik meg (↪30.o.).

(BL.32) $\text{COMPARABLE}(\mathbf{P}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o(\mathbf{P}(m,n) \rightarrow (\mathbf{P}(m,o) \vee \mathbf{P}(o,n)))$

$\mathbf{P}(m,n)$ preferenciareláció **összehasonlítható**

Az összehasonlíthatóság elvének felvétele egyben azt is jelentette, hogy von Wright nem írta elő a – sokat vitatott – tranzitivitás tulajdonságának fennállását, ami viszont a két axiómából már levezethető vált. Ez az alábbi módon látható be:

1) $\mathbf{P}(n,o) \rightarrow \neg\mathbf{P}(o,n)$
2) $(\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow (\mathbf{P}(m,n) \wedge \neg\mathbf{P}(o,n))$
3) $(\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow ((\mathbf{P}(o,n) \vee \mathbf{P}(m,o)) \wedge \neg\mathbf{P}(o,n))$
4) $(\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow (\mathbf{P}(m,o) \wedge \neg\mathbf{P}(o,n))$
5) $(\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow \mathbf{P}(m,o)$

Első lépésben felvesszük a $\mathbf{P}(n,o)$ aszimmetrikusságát (1. axióma). Második lépésben a kondicionális mindkét oldalát kibővítjük ugyanazzal a $\mathbf{P}(m,n)$ taggal (Propozicionális Logika: konjunkció bevezetése). Harmadik lépésben a kondicionális jobb oldalán levő $\mathbf{P}(m,n)$ helyébe behelyettesítjük az összehasonlíthatóság képletének jobb oldalát (2. axióma). Negyedik lépésben elhagyjuk a kondicionális következményéből a jobb oldali $\neg\mathbf{P}(o,n)$ összetevőt (PL: konjunkció eliminálása). Ezzel a tranzitivitás képletéhez jutunk. Megállapítható tehát, hogy a $\mathbf{P}(m,n)$ szigorú preferenciarendezés aszimmetrikus, irreflexív és tranzitív. Ezután von Wright bevezette a $\mathbf{J}(m,n)$ (gyenge) indifferenciarelációt (valójában az összehasonlíthatatlansági relációt (↪83.o.), és az eddig axiómák és tételek alapján levezette, hogy $\mathbf{J}(m,n)$ reflexív, tranzitív és szimmetrikus reláció. A preferenciarendezés és a gyenge indifferencia reláció segítségével pedig kimondta az (érték)-ekvivalencia elvét, amely az alábbi háromféle állapot lehetőségét tételezi:

(BL.33) $\mathbf{J}(m,o) \wedge \mathbf{J}(o,n) \rightarrow \mathbf{J}(m,n)$
$\mathbf{P}(m,n) \wedge \mathbf{J}(m,o \rightarrow \mathbf{P}(o,n))$
$\mathbf{P}(m,n) \wedge \mathbf{J}(n,o) \rightarrow \mathbf{P}(m,o)$

P-re és **J**-re érvényes az **(érték)-ekvivalencia elve** (HUN) – principle of (value)-equivalence (ENG)

Az ekvivalencia elve az axiómákból levezethető következmény, de önmaga nem axióma. Von Wright az aszimmetricitáson és az összehasonlíthatóságon túl több axiómát már nem definiált, tehát az 1963-as modelljéhez képest jóval gyengébb axiómarendszert javasolt. Úgy tűnik, hogy leginkább Bengt Hansson kritikája hatott leginkább von Wrightre, mert az átalakított modellje sokkal jobban hasonlít Hansson javaslatához, mint akár a saját korábbi, akár Chisholmék elképzeléseihez.

Bár von Wright sosem tartotta önálló axiómának a kontrapozíció elvét (↬39.o.), de azért megemlítette azt az új tanulmányában. Ezt az elvet Halldén javasolta felvenni a preferencia modelljébe [205]. Von Wright [566] nem építette be ugyan az új axiómarendszerébe, de az a tény, hogy alkalmazta, arra utal, hogy úgy gondolta, a preferenciára vonatkozóan érvényesnek lehet tartani. Ezt az elvet Chisholm és Sosa támadta (akkor elsősorban mint Halldén javaslatát), de hogy miért és hogyan kritizálták a kontrapozíció elvét, azt később mutatom be (↬96.o.). Egyelőre álljon itt magának az elvnek a formulája.

(BL.34) $\forall m \forall n(\mathbf{P}(m,n) \leftrightarrow ((\mathbf{P}(\neg n, \neg m)))$

P preferenciarelációra érvényes a **kontrapozíció elve**

Az 1972-es tanulmányában von Wright nagy teret szentelt annak, hogy megmutassa, miként lehet leírni a lehetséges világ szemantika alapján a ceteris paribus preferencia fogalmát. Erre a kérdésre külön fejezetben térek ki (↬111.o.), ezért itt nm tárgyalom. A cikke végén nekifutott újra annak a feladatnak, hogy hogyan lehet az olyan monadikus fogalmat, mint a ‘jó’ vagy a ‘rossz’ kategóriája definiálni a ‘jobb’ fogalom segítségével. Von Wright az előbbi kettőt érték-abszolút, az utóbbit összehasonlító-érték fogalomnak tartotta, és másként definálta őket az 1963-as megoldásához képest [566].

Von Wright felvetette, hogy a jó és a rossz fogalmának meghatározásához G.E. Moore üres világ (empty world) fogalmára lehetne támaszkodni, csak ehhez tisztázni kéne, mit is jelent pontosan ez a fogalom. Von Wright javaslata az volt, hogy az üres világ fogalmát az **ellentmondással** (↬15.o.) lehetne definiálni, mert az logikailag lehetetlen világ, ahol semmi sem igaz és semmi sem hamis. Ezt a világot a lehetséges világok sorában a nulladiknak tekintve (0-világ) és w_0-val jelölve olyan fogalomhoz jutunk von Wright szerint, amely alapján mondhatjuk, hogy jó az a w_x világ, amelyre igaz, hogy $\mathbf{P}(x,0)$, rossz az a w_y világ, amelyre igaz, hogy $\mathbf{P}(0,y)$. Mindezt a totális állapotokra írhatnánk így. Egy tetszőleges egyedi m állapotra vonatkoztatva a jót és a rosszat a következőképpen határozhatnánk meg: $\mathbf{P}_C(m,0)$ a jó, $\mathbf{P}_C(0,m)$ pedig a rossz definíciója.

Preferenciastruktúra

Az eddigiekből már kiderülhetett, de tézisszerűen még nem mondtam ki, hogy a preferencia jelenségét nem egyetlen, hanem inkább több preferenciareláció segítségével írhatjuk le igazán. Első nekifutásra tán még tűnhetne úgy, hogy a preferenciarendezés fogalma jelenti

a jelenség lényegét, de ez nem így van. A különböző preferencialogikái már abba is különböznek, hogy a szigorú vagy a gyenge preferenciarendezést teszik-e az elmélet középpontjába. De kiderült az is, hogy az indifferencia vagy az összehasonlíthatatlnasági relációknak is olykor komoly szerepe van. Az elméletek ezért mozdultak el abba az irányba, hogy ne egyik vagy másik preferenciarelációról beszéljünk, hanem preferenciastruktúráról. Ez annyit jelent, hogy egyfelől általában vett **preferenciarelációról** beszélünk, és megkülönböztetjük egymástól ennek típusait (**R** gyenge és **P** erős rendezés, **I** indifferencia, **J** összehasonlíthatatlanság)[40], másfelől ezeket a preferenciareláció-típusokat, valamint a köztük értelmezett műveleteket egy struktúrába összefogva beszélünk **preferenciastruktúráról**. Kiindulásként vegyünk egy A (tartó)halmazon értelmezett $(\mathbf{P}, \mathbf{I}, \mathbf{J})$ preferenciastruktúrát, és nézzük meg, milyen tulajdonságokat definiálhatunk ezen a struktúrán [451].

(BL.35) $\textsc{irreflexive}(\mathbf{P}) \wedge \textsc{reflexive}(\mathbf{I}) \wedge \textsc{irreflexive}(\mathbf{J})$

P és **J** irreflexív, **I** reflexív

(BL.36) $\textsc{asymmetrical}(\mathbf{P}) \wedge \textsc{symmetrical}(\mathbf{I}) \wedge \textsc{symmetrical}(\mathbf{J})$

P aszimmetrikus, **I** és **J** szimmetrikus

(BL.37) $\mathbf{P} \cap \mathbf{I} = \varnothing \wedge \mathbf{P} \cap \mathbf{J} = \varnothing \wedge \mathbf{I} \cap \mathbf{J} = \varnothing$

P, **I** és **J** kölcsönösen kizárják egymást

(BL.38) $\mathbf{P} \cup \mathbf{P}^{-1} \cup \mathbf{I} \cup \mathbf{J} = A \times A$

P és a konverze ($\mathbf{P}^{-1}$), **I** és **J** együttesen kiadják a tartóhalmazon vett egész relációt

Az első két tétel azokat a relációtulajdonságokat rendeli a relációkhoz, amelyeket már korábban megismertünk. Fontos összefüggést fejez ki a harmadik tétel, miszerint a három alapreláció diszjunktnak tekinthető egymáshoz képest. Végül a negyedik tételből kiderül az is, hogy ha a három alaprelációhoz hozzávesszük még a szigorú rendezés konverzét, akkor a négy reláció teljesen lefedi az tartóhalmazon értelmezhető teljes relációt. Mivel bizonyítható az is, hogy az **P** rendezés konverze is diszjunkt az indifferencia és összehasonlíthatatlansági relációkkal, így kijelenthető, hogy a $(\mathbf{P}, \mathbf{P}^{-1}, \mathbf{I}, \mathbf{J})$ preferenciastruktúrában a négy reláció teljes partíciót alkot.[41] Sem a hármas, sem a négyes relációstruktúrában nem szerepel az **R** gyenge rendezési reláció. Ezt azonban könnyen pótolhatjuk az alábbi összefüggés alapján.

(BL.39) $\mathbf{R} = \mathbf{P} \cup \mathbf{I}$

A új rendezési reláció segítségével további összefüggéseket is megállapíthatunk a hármas struktúra elemei és a gyenge rendezés között. Ehhez használnunk kell a konverzreláció (↬38.o.), a duálreláció (↬40.o.) és a komplementer reláció (↬40.o.), valamint a teljesség vagy erős összefüggőség (↬29.o.) és az üres reláció (↬28.o.) fogalmát. Először a hármas struktúra elemeit írhajuk fel a gyenge rendezési reláció segítségével, majd a gyengerendezés teljességére adhatunk meg egy új feltételt.

[40] Tehát – figyelem – ezeket mind preferenciarelációnak nevezzük.

[41] A teljes partíció megállapításához szükséges, hogy a komponensek diszjunktak legyenek egymáshoz képest. A négyes preferenciastruktúra negyedik eleme itt a $\mathbf{P}^{-1}$ konverzreláció. Általában véve nem igaz, hogy egy reláció konverze diszjunkt lenne az alaprelációjával, de mivel itt **P** aszimmetrikus, ezért ebben az esetben fenáll ez a feltétel.

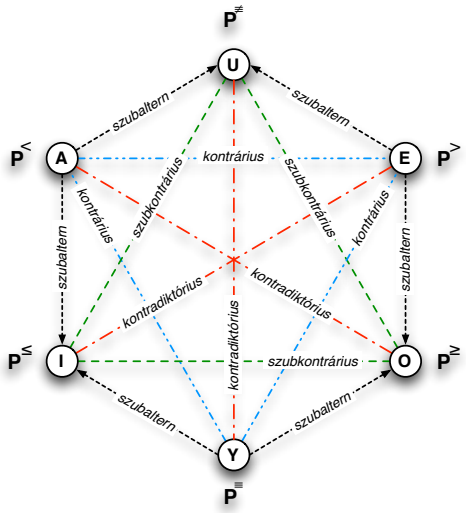
(BL.40) $(\mathbf{P}, \mathbf{I}, \mathbf{J}) = (\mathbf{R} \cap \mathbf{R}^{-d}, \mathbf{R} \cap \mathbf{R}^{-1}, \overline{\mathbf{R}} \cap \mathbf{R}^{-d})$

az (**I**, **J**, **R**) relációstruktúra elemei az **R** segítségével kifejezve

(BL.41) $\text{STRONGLY_CONNECTED}(\mathbf{R}) \leftrightarrow \mathbf{J} = \underset{\perp}{\mathbf{E}}$

az **R** gyenge preferenciarendezés erősen összefüggő, ha **J** üres

A bemutatott összefüggések alapján belátható az is, hogy a preferenciafogalmakra ráhúzhatjuk a logikai hatszög sematizmusát.



A preferenciahatszögben nem szerepel az összehasonlíthatatlansági reláció, viszont szerepel az indifferencia kontradiktorikus ellentéte, nevezhetnénk ezt differencia relációnak is, ha már nem lenne ilyen néven egy különbözőségi relációnk (↫28.o.). Ennek az az oka, hogy a preferenciareláció fogalmára lehet a logikai hatszöget "ráfeszíteni", ami viszont még nem adja ki a teljes relációt (↫27.o.). A preferenciareláció alá sorolhatjuk a fenti hat preferenciareláció-típust, és ezek közül három-három reláció mindig kiadja teljes preferenciarelációt. Igaz például a következő összefüggés: $\mathbf{P}^{>} \cup \mathbf{P}^{\equiv} \cup \mathbf{P}^{<} = \mathbf{P}$. Viszont az univerzális relációt csak úgy kaphatjuk meg, ha a preferenciarelációhoz hozzáadjuk még az összehasonlíthatatlansági relációt is.

Preferencia és reflexivitás

A preferenciaelméleteken belül nem igazán szokás reflektálni a reflexivitás kérdésére. Ez talán magyarázható azzal, hogy a reflexivitás részben összefügg olyan más tulajdonságkörökkel (például a szimmetricitással, az összekapcsoltsággal), amelyeket viszont mindig sokkal nagyobb figyelemmel kísértek. Az irreflexivitás több más relációtulajdonságból is következik. Például az aszimmetrikus reláció vagy az összefüggő reláció mindig irreflexív (↫33.o.). Von Wright sohasem vette fel az axiómarendszerébe a reflexivitás tulajdonságát, de több helyen is kategórikusan kijelentette, hogy az intrinzikus preferenciának irreflexívnek kell lennie mondván: „egy állapot sem lehet jobb önmagánál" [566, 147]. A reflexivitás direkt tárgyalását tekintve az egyetlen kivételt talán az jelenti, hogy a preferenciastruktúrák tárgyalásakor felveszik az axiómák közé a preferenciarelációk reflexivitási tulajdonságait.

Preferencia és szimmetricitás

Az aszimmetrikusság feltételezése nagyon erős, sokan nem is fogadják el. A von wrighti axióma elfogadása azt jelenti, hogy két világállapot összehasonlításakor mindig valamelyiket szigorúan jobbnak találjuk a másiknál. Azok, akik ezt a tulajdonság tételezik, azok az **szigorú preferenciát** teszik az elméletük középpontjába. Ezzel szemben azok, akik a **gyenge preferencia** fogalmát tartják alapvetőnek, és arra alapozzák elméletüket, azok a gyenge preferenciához szükséges **antiszimmetrikus** tulajdonság (↬29.o.) létezését írják elő. Bármelyiket is vesszük fel a modellünkbe kiindulásként, könnyen levezethető az egyik fogalom a másikból, ezért nagy jelentősége nincs választásnak. Mivel ezt korábban még nem rögzítettük, érdemes itt felvenni a tételek közé a gyenge preferencia antiszimmetrikusságát.

(BL.42) ANTISYMMETRIC(**R**)

R gyenge preferenciarendezés **antiszimmetrikus**

A szimmetrikusság egyik preferenciarendezésnél sem jöhet szóba, viszont – ahogy azt már korábban rögzítettem – mind a közömbösségi, mind az összehasonlíthatatlansági reláció szimmetrikus.

Preferencia és összefüggőség

Von Wright nem foglalkozott a teljesség tulajdonságával, a Chisholm és Sosa páros sem, Bengt Hansson viszont a totalitás tulajdonságát felvette az axiómái közé (↬87.o.). Bár Bengt Hansson elméletében fontos szerepe van a totalitás tételezésének,[42] ettől még Sven Ove Hansson igaza van, amikor azt állítja, hogy preferencialogikában általában nem követelik meg ezt a tulajdonságot. Meg is magyarázta, miért. Azért nem, mert ez a feltétel nagyon erős elvárást teremt, ami sokszor nem vagy csak nagyon nehezen teljesíthető. De miért is kéne teljesülnie? Ha valaki preferálja m-t mind n-nel, mind o-val szemben, akkor tud választani, miért kéne foglalkoznia azzal, hogy n és o közül melyiket preferálja inkább. Még konkrétabban: ha autót akarunk vásárolni, és két lehetőség között ingadozunk, akkor ez ezek közti döntéshez nincs szükségünk arra, hogy állást foglaljunk a további húsz kocsivásárlási lehetőséggel kapcsolatban. Elméleti szempontokból azonban sokat jelent, ha valahol kimondhatjuk a teljesség fennállását, ezért a formális elméletek szeretik ezt a tulajdonságot (lásd Bengt Hansson).

Sven Ove Hansson röviden kitér arra a kérdésre, hogy mit lehet tenni, ha egy halmazon nem teljes a rendezés (incomplete), és teljesé akarjuk azt tenni. S.O. Hansson szerint három kimenet lehetséges: vagy egyértelműen megoldható (unique resolvable), amikor egyetlen elrendezést kapunk, vagy többféleképp megoldható (multiply resolvable), amikor többféle rendezést kapunk, vagy megoldhatatlan (irresolvable), amikor nem jutunk teljes rendezéshez. Utóbbira példa: választani a gázai piramis lerombolása vagy a pandák kihalása között. Az első két eset S.O. Hansson összehasonlíthatatlannak (incomparable), a harmadikat inkommenzurábilisnak (incommensurable) nevezi [219].

[42] Amikor nem teljesül az összefüggőség (totalitás) tulajdonsága, akkor a preferenciarendezést Bengt Hansson csak részlegesnek nevezi, és ezen túl is igaz, hogy a totalitás tulajdonságára több fontos tételében is erősen támaszkodik [213].

Preferencia és a disztribúció elve

A negyedik von wrighti axióma, a disztribúció elve (↳83.o.) azzal a kérdéssel kapcsolatos, hogy miként működik a preferencia a diszjunkció esetén, vagyis akkor, amikor úgy kell a dolgokat összehasonlítani, hogy az állapotok diszjunktív kapcsolatban lehetnek egymással. Von Wright erre azt a példát adta, hogy a munkahelyén valakinek döntenie kell arról, hogy mit preferál jobban: a fizetésemelést (m) vagy hosszabb éves szabadságot (n) a napi munkaidő csökkentéshez (o) képest. A kérdés, hogy mit lehet mondani az ilyen összetett állapotokra vonatkozóan ($(m \vee n)$ az o-val szemben). Von Wright ilyen-olyan feltételezések és levezetések után hozta ki azt formulát, amit végül axiómának minősített.

(BL.43) $\forall m \forall n \forall o \forall p(\mathbf{P}(m \vee n, o \vee p) \leftrightarrow (\mathbf{P}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{P}(m \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p)$
$\wedge\ \mathbf{P}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge o) \wedge \mathbf{P}(n \wedge \neg o \wedge \neg p, \neg m \wedge \neg n \wedge p)))$

Nyugodtan mondhatjuk, hogy sokadik ránézesre sem tűnik intuitívnek a formula. Ebben a formájában nehéz befogadható értelmezést adni neki. Pedig van. Több más szerző mellett Patrick Girrard is újraértelmezte, leegyszerűsítette az eredeti von wrighti formulát és megmutatta, hogyan azt le lehet vezetni a konjunktív expanzió elvének és a jóval egyszerűbb disztributivitási tulajdonságoknak a felhasználásával [192]. Az átalakítás eredményeként a von wrighti formula két egyszerűbb és érthetőbb formulára visszavezethető.

(BL.44) $\forall m \forall n \forall o(\mathbf{P}(m \vee n, o) \leftrightarrow (\mathbf{P}(m, o) \wedge \mathbf{P}(n, o)))$

P preferenciareláció a diszjunkcióra **jobbról konjunktíve disztributív**

(BL.45) $\forall m \forall n \forall o(\mathbf{P}(m, n \vee o) \leftrightarrow (\mathbf{P}(m, n) \wedge \mathbf{P}(m, o)))$

P preferenciareláció a diszjunkcióra **balról konjunktíve disztributív**

A fenti képletek alapján már az is megállapítható, hogy a von Wright által megfogalmazott disztribúciós axiómák meglepően közel vannak Bengt Hansson axiómáihoz (↳87.o.). Az egyszerűsített von wrighti formulákhoz képest a hanssoni disztribuciós feltételek csak annyiban gyengébbek, hogy amíg az utóbbiak csak az egyik irányban érvényesek, tehát csak kondicionális művelet kapcsolja össze a két tagmondatot, addig az előbbi esetben kétirányú kapcsolatot ír le a bikondicionális logikai kapcsoló.

(BL.46) $\forall m \forall n \forall o((\mathbf{R}(m, n) \wedge \mathbf{R}(m, o)) \rightarrow \mathbf{R}(m, n \vee o)))$

(BL.47) $\forall m \forall n \forall o((\mathbf{R}(m, o) \wedge \mathbf{R}(n, o)) \rightarrow (\mathbf{R}(m \vee n, o)))$

Persze azért van egy másik különbség is közöttük, csak azt talán kicsit nehezebb észrevenni. Különböznek abban, hogy amíg von Wright a $\mathbf{P}(m, n)$ szigorú preferenciára, addig Bengt Hansson az $\mathbf{R}(m, n)$ gyenge preferenciára mondta ki a disztributivitási axiómáit. Tanulságos egyébként az a mód is, ahogy Bengt Hansson eljutott az axiómáihoz. Először értékelte az erős disztribúciós axiómákat, majd meggyőző példákkal igazolta, hogy ezek elfogadása intuícióinkkal ellentétes következményekkel járna [213]. Első lépésben Hansson feltette, hogy a következő formulát kellene elfogadnunk.

(BL.48) $\forall m \forall n \forall o \forall p(\mathbf{P}(m \vee n, o \vee p) \leftrightarrow (\mathbf{F}(m, n, o, p) \veebar \mathbf{F}(n, m, o, p) \veebar \mathbf{F}(m, n, p, o) \veebar$
$\mathbf{F}(n, m, p, o)))$

A formula jobb oldalán látható négy argumentumú **F**-operátor azt fejezi ki, hogy az argumentumokba írt négy alternatíva az adott preferenciasorrendbe rendezhető. A formula intuitívnak tűnik, hiszen annyit fejez ki, hogy az első két alternatív diszjunkcióját jobbnak értékeljük a második két alternatíva diszjunkciójához képest, akkor az egyenkénti összehasonlítás során az első két alternatíva előbbre kell, hogy kerüljön a második két alternatívánál (illetve fordítva). Másként: ha a bal oldalon kifejezett preferenciarendezés fennáll, akkor a jobb oldali képlet négy komponense közül az egyiknek kell teljesülnie. Valóban úgy tűnik, hogy ez a feltételezés elfogadható. B. Hansson azonban a fenti formulából levezet egy abból következő tételt, amiről viszont egy ellenpéldával bebizonyítja, hogy vállalhatatlan értelmezést is lehetővé tesz. A levezetett tétel így néz ki:

(BL.49) $\forall m \forall n \forall o(\mathbf{I}(m,n) \rightarrow \mathbf{I}(m \vee o, n \vee o))$

A képletben az $\mathbf{I}(m,n)$ indifferenciareláció szerepel (amit Bengt Hansson $\mathbf{S}(m,n)$ 'same as' relációnak nevez). Az alapfeltételezésből logikailag következő tétel azonban túl erős, és vállalhatatlan következményekkel jár. Ezt Hansson a következő példával szemlélteti. Tegyük fel az alábbiakat.

m	valaki megnyeri a fődíjat az ötös lottón
n	valaki megnyeri a fődíjat a hatos lottón
o	valaki nyer egy kettest az ötös lottón
$m \vee o$	valaki megnyeri a fődíjat és nyer egy kettest az I. lottón
$n \vee o$	valaki megnyeri a fődíjakat mind az ötös, mind a hatos lottón

Nyilván nem kell sokáig magyarázni, hogy amíg az m és n állapotokra mondhatjuk, hogy ugyanolyan jók (vagyis indifferensek egymással szemben), addig az $(m \vee o)$ és az $(n \vee o)$ "összetett alternatívák" messze nem tekinthetők egyenrangúaknak. Ezért Hansson a kezdeti feltételezést elvetette. A következő lépésben keresett egy gyengébb formulát abban reménykedve, hogy az már ki fogja elégíteni az intuícióinkat. Az addigi axiómák és tételek alapján levezetett egy következő tételt, ami így nézett ki:

(BL.50) $\forall m \forall n \forall o \forall p((\mathbf{I}(m,n) \wedge \mathbf{I}(o,p)) \rightarrow \mathbf{I}(m \vee \neg n, o \vee \neg p))$

A formula úgy értelmezhető, hogy ha van két-két állapot (m és n, illetve o és p), amelyek páronként egyforma értékesek ($\mathbf{I}(m,n)$ és $\mathbf{I}(o,p)$), akkor a két állapotpárból képzett két-két speciális diszjunktív állapot ($(m \vee \neg n)$ és $(o \vee \neg p)$) is ugyanolyan értékűek. Azonban ez a tétel sem maradhatott meg, mert Hanssonnak erre is akadt ellenpéldája. A történet részben megint a lottózásról szól, de most más alternatívák kerülnek a lottónyeremények mellé. Ezeket mutatom be a következő táblában.

m	valaki 100 millió forintot nyer az ötös lottón
n	valaki 100 millió nyer a hatos lottón
o	valaki meghal egy balesetben, mert elüti egy zöld busz
p	valaki meghal egy balesetben, mert elüti egy kék busz

Az első két alternatíva nyilván ugyanannyit ér az ember számára, és igaz ez a második kettő alternatívára is. De ezek konjunkcióit egymással összevetve már egyáltalán nem áll fent az értékegyezés.

Ezek után vezette be Hansson a már ismertetett axiómáit, amelyeket "hivatalosan" is axiómaként minősített. Ezután azonban, miután a disztribúciós axiómákból kiindulva további tételeket vezetett le, – némiképp meglepő módon – maga hozott példát arra, hogy még ezek a legyöngített formulák is képesek az elvárásainkkal ellentétes eredményeket produkálni.[43] A példa szerint valaki távol az otthonától a családjától kap leveleket. Tegyük fel, hogy a következő hírek érkeznek hozzá.

m	a kutyája jól van
n	a fia jól van
o	a lánya jól van
p	a felesége meghalt autóbalesetben

A kérdés az, vajon hogyan rendezi el ezeket a híreket akkor, ha diszjunkciót is megengedünk közöttük. Egy furcsa, de logikailag nem támadható kiindulópontot választva Hansson bemutat egy logkailag korrekt levezetést, amelynek a végén az intuíciónkkal ellentétes eredményhez jutunk. A formulákkal kifejezett lépések a következők.

1) $\mathbf{R}(m, n \vee (o \wedge p))$
2) $\mathbf{R}(m, (n \wedge p) \vee o))$
3) $\mathbf{R}(m, (n \vee (o \wedge p)) \vee ((n \wedge p) \vee o))$
4) $\mathbf{R}(m, (n \vee (o)) \vee ((n) \vee o))$
5) $\mathbf{R}(m, (n \vee o))$

Kiindulásként két preferendezést állított fel Hansson diszjunktív állapotok között. A fenti első sor ($\mathbf{R}(m, n \vee (o \wedge p))$) azt fejezi ki, hogy jobb az a hír, mely szerint 'a kutyája jól van' ahhoz az összetett hírhez képest, hogy 'a fia jól van (n) vagy a lánya jól van, de a felesége meghalt' ($o \wedge p$). A rossz hír jelenléte a diszjunkció jobb oldali tagjában az egész jobb oldali relátumot ($n \vee (o \wedge p)$) diszpreferálta. Ugyanez igaz a második sorra is a megfelelő módosításokkal. A harmadik sor formulájához az első disztribúciós axióma alkalmazásával juthatunk, amikoris az axióma képletébe behelyettesítjük a fenti két sort, mint az axióma bal oldali konjunkciójának két elemét. A harmadik formulából úgy kapjuk meg a negyediket, hogy kétszer alkalmazzuk a belső konjunkciókra a konjunkció eliminálásának műveletét (↰21.o.). Ennek eredményeként elhagyhatjuk a belső konjunkciók p-t tartalmazó tagjait. A negyedik sor formuláját sima propozíciológikai egyszerűsítésekkel "kitisztíthatjuk", így jutunk az ötödik formulához. Ez viszont olyat állít, ami intuícióellenes, hiszen a $\mathbf{R}(m, (n \vee o))$ formula annyit jelent, hogy az illető jobban örülne annak a hírnek, hogy 'jól van a kutyája' ahhoz képest, hogy 'jól van a fia vagy a lánya'. Feltételezhetjük ugyan, hogy lehetnek ilyen beállítódású személyek is, de semmiképpen sem ezt az attitűdöt tartjuk általánosnak, ezért ezt a következtetést mint kontraintuitív végeredményt nyugodtan "elutasíthatjuk". Bengt Hansson rezignált végkövetkeztetése ezekután csak az, hogy a két disztribúciós axiómát nem lehet feltétel nélkül alkalmazni.

[43] Hansson egy lábjegyzetben megjegyezte, hogy Sven Danielssonnak tulajdonítja bemutatott példáját.

Preferencia és a kontrapozíció elve

A témakör felvezetésekor említettem, hogy Sören Halldén felvette a posztulátumai közé a kontrapozíció elvét (↰89.o.). Von Wright az 1963-as esszéjében még nem beszélt róla, de az 1972-es cikkében már igen, és noha az elvet nem vette fel az új axiómarendszerébe, az intuícióinkkal megegyezőnek és ezért – legalább részben – elfogadhatónak tartotta. Tette ezt annak ellenére, hogy előbb Chisholm és Sosa szerzőpár [98], majd Bengt Hansson is meggyőző példákat hozott arra, hogy feltétel nélkül nem szabad elfogadni ezt az elvet [213]. Chisholm és Sosa ellenpéldája nem annyira a hétköznapi élet problémateréből merített, de ettől még a belőle kibontható kellemetlen logikai következmény fel- és elismerhető. Két állítást mutattak be arra, hogyan hasonlítsuk össze az állításokat, illetve a tagadásaikat egymással.

m	vannak boldog kócsagok
n	vannak kövek
$\neg m$	nincsenek boldog kócsagok
$\neg n$	nincsenek kövek

Chisholmék szerint a $\mathbf{P}(m,n)$ ítélet elfogadható, vagyis jobb az az állapot, amiben boldog kócsagok vannak ahhoz képest, amiben "csak" kövek vannak, viszont az állítás kontraponáltja, vagyis a $\mathbf{P}(\neg n,\neg m)$ ítélet szerintük már nem igazán fogadható el kérdezvén: "miért lenne jobb állapot az, amikor nincsenek kövek azzal szemben, amikor nincsenek boldog kócsagok". A két semmi se nem jobb, se nem rosszabb egymással szemben.

Jobban megérthetjük a Chisholmék példája mélyén rejlő problémát, ha felidézem a Bengt Hansson által bemutatott esetet is [213]. Képzeljük el, hogy valaki valamilyen versenyen indul, ahol lehet első és lehet további díjakat nyerni. Tekintsük a következő alternatívákat.

m	valaki első díjat nyer
n	valaki valamilyen díjat nyer
$\neg m$	valaki nem nyer első díjat
$\neg n$	valaki semmilyen díjat sem nyer

Első lépésként nyugodtan feltételezhetjük, hogy $\mathbf{P}(m,n)$ fennáll, hiszen az első díj többet ér bármilyen más díjnál. Ennek a konkrét relációnak a $\mathbf{P}(\neg n,\neg m)$ kontraponáltja viszont már nem igaz, hiszen ahhoz képest, hogy az ember nem nyeri meg az első díjat (ami itt együtt jár azzal, hogy semmilyen díjat nem szerez), bármilyen díj elnyerése jobb lehet, tehát a semmilyen díj elnyerésének alternatívája nem jobb az első díj el nem nyerésének alternatívájához képest. Ebben a történetben már egyértelműen látszik, hogy az ellentmondást az okozza, hogy a két alternatíva (m és n) nem független egymástól. Ez azt jelenti, hogy sem maguk, sem a tagadásaik nem teljesen diszjunktak, ami azért okoz problémát, mert ez a helyzet lehetetlenné teszi köztük az egyértelmű összehasonlítást.

Mielőtt továbblépnék a megoldás irányába, előtte még fontosnak tartom jelezni, hogy ez a probléma nemcsak a preferenciarendezés sajátja. S.V. Hansson arra is mutatott példát, amikor a kontrapozíció az indifferenciareláció esetében is gondot okoz [219, 349]. Tekintsük a következő helyzetet (illetve azok tagadásait).

m	A-nak két példánya van Rousseau Társadalmi szerződéséből
n	A-nak egy példánya van Rousseau Társadalmi szerződéséből
$\neg m$	A-nak nincs két példánya Rousseau Társadalmi szerződéséből
$\neg n$	A-nak nincs példánya Rousseau Társadalmi szerződéséből

Egy ilyen helyzetben – mondja S.V. Hansson – a két alternatíva m és n indifferens lehet A számára, hiszen egyszerre úgyis csak egy könyvet olvashat az ember.[44] Mondhatjuk tehát, hogy $\mathbf{I}(m,n)$ fennáll. De nem mondatjuk ezt az alternatívák kontraponáltjára, hiszen $\neg n$ esetén A nem olvashatja a könyvet, míg $\neg m$ esetén van rá esélye, hiszen attól, hogy nincs két példány a könyvből, egy még lehet. Már jeleztem, hogy mi okozza itt a problémát: ezek az ellentmondások akkor keletkeznek, amikor az alternatívák nem teljesen függetlenek egymástól. Sven Ove Hansson ezért azzal javasolta kiegészíteni a kontrapozíció elvét rögzítő axiómát, hogy abból ki kell zárni az alternatívák közötti logikai inkonzisztencia lehetőségét. Ezt a kizáró feltételt az alábbi módon fejezhetjük ki.

(BL.51) $\forall m \forall n(\mathbf{P}(m,n) \leftrightarrow (((m \wedge \neg n) \neq \bot \wedge (\neg m \wedge n) \neq \bot) \rightarrow \mathbf{P}(\neg n, \neg m)))$

ellentmondásmentes kontrapozíció elve (HUN) – principle of contradiction-free contraposition (HUN)

Ugyan a fenti tétel a szigorú preferenciarendezésre vonatkozik, de ugyanúgy ki lehet mondani az indifferenciarelációra nézve is. Ezzel a szűkítéssel már elkerülhetjük a kontrapozíció elvének ellentmondásait – gondolhatnák, de nem így van. Annak bizonyságára, hogy semmi sem olyan egyszerű, mint ahogy látszik, a Sven Ove Hansson által javasolt, fenti szűkítve-pontosító definíció után bemutatok egy példát, amely ellentmondani látszik az ellentmondásmentes kontrapozíció elvének [216]. Képzeljük el az alábbiakat.

m	valaki felajánl 100 dollárt jótékony célokra
n	valaki felajánl 50 dollárt jótékony célokra
$\neg m$	valaki nem ajánl fel 100 dollárt jótékony célokra
$\neg n$	valaki nem ajánl fel 50 dollárt jótékony célokra

Ebben a helyzetben nyilván fennáll a $\mathbf{P}(m,n)$, de nehéznek tűnik elfogadni a $\mathbf{P}(\neg n, \neg m)$ kontrapontált igazságát. Miért lenne jobb az egyik semmi a másik semminél? Hansson logikai érveket is hoz amellett [216], hogy bár furcsa, de mégiscsak el kéne fogadnunk a kontrapozíció érvényességét ebben az esetben is. Én csak egy másik példát hoznék, hogy pszichológiailag jobban elfogadhatóvá váljék mindez. Képzeljük el, hogy lottózunk, és két alternatívát kell értékelnünk, valamint képzeljük még azt is ide, hogy mindkét alternatívát megálmodtuk, és felébredve konstatáltuk, hogy a nyeremény csak álom volt.

m	nyerünk 2 milliárd 5 forintot a lottón
n	nyerünk 1429 forintot a lottón
$\neg m$	nem nyerünk 2 milliárd 5 forintot a lottón
$\neg n$	nem nyerünk 1429 forintot a lottón

[44] Fontos, hogy csak erre az egy szempontra figyeljünk a példában, mert ha a két könyvből az egyiket el lehetne adni és új könyvet lehetne érte venni, akkor a "két-könyves" állapot nyilván jobbnak számítana.

Nyilvánvaló, hogy fenntarthatjuk a $\mathbf{P}(m,n)$ igazságát. Ezek után gondoljunk bele abba, melyik ébredés utáni felismerés lehet fájdalmasabb a két negált alternatíva közül. A példa alapján talán már jobban elfogadható, hogy az álmodott (képzelt) hatalmas nyeremény nem teljesülése iránt nagyobb fájdalmat érzünk, mint a kis nyeremény "elvesztése" miatt – még akkor is, ha mindkét esetben ugyanazt a semmit "nyertünk", vagyis jobban preferáljuk a kisebb fájdalmat (a kisebb nyeremény elmaradása miatt), mint a nagyobb fájdalmat (a nagyobb nyeremény elmaradása miatt).

Talán kicsit nehezebb elfogadni az igazságát, de úgy tűnik, hogy az előző két példa csak látszólag mond ellent az ellentmondásmentes kontrapozíció törvényének. De utóbbi elv bevezetése még nem ad magyarázatot Chisholm és Sosa ellenpéldájára. Chisholmék boldog kócsagos történetében benne rejlő látszólagos ellentmondást akkor oldhatjuk fel, ha már többet tudunk a következő fontos preferencialogiai axiómáról, a konjunktív expanzió elvéről. Sören Halldén, Bengt Hansson, Sven Ove Hansson is hivatkozott arra, hogy a kontrapozíció és a konjunktív expanzió elvei között logikai azonosság van [213]. Ezt egyébként könnyű belátni. A konjunktív expanzió elvéből kiindulva lépésről lépésre el lehet jutni a kontrapozíció elvét kimondó formuláig.

(BL.52) a konjuktív expanzió és a kontrapozíció elve logikailag azonos

1) $\mathbf{P}(m,n) = \mathbf{P}(m \wedge \neg n, \neg m \wedge n)$
2) $\mathbf{P}(m,n) = \mathbf{P}(\neg n \wedge m, n \wedge \neg m)$
3) $\mathbf{P}(m,n) = \mathbf{P}(\neg n \wedge \neg\neg m, \neg\neg n \wedge \neg m)$
4) $\mathbf{P}(m,n) = \mathbf{P}((\neg n) \wedge \neg(\neg m), \neg(\neg n) \wedge (\neg m))$
5) $\mathbf{P}(m,n) = \mathbf{P}((\neg n), (\neg m))$
6) $\mathbf{P}(m,n) = \mathbf{P}(\neg n, \neg m)$

Az első sor mutatja a konjunktív expanzió elvét. A második sorban csak megcseréltük a konjunkciók összetevőinek sorrendjét. A harmadik lépésben betettünk egy-egy kettős tagadást, ami – mint tudjuk – nem változtat az igazságértékeken. A negyedik sorban zárójellel határoltuk el mindegyik változó negáltját, és az így képzett "új" változókra (tehát a $\neg m$-re és az $\neg n$-re) nézve újra alkalmaztuk a konjuktív expanzió elvét ("visszafele"). Így kaptuk az ötödik sor formuláját, amiből már csak a fölösleges zárójeleket kellett elhagyni, hogy a hatodik sorban láthassuk a kontrapozíció törvényének képletét.

Ez alapján mindazt, amit a kontrapozícióval kapcsolatban bemutattam, elmondhatnám a konjunktív expanzió elvével kapcsolatban is. Még azt is mondhatnánk, hogy szükségtelen a konjunktív expanzió elvének önálló tárgyalása, de ez nem így van. Látni fogjuk, hogy további, eddig még nem tárgyalt problémákkal kerülünk szembe a következő fejezetben. Logikai szempontból valóban nem lenne szükség a konjuktív expanzió elvének elemzésére, de gyakorlati megfontolások miatt mégis érdemes megtenni ezt. Egyszerűen azért, mert a logikailag ekvivalensnek tekinthető "jelenségekre" másfajta nézőpontból közelítünk, és ez – legalább részben – másfajta kép kialakításához, másfajta összefüggések felismeréséhez segíthet minket.

Preferencia és a konjunktív expanzió elve

Von Wright harmadik axiómája, a konjunktív expanzió elve (↬83.o.) az (állapot)változások és a preferencia között állapít meg összefüggést. Von Wright abból indult ki, hogy a két állapot (m és n) együttes figyelembe vételével négyféle, egymást kölcsönösen kizáró világállapot lehetséges:

$$(m \wedge n) \veebar (m \wedge \neg n) \veebar (\neg m \wedge n) \veebar (\neg m \wedge \neg n)$$

és ezek bármelyikének fennállását feltételezzük, mindegyik esetben négyféle változás lehetséges (minden marad a régiben, az egyik állapot marad, a másik változik, az egyik marad, a másik változik, valamint mindkettő változik). Bármelyik ilyen kettős-állapotban vagyunk is, mindegyik esetben az a változás lehet kívánatos a $\mathbf{P}(m,n)$ preferenciával rendelkező ágens számára, amelyik az $(m \wedge \neg n)$ állapotot eredményezi a $(\neg m \wedge n)$ helyett. Ez mind a négy esetben más és más transzformációt kíván meg, de a változás tartalma mindig az előbbi kell legyen.

Ugyanezt a problémát azonban más irányból is meg lehet közelíteni. Fel lehet ugyanis tenni azt a kérdést, hogy mi is történik valójában a preferenciáink kinyilvánításakor. Azt tudjuk, hogy ilyenkor állapotokat hasonlítunk össze. Ez mit jelent? Képzeljünk el egy gyereket [219], akinek azt mondjuk, hogy: — *Jobb az, ha van egy macskánk, mint egy kutyánk*. Ha erre a kijelentésre a kisgyerek azt válaszolja, hogy: — *Nem! Az a jobb, ha van egy kutyánk, ha van már egy macskánk is*, akkor ennek hallatán azt feltételezzük, hogy a gyerek nem érti még az összehasonlítás (és a választás) lényegét. Az összehasonlítás (választás) ugyanis feltételez valamiféle diszjunkt-ságot az alternatívák között abban az értelemben, hogy kizárja a két állapot együttes kívánását. Ilyenkor – az egymást kölcsönösen kizáró – a 'kutya-de-nem-macska' vagy a 'macska-de-nem-kutya' állapotokat (alternatívákat) kell összehasonlítani. Nos, pontosan ez fejezi ki a konjunktív expanzió elve. Ezért javasolta már Halldén is, aztán von Wright is az elv elfogadását. Sokan hivatkoztak Héctor-Neri Castañeda véleményére, miszerint ezt az elvet fel lehet fedezni már Szent Pálnak a házasságról szóló mondatában is.

> „De ha magukat meg nem tartóztathatják, házasságban éljenek: mert jobb házasságban élni, mint égni." (Pál Apostolnak Korinthusbeliekhez írt első levele, 7.9)

Vagyis – ahogy Castañeda értelmezi – „jobb házasságban élni és nem égni, mint nem házasságban élni és égni." Ha az összehasonlítást (választás) így értelmezzük, akkor szükségesnek látszik a konjunktív expanzió elve. Ezzel azonban nem mindenki van így. Chisholm és Sosa például elutasította ezt az elvet [98]. A kócsagos példájukat kicsit átalakítva a következő alternatívákat ajánlották összehasonlításra.

m	vannak kövek
n	vannak boldogtalan kócsagok
$m \wedge \neg n$	vannak kövek és vannak boldog kócsagok
$\neg m \wedge n$	nincsenek kövek és nincsenek boldog kócsagok

Chisholmék szerint kijelenthetjük, hogy az $m \wedge \neg n$ alternatíva jobbnak mondható, mint a $\neg m \wedge n$ állapot, ellenben az m-t inkább közömbösnek érezzük az n-hez képest. Még inkább érezhetjük az ellentmondást abban a példában, amit Chisholmék egy másik cikkükben adtak meg [97]. Ebben a következő állapotokat kell összehasonlítani.

m	nincs három boldog görög
n	van két boldogtalan római
$m \wedge \neg n$	nincs három boldog görög és nincs két boldogtalan római
$\neg m \wedge n$	van három boldog görög és van két boldogtalan római

Az egyszerű alternatívák összehasonlításából az jön ki, hogy **P**(m,n) igaz, hiszen a két boldogtalan római léte rosszabb, mint a boldog görög nem létezése. Ezzel szemben a boldog görögök és boldogtalan rómaiak nem létezésével szemben jobb a három boldog görög létezése még a két boldogtalan rómaival együtt is, mivel az boldogságok mérlege még így is pozitív.[45]

Sven Ove Hansson ennek a problémának a megoldására is ugyanazt javasolja, mint a kontrapozíció esetében: pontosítani, szűkíteni kell a konjunktív expanzió formuláját. Azért kell ezt tenni, mert itt is ugyanaz a probléma oka, az ellentmondásos állapotok kialakulása. Szemléltetésként bemutat egy olyan példát, amiben a következő két alternatíva között kell dönteni.

m	nyerek 100 millió forintot a lottón, és tartozom valakinek 2000 forinttal
n	tartozom valakinek 2000 forinttal
$m \wedge \neg n$	nyerek 100 millió forintot a lottón, és tartozom valakinek 2000 forinttal, és nem tartozom valakinek 2000 forinttal

Ebben a helyzetleírásban jól látszik az a probléma, hogy $m \wedge \neg n$ állapotok között ellentmondás van. A logikai ellentmondás, az elméleti gondok ellenére természetesen meg tudjuk oldani az ilyen helyzeteket, igenis tudunk dönteni ilyenkor is, mégpedig úgy, hogy az összehasonlítás ebben az esetben más lesz: nem $m \wedge \neg n$ és $\neg m \wedge n$, hanem m és $\neg m \wedge n$ alternatívák közül választunk. Éppen ezért javasolja azt Hansson, hogy ezt a "pontosítást" emeljük be a preferencia modelljébe. Ekkor a következő formulához jutunk.

(BL.53) $\forall m \forall n (\mathbf{P}(m,n) \leftrightarrow (m \wedge \neg n = \bot \rightarrow \mathbf{P}(m, \neg m \wedge n)) \underline{\vee} (m \wedge \neg n \neq \bot \rightarrow \mathbf{P}(m \wedge \neg n, \neg m \wedge n)))$

ellentmondásmentes konjunktív expanzió elve (HUN) – principle of contradiction-proof conjunctive expansion (ENG)

Ez az ellentmondásmentes szűkítés hasonlít a kontrapozíció elve esetében alkalmazott megoldáshoz, amit már nyilván nem tarthatunk véletlennek annak ismeretében, hogy a konjunktív expanzió és a kontrapozíció elvei logikailag ekvivalensnek tekinthetőek. Kiderült azonban, hogy ez a szűkítés még mindig nem elég. Sven Ove Hansson mutatott egy olyan példát, ami miatt az ellenmondásmentes konjunktív expanzió elvét még tovább kellett pontosítani [219]. Lássuk a példát!

m	valaki utazik a Holdra
n	valaki űrhajóval utazik
$m \wedge \neg n$	valaki utazik a Holdra és nem űrhajóval utazik
$\neg m \wedge n$	valaki nem utazik a Holdra és űrhajóval utazik

Itt más típusú probléma jelentkezik, mint a korábbiakban. A táblázatban feltüntetett mind a négy állapot lehetséges logikailag. Jelenlegi tudásunk és felkészültségünk szerint azonban az $m \wedge \neg n$ állapot technikailag nem elképzelhető, így minden olyan összehasonlítás, amelyben ez az alternatív szerepel, értelmetlennek tűnik, nem megengedhető. De ez a probléma orvosolható, mégpedig úgy, hogy az ellentmondásmentesség követelményét ki kell terjeszteni erre az esetre is. Nemcsak a logikai ellentmondások lehetőségét kell tehát

[45] röviden említeni Chisholmék hedonisztikus szemléletét, és ami ebből fakad.

kizárni, hanem azt is, hogy szemantikailag értelmetlen alternatívákat kelljen összehasonlítani valami mással. Ezt a további szűkítést **ellentmondásbiztos** megoldásnak nevezem, szemben a korábbi ellentmondásmentes szűkítéssel, amely "csak" a logikai ellentmondásokat zárta ki a formulákból.

(BL.54) $\forall m \forall n(\mathbf{P}(m,n) \leftrightarrow (m \wedge \neg n = ??????? \rightarrow \mathbf{P}(m, \neg m \wedge n)) \underline{\vee} (m \wedge \neg n ?????? \rightarrow \mathbf{P}(m \wedge \neg n, \neg m \wedge n)))$

ellentmondásbiztos konjunktív expanzió elve (HUN) – principle of contradiction-proof conjunctive expansion (ENG)

Ha ezt a kiegészítést megtettük, akkor természetesen ezt a második szűkítést érdemes elvégeznünk a kontrapozíció elvére vonatkozóan is. Így az alábbi módosított formulához jutunk.

(BL.55) $\forall m \forall n(\mathbf{P}(m,n) \leftrightarrow (((m \wedge \neg n) \neq ????? \wedge (\neg m \wedge n) \neq ????) \rightarrow \mathbf{P}(\neg n, \neg m)))$

ellentmondásbiztos kontrapozíció elve (HUN) – principle of contradiction-proof contraposition (HUN)

A konjunktív expanzió és a kontrapozíció elveinek kétszeres pontosítása után visszatérhetünk Chisholm és Sosa legelső kócsagos példájához (↰96.o.). Sven Ove Hansson szerint ha a példára kétféleképpen is alkalmazzuk a konjunktív expanzió elvét (azaz a kontrapozíció bal- és jobb oldalára egyaránt rávetítjük a konjunktív expanzió képletét), akkor belátható, hogy Chisholmék példája nem tartalmaz ellentmondást – akkor sem, ha a "hétköznapi intuíciónk" számára ez nehezebben befogadható [219].

Preferencia és választás

Bár a von wrighti konjuntív expanzió elvével kapcsolatban írt tanulmányt R.E. Jennings [261], mégsem az elvről szóló fejezetben foglalkozom Jennings véleményével, hanem a választás és preferencia kapcsolatát taglaló részben. Ennek az az elsődleges oka, hogy Jennings nem a korábban kifejtettek értelemben szólt hozzá a konjunktív expanzió elvéhez, hanem csak azt állította, ez tulajdonság nem a preferenciarelációt, hanem a **választás** (choice) jelenségét jellemzi. Jennings szerint ha m-t preferálom n-nel szemben, akkor az azt is kell, hogy jelentse, hogy van valamilyen tapasztalatom mind az m-mel, mind az n-nel kapcsolatban. Valaminek a megtapasztalása azonban még nem jelenti automatikusan azt is, hogy tudnánk, milyen lehet annak a valaminek a hiánya, ezért erre nem is hivatkozhatunk az összehasonlítások során. Az explikáció érdekében Jennings elkülönített kétféle preferenciafogalmat, és ezeket $\mathbf{P}^j$-preferencia és $\mathbf{R}^j$-preferencia kifejezésekkel jelölte.[46] Mindkét esetben ugyanazt a helyzetet értelmezzük:

valaki az m alternatívát preferálja n-nel szemben

Azonban amíg a $\mathbf{P}^j$-preferenciával azt az értelmezést fejezzük ki, hogy:

valakinek jobban tetszik az m, mint az n-t

addig $\mathbf{R}^j$-preferencia segítségével azt reprezentáljuk, hogy:

[46] Jennings saját relációi jelölésében természetesen nem használ felső indexet, azt csak én teszem itt ki, hogy nem legyen keveredés a már bevezetett **P** és **R** relációk és Jennings relációi között.

valaki inkább szeretné *m*-t, mint az *n*-t

Jennings szerint a két értelemzés abban különbözik egymástól, hogy amíg az előbbiben nincs semmilyen akarati tényező, addig az utóbbiban van, és így az előbbi valóban a preferenciát mint tetszést, az utóbbi pedig inkább a preferenciát mint választást (choice) reprezentálja. Ez a különbség megnyilvánul abban (is), hogy a preferencia esetében nem érvényes a konjunktív kiterjeszthetőség elve, míg a választás esetében igen. A kétféle reláció további két fontos szempontból is eltér egymástól. Egyfelől különböznek a magyarázatokban betöltött szerepük tekintetében, másfelől más előfeltételekkel élünk akkor, amikor a kétféle relációt tulajdonítjuk valakinek. A $\mathbf{P}^j$-preferencia és $\mathbf{R}^j$-preferencia aszimmetrikus viszonyban állnak egymással abban az értelemben, hogy az utóbbiakat mindig megmagyarázhatjuk az előbbiekre hivatkozva, de ez a lehetőség fordítva már nem áll fenn. Jennings példájával élve arra a kérdésre, hogy:

— Miért kéne neked inkább a narancs, mint mandarin?

adhatunk olyan magyarázatot, hogy egyszerűen hivatkozunk arra:

— Mert jobban szeretem a narancsot a mandarinnál.

de arra kérdésre, hogy:

— Miért szereted jobban a narancsot a mandarinnál?

nem tudunk értelmesen azzal válaszolni, hogy utalunk a vonatkozó $\mathbf{R}^j$-preferenciánkra. A másik különbség a jenningsi $\mathbf{P}^j$- és $\mathbf{R}^j$-preferenciák között abban van, hogy milyen fokú és milyen típusú magyarázatot fogadunk el, amikor meg akarjuk indokolni őket. A $\mathbf{P}^j$-preferencia esetén az előző kérdésre elfogadjuk az alábbi választ:

— Hát, csak.

Ebben az esetben a magyarázat lényege az, hogy egyszerűen hivatkozhatunk a létező $\mathbf{P}^j$-preferenciánkra: “ezt szeretjük jobban”. Itt nincs racionális magyarázat, nincs érvelés, nincs megokolás, nincs másra hivatkozás, csak a preferencia “puszta” deklarálása. Ezzel szemben a $\mathbf{R}^j$-preferencia esetén mindig kell valamilyen indoklás, valamilyen okra való hivatkozás. Ekkor a ‘Miért?’ kérdésre nem fogadjuk el az “önmagára hivatkozó” indoklást. Az alábbi kérdésre adott választ nem érezzük “teljesnek”, elfogadhatónak.

— Miért kéne neked jobban fahéj, mint gyömbér?
— Hát, csak kéne.

Ilyenkor nem elég a ‘csak’ válasz. Ha mégis ezt feleljük, akkor azt úgy fogják értékelni mások, hogy nem akarjuk felfedni a preferenciáinkat, esetleg egyfajta deklarációként. Ha ugyanis azt mondjuk:

— Nekem jobban kéne most fahéj, mint gyömbér.

ezt inkább úgy lehet érteni, hogy nem a preferenciánkat fejeztük ki, hanem “kihirdettük”, közöltük a választásunkat, de nem értelmezhetjük úgy, hogy ezzel magyarázatot adtunk volna a ‘miért’ kérdésre. A választásainkank mindig (de legalábbis az esetek többségében) van indoka, sőt, nagyon sok esetben ezek az indokok pont saját belső $\mathbf{P}^j$-preferenciánk. A kétféle preferencia közti harmadik különbségről már volt szó: $\mathbf{P}^j$-hez kell tapasztalat az alternatívákról, $\mathbf{R}^j$-hez nem feltétlen szükséges. Jennings szerint mondhatjuk, hogy:

— Inkább szeretnék a pokolra kerülni, mint a mennyországba.

de – tényleges tapasztalat hiányában – nem mondhatjuk, hogy:

— Nekem jobban tetszik a mennyország, mint a pokol.

Jennings ezek után egy formalizálási javaslatot is bemutat az újonnan bevezetett fogalomra, a $\mathbf{R}^j$-preferenciára. Kiindulásként rögzíti, hogy a $\mathbf{P}^j$ reláció argumentumai állapotok, az $\mathbf{R}^j$ reláció argumentumai cselekvések (és csak kényelmi, rövidítési okok miatt használjuk ugyanazokat a jeleket az argumentumok változóinak). Abból indulhatunk ki, hogy ha $\mathbf{R}^j(m,n)$ igaz, vagyis valaki inkább tenné m-t, mint a n-t, akkor erre mondhatjuk azt is, hogy választania kell az alternatívák közül. Az m és n közti választás az exkluzív diszjunkció (biszubtrakció) műveletével fejezhető ki (↰16.o.), vagyis írhatjuk, hogy $m \underline{\vee} n$. Az exkluzív diszjunkció képlete: $(m \vee n) \wedge \neg(m \wedge n)$, ahol a formula második tagjával zárjuk ki a két alternatíva együttes fennállását, és ezzel biztosítjuk az alternatívák diszjunktságát. Ezt a formulát viszont – a disztribúciós szabály segítségével – átalakíthatjuk a következő alakra: $(m \wedge n) \vee (\neg m \wedge n)$. Ha ezek után az m és n közötti választás tevékenységét $\#(m,n)$ szimbólummal jelöljük, akkor vonatkoztathatjuk rá az előző összefüggést.

(PL.56) $\quad \textsc{choice}(m,n) \equiv \#(m,n)$

(PL.57) $\quad \#(m,n) \leftrightarrow \#((m \vee n) \wedge \neg(m \wedge n)) \equiv \#(m \wedge \neg n, \neg m \wedge n)$

A választás itt azt jelenti, hogy m-t választja. A választás mint exkluzív diszjunkció tételezésének van egy fontos következménye: ha a két alternatíva valamelyikének létezése maga után vonja a másik létezését, akkor nem beszélhetünk választásról, hiszen ekkor a két alternatíva nem, hogy nem zárja ki egymást, de az egyik feltételezi a másikat. Jenning szerint a cselekvéseinkre kétféle módon adgatunk magyarázatot: létezik unális (unal), illetve duális (dual) leírás, magyarázat. A $\mathbf{P}^j$- és $\mathbf{R}^j$-preferenciák a cselekvéseink duális magyarázatában adhatnak segítséget.

K_1 — Miért csináltad x-t?
V_{11} — Azért, mert x-t akartam csinálni.
V_{12} — Azért, mert szeretem x-t csinálni.
K_2 — Miért csináltad x-t az y helyett?
V_{21} — Azért, mert inkább tenném x-t, mint y-t.
V_{22} — Azért, mert jobban szeretem csinálni x-t, mint y-t.

A kérdésekre adott válaszok magyarázóereje eltérő. A V_{11} és V_{21} válaszok üresek, nem érezzük elfogadható indoklásoknak őket. Ezzel szemben érvényesnek, elfogadhatónak mondható a V_{12} és V_{22} válasz, amikor egyszerűen hivatkozunk a preferenciáinkra (unális vagy duális módon). Fontos észrevennünk, hogy a V_{12}, illetve a V_{22} válaszok a $\mathbf{P}^j$-, illetve $\mathbf{R}^j$-preferenciának felelnek meg. Az egyik fontos különbség pedig az közöttük, hogy a $\mathbf{P}^j$-preferencia "önmagát magyarázza", nincs szüksége "külső" indoklásra, érvelésre, az $\mathbf{R}^j$-preferencia viszont önmagában nem áll meg, mindig szüksége van "külső" érvekre, indokokra.

„My reasons for wanting to do x are my reasons for doing x. Similarly, the reasons why I would rather do x than y are my reasons for doing x rather than y. ... Thus the relation between liking to do x and doing x is different from the relation between wanting to do x and doing x. There is a parallel difference between the relation between having a P-preference of doing x to doing y and doing x rather than y, and the relation between having

> the R-preference (do x) R (do y) and doing x rather than y. Our likes provide reasons for and explanations of both our wants and our actions. Our P-preferences provide explanations of both our R-preferences and our choices." [261, 565]

Jennings megkülönböztetésének lényegét azután lehet csak igazán megérteni, miután elemeztük a választás (↪168.o.), valamint az intencionalitás problémakörét, különösen a vágy és akarat fogalmait (↪**??**.o.).

A preferenciákat vizsgálva von Wright felvetett még egy további tipizálási lehetőséget is, amikor elkülönítette azokat helyzeteket, amelyekben van **kockázat** (risk) azoktól, amelyekben nincs. Még azt is javasolta (igaz, nem indokolja meg, miért), hogy a kockázatmentes (risk-free) preferenciakinyilvánításokat nevezhetnénk **tiszta preferenciáknak** (pure preferences), míg a kockázatot is magába foglaló preferenciarendezésekre használhatnánk a **preferenciális választások** (preferential choices) kifejezést. Ha egy ponton módosítunk von Wright korábbi példáján azt feltételezve, hogy a munkavállalónak úgy kell döntenie a munkahelyén, hogy a fizetésemelést (m) vagy hosszabb éves szabadságot (n) választja inkább a napi munkaidő csökkentéshez (o) képest, hogy az első alternatíva diszjunkciójában (m vagy n) szereplő két összetevő közül csak az egyik teljesülhet, viszont nem előre lehet tudni, hogy melyik. Ha eközben igaz az is, hogy az alternatívák páronkénti összehasonlítás alapján az illető jobbnak tartja a fizetésemelést (m-t) a rövidebb munkaidőnél (o-nál), viszont utóbbit jobbnak ítéli a hosszabb szabadságnál (n-nél), akkor a döntési helyzetbe kockázat "kerül". Ha a munkavállaló az ($m \vee n$) opciót választja, akkor – a döntése után – kiderülhet, hogy a legpreferáltabb állapothoz jut, de az is lehet, hogy a legkevésbé kedvelt alternatívához. Ha viszont a második alternatívát választja, akkor biztosan nem érheti el a legkívánatosabb állapotot, amire adott valószínűsége pedig lett volna az első alternatíva választása esetén. Már ebből a példából is látszik, hogy a preferenciális választásokra nem igazak a disztributivitási kritériumok, ami miatt von Wright rögzítette is, hogy esszéjében csak a tiszta preferenciákkal foglalkozik.

Ha fel akarjuk venni ezt a megkülönböztetést az ontológiánkba, akkor definiálni kell a kockázat fogalmát, ami alapján már könnyen meg lehet határozni a most bemutatott kétféle preferenciafogalmat is.

(BL.58) $\text{RISK} \stackrel{\text{def}}{=} m|p_m, n|p_n$

(BL.59) $\text{PURE_PREFERENCE} \equiv \text{PREFERENCE} \stackrel{\text{def}}{=} \mathbf{P}(m|p_m, n|p_n) \wedge p_m = p_n = 0$

(BL.60) $\text{PREFERENTIAL_CHOICE} \equiv \text{CHOICE} \stackrel{\text{def}}{=} \mathbf{P}(m|p_m, n|p_n) \wedge p_m \neq 0 \wedge p_n \neq 0$

Kockázatmentes helyzetekben a tiszta preferenciáinkat nyilváníthatjuk ki, ezeket tarthatjuk az igazi, külső befolyástól (valószínűségekre alapozott megfontolásoktól) mentes értékkifejeződéseknek. Az intrinzikuskus preferencia kritériumrendszerét erre a típusra dolgozta ki von Wright. A **kockázatterhes helyzetekben**, amikor a különböző valószínűségekkel ellátott állapotok között kell választanunk, más logika mentén működik az ember.

Ceteris paribus preferencia

Az ötödik axiómát von Wright még az amplifikáció elveként vezette be (↳83.o.), de később (és mindenhol máshol) ezzel a tulajdonsággal már a ceteris paribus preferenciát írta le (hogy aztán az 1972-es tanulmányában már holisztikus preferenciának hívta). A lehetséges terminusok közül végül a ceteris paribus preferencia kifejezés vált elterjedtté, ezt használom én is.

Bengt Hansson kritizálta az amplifikáció elvét mondván, hogy ellentmondáshoz vezethet, ha az alternatívák nem függetlenek egymástól. Könnyen szemléltethető mindez, ha feltételezzük, hogy $o = m \wedge n$. Ezt az egyenlőséget behelyettesítve az amplifikáció elvének formulájába, majd a fölösleges belső konjunkciókat kiiktatva, illetve a belső konjunkciók szimmetriáját kihasználva a következő képletet kapjuk:

$$
\begin{aligned}
&1)\ \forall m \forall n \forall o(\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge o, n \wedge o) \wedge \mathbf{P}(m \wedge \neg o, n \wedge \neg o)))\\
&2)\ \forall m \forall n \forall o(\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge m \wedge n, n \wedge m \wedge n) \wedge \mathbf{P}(m \wedge \neg(m \wedge n), n \wedge \neg(m \wedge n))))\\
&3)\ \forall m \forall n \forall o(\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge n, n \wedge m) \wedge \mathbf{P}(m \wedge \neg(m \wedge n), n \wedge \neg(m \wedge n))))\\
&4)\ \forall m \forall n \forall o(\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge n, m \wedge n) \wedge \mathbf{P}(m \wedge \neg(m \wedge n), n \wedge \neg(m \wedge n))))
\end{aligned}
$$

A negyedik sorban már látszik az ellentmondás, hiszen a bikondicionális jobb oldalán álló konjunkció első tagja ($\mathbf{P}(m \wedge n, m \wedge n)$) ellentmondásban van von Wright **P**-relációjával, ami – lévén szigorú preferenciarendezés – aszimmetrikus. Ezt persze korrigálni lehet a korábban már többször alkalmazott ellentmondásbiztos klauzula beépítésével az elv formulájába. Ez tehát "kezelhető" problémát jelent. És nem is lényeges ponton "támadja" a ceteris paribus preferencia fogalmát. Von Wright számára nem ezért volt fontos ez a fogalom.

Amikor preferenciáról beszélünk, akkor két állapotot hasonlítunk össze. Feltehető a kérdés, hogy amikor így döntünk, akkor vajon figyelembe vesszünk-e más állapotokat, illetve állapotváltozásaokat, és ha igen, akkor mennyit? Mi van akkor, ha figyelünk arra is, hogy nem csak két, de több állapot is fennállhat akkor, mikor éppen azt a két állapotot hasonlítunk össze, és nemcsak az éppen összehasonlított állapotok változhatnak, hanem bármi más is. Akkor érthetjük meg, hogy miért is fontos ez a kérdés, ha von Wright harmadik axiómájára gondolunk, miszerint preferálni m-t n-nel szemben ekvivalens azzal, hogy a jelen helyzethez képest preferáljuk azt a változást, ami az $(m \wedge \neg n)$ helyzetbe visz minket ahhoz a változáshoz képest, ami az $(\neg m \wedge n)$ helyzetet eredményezi. Ebben az interpretációban már sokkal érthetőbb, miért lehet feltenni azt a kérdést, hogy ilyen esetekben miért nem veszünk figyelembe más változásokat, más helyzeteket is.

A kérdésre válaszul von Wright háromféle lehetőséget különít el egymástól: az abszolút, a feltétel-nélküli és feltételes preferencia kinyilvánításának helyzetét. Amikor az egyén az $(m \wedge \neg n)$ állapotot eredményező változást nemcsak a $(\neg m \wedge n)$, de bármely más állapotot eredményező változáshoz képest is preferálja, akkor **abszolút preferenciáról** (absolute preference) beszélhetünk. Ilyen esetekben ha egy világállapot tartalmazza a $(\neg m \wedge n)$ állapotot, akkor azt ez az egyén mindenképpen jobban tartja az összes többi világállapotnál, amelyben $(\neg m \wedge n)$ is benne van. Az ilyen abszolút preferenciából azonban egy ember (egy adott pillanatban) csak eggyel rendelkezhet (ezt a tézist von Wright röviden le is vezeti, amit itt nem mutatok be), ezért ezt a preferenciatípust von Wright elhanyagolhatónak véli (mivel a gyakorlatban érdektelen).

A **feltételes preferencia** (conditional preference) fogalmát úgy értelmezi von Wright, hogy ilyen esetekben két állapot összehasonlítása függ egy harmadik állapottól. Azt a

példát adja erre, hogy ha valakinek pénzre van szüksége és el kell adnia egy ruhadarabját, akkor három ruhadarabot (trikót, cipőt és nadrágot) feltételezve másként értékelheti azt a kérdést, hogy melyiket adja el a trikó (m) és a cipő (n) közül attól függően, hogy mit gondol a nadrág (o) eladásáról. Ha nem kell (nem akarja) eladni a nadrágot, akkor preferálja a cípőt a trikóval szemben (mert az "megy" a nadrághoz), vagyis az utóbbit fogja eladni. Viszont ha el kell (el akarja) adni a nadrágot (mert, mondjuk, kinőtte), akkor már inkább a trikó megtartását választja a cipővel szemben (mert nadrág nélkül úgysem használhatná).

A feltétlen vagy **feltétel-nélküli preferenciáról** (unconditional preference) akkor beszélhetünk, amikor feltételezhetjük, hogy a világ nem változik más állapotait tekintve, tehát csak a m és n állapotokat érintő változások történnek. E helyzet leírására szokták használni a **ceteris paribus preferencia** kifejezést.[47] A von wrighti preferencialogika a feltétel-nélküli, ceteris paribus preferenciákról szól. A feltétel-nélküliség mozzanatát von Wright az ötödik axiómájával definiálta. Ez a tétel egy elsőrendű logikai szintaxison alapuló formula igazságát tételezi. Azóta kiderült, hogy a ceteris paribus elvet igazából a modális logika segítségével lehet jól megragadni. Von Wright megoldási javaslata után sokáig az volt a kérdés, mi lesz a jó megoldás erre. Sven Ove Hansson már egy 1995-ös cikkében is tett egy javaslatot a ceteris paribus preferencia formalizálására [218], amin aztán a 2001-es nagy preferenciatanulmányában módosított [219], de azt a megoldást, amit egyre szélesebb körben fogadnak el a szakirodalomban, mások dolgozták ki. Ezt viszont csak egy későbbi fejezetben mutatom be (↬111.o.).

Preferencia és tranzitivitás

A tranzitivitás feltételét is sokan vitatják arra hivatkozva, hogy a tényleges társadalmi gyakorlatban sok olyan példát találhatunk, amikor nem érvényesül a tranzitivtás. Von Wright is tudatában van a tranzitivitással szembeni kételyeknek, de azt válaszolja, hogy a tranzitivitás ellenében felhozott ellenpéldák szerinte mindig az extrinzikus preferenciákra vonatkoznak, amikor valami hatására időközben megváltozik valakinek a preferenciarendezése (hiszen, mondja von Wright, a tranzitivitás mérése időbeli folyamat, ami alatt változhat az extrinzikus preferencia). Szerinte az intrinzikus preferenciákra fenntartható a tranzitivitás kritériuma.

Még ha sokan vannak, akik túl erősnek tartják, ezért nem fogadják el a preferencia tranzitivitási feltételét, a tranzitivitás tulajdonságának fontosságát senki sem tagadja. Nélkülözhetetlenségét jelzi az a tény is, hogy sokféle módon próbáltak gyöngíteni a tranzitivitás kritériumán, és új tranzitivitás-közeli fogalmakat vezettek be. Ilyenek előfordultak a relációtulajdonságok között is (↬28.o.), emlékezzünk csak például a kvázi-tranzitivitás vagy az aciklusosság fogalmára. A relációtulajdonságokkal való minősítésen túl úgy nyerhetünk további tranzitivitás-közeli fogalmakat, hogy több relációt vonunk be a definíciókba. Ilyeneket definiáltuk a relációelméleti fejezetben, például az IP- vagy a PI-tranzitivitás fogalmait (↬32.o.). Mivel akkor még nem vezettük be a különböző preferenciarelációkat, így egyetlen relációra támaszkodva kellett megadnunk a definíciókat. Most már egyszerűsíthetünk a korábbi formulákon, ha az $\mathbf{R}$ gyenge preferencia fogalmából képzett $\mathbf{P}$ szigorú preferencia, $\mathbf{I}$ indifferenciarelációk meghatározásait is figyelembe vesszük.

(PL.61) $\quad \text{PP-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow \mathbf{P}(m,o))$

[47] A 'ceteris paribus' latin kifejezés jelentése: 'minden más változatlan'.

(PL.62) $\text{IP-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{I}(m,n) \wedge \mathbf{P}(n,o)) \rightarrow \mathbf{P}(m,o))$

(PL.63) $\text{PI-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{P}(m,n) \wedge \mathbf{I}(n,o)) \rightarrow \mathbf{P}(m,o))$

(PL.64) $\text{II-TRANSITIVE}(\mathbf{R}(m,n)) \stackrel{\text{def}}{=} \forall m \forall n \forall o((\mathbf{I}(m,n) \wedge \mathbf{I}(n,o)) \rightarrow \mathbf{I}(m,o))$

A későbbiekben még használjuk ezeket a tranzitivitási tulajdonságokat, illetve a különböző tézisekben, levezetésekben nem egyszer lehet hivatkozni azokra az összefüggésekre, amelyek a tranzitivitás fogalomkörébe tartozó tulajdonságok között állapíthatók meg. Fel lehet tenni itt a kérdést, miért van szükség ilyen sokféle fogalomra a tranzitivitás fogalombokrán belül. Azért, mert a racionális emberi cselekvések magyarázatában igyekszünk kizárni az egymás után kimondott preferenciaállítások sorából a ciklikusság megjelenését. De miért is baj a ciklikusság? Erre a kérdésre Amos Tversky pénznyelő (money-pump) példájával szokás válaszolni. Ha nem követelnénk meg a tranzitivitást, akkor a viselkedésünket ugyanúgy irracionálisnak lehetne minősíteni, mint Tversky példaemberének cselekvését.

> „A tranzitivitás a racionális visekedés egyik alapvető és talán legszigorúbb elve. Mondhatjuk azt is, hogy aki megsérti a tranzitivitás elvét, az voltaképpen 'pénznyelőként' működik. Tegyük fel, hogy valaki preferálja y-t x-hez, z-t y-hoz, valamint x-t z-hez képest. Ekkor megalapozott lehet az a feltételezésünk, hogy az illető hajlandó fizetni azért, hogy kicserélhesse először x-t y-ra, aztán y-t z-re, végül z-t x-re. Ám ha így tesz, a végére ugyanoda jut, ahonnan indult, csak épp kevesebb lesz a pénze." [548, 45]

A preferenciarelációkat (szigorú rendezés, gyenge rendezés, indifferencia) egymás után többször is lehet alkalmazni, vagyis sorba lehet őket kötni. Ilyen esetekben egyszerűsíthetünk a jelölésen, ha az egymást követő diadikus viszonyokat lineaárisan "összekötjük":

(BL.65) $\mathbf{P}(m,n) \wedge \mathbf{P}(n,o) \wedge \mathbf{I}(o,p) \equiv PPI \equiv (m > n) \wedge (n > o) \wedge (o \mathrel{|||} p) \equiv m > n > o \mathrel{|||} p$

Bármennyire is fontosnak és szükségesnek érezzük a tranzitivitást mind a preferencia, mind a racionális vilsekedés meghatározásához, az intranzitivitás, a ciklikusság gyakran tetten érhető a cselekvéseinkben. Sven Ove Hansson szerint négyféle tipikus tranzitivitási kör létezik, és az elemzésekben felhozott legtöbb példa ezek valamelyikébe tartozik [219]. A továbblépés előtt érdemes bizonyos jelölési konvenciókat bevezetni.

$\mathbf{I}^n\mathbf{P}$-CYCLE	egymás után n-szer **I**-, majd egy **P**-reláció követi egymást $(n \geq 2)$
IPIP-CYCLE	egy **I**-, egy **P**-, **I**-, **P**-reláció követi egymást
IPP-CYCLE	egy **I**-, majd két **P**-reláció követi egymást
$\mathbf{P}^n$-CYCLE	egymás után n-szer **P**-reláció követi egymást $(n \geq 3)$

Az $\mathbf{I}^n\mathbf{P}$-ciklus szemléltetésére a R. Duncan Luce példáját hívom segítségül [313]. Tegyük fel, hogy valaki jobban szereti a kávét egy kockacukorral inni, mint öttel (ami azt is jelenti, hogy képes megkülönböztetni az egy-cukros kávét az öt-cukrostól). Ezután képzeljünk el egy kísérletet, amelyben 401 csésze kávét kell összehasonlítania úgy, hogy az egymást követő i. kávékba rendre $\left(1+\frac{i}{100}\right)x$ gramm cukrot tesznek (ahol x a kockacukor súlya). Nyilvánvaló, hogy ilyen kicsit mennyiségi különbségek mellett senki sem képes megkülönböztetni az egymást követő kávékban levő cukor mennyiségét, és így a kávé ízét ugyanolyannak érzi, ezért minden egymást követő kávépárost egyformának ítél (közömbös velük szemben). Ezt kell mondania az első összehasonlítástól kezdve az utolsóig. A

közömbösségi reláció tranzitivitását feltételezve viszont bármelyik két kávéra is igaz kell legyen az, hogy ugyanolyannak tartja őket a kísérleti alany. Ezzel azonban az indifferens ítéletek sorának végére jutva ellentmondás keletkezik, hiszen az alany az első összehasonlítási körben az egy-cukros kávéhoz képest rosszabbnak tartotta az öt-cukros kávét, míg a második – szomszédos páros – összehasonlítási körben az 1. és a 401. csészében levő kávékat ugyanolyannak ítélte meg, holott bennük ugyanannyi cukor (egy, illetve öt kockányi) van, mint az első körös összehasonlításban.

Az $\mathbf{I}^n\mathbf{P}$-ciklus legegyszerűbb változata az **IIP**-ciklus, amelyre Sven Ove Hansson Frank Restle 1961-es könyvéből idéz egy példát [219]. Restle olyan esetet ír le, amelyben valakinek három alternatíva közül kell választania úgy, hogy a három összehasonlításból kettő nehéz, egy pedig könnyű döntést jelent. A alternatívák a következők:

m	utazás Kaliforniába
n	utazás Floridába
o	utazás Kaliforniába + egy alma

Ebben a helyzetben mind az első és második, mind a második és harmadik opció közötti választás nehéz döntés (a plusz alma elhanyagolható előnyt jelent ebben az összehasonlításban, a döntés lényege ettől nem változik). Bármilyen kicsi és az alma értéke az utazások értékéhez képest, az első és harmadik opció öszehasonlításakor mégiscsak többletet jelent az utóbbi számára, tehát itt könnyen döntést lehet hozni. Igaz tehát, hogy **I**(m,n), **I**(n,o), illetve **P**(m,o), vagyis teljesül az **IIP**-ciklus. Lényegében ugyanezt a helyzetet vizsgálta Dan Ariely is abban az empirikus kísérletében, amelyben a résztvevőknek egy római és egy párizsi utazás között kellett választani, miközben harmadik opcióként felajánlották számukra a római utat egy ingyenreggelivel együtt [21], bár Ariely kérdésfeltevése és megközelítésmódja nem igazán az intranzitív ciklusok vizsgálatához kötődik. Az általa levont tanulság azonban itt is hasznos. A kísérletben a résztvevők gyakran választották az ingyenreggelivel járó római utat, és Ariely úgy értelmezi ezt a helyzetet, hogy mindez azért történt így, mert a harmadiknak kínált csali ajánlat segített "kimenekülni" abból a helyzetből, amelyben nem tudott igazából dönteni az ember két egyenragú opció között. Erre a kérdésre később, a racionalitás tárgyalásakor még visszatérek.

Folytatva az intranzitivitási ciklusok tárgyalását érdemes Restle korábbi példáján módosítanunk kicsit. Ha az utazással kapcsaolatos választási lehetőségeket kiegészítjük egy további alternatívával, akkor olyan helyzetethez jutunk, amellyel az **IPIP**-ciklust szemléltethetjük. Vegyük a következő négy alternatívát:

m	utazás Kaliforniába
n	utazás Floridába
o	utazás Floridába + egy alma
p	utazás Kaliforniába + egy alma

A négy választási lehetőségre nézve igaz lehet az, hogy az egyén nem tud dönteni Kalifornia és Florida között, de a plusz alma miatt a harmadik opciót választja a másodikkal szemben, aztán ha egy-egy almával "növelik" két város vonzerejét, akkor továbbra is döntetlen marad köztük, viszont az elsőhöz képest a negyedik alternatíva többet ér az

ajándék alma miatt. Tehát fennállnak a következő rendezések: **I**(m,n), **P**(n,o), **I**(o,p), illetve **P**(m,p). Ez azt jelenti, hogy ez a példa már az **IPIP**-ciklus jelenségébe tartozik.

Az **IPP**-ciklusra Hansson Yew-kwang Ng példáját idézi [377]. Tegyük fel, hogy a döntési feladat abból áll, hogy papírlapcsomagok közül kell választani kétféle szempont alapján. A vastagságot tekintve az első csomag papírja vastagabb, mint a másodiké, de az első és harmadik, valamint a második és harmadik csomag lapvastagságai között nincs érezhető különbség. Ekkor igaz a következő: $\mathbf{P}^v(m,n), \mathbf{I}^v(n,o), \mathbf{I}^v(o,p)$. A második szempont szerint, a papír fehérsége alapján a második csomag papír fehérebb a harmadiknál, de az első és harmadik, illetve az első és második között nem lehet különbséget tenni. Ebben a dimenzióban a következő rendezések élnek: $\mathbf{I}^f(m,n), \mathbf{P}^f(n,o), \mathbf{I}^f(o,p)$. Ha a két szempontot együttesen figyelembe véve akarunk döntést hozni, és azt a papírt keresnénk, amelyik a legvastagabb és legfehérebb, akkor gondba kerülnénk, mert a preferenciáink ciklusosak. Az vastagság dimenziójában az m-t, a fehérség szempontja szerint az y-t kéne választanunk, de a kettős szempont szerint nem lenne "győztes".

A negyedik ciklustípus, a $\mathbf{P}^n$-ciklus, amire Warren S. Quinn önkínzó önkéntes példáját hozhatjuk fel [424]. Quinn gondolatkísérletében olyan speciális szerkezetet szerelnek egy önkéntesre, amellyel fokozatosa, kis lépésekben lehet növelni az áramütés nagyságát. Az eszköz 1000 fokozatot képes kezelni. Az egyes fokozatokban okozott áramütés, illetve az így kiváltott fájdalom nagysága közötti különbséget nem lehet érezni, de minél távolabb van egymástól két fokozat, annál inkább érzékelhetővé válik a különbség, egyre nő a fájdalomérzet. A skála végén már elviselhetetlen mértékű fájdalmat képes kiváltani az eszköz. A kísérlet szereplője minden héten dönthet arról, hogy egy fokozattal növeli-e a szintet, vagy kiszáll a játékból. Minden növelésért jutalmat kap. A kísérletben minden egyes lépésre igaz, hogy az önkéntes jobbnak tartja az adott fokozatot a megelőzőnél, hiszen a két szomszédos fokozatban nem érzékelhető különbség az áramütések, illetve az okozott fájdalmak nagysága között. Ám minél messzebbre jut az első lépéstől, annál biztosabban mondhatjuk, hogy a "sokadik" fokozatot már nem találja jobbnak a legelsőhöz képest (hiszen ekkor már érzékelhetően nagy különbség van a fájdalmak mértékében). Azt kell mondanunk tehát, hogy a lépésenkénti összehasonlítás alapján ez a preferencia-rendezés intranzitív. Ha viszont az egyes fokozatokat "összevonjuk", és nagyobb intervallumokat veszünk figyelembe, akkor megváltozik a helyzet. Ha például a 0., az 500. és az 1000. lépést hasonlítjuk össze, akkor egyértelmű, hogy az egyén azt fogja mondani, hogy preferálja az 500. fokozatot az 1000.-hez képest, a 0.-at az 500.-hoz képest, valamint a 0.-at az 1000.-hez képest, vagyis a rendezése ekkor és így tranzitívvá válik. Quinn kísérletete $\mathbf{P}^{1001}$-ciklusú intranzivitást mutat be, hiszen 1001 lépésen keresztül, minden egymást követő fokozatban **P**-rendezésről van szó, míg a sorozat két végpontja között már nem áll fent ez a kapcsolat. Quinn példája persze nem igazán életszerű, de hogy ezzel ő is tisztában van, azt szépen jelzi az a tény, hogy a fenti gondolatkísérletén túl említést tesz teljesen hétköznapi jelenségekről is, amelyekben ugyanazt a jelenséget tapasztalhatjuk, mint az önkínzó önkéntes példájában. Minden egyes evésünket felfoghatunk úgy, mint egy kielégüléssel, élvezettel járó apró lépést, és egyik evéstől sem nő meg látható módon a súlyunk, de ha kicsit "messzebbről" vagyis nagyobb időszeletben nézünk a folyamatra, akkor már észrevehetjük a hízásunkat. De ugyanez igaz minden egyes cigi elszívása okozta élvezet és az esetleg megjelenő tüdőrák fájdalma vagy a mindennapi lustálkodásaink öröme és az elpocsékolt élet feletti kesergés között feszülő ellentmondásra is [424].

A bemutatott ciklusos jelenségekből két mozzanatot emelnék ki. Egyfelől a példák-

ban láthatóan gyakran van fontos szerepe annak a ténynek, hogy az egyes lépések (alternatívák) között nm tudunk igazán különbséget tenni. Ilyen esetekben a **diszkriminációképességünk** alacsony foka (vagy másként: a diszkriminációképtelenségünk) magyarázhatja meg a jelenséget. Másfelől a fenti példák között "megjelent" egy olyan új jelenség is, amilyennel még nem foglalkoztunk eddig. Yew-kwang Ng példájában ugyanis egyértelműen az okozta a "problémát", hogy először két rendezési szempont szerint kellett (lehetett) ítélkezni, majd ezek után és ezek alapján kellett a végső eredményt kihozni. A két említett mozzanat nem teljesen független egymástól.[48] Az elméleti kapcsolat talán nem szükségszerű, de történetileg biztosan összefonódott a két dolog. Amikor Amos Tversky elemezte az intranzitivitás problémáját [548], és a cikkében felvetette a diszkrimináció problémakörét mint fontos magyarázó tényezőt, akkor – többek között – R. Duncan Luce-ra is hivatkozott. Utóbbi volt az, aki bevezette a szemirendezés (semiorder) fogalmát [313], ami egy preferenciarelációból és egy indifferencia-relációból áll az alábbi feltételek mentén.

(BL.66) SEMIORDER $\stackrel{\text{def}}{=}$

1) $\forall m \forall n (\mathbf{P}(m,n) \veebar \mathbf{P}(n,m) \veebar \mathbf{I}(m,n))$
1) $\forall m (\mathbf{I}(m,m))$
3) $\forall m \forall n \forall o \forall p ((\mathbf{P}(m,n) \wedge \mathbf{I}(n,o) \wedge \mathbf{P}(o,p)) \rightarrow \mathbf{P}(m,p))$
4) $\forall m \forall n \forall o \forall p ((\mathbf{P}(m,n) \wedge \mathbf{P}(n,o) \wedge \mathbf{I}(n,p)) \rightarrow \neg(\mathbf{I}(m,p) \wedge \mathbf{I}(o,p)))$

szemirendezés **P** a preferenciareláció, **I** az indifferenciareláció

Tversky megjegyezte, hogy szemirendezés a lexikografikus rendezésnek egyik típusát is jelenti egyben. Észre kell azonban vennünk, hogy ez a fogalom egy nagyon fontos ponton eltér az eddig rendezési fogalmainktól. Itt ugyanis nem egy, hanem két rendezési fogalomról (egy preferencia- és egy indifferenciarelációról) van szó, ami az "egyszerű" rendezésekhez képest egy eddig nem látott problémához vezet: ahhoz, hogy a kétféle rendezési relációtól végleges (valamiféle együttes) eredményt várhassunk el, az kell, hogy a "belső" relációk között is megállapítsunk valamilyen viszonyrendszert. Csak ennek létezésekor (és ennek ismeretében) lehetséges ugyanis az egyesített végeredmény elérése. A probléma kezelése, megoldása azonban már kivezet minket a preferencialogika szűkebb értelemben vett területéről, ezért itt most nem foglalkozom vele. Később természetesen még visszatérek erre a kérdésre.

Preferencia és modális logika

von Wright 1972-es cikkében már két lehetséges világ értelmezést kínál az intrinzikus preferencia fogalmára [566, 146-147]. Lehetéges világ szemantika alapján: m-világ (illetve n-világ) azt jelenti, hogy abban a világban m (illetve n) fennáll.

C_i-világok (amikben nem áll fent mindkettő??????

Def$_1$: m $\forall$-preferált n-nel szemben C_i-világ feltételei közepette, ha minden olyan C_i-világ, amely egyben m-világ, de nem n-világ, preferált minden olyan C_i-világgal szemben, amely egyben n-világ, de nem t-világ

[48] A kétféle szempont, és a köztük teremtendő "összehangolási kényszer" egyébként megjelenik Quinn példájában is.

Def$_2$: m $\exists$-preferált n-nel szemben C_i-világ feltételei közepette, ha néhány olyan C_i-világ, amely egyben m-világ is preferált minden olyan C_i-világgal szemben, amely nem n-világ, egyben n-világ, de nem létezik olyan C_i-világ, amely t-világ is egyben, és preferált minden olyan C_i-világgal szemben, amely nem m-világ.

az előzmény Craig Boutilier preferencia [69]

Semantically, a goal can be viewed as a set of possible worlds, those states of affairs that satisfy the goal proposition. Intuitively, if we ignore considerations of ability, the set of goal worlds should be those considered most desirable by an agent (or its designer). To achieve all goals is to ensure that the actual world lies within this desirable set.

Ceteris paribus preferencia modális logikai alapon

Korábban már bevezettem a ceteris paribus preferencia fogalmát (↰105.o.), és jeleztem azt is, hogy a fogalom formális leírását modális logikai alapon lehet jól megoldani. Most érkeztünk el arra a pontra, amikor bemutathatom, hogyan lehet ezt a formalizálást elvégezni. Kiindulásként idézzük fel az amplifikáció elvének formuláját, amit von Wright összekapcsolt a ceteris paribus preferenciával (a formulából itt elhagyjuk a kvantorokat).

$$\mathbf{P}(m,n) \leftrightarrow (\mathbf{P}(m \wedge o, n \wedge o) \wedge \mathbf{P}(m \wedge \neg o, n \wedge \neg o))$$

A ceteris paribus elv azt akarja megragadni, hogy a preferenciáinkat úgy fejezzük ki két állapot között, hogy közben minden más állapotot változatlannak ítélünk meg. Először érdemes röviden értelmezni, mit is jelent ez a kijelentés. Az összehasonlítandó állapotok összességét nevezhetjük a tárgyalási univerzumnak (universe of discourse). Ebben a modellben a ceteris paribus elvet úgy kell értelmeznünk, hogy ha két állapotot (m-t és n-t) vetünk össze, akkor minden o állapotot, alternatívát a tárgyalási univerzumban, amelynek nincs direkt relevanciája az értékelés során, konstansnak tekintünk az összehasonlítási folyamat egészében. Vegyünk egy példát! Tegyük fel, hogy az esőköpenyt (r) preferáljuk az esernyővel (u) szemben uPr miközben harmadik potenciális alternatívaként ott van az az állapot is, hogy csizmát viselünk (b). Ekkor a tényleges, aktuális preferenciánkat így írhatjuk le: $r \wedge \neg b P u \wedge b$. Megváltozna az értékelésünk, ha elvesztenénk a csizmánkat, mert ez a tény átfordítaná a preferenciánkat, és az esőkabátot preferálnánk az esernyővel szemben, de az amplifikációs szabály elfogadása kizárja ezt a lehetőséget. A ceteris paribus elv érvényessége azt jelenti, hogy a csizma (r alternatíva) léte (vagy nem léte) nem befolyásolhatja a döntést, tehát mindvégig konstansnak kell tartanunk a hatását, és ezért el kell tekintenünk tőle.

Doyle-Wellman 1991 Desire = ceteris paribus preference? Hic et nun itt es most preferencia????

Ceteris paribus elv a kontextualis érvényességet, a preferenciák relativitását fejezi ki

Preferencia: állapotoktól propozíciókig

Fontos előzményként kell tekintenünk Jon Doyle és Michael Wellman cikkére, amelyben a szerzők először bontották fel két összetevőre a ceteris paribus preferenciák reprezentálásának feladatát [133]. Doyle és Wellman abból a praktikus szempontból fontos kérdésre keresste a választ, hogy miként lehetne a propozíciók között összehasonlításokat (és ezáltal preferenciarendezéseket) tenni. Doyle-ék számára azért volt fontos ez a kérdés, mert

von Wright állapotok közti preferenciákról beszélt, amiket individuális propozíciókkal mindig le lehetett írni, de amikor a mérnökök a számítógépes döntéstámogató rendszerek fejlesztése során a céltételezésekben, eszközkiválasztásokban megbúvó preferenciákat is modellezni akarták, akkor hamar kiderült, hogy nem járható út az, amikor a preferenciákat az individuumok szintjén adják meg a rendszer számára, mert végtelen sok adat összegyűjtésére, betáplálására és számítására lenne szükség. Ehelyett persze kínálta magát az a lehetőség, hogy az individuális szintű adatok helyett általánosabb szintű adatokat kezeljenek, de itt újfajta gondok jelentkeztek.[49]

Az individuumok összehasonlítása "egyszerű". Ha Béla magasabb, mint Kati, vagy Hans jobban szereti a magasabb fizetést az alacsonynál, akkor erre könnyű képleteket, számítógépes programokat írni. De mi van akkor, ha individuumok valamilyen osztályát kell összehasonlítani. Például egy ruházati cég számára fontos lehet az olyan megállapítás, hogy a dánok magasabbak és soványabbak, mint a szicíliaiak, mert ennek ismeretében más méretű ruhákkal megy be a két piacra. Amikor ilyenfajta döntéseket kell hozzunk, akkor persze mindig találhatunk kivételeket, de ezt tudva is törekszünk arra, hogy ne individuumokra, hanem individuumok egész osztályára tudjunk összehasonlító állításokat tenni. Vannak tehát az individuumok (ahol könnyen tudunk összehasonlításokat tenni), és vannak az individuumok osztályát leíró propozíciók, ahol – egyelőre – még nem megoldott az összehasonlítás.

Első lépésként megpróbálhatjuk az individuumokra vonatkozó összehasonlítást (preferenciarendezést) az individuumok valamely osztályára vonatkozó összehasonlítássá alakítani a **sima felemelés** (simple lifting) módszerével. Tegyük fel, hogy az alábbi egyedeink vannak egy adott Ω halmazban: $m, n \in \Omega$, és ezek a következő osztályokba tartoznak: $m \in \phi, n \in \psi$. Ekkor nyilván igaz, hogy: $\phi, \psi \subseteq \Omega$. Ekkor az egyedek közti a már jólismert preferenciarendezési műveleteket értelmezhetjük: a $\geq$ legalább olyan jó, a $>$ szigorúan jobb és az $\sim$ indifferens relációt. Ezek alapján – a sima felemelés módszerével – definiálhatjuk az osztályok (propozíciók) közti rendezéseket is.

(BL.67)
1) $\phi \geq \psi \stackrel{\text{def}}{=} \forall m \in \phi \forall n \in \psi (m \geq n)$
2) $\phi > \psi \stackrel{\text{def}}{=} \forall m \in \phi \forall n \in \psi (m > n)$
3) $\phi \sim \psi \stackrel{\text{def}}{=} \forall m \in \phi \forall n \in \psi (m \sim n)$

Ezzel a megoldással azonban komoly gondok adódnak, amit akkor érthetünk meg igazán, ha egy konkrét példán keresztül szembesülünk az elképzelés nem kívánt következményeivel. Ehhez először definiálnunk kell pár további feltételt. Elsőként rögzítenünk kell azt, hogy az összehasonlítandó szempontokat (propozíciókat) logikailag függetlennek kell tartani egymástól. Ha ezt nem követelnénk meg, akkor az összehasonlítások során logikai ellentmondásokhoz juthatnánk (erről már volt szó korábban).[50] A logikai függetlenséghez az szükséges, hogy a propozíciók és komplementereik összes lehetséges metszetében legyen előfordulás, a dolgok szemléltetéséhez elégséges a szemifüggetlenség – némileg gyengébb – fogalma is, amikor nem kell az összes lehetséges kombinációra megkövetelni a nem-ürességet.

[49] Doyle és Wellman nem igazán választotta el a preferenciarendezéseket a másfajta rendezésektől, ezért sokszor egyszerűen csak összehasonlításokról beszéltek a cikkükben. Ez a mondanivalójuk érvényességén semmit sem változtat, és ha nem zavaró, olykor én is átveszem a nyelvezetüket. Ebben a kontextusban tehát az összehasonlítás műveletét a preferenciarendezés értelmében használjuk, még akkor is, ha nyilvánvalóan másfajta – például magasság szerinti – összehasonlításról van szó.

[50] Megkeresni és hivatkozni a megfelelő helyre !!!!!!!!!!

(BL.68) $\forall i \forall j(\phi_i \cap \phi_j \neq \varnothing \wedge \phi_i \cap \overline{\phi}_j \neq \varnothing \wedge \overline{\phi}_i \cap \phi_j \neq \varnothing \wedge \overline{\phi}_i \cap \overline{\phi}_j \neq \varnothing)$

logikai függetlenség (HUN) – logical independence (ENG)

(BL.69) $\forall i \forall j(\phi_i \cap \overline{\phi}_j \neq \varnothing \wedge \overline{\phi}_i \cap \phi_j \neq \varnothing)$

logikai szemifüggetlenség (HUN) – logical semi-independence (ENG)

A logikai szemifüggetlenség alapján definiálhatunk háromféle legnagyobb rendezési fogalmat, amiket a következő formulákkal fejezhetünk ki.

(BL.70)
1) $\forall i((\phi_i \geq \overline{\phi}_i \wedge \forall k \forall l(\phi_k \overline{\phi}_l \neq \varnothing)) \rightarrow \forall m \in \phi_i \forall n \in \overline{\phi}_i(m \geq n))$
2) $\exists i(\phi_i \geq \overline{\phi}_i \rightarrow \forall m \in \phi_i \forall n \in \overline{\phi}_i(m \geq n))$
3) $\exists i \exists j(\phi_i \geq \overline{\phi}_i \rightarrow \forall m \in \phi_i \forall n \in \overline{\phi}_j(m \sim n))$

(no comparatives theorem)

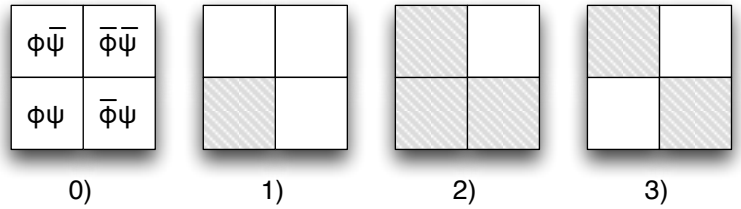
A legnagyobb minősítésnek fent meghatározott három fokozatának Doyle-ék a következő nveket adták:

- maximálisan legnagyobb (maximal greatness)
- közepesen legnagyobb (intermediate greatness)
- minimálisan legnagyobb (minimal greatness)

Jobban megérthetjük mind a fenti definíciókat, mind a bemutatni kívánt problémát, ha konkrét Doyle és Wellman a következő példát hozta. Vegyünk két szempont, a fizetést és a napi munkaidő hosszát, és ezek alapján ítéljük meg a választható munkahelyeket. Külön-külön mondhatjuk, hogy a magas fizetés jobb, mint az alacsony, illetve a rövidebb munkaidő jobb, mint a hosszabb, ezek alapján egyértelmű összehasonlításokat tudunk tenni. De gondba kerülünk, ha a két szempontot egyszerre akarjuk figyelembe venni. Egy ábra segítségével megvilágíthatjuk, miért. Mivel egy szemponton belül kizáró vagylagos választható lehetőségek vannak, az alternatívák így egymás ellentettjeinek, komplementereinek számítanak. Ezt figyelembe véve jelöljük a választási lehetőségeket a következő módon.

ϕ	– magasabb fizetés	$\overline{\phi}$	– alacsonyabb fizetés
ψ	– rövidebb munkaidő	$\overline{\psi}$	– hosszabb munkaidő

Mivel a két szempont független egymástól, ezért a szemifüggetlenség feltétele fennáll. Ekkor a $(\phi\psi, \phi\overline{\psi}, \overline{\phi}\psi, \overline{\phi}\overline{\psi})$ kompozíciósorozat teljes partíciót alkot, vagyis egymást kölcsönösen kizárják. Ezt a következő módon ábrázolhatjuk.



A választható konkrét munkahelyeket mindig besorolhatjuk az ábra 0) jelű négyzetének valamelyik negyedébe, az 1)-2)-3) jelű négyzetek pedig azt mutatják, hogy a korábban definiált három rendezési reláció melyik negyede(ke)t jelöli ki a legjobbnak. Az 1)

négyzet jelöli a maximálisan legnagyobb rendezés eredményét, amikor azokat a konkrét egyedeket soroljuk ide, amelyek mindkét szempont szerint jobbak voltak. Ezek azok a munkahelyek, amelyek magas fizetést adnak és rövid munkaidejűek (vagyis mindkét szempont szerint jobbak). A 2) négyzet a közepesen legnagyobb rendezésnek felel meg. Ez az összehasonlítás azokat az individuumokat minősíti jobbnak, amelyek legalább egy szempont szerint jobbak a többinél. Ilyenek lesznek mindazok a munkahelyek, amelyek vagy magasabb fizetést, vagy rövidebb munkaidőt kínálnak, vagy mindkettőt egyszerre nyújtják. Ez jóval gyengébb feltétel az előzőhöz képest, így ennek a kritériumnak sokkal több munkahely képes megfelelni. Az ábra 3) jelű négyzete mutatja, hogy mit adna vissza a minimálisan legnagyobb rendezés. Ekkor a két egyenlően legnagyobb állapotot jelent az, amikor az egyik szempont szerint jobbak, a másik szerint rosszabbak az alternatívák, illetve fordítva.

Ezzel a rendezési hármas javaslattal az a legnagyobb probléma, hogy nem ad módot mindenféle más összehasonlításra. Nem kerülhetünk olyan helyzetbe például, amikor a magasabb fizetésű és hosszabb munkaidejű munkahelyet preferálnánk az alacsonyabb fizetéssel és rövidebb munkaidővel szemben. Márpedig olykor kívánatos lenne. Például akkor, amikor a ceteris paribus elvet szeretnénk modellezni. Valahogyan gyöngíteni kell a feltételeken, hogy be tudjuk vezetni a ceteris paribus elvet. Doyle és Wellman újítása az volt ezen a ponton, hogy bevezették a kontextuális összehasonlítás (contextual comparison) fogalmát, amit elvezethet minket a ceteris paribus összehasonlítás lehetőségéhez. Ha azt mondjuk, hogy két individuum 'ugyanaz, miközben minden más változatlan' (same, other things equal), akkor ez két lépésben definiálható. Elsőként bizotsítani kell azt, hogy valamik ugyanazok, amit az individuumok közötti ekvivalenciával ragadhatunk meg, majd azt kell valahogy megfogni, hogy minden más változatlan. Ehhez kell a kontextuális ekvivalencia fogalma, amire már "ráülhet" a ceteris paribus összehasonlítás (comparison ceteris paribus). A kontextuális ekvivalenciát Doyle-ék az alábbi módon definiálták.

(BL.71) $\text{CONTEXTUAL_EQUIVALENCE}(m,n) \equiv m \overset{c}{\equiv} n \text{ mod } \phi, \psi, \ldots \overset{\text{def}}{=} \wp(\wp(\Omega)) \mapsto \varepsilon(\Omega)$

kontextuális ekvivalencia (HUN) – contextual equivalence (ENG)

Az individuumok közti ekvivalenciarelációk segítségével ragadhatjuk meg a propozíciókat, hiszen a propozíciók (vagy azok komplementerei) alá sorolható egyedek mind ekvivalensek egymással ebben a minőségükben. A kontectuális ekvivalencia fogalmának segítségével már definiálni lehet a gyengén nagyobb ceeteris paribus rendezési relációt.

(BL.72) $\text{WEAKLY_GREATER_CP}(\phi,\psi) \equiv \phi \trianglerighteq \psi \overset{\text{def}}{=} \forall m \in \phi\overline{\psi} \forall n \in \overline{\phi}\psi(m \succeq n \wedge (m \overset{c}{\equiv} n \text{ mod } \psi\overline{\psi}, \overline{\phi}\psi))$

gyengén nagyobb ceteris paribus (HUN) – weakly greater ceteris paribus (ENG) – gyengén jobb ceteris paribus (HUN)

Miért is érdekes ez? A definíció azokat az individuumokat hasonlítja össze, amelyekre fennáll, hogy $\phi\overline{\psi}$-be, illetve a $\overline{\phi}\psi$-be tartoznak, nem pedig azokat, amelyekre az lenne igaz, hogy a ϕ-be, illetve a ψ-be tartoznának. Azért értelmes ez így (és nem értelmes másként), mert az az összehasonlítás, amely során mindkét individuum bene lenne ϕ-ben, nem tudna köztük releváns módon különbséget tenni csak az alapján, hogy kielégítik-e ψ-t (hiszen ϕ-t mindketten kielégítik). Mivel fordítva is hasonlót mondhatnánk, vagyis ha az összehasonlítandó egyedek benne lennének a ψ-ben, akkor azokat csak a ϕ alapján lehetne

összehasonlítani. Ezért csak a definícióban alkalmazott összehasonítási mód végezheti el a két propozíció közti összehasonlítást.

(BL.73) $\text{STRICTLY_GREATER_CP}(\phi, \psi) \equiv \phi \rhd \psi \stackrel{\text{def}}{=} \phi \trianglerighteq \psi \wedge \neg(\psi \trianglerighteq \phi)$

erősen nagyobb ceteris paribus (HUN) – strictly greater ceteris paribus (ENG) – szigorúan jobb ceteris paribus (HUN)

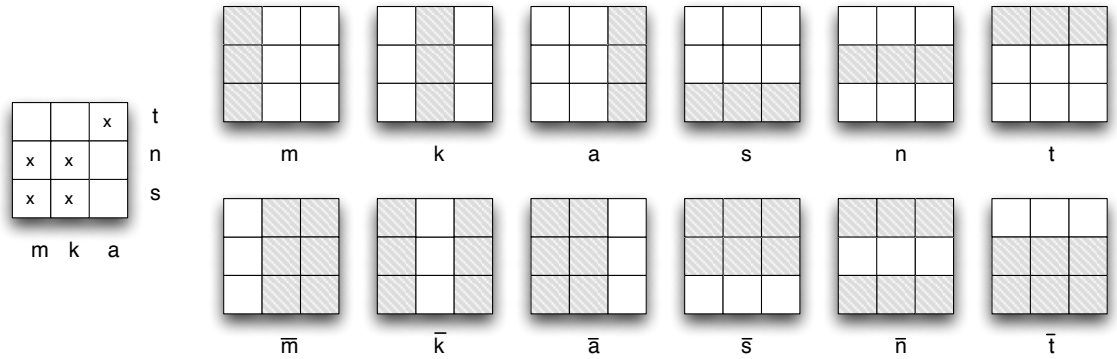
(BL.74) $\text{EQUALLY_GREAT_CP}(\phi, \psi) \equiv \phi \triangleq \psi \stackrel{\text{def}}{=} \phi \trianglerighteq \psi \wedge \psi \trianglerighteq \phi$

egyenlően nagy ceteris paribus (HUN) – equally great ceteris paribus (ENG) – ugyanolyan jó ceteris paribus (HUN)

Nézzünk meg egy újabb példát: hasonlítsuk össze az embereket a testmagasságuk, illetve testsúlyuk alapján. Soroljuk be úgy az embereket, hogy mindkét dimenzióban legyen három-három fokozatunk. A tulajdonságfokozatokat, illetve ezek komplementerét mind egy-egy proporzíciónak tarthatjuk, és az alábbi módon jelölhetjük.

m = magas	$\overline{m}$ = közepes + alacsony	s = sovány	$\overline{s}$ = normál + túlsúlyos
k = közepes	$\overline{k}$ = magas + alacsony	n = normál	$\overline{n}$ = sovány + túlsúlyos
a = alacsony	$\overline{a}$ = magas + közepes	t = túlsúlyos	$\overline{t}$ = sovány + normál

A két tulajdonság fokozatait preferenciasorrendbe rakhatjuk, a magasság szerint a magastól az alacsony felé, a súly szerint a soványtól a túlsúlyos felé csökkenő rendezéssel. A bevezetett jelölésünket használva legyen érvényes a következő két sorrend: $m \trianglerighteq k \trianglerighteq a$, illetve $s \trianglerighteq n \trianglerighteq t$. Végül tételezzük fel még azt, hogy van 5 személyünk, akikre vonatkozóan a következők igazak: van egy magas és sovány, egy közepesen magas és sovány, egy magas és normál súlyú, egy közepesen magas és normál súlyú, végül egy alacsony és túlsúlyos emberünk. Mindezeket az információkat a következő ábra segítségével szemléltethetjük vizuálisan.



A kérdés, hogy mit mondhatunk a ceteris paribus preferenciákról ebben az esetben. Nos, két állítást tehetünk. Mondhatjuk, hogy egyfelől a magas + sovány gyengén jobb ceteris paribus a magas + normál súlyúhoz képest, másfelől a közepes + sovány gyengén jobb ceteris paribus a közepes + normál súlyúhoz képest, vagyis: $m \wedge s \geq m \wedge n$, illetve $k \wedge s \geq k \wedge n$.

Johan van Benthem, Patrick Girard és Olivier Roy szerint a feltétel-nélküliséget modális logikai alapon kell inkább megragadni, ezért kidolgozták a ceteris paribus preferencia $\mathscr{L}_{\wp}$ modális nyelvét és modeljét [554]. Benthenék tanulánya rövid idő alatt komoly szakmai elismeréseket szerzett. Sokan megkerülhetetlennek minősítették a cikket – megédemelten. A tanulmány gondolatmenetét azonban nem mutatom be itt, mert egyfelől túl hosszú lenne, másfelől nagyon komoly modális logikai felkészültség és tapasztalat szükséges hozzá. Meghyn Bienvenu, Jérôme Lang és Nic Wilson kidolgoztak egy olyan általános modális nyelvet és modellt a preferenciára, amely a filozófusok által kidolgozott preferencialogikákat, és mesterséges intelligencia világán belül kifejlesztett preferencia-reprezentációs nyelveket egyaránt képes integrálni [54]. Ennek – más egyéb okok mellett – azért érezték szükségét, mert figyelemmel akartak lenni az adott nyelven történő gépi következtetések számításigényeire is.

A preferencia általános fogalmára nézve különösen igaz, de még az intrinzikus preferenciára vonatkoztatva is kijelenthető, hogy a preferenciáink nem statikusak. Az extrinzikus preferencia együtt változhat azzal, ahogy valamilyen hitünk, ismeretünk megváltozik a világról. Ez a hatás az intrinzikus preferencia esetében is jelentkezhet, bár nyilván lassabban, kevésbé reflektált módon. De a preferenciáink másfajta külső hatásnak is ki vannak téve. Kaphatunk parancsokat, megfogalmazhatnak morális elvárásokat velünk szemben, új információhoz juthatunk, új összefüggéseket ismerhetünk fel, javaslatokat, tanácsokat adhatnak nekünk, melyek eredményeként változhatnak a világra vonatkozó értékeléseink. Johan van Benthem és Fenrong Liu cikke azt elemzi [555], hogy a preferenciáink hogyan változnak a javaslatok (suggestion), parancsok (order) hatására miközben különbséget tesznek ulyan fogalmak között mint a frissítés (update) és szintemelés (upgrade) között. Fenrong Liu 2011-ben könyvet jelentetett meg a preferenciák változásairól, dinamikájáról [311].

1.12 Deontikus logika

A deontikus logika (vagy imperatív logika) az előírások, kötelezések, tiltások, megengedések logikai formalizálásával foglalkozik. Nem célom bemutatni itt a deontikus logika kialakulásának, fejlődésének történetét, nem akarok foglalkozni az elmélet legfontosabb eredményeivel sem. A témáról szóló könyvek, tanulmányok közül íme egy elfogult és teljesnek messze nem mondható lista: [446], [563], [302], [10], [215], [214], [90], [91], [274], [471], [320], magyarul olvasható Solt Kornél munkája [504]. A cselekvéselmélet számára a deontikus logika mint a modális logika egyik ága a leginkább használható. Azért van ez így, mert az előírások mindig valamilyen cselekvésre vonatkoznak, tehát a cselekvéselmélet központi fogalma, a cselekvés van a deontikus logika fókuszában is. A modális logika különböző kalkulusait meghatározó sémák közül a (D) séma az (↪63.o.), amit a deontikus logikához kapcsolhatunk. A deontikus logikában a két modális operátor, a 'szükségszerű' és a 'lehetséges' átértelmezhető úgy, hogy 'kötelező', illetve 'megengedett' legyen a jelentésük. Ekkor a (D) sémát úgy lehet interpretálni, hogy 'ha valami kötelező, akkor az megengedett is egyben'. A deontikus logika mint a kötelezések logikája 'arról szól', hogy mit tudunk mondani a jövőről, pontosabban – a jövőre vonatkozó – kötelezések és megengedések világáról. Minden kötelezés, tiltás, megengedés a "jövőre irányul", hiszen az ilyen helyzetek ágensei (vagy legalábbis egy részük) azzal a várakozással fordul az előírások felé, hogy azok a jövőben teljesülni fog-

nak. Ez a jövőre vonatkozás adja a deontikus logika egyik specialitását, hiszen ez azt jelenti, hogy ez a logikai szakterület nem tényekkel foglalkozik. Ettől még sok minden fontos felismerést, tudást köszönhetünk neki, de azt a hiányosságát, hátrányát mindenképpen rögzíteni érdemes, miszerint ez a diszciplina nem tud azonban semmi értelmeset mondani arról, hogy mikor, miért valósulnak meg azok a cselekvések, amelyekre az előírások irányulnak.[51]

Deontikus operátorok

A deontikus logika első elméletében von Wright négy deontikus operátorról írt [567], és az utána következő szerzők is kitartottak amellett, hogy a deontikus operátorok a logikai négyszögre "húzhatók rá". Persze, ebben a diskurzusban is "felbukkantak" ugyanazok a problémák, amelyek a logikai négyszög más területeken való alkalmazásakor jelentkeztek (↰47.o.). A **kötelező** és a **tiltott** univerzális kategóriákkal nem nagyon voltak problémák, de bizonytalanság, pontatlan fogalmi határok jellemezték a partikuláris kategóriákat. A **megengedett** értelmezhető volt mind az állító, mind a tagadó partikuláris kategóriákra, bár intuitívabbnak tűnt az állító jellegű használata, és olykor "nehézséget okozott" a négyszög (O)-sarokjának megnevezése, illetve pontos értelmezése. A fontosságuk miatt sok nyelvi forma termelődött ki a kategóriák jelölésére. Íme egy – biztos nem teljes – lista magyarul:

kötelező	megparancsolt, kellő, támogatott, obligát, ajánlott, muszáj, szükséges, nem mellőzhető, nem lehet nem megtenni
tilos	tiltott, nem lehet, nem lehetséges, nem szabad, nem megengedett, megengedhetetlen, tűrhetetlen, nem tűrt, elfogadhatatlan, kötelező tartózkodni
megengedett	szabad, lehet, lehetséges, jogos, nem tilos, nem tiltott, tűrt, elfogadott, tolerált
mellőzhető	szabad, lehet, lehetséges, jogos, nem kötelező, elhagyható, elmulasztható, eltekinthető, eltekintett, felmentett, tolerált

A használt kifejezések listájából látszik, hogy a természetes nyelvi szintre (is) jellemző a terminológiai túlhasználat, a pongyolaság. A nyelv rugalmassága miatt egyes kategóriák több fogalomhoz köthetően is feltűnnek, a tagadás különböző nyelvi formáin keresztül többszörösen újrahasznosítjuk az alapterminusainkat. A terminológiai gazdagság (és pongyolaság) vélhetőleg minden nyelven tetten érhető. Angolul is "sokszínűséget" tapasztalhatunk (és ez a lista is nyilván bővíthető lenne):

obligatory	commanded, duty, compulsory, required
forbidden	impermissible, banned, prohibited, prohibition, disallowed
permitted	permissible, allowed, right, free, claim
omissible	non-obligatory, no-duty, duty-free, gratuitous, exemption

Csakúgy, mint a kategoriális logika bármely más területén, a legnagyobb problémát itt is a partikuláris kategóriák jelentetik. A **megengedett** fogalmát ugyanis a hétköznapi/jogi kommunikáció során többféle értelemben is használjuk. Induljunk ki Jan C. Joerden szemléletes példájából [263, 204]. Két mondatban is ugyanazt a 'megengedett' terminust használjuk, miközben más és más jelentés tulajdonítunk neki.

(1) Ha valaki egy közúti baleset helyszínére ér, megengedett számára, hogy segítsen.

(2) Ha valaki megszerezte a vezetői engedélyét, megengedett számára, hogy gépjárművet vezessen.

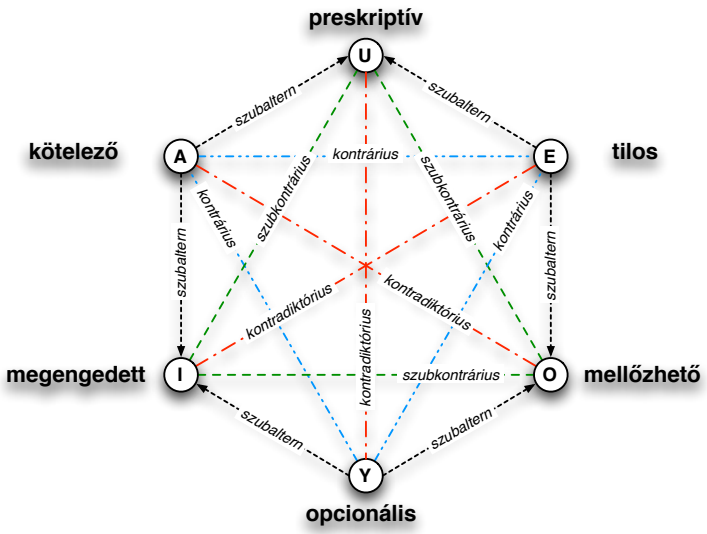
[51] Erről a kérdésről bővebben írtam: [516].

Az első mondatra azonnal rá lehet vágni, hogy a baleset helyszínén nemcsak megengedett, de kötelező a segítség. Ez persze akkor azt is jelenti, hogy ebben az esetben a kötelezettség maga után vonja a megengedés aktusát, amiből következik, hogy a két fogalom között semmiképpen sincs ellentétes viszony. A második mondatban azonban más a helyzet. Ekkor ugyanis a kötelezettséget épp', hogy szembenállónak érezzük a mondatba foglalt megengedéssel szemben. Arról pedig nyilvánvalóan szó sincs, hogy a jogosítvány megszerzésével párhuzamosan kötelességünk lenne autót vezetni. Ebben az esetben a kötelezés már kontradiktórikus viszonyban áll a megengedéssel. Joerden két példája mellé odatehetek még egy harmadikat is (ami az első kettőhöz képest megintcsak – részben – eltérő jelentéssel rendelkezik).

(3) Megengedett, hogy ne szavazzon valaki a parlamenti választásokon.

Ez a megengedés azt fejezi ki, hogy valami (itt: a szavazás cselekedete) nem kötelező, ami viszont logikailag "megengedi" azt az értelmezést, hogy ez a valami, vagyis a szavazás akár tilos is lehet.

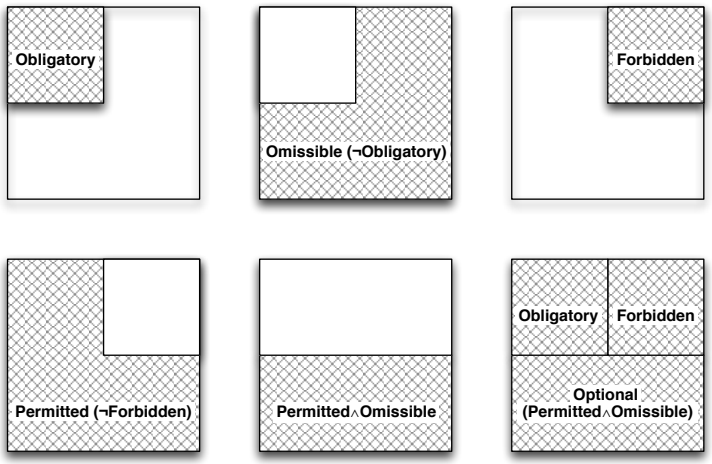
Van tehát három példamondatunk, amelyekben három – egymást részben átfedő, de nem teljesen azonos – jelentés mellett használjuk ugyanazt a nyelvi jelet, a 'megengedés' terminusát. A probléma tisztázásának elvi lehetőségét a logikai hatszög eredményeinek alkalmazása nyújthatja ezen a területen is. A logikai hatszög "felfedezése" után G. Kalinowski vezette be a deontikus hatszög fogalmát.[52] Az univerzális deontikus kategóriák, a **kötelező** (OBLIGATORY) és a **tiltott** (FORBIDDEN) fölöttes kategóriájaként be kell vezetnünk a **preskriptív** (PRESCRIPTIVE) fogalmát, ami a deontikus hatszög (U)-sarkába "kerül". A két partikuláris deontikus kategóriára használhatjuk a **megengedett** (PERMITTED), illetve az **mellőzhető** (OMISSIBLE) fogalmakat, és ezek fölöttes fogalmaként az (Y)-sarokba betehetjük az **opcionális** (OPTIONAL) kategóriát.



A deontikus logikai hatszög használata ugyanazért kívánatos, mint amiért a logikai hatszöget "szeretjük": egyértelmű kapcsolatok teremthetők így a hatszög kategóriái között

[52] G. Kalinowski, La logique des normes. PUF, Paris (1972), idézi: [364]. Jörg Hansen szerint a deontikus logikai négyszöget Arthur N. Prior alkalmazta először 1955-ben [212].

(**kontradiktórius**, **kontrárius**, **szubkontrárius**), illetve olyan kategoriális felosztáshoz juthatunk, amelyre már érvényesül a JEPD-elv (ami a logikai négyzet elemeire még nem teljesült). Utóbbi szempont (elvárás) nagyon fontos mozzanat. Már Jeremy Bentham is törekedett arra, hogy a JEPD-elvnek megfelelő felosztást használjon a jogi relációk között [308, 63], ezért alkalmazta azt a kategóriahármast (**kötelezés**, a **tiltás** és a **nem tiltás és nem kötelezés**), ami teljesen egybevág a deontikus logikai hatszög kontrárius viszonyban levő fogalmainak csoportjával. Ugyanezt az összefüggést szemléltethetjük a deontikus logikai operátorok egymáshoz való viszonyát bemutató ábrán is:



Az ábra szépen mutatja, hogy bár a két univerzális fogalom, a **kötelező** (OBLIGATORY, O) és a **tiltott** (FORBIDDEN, F) egymást kizáró kategóriák, a két partikuláris kategória, a **megengedett** (PERMITTED, P) és a **mellőzhető** (OMISSIBLE) egyszerre vannak átfedésben egymással, illetve valamelyik preskriptív fogalommal. Ezek az átfedések okozzák a deontikus operátorok kezelésének nehézségeit, pongyolaságát. A JPED-elv viszont fennáll a két preskriptív kategória (kötelezés és tiltás), valamint a két partikuláris kategória konjunkciója, tehát az **opcionális** (OPTIONAL) kategória között. Természetesen a deontikus logikai hatszög elfogadása, alkalmazása nem ígér csodát. Bevezetése "mindössze" azt az előnyt ígéri, hogy sok esetben azonosítani tudjuk, hogy valamely konkrét nyelvi megnyilatkozáshoz melyik fogalmat (logikai kategóriát) kell hozzárendelnünk. A fejezet első felében bemutatott konkrét mondatainkra visszautalva: az első példa a deontikus logikai hatszög (I)-sarkába, a második az (Y) pontjába, a harmadik pedig az (O) esetbe tartozik.

Sztenderd deontikus logika

A korábban bemutatott modális axiómák közül a (D) deontikus axióma (↰63.o.) jelenti a kulcsot a **sztenderd deontikus logika** (Standard Deontic Logic, SDL) felépítéséhez. A deontikus logikai séma definiálásakor a szükségszerű, illetve a lehetséges ($\Box$, illetve $\Diamond$) modális operátorok helyett ezek deontikus megfelelőit ($\underline{\mathbf{O}}$, illetve $\underline{\mathbf{P}}$) használhatjuk. A deontikus operátorok hatókörében levő A szimbólum nem egy cselekvés neve (ahogy azt von Wright még az első modelljében tételezte), hanem egy olyan propozíció jele, amely propozíció azt írja le, hogy az adott cselekvést végrehajtották (ezt az értelmezést Prior javasolta először). A deontikus logikai szakirodalomban elterjedt felfogás szerint tehát

$\underline{\mathbf{O}}(A)$ interpretációja: 'kötelező, hogy az a helyzet megvalósuljon, amelyet az A leíró mondattal fejezünk ki' [212]. A továbblépéshez rögzítenünk lell a deontikus logika nyelvét.

(DL.0) $\mathscr{L}_{DL}(\mathscr{L}_{QL}, \underline{\mathbf{O}})$, ahol

$\mathscr{L}_{QL}$	propozícionális logika
$\underline{\mathbf{O}}$	kötelező deontikus logikai operátor

A deontikus logika modelljét – Jörg Hansen leírását követve [212] – a következő axiómasémák segítségével építhetjük fel, amihez hozzávesszük még az $\underline{\mathbf{F}}$ tiltás és a $\underline{\mathbf{P}}$ megengedés operátorának definícióját.

(DL.1) $\textsc{permitted}(A) \equiv \underline{\mathbf{P}}(A) \stackrel{\text{def}}{=} \neg\underline{\mathbf{O}}(\neg A)$

megengedett az A cselekvés akkor és csak akkor, ha nem kötelező tartózkodni a végrehajtásától.

(DL.2) $\textsc{forbidden}(A) \equiv \underline{\mathbf{F}}(A) \stackrel{\text{def}}{=} \underline{\mathbf{O}}(\neg A)$

tiltott az A cselekvés akkor és csak akkor, ha kötelező tartózkodni a végrehajtásától.

(DL.3) Ha $\vdash A \leftrightarrow B$, akkor $\vdash \underline{\mathbf{O}}(A) \leftrightarrow \underline{\mathbf{O}}(B)$

(Ext) extenzionalitás elve (principle of extensionality): ha A kötelező ($\underline{\mathbf{O}}(A)$), akkor minden vele ekvivalens B propozíció is kötelező ($\underline{\mathbf{O}}(B)$).

(DL.4) $\underline{\mathbf{O}}(A \wedge B) \rightarrow (\underline{\mathbf{O}}(A) \wedge \underline{\mathbf{O}}(B))$

(M) monotonitás elve (principle of monotony): ha A és B együttesen kötelező, akkor egyenként is az.

(DL.5) $(\underline{\mathbf{O}}(A) \wedge \underline{\mathbf{O}}(B)) \rightarrow \underline{\mathbf{O}}(A \wedge B)$

(C) kumulativitás elve (principle of cumulativity): ha A és B kötelező, akkor ketten együtt is azok (ez a monotonitás elvének megfordított fennállását fejezi ki).

(DL.6) $\underline{\mathbf{O}}(A) \rightarrow \underline{\mathbf{P}}(A)$

(D') deontikus axióma (deontic axiom): ha az A cselekvés kötelező ($\underline{\mathbf{O}}(A)$), akkor megengedett is egyben ($\underline{\mathbf{P}}(A)$).

(DL.7) $\underline{\mathbf{O}}(A \vee \neg A)$

(N) deontikus szükségességi szabály (deontic necessitation rule): ha valami tautológia, akkor az kötelező.

Hansen megadta azt is, hogyan teremthető kapcsolat von Wright deontikus logikájával. Az (M) és (C) axiómák alapján levezethető von wright-i **disztribúció elv**, miszerint: $(\underline{\mathbf{P}}A \vee \underline{\mathbf{P}}B) \rightarrow \underline{\mathbf{P}}(A \vee B)$. A (D') séma ekvivalens pedig a von wright-i **megengedés elvével** (principle of permission), ami így néz ki: $\underline{\mathbf{P}}A \vee \neg\underline{\mathbf{P}}A$. Mindezek alapján definiálhatunk egy deontikus nyelvet és egy arra épülő deontikus rendszert.

DEONTIKUS LOGIKAI NYELV (THE LANGUAGE OF DEONTIC LOGIC)

A deontikus logika $\mathscr{L}_{DL}$ ábécéje megegyezik a propozícionális logika $\mathscr{L}_{PL}$ nyelvével, amely tartalmazza még az '$\underline{\mathbf{O}}$' deontikus operátort. Az $\mathscr{L}_{DL}$ nyelvet ugyanúgy definiáljuk, mint a $\mathscr{L}_{PL}$ propozícionális logika nyelvét annyi többletszabály tételezésével, hogy ha $A \in \mathscr{L}_{DL}$, akkor $\underline{\mathbf{O}}(A) \in \mathscr{L}_{DL}$ is igaz. A $\underline{\mathbf{P}}(A)$ megengedés operátorával rövidíthetjük a $\neg\underline{\mathbf{O}}(\neg A)$, illetve a $\underline{\mathbf{F}}(A)$ tiltás operátorával $\underline{\mathbf{O}}(\neg A)$ kifejezéseket.

SZENDERD DEONTIKUS LOGIKAI RENDSZER (THE SYSTEM SDL)

> Az *SDL* az a legkisebb halmaz, amely tartalmazza a $\mathscr{L}_{DL}$ nyelv öszes tautológiapéldányát, az (M), (C), (D) és (N) axiómasémákat, valamint zárt az (Ext) és a modus ponens szabályaira nézve.

Megemlítem, de mivel nem használom a modellépítésben, nem mutatom be azt a továbbfejlesztési irányt, amely a deontikus logikai hatszög további kiterjesztéseként tekinthető: Alessio Moretti, The Geometry of Standard Deontic Logic [364].

A deontikus logika redukciója az alethikus logikára

Egymástól függetlenül, szinte teljesen egyidőben, lényegét tekintve ugyanazt a redukciós programot fogalmazta meg Stig Kanger [273] és Alan Ross Anderson [8], [9], [11]ENG)-[12]HUN). Anderson H.B. Bohnert 1945-ből származó javaslatát formalizálta, amely szerint a deontikus logikát vissza lehet vezetni az altehikus logikára úgy, hogy a deontikus operátorokat a szankció, szankcionálás fogalmára hivatkozva definiáljuk. Ehhez a normál alethikus logika elméletére (kalkulusára), az alethikus logika **szükségszerű**, illetve **lehetséges** ($\Box$, illetve $\Diamond$) operátorára, valamint egy új kategóriára, az előírt cselekvés végre nem hajtásakor kiszabandó S **szankció** fogalmára van szükség (Anderson olykor szankciólogikaként hivatkozott az új logikára). Az új elemként bevezetett fogalmat később elnevezték **prohairetikus konstansnak** (prohairetic constant), leginkább azért, mert Kanger is hasonló megfontolásokat követve szintén egy új fogalom felvételét javasolta az alethikus logika sémája mellé, de ő nem a szankciót, hanem a morál által előírt Q cselekvést (pontosabban az azt leíró propozíciót) emelte be a modelljébe. A két javaslat teljesen ekvivalens egymással [212], mégha a kötelezés deontikus operátorának definíciója el is tér egymástól a két esetben:

(DL.8) $\mathbf{O}_Q(A) \equiv \mathbf{O}(A) \equiv \text{OBLIGATORY}(A) \equiv \mathbf{O}A \stackrel{\text{def}}{=} \Box(Q \to A)$

(DL.9) $\mathbf{O}_S(A) \equiv \mathbf{O}(A) \equiv \text{OBLIGATORY}(A) \equiv \mathbf{O}A \stackrel{\text{def}}{=} \Box(\neg A \to S)$

Ezt a redukciót természetesen alá lehet támasztani megfelelő axiómákkal is. A két esethez az alábbi formulákat lehet hozzárendelni:

(DL.10) $\Diamond Q$ $\quad$ (D_Q)

($\mathbf{D}_S$) a formula interpretációja az lehet, hogy ‘lehetséges, hogy S hamis’, vagy másként: ‘a szankcionálás kontingens, nem szükségszerű’,

(DL.11) $\Diamond\neg S$

($\mathbf{D}_Q$) a formulát pedig értelmezhetjük úgy, miszerint ‘lehetséges, hogy Q igaz’.

(DL.12) $\Box(A \to B) \to (\Box A \to \Box B)$

(K) a modális logikai séma azt követeli meg, hogy ha egy kondicionális szükségszerű, akkor a szükségszerű előtag materiálisan implikálja a szükségszerű utótagot.

(DL.13) $\vdash A \to \vdash \Box A$

(RN) ha az A a nyelv eleme, akkor a $\Box A$ kifejezés is az.

Az eddigiek alapján természetesen már sorban definiálhatjuk a deontikus operátorokat:

(DL.14) $\underline{\mathbf{O}}_S(A) \equiv \underline{\mathbf{O}}(A) \equiv \textsc{obligatory}(A) \equiv \underline{\mathbf{O}}A \stackrel{\text{def}}{=} \Box(\neg A \to S)$

(DL.15) $\underline{\mathbf{P}}_S(A) \equiv \underline{\mathbf{P}}(A) \equiv \textsc{permitted}(A) \equiv \underline{\mathbf{P}}A \stackrel{\text{def}}{=} \neg\Box\neg(A \wedge \neg S)$

(DL.16) $\underline{\mathbf{F}}_S(A) \equiv \underline{\mathbf{F}}(A) \equiv \textsc{forbidden}(A) \equiv \underline{\mathbf{F}}A \stackrel{\text{def}}{=} \Box\neg(A \wedge \neg S)$

(DL.17) $\underline{\mathbf{N}}_S(A) \equiv \underline{\mathbf{N}}(A) \equiv \textsc{omissible}(A) \equiv \underline{\mathbf{N}}A \stackrel{\text{def}}{=} \neg\Box(\neg A \to S)$

(DL.18) $\underline{\mathbf{D}}_S(A) \equiv \underline{\mathbf{D}}(A) \equiv \textsc{prescriptive}(A) \equiv \underline{\mathbf{D}}A \stackrel{\text{def}}{=} \neg\Box(\neg A \to S)$

(DL.19) $\underline{\mathbf{A}}_S(A) \equiv \underline{\mathbf{A}}(A) \equiv \textsc{optional}(A) \equiv \underline{\mathbf{A}}A \stackrel{\text{def}}{=} \neg\Box(\neg A \to S)$

Érdemes némi magyarázatot adni arra, miért pont az adott rövidítéseket használták. Nos, az 'optional' az 'Alternative', a 'prescriptive' a 'Dominated', az 'omissible' pedig a 'Non-obligatory' szavak kezdőbetűjére utal. Az andersoni és kangeri megközelítések közti különbséget pedig az fejezi ki, hogy amíg Kanger egy ideális világor feltételezett, addig Anderson egy szubideális világot tett a magyarázatának kiindulópontjába. Bár itt nem jogelmélettel foglalkozunk, mégis érdemes felidézni a híres jogtudós, Hans Kelsen jogfelfogását – egy másik híres jogtudós, H.L.A. Hart interpretálásában:

> „A jog az az elsődleges norma, ami szankciót állapít meg. Nincs olyan törvény, ami tiltja az emberölést, csak olyan törvény van, ami arra utasítja a hivatalos személyeket, hogy bizonyos feltételek esetén bizonyos szankciókat alkalmazzanak azokra, akik megölnek valakit." [228, 50]

Az idézetből ugyan nem derül ki egyértelműen, de tény, hogy Kelsen a jog fogalmát a szankció fogalmából eredeztette, amit más jogtudósok vitattak ugyan, de ez számunkra itt nem lényeges. Az viszont fontos, hogy pontosan ezt a kelseni megközelítést reprodukálta a logika nyelvén Anderson és Kanger deontikus logikai redukciója. Peng-Hsiang Wang – Kangerék után jó ötven évvel – arra vállalkozott, hogy tisztázza, összeegyeztethető-e Kelsen jogfelfogása és Anderson deontikus logikai elmélete [573]. A válasza az lett, hogy igen, de ehhez részlegesen át kell alakítani az andersoni formalizmust az alapján, ahogy von Wright különbséget tesz a kétfajta **Ought** kategória, a deontikus és technikai kötelezőség között. Nem mutatom be, csak az érdekesség kedvéért említem meg, hogy J.J.C. Meyer az andersoni alapokat integrálta a dinamikus logikával, és így a szokásostól eltérő felépítésű deontikus logikát dolgozott ki [353].

Kellellentétes parancs paradoxona

Arthur N. Prior fedezte fel elsőként a materiális implikáció analógiájaként megteremtett deontikus logikai implikációval kapcsolatos paradoxont [422]. Tanulmányában – von Wright pár évvel korábban kidolgozott deontikus rendszerére utalva – Prior először rögzíti azt a nyilvánvaló tényt, hogy egyfelől az implikáció átírható konjunktív alakba, másfelől a kötelezés traszformálható megengedésbe, így az alábbi átalakítások érvényesek:

$$\underline{\mathbf{O}}(A \to B) = \underline{\mathbf{O}}(\neg A \wedge B) = \neg\underline{\mathbf{P}}\neg(\neg A \wedge B) = \neg\underline{\mathbf{P}}(A \wedge \neg B)$$

Ezután Prior felidézi von Wright deontikus rendszerének egy törvényét, miszerint ha egy cselekvés tiltott, akkor az tiltva van bármely más cselekvéssel konjunkcióba állítva, de tiltva van bármely más cselekvéstől való tartózkodással együtt is. Ezt formálisan így lehet kifejezni:

$$(\neg \underline{\mathbf{P}}A \to \neg \underline{\mathbf{P}}(A \wedge B)) \wedge (\neg \underline{\mathbf{P}}A \to \neg \underline{\mathbf{P}}(A \wedge \neg B))$$

A továbbiakban a formula második összetevője az érdekes Prior számára, mivel abban el lehet végezni egy helyettesítést a deontikus implikáció fenti összefüggése alapján. Ekkor kapjuk a következő formulát:

$$\neg \underline{\mathbf{P}}A \to \underline{\mathbf{O}}(A \to B)$$

Ennek az értelmezése során meglepő következtetéshez juthatunk. A formula ugyanis úgy interpretálható, hogy ha valami tiltva van ($\neg \underline{\mathbf{P}}A$), akkor ha valaki mégis megteszi azt (A-t), akkor meg kell tenni bármi mást is (B-t). Prior ezt azzal példázza, hogy ha tiltott lopni, de valaki mégis lop, akkor – például – a házasságtörést is el kell követnie. Ez ellenkezik a normákkal kapcsolatos intuíciónkkal. Ezt a problémát nevezték el Prior-paradoxonnak.

Chisholm tovább gondola a problémát, és 1963-ban írt cikkében új fogalmat vezetett be [94]. A logikában akkor már régóta használt **tényellentétes** vagy kontrafaktuális (contrafactual) kifejezés mintájára megalkotta a **kellellentétes** vagy kontranormált (contrary-to-duty) kötelesség fogalmát.[53] Ezen azokat a helyzeteket értette, amelyekbe akkor kerülünk, miután megszegtük valamely kötelesség előírását, egy újabb – feltételes – kötelességhez kell igazodnunk. Vétségeink újabb kötelmeket szülnek – mondja Chisholm. Az igazi problémát azonban az a tény jelenti, hogy az ilyen helyzetek paradoxont rejtenek magukban. Chisholm a következő példával szemlélteti mindezt.

1) Kötelező, hogy valaki átmenjen segíteni a szomszédjának.
2) Kötelező, hogy ha valaki átmegy segíteni, akkor előtte szólnia kell, hogy megy.
3) Kötelező, hogy ha valaki nem megy segíteni, akkor tilos előtte szólnia, hogy megy.
4) Valaki nem megy át.

Az 1) és 4) mondatok egyértelműen leírhatók, de a 2) és 3) propozíciókat kétféleképpen is megragadhatjuk. Ebből következik, hogy ezt a négy mondatot négyféleképpen is formalizálhatjuk, bár egyik sem lesz tökéletes. A négyféle leírási mód így néz ki:

$$\begin{array}{rcccc}
\text{i)} & \underline{\mathbf{O}}(A) & \underline{\mathbf{O}}(A \to B) & \neg A \to \underline{\mathbf{O}}(\neg B) & \neg A \\
\text{ii)} & \underline{\mathbf{O}}(A) & \underline{\mathbf{O}}(A \to B) & \underline{\mathbf{O}}(\neg A \to \neg B) & \neg A \\
\text{iii)} & \underline{\mathbf{O}}(A) & A \to \underline{\mathbf{O}}(B) & \underline{\mathbf{O}}(\neg A \to \neg B) & \neg A \\
\text{iv)} & \underline{\mathbf{O}}(A) & A \to \underline{\mathbf{O}}(B) & \neg A \to \underline{\mathbf{O}}(\neg B) & \neg A
\end{array}$$

Akármelyik leírási kísérletet vizsgáljuk meg, mindegyik esetről kiderül, hogy vagy inkonzisztens a négy formula, vagy az egyik kijelentés levezethető valemelyik másból, tehát nem függetlenek a mondatok egymástól [535], [212]. Márpedig az intuíciónk szerint a fenti négy mondat egymástól független és konzisztens.

Korábban már szó esett arról, hogy a deontikus logikára nem érvényes a (T) modális kalkulus (↰64.o.). Ebben az esetben viszont mondhatjuk azt is, hogy a formula kontraponáltja sem lehet igaz, vagyis:

[53] Magyar nyelven egyetlen forrást találtam, amely hivatkozott a contrary-to-duty kategóriájára. Szolcsányi Tibor a ‘kötelezettség-ellentétes norma’ kifejezést vezette be a ‘contrary-to-duty’ fordításaként [535]. Tartalmilag megfelelő ez a kifejezés is, de túl hosszúnak találom, ezért kitartok saját javaslatom mellett.

$$\nvdash \neg A \rightarrow \neg \underset{\sim}{\mathbf{O}} A$$

Természetes nyelven ezt úgy fejezhetjük ki, hogy ha nem áll fent egy A propozícióval leírt cselekvés, abból még nem következik az, hogy ne lenne az a cselekvés kötelező. Ez teljesen egybevág azzal a mindennapos tapasztalattal, hogy a normákat nem egyszer megsértik az emberek, de ettől a normák még nem vesztik el az érvényességüket.

Diadikus deontika

A kellellentétes vagy általában a feltételes parancsokkal kapcsolatos problémák megoldásának reményében dolgozták ki a diadikus deontikát, vagyis a **diadikus deontikus logikát**, amit – a monadikus deontikával szemben – képessé tettek arra, hogy a feltételes kötelességet (conditional obligation) kezelni lehessen vele. Von Wright már 1956-ban javasolta, hogy be kellene vezeti a diadikus deontikus operátorokat (dyadic deontic operator), amelyek alkalmazásával kivédhetővé válnak a jelzett gondok [562]. A diadikus formulát a következő alakban javasolta felírni:

(DL.20) $\underset{\sim}{\mathbf{O}}_D(A/B) \equiv \underset{\sim}{\mathbf{O}}(A/B)$

A formulát úgy kell értelmezni, hogy az A-val jelölt cselekvés kötelező feltéve, hogy a B feltétel érvényes. A diadikus deontikus logikai operátorok segítségével természetesen ki lehet fejezni a monadikus operátorokat is. Ehhez a feltételként egy tautológiát kell megadnunk, aminek eredményeként az operátorunk feltétel nélkülivé "változik":

(DL.21) $\underset{\sim}{\mathbf{O}}_m(A) \equiv \underset{\sim}{\mathbf{O}}(A) \equiv \underset{\sim}{\mathbf{O}} A \stackrel{\text{def}}{=} \underset{\sim}{\mathbf{O}}(A/\top)$

ahol $\underset{\sim}{\mathbf{O}}_m$ a monadikus operátort, $\top$ a tautológiát jelöli. Von Wright megoldása a jó irányba tet lépésként tekinthető, azonban ő még nem épített fel egy teljes diadikus logikát erre a feltételes formulára. Bengt Hansson volt az, aki egy 1969-es tanulmányában rendszerszerűen dolgozta ki a diadikus deontikus logika (Dyadic Deontic Logic, DDL) teljes struktúráját [214], így őt tekinthetjük a DDL kidolgozójának. Hansson rendszerét pár évvel később axiomatizálta Wolfgang Spohn [507], és azóta sokan mások is hozzájárultak az új deontika építéséhez [20].

A diadikus deontikákra azért van szükség – persze sok minden más mellett –, mert a normák nem mindig vezethetők vissza egyértelműen a jó és a rossz szembeállítására. Le lehet-e lőni a terroristák által eltérített, utasokkal teli repülőgépet, ha tudni lehet, hogy még nagyobb kárt akarnak okozni vele a terroristák? Ezer ilyen kérdést produkál az életünk. A diadikus logika segítségével – mégha a gyakorlati életben megoldani nem is tudjuk az effajta problémákat – legalább elméletileg meg tudjuk ragadni őket úgy, hogy közben mentesülünk a logikai paradoxonoktól.

Mint minden más logikában, úgy a deontikus logikán belül is vannak paradoxonok. Ezek jelentős része a klasszikus deontikus logikában bukkant fel, és részben ennek a ténynek köszönhető, hogy az idők során kidolgoztak olyan nem-klasszikus deontikákat, amelyek létrehozásával pont az effajta ellentmondásokat akarták kiküszöbölni. A **Jörgensen dilemma** abban az értelemben még nem is sorolható a paradoxonok körébe, hogy nem klasszikus logkai paradoxont ír le, hanem azt a kérdést "feszegeti", hogy miként lehet egyáltalán a kijelentések világából [268]. A deontikus logikai paradoxonok közül talán

az első lehetett a felfedezőjéről elnevezett **Ross paradoxon** [446]. A klasszikus példa szerint abból a felszólításból, hogy *Küldd el a levelet vagy égesd el!*, a klasszikus deontikus logika alapján levezethető az a parancs is, hogy *Égesd el a levelet!*, ami nyilván ellenkezik az intuíciónkkal. Meglepő paradoxont produkál a **szabad választás engedélye** (free choice permission) [319]. Vegyünk a következő példát. *Fred vagy Bill jöhet.* Ez az engedély ekvivalens azzal a mondattal, hogy: *Fred jöhet, és Bill jöhet.* Ez pedig furcsa, mert egy diszjunkcióból konjunkciót kaptunk, amikor az engedélyezés deontikus hatókörét kiterjesztetük a belső diszjunkció tagjaira. Jennings kiterjesztette ezt a problémát úgy, nem csak az engedély hatályába tartozó cselekvésekről lehet szó, de a cselekvőkre vonatkozó diszjunkció is okozhat gondot [260, 263]. Más szóval Jennings szerint akkor is jelentkezhetnek hasonló problémák, ha elhagyjuk a deontikus logika világát. Bizonyságul erre további két példát is bemutat. Először az összehasonlítások világából hoz egy mondatot: *Pierre fontosabb mint William vagy Réné.*, aminek kiterjesztő átalakítása meghagyja az eredeti mondat jelentését: *Pierre fontosabb, mint William, és Pierre fontosabb, mint Réné.* Jennings másik példája minden eddiginél egyszerűbb, leírása nem igényel semmi különös logikát, de a probléma ebben az esetben is jelentkezik. Vegyük az alábbi mondatot: *Ha Laurie vagy Alison távozik, akkor távozom én is.* Ez a mondat ugyanúgy egészíthető ki, mint az eddigiek: a kiegészítés közben a konjunkciót kell használjuk. Az átalakított mondat így néz ki: *Ha Laurie távozik, akkor távozom én is, és Alison távozik, akkor távozom én is.* A példák alapján elfogadható Jennings érve, miszerint igazából nem deontikus logika problémájáról, hanem valami másról van itt szó. Kétféle megoldási lehetőséget is bemutat, hogy aztán az egyik mellett letegye a voksát. Az egyik értelmezés szerint ezekben a mondatokban azért alkalmazunk 'vagy'-ot az 'és' helyett, mert az utóbbi használata egyértelműen a konjunktív jelentést írná elő. Ha ugyanis ezt a mondatot olvassuk: *Ha Laurie és Alison távozik, akkor távozom én is.* akkor egyértelműen arra gondolunk, hogy a feltétel két tagja egyszerre fent kell, hogy álljon a következmény létrejöttéhez, vagyis akkor távozom én is, ha már mindketten távoztak. Az eredeti mondatban pedig ez nem volt benne. Mivel a természetes nyelv sem oldja meg egyértelműen ezt a problémát, ezért voltaképpen minden formalizálási kísérlet gondban van.

Az "igazi" deontikus logikai paradoxonok közé tartozik a **jó szamaritánus paradoxona** (Good Samaritan's Paradox), aminek a lényegét először A.N. Prior írta meg [422], [423]. A paradoxon arról szól, hogy ha létezik az a parancs, hogy segítenünk kell a nélkülöző vagy a megtámadott embernek, akkor abból – SDL alapon – levezethető az, hogy léteznie kell nélkülöző és megtámadott embereknek. Igazából Lennart Åqvist cikke tette ismertté a paradoxont [19], és ő volt az is, aki összekötötte a kérdést a kellellentétes parancsok problémájával. A paradoxon feloldható a diadikus denotikus logika (DDL) segítségével: [212, 66]. A **szelíd gyilkos paradoxona** (gentle murderer' paradox) James William Forrester nevéhez köthető [173], és röviden úgy összegezhető, hogy *(i)* ha van egy feltételes parancsunk, miszerint "ha gyilkosz, legalább gyengéden (értsd: minél kisebb fájdalommal) tegyed", illetve tudjuk, hogy egyfelől *(ii)* valaki gyilkosságot követ el, másfelől *(iii)* ha valaki gyengéden gyilkolt, akkor gyilkolt is egyben, akkor mindezekből levezethető az, hogy *(iv)* annak a valakinek gyilkolnia kellett. Ez a paradoxon is feloldható a DDL segítségével [212, 66]. Bemutatás nélkül hivatkozom még az E.J. Lemmon által megfogalmazott **Platon dilemmájára** (Plato's Dilemma) [301], az A.N. Priorhoz köthető **elköteleződés paradoxonára** (paradox of commitment) [422], valamint a Sven Ove Hansson által explikált **tudó** (knower) paradoxonára [219], [543], illetve **bosszúálló paradoxonára** (re-

venger's paradox) [217].

A deontikus logika fejlődéséhez nagyban hozzájárult az a tény, hogy az 1970-es években a téma iránt érdeklődő mérnökökből, filozófusokból, logikusokból kialakult egy közösség. A közösséget magát, valamint a kétévente megrendezett konferenciáikat szimbolikusan jelző ΔEON kifejezésre utalva Δ-közösségnek nevezem őket. Gondolkozásuk fejlődéséről, az érdeklődési fókusz eltolódásairól, a legfontosabbnak tartott kérdések változásairól, a tíz legfontosabb probléma bemutatásáról több tanulmányt is tudok ajánlani az érdeklődők számára[212], [63], [211]. Az egyik cikk az egymást követő évtizedekben az alábbi címszavak álltak a deontikus logikával foglalko szakemberek fókuszában.

szakasz	rendszer	fő érdeklődési kör
50s	monadikus modális logika	**O̲** és **P̲** viszonya
60s	diadikus modális logika	kapcsolat **O̲** és a tények, normasértések között, szubidealitás és optimalitás, CTD
70s	temporális deontikus logika	**O̲** és az idő viszonya
80s	cselekvési deontikus logika	**O̲** és a cselekvés viszonya
90s	támadható deontikus logika	dilemmák, CTD
00s	felszólítások, normatív rendszerek	Jörgensen dilemma

*

A deontikus logikával kapcsolatban indulásától fogva sokakban megfogalmazódott az a remény, hogy ez a diszciplina segítheti majd a normatív, jogi szövegek gépi elemzésében, az ilyen szövegeken belüli inkonzisztenciák megtalálásában. Ezt az ígéretét azonban az évtizedeken keresztül tartó folyamatos fejlődése ellenére sem tudta beteljesíteni ez a tudományterület. Nyilván nem szabad, nem lehet egyetlen tényezővel megmagyarázni mindezt, de egy magyarázó lehetőséget azért kiemelnék a sok közül. A deontikus logika eddig bemutatott elméletének komoly hiányossága a személytelensége, imperszonális jellege. Kötelezésekről, tiltásokról, megengedésekről van szó, de úgy, hogy közben nem tudjuk, hogy ezek kire, kikre irányulnak. Egész pontosan nem is igazán az a baj, hogy nem tudjuk, kikre vonatkoznak az előírások, vagy kik adják ki ezeket, bár azért ezek is fontos kérdések, hanem sokkal inkább a megközelítés mondanikus jellege a probléma. A személytelenséget az imperszonális jelző alkalmazásával ragadhatjuk meg. Az **imperszonális deontikában** "nincsenek" személyek, így nem lehet kapcsolatot teremteni a deontikus operátorok és az ágensek között, emiatt nem lehet az ágencia egész problémakörét sem kezelni. De még nagyobb baj, hiány az, hogy ez a deontika nem képes megragadni azt a tényt, hogy a társadalmi cselekvésekben megnyilvánuló normativitás lényegét tekintve relációs jellegű, aminek kezelésére relációs fogalmakra, és ezért **perszonális deontikára** lenne szükség. A társadalmi cselekvés formális modelljének felépítésében kulcsszerepet szánok a perszonális deontika elméletének, amit a későbbi fejezetekben mutatok majd be.

*

A deontikus logika témakörök bemutatása után két érdekességről tennék említést. Az egyik a túlteljesítés, a **szupererogáció** kérdése. Urmson egy 1958-as cikkében vetette fel azt a problémát, hogy a deontikus logikában bevezetett modalitásokkal leírhatjuk ugyan

a kötelességhez, a megengedéshez és a tiltáshoz igazodó cselekvéseket, ám ezzel a fogalomhármassal nem lehet minden morális aktust megragadni. A *szent* és a *hős* cselekvése kilóg ebből a klasszifikációból [551](eng)-[552](hun). Bizonyos emberek akkor is teljesítik kötelességeiket, amikor a többiek nem. A többség gyengeségből vagy gyávaságból nem teszi meg, ami kötelező lenne számára, csak kevesek tudnak erősek maradni. A szent ellenáll a kísértésnek, a hős legyőzi a félelmét, de mindkettő megcsinálja, amit morális kötelességének tart (saját maga és a környezete). A különbség annyi csak köztük, hogy miben győzik le magukat, minek állnak ellen: a szent a vágyainak, a hős a félelmeinek. Urmson tanulmányának megjelenése után Joel Feinberg is hozzászólt a kérdéshez[159], az első – formális szempontokat is figyelembe vevő –elméletformálási próbálkozások Roderick Chisholm nevéhez fűződnek [95], [96], [97]. A szupererogáció fogalmára hivatkozva Paul McNamara a deontikus hatszög bővítésére, a supererogation modellbe emelésére tett javaslatot [336], [335], mint ahogy ezt tette Jan C. Joerden is, aki szerint a deontikus hatszög helyett inkább deontikus tízszöget (Deontological Decagon) kellene használnunk [263]. A szupererogáció fogalmát már cask azért is fontosnak tartom, mert elképzelhetőnek tartom, hogy segítével tisztázni lehet a jézusi etika legfontosabb tézisének, az aktív szeretet parancsának lényegét.

A deontikus logikában nem nagyon esik róla szó, ezért önmagában is érdekes lehet az iszlám kultúra szankcionálási gyakorlatának bemutatása, de különös jelentőséget adhat, egyfajta ellenpontot jelenthet az a tény, hogy az elméletileg létező, de az európai kultúrkör szankcionálási gyakorlatából hiányzó elemek itt megjelennek.

modalitás	szankció elvégzésekor	szankció mellőzésekor
megparancsolt	nincs semmi	büntetés
ajánlott	jutalom	nincs büntetés
tiltott	büntetés	nincs semmi
helytelenített	nincs büntetés	jutalom
közömbös	nincs semmi	nincs semmi

Chapter 2

Társadalmi cselekvés

„There is only one social science.” Jack Hirshleifer [237]

A népi bölcsesség szerint a nevelés során két mondatot ismételgetünk folyamatosan: *Kezdd már el!* és *Hagyd már abba!*

„Aki felismeri a tettben a tétlenséget és a tétlenségben a tettet, az olyan ember a bölcs, az olyan az önmagát fegyelmező, bármely tettet végez.” (Bhagavad Gítá, IV.18.)

Az társadalmi cselekvés magyarázatában központi szerepet kell tulajdonítanunk az intencionalitás kategóriájának, így ez utóbbi tárgykörébe tartozó fogalmak formalizásához szükségünk nyelvet az intencionalitás nyelvének fogom nevezni. Mind a cselekvések, mind az intencionalitás alá sorolható további kategóriák, mint a vélekedés, vágy, szándék, választás formális leírásához ezt a nyelvet használom.

(IL.0) $\mathscr{L}_{JAL}(\mathscr{L}_{PL}, \mathbb{S}, \mathbb{A}, \Phi, \mathbb{C})$, ahol

$\mathscr{L}_{ML}$	modális logika
$\mathbb{S}$	világállapotok halmaza
$\mathbb{A}$	ágensek halmaza
Φ	vélemények halmaza
$\mathbb{C}$	választások halmaza

A fejezetet a cselekvés fogalmának értelmezésével kezdem, majd az intencionalitás jelenségén belül bemutatom hogyan lehet jellemezni a késztetéseink, vágyaink, szándékaink, választásaink rendszerét.

2.1 Cselekvés

A cselekvéselmélet központi kategóriája a társadalmi cselekvés. Ezt a fogalmat Max Weber is a kategóriatana középpontjába helyezte, és Weber meghatározása azóta is igazodási

pontot jelent sok társadalomtudós számára. Én is ebből az értelmezésből fogok kiindulni. A weberi cselekvésfogalmát felidézve azonban látni fogjuk, hogy ez a meghatározás nem is olyan egyszerű, és komolyabb explikációs munkát kell belefektetni, hogy pontos definíciót adhassunk. A társadalmi cselekvés fogalmának weberi értelmezését akkor érthetjük meg pontosan, ha követjük azt az logikát, ahogyan Weber összekapcsolja egymással a cselekvéselmélet négy alapfogalmát, a **viselkedés**, a **cselekvés**, a **társadalmi cselekvés** és a **társadalmi kapcsolat** kategóriáját. A viselkedés fogalmát explicit módon ugyan nem definiálja Weber, de a cselekvés fogalmának meghatározásából kibontható, mit gondol róla. A **cselekvés** fogalmát Weber az alábbiak szerint határozta meg.

> „... »cselekvésnek« nevezünk minden emberi viselkedést (mindegy, hogy külső vagy belső ténykedésről, valaminek az elmulasztásáról vagy eltűréséről van-e szó), ha és amennyiben a cselekvő, illetve a cselekvők valamilyen szubjektív értelmet kapcsolnak vele össze." [575, 38]

Kezdjük a definíció egyik könnyebben értelmezhető elemével: a cselekvéshez kellenek **cselekvők**, cselekvésre képes ágensek. Hogy mit tarthatunk ágensnek, azon belül pedig mit tarthatunk emberi cselekvőnek, azt itt nem tárgyalom. Elegendőnek tartom azt állítani, hogy a cselekvők azok az ágensek, akik képesek cselekvésre, vagyis a cselekvés minőségére vezetem vissza a cselekvő fogalmát. Lépjünk tovább a weberi definíció másik következő elemére. A definícióban benne van egyfelől a viselkedés fogalmának implicit meghatározása, másfelől a viselkedéshez képest a cselekvés fogalmában benne levő többlet explicit megadása. Viselkedés lehet valamilyen 'ténykedés' (tevés), 'mulasztás' vagy 'tűrés' (tartózkodás), illetve mindezek lehetnek 'belső' vagy 'külső' viselkedések. Viselkedés lehet, ha valaki énekel, eszik, sétál, olvas, becsukja az ajtót, gondolkozik, segít valakin, adót fizet vagy nem fizet adót. A viselkedés (magatartás) a világ lehetséges állapotai között lehetséges transzformációkkal hozható kapcsolatba. Ha az asztalon van egy virsli és én megeszem, akkor a viselkedésem eredményeként a következő világállapotban már nem lesz a virsli az asztalon (a gyomromban lesz). A viselkedésnek ezt az értelmezését persze Weber nem bontja így ki, de megteszik ezt helyette mások. Hogy hogyan, azt kicsit később fogom bemutatni.

A viselkedés és a világállapot-transzformációk összekapcsolásával csak annyi a "baj", hogy ezt a meghatározást kiterjeszthetjük minden élőlényre vonatkozóan. Nem véletlen ezért az, hogy Weber még hozzátesz valamit a cselekvés fogalmának definíciójához. Szerinte ugyanis a cselekvésről csak akkor beszélhetünk, ha a "cselekvő szubjektív értelmet kapcsol" a viselkedéséhez. A cselekvés kategóriája ezáltal egyfelől a viselkedés fogalmának generikus alárendeltje lesz, másfelől a "szubjektív értelem" minőségével egy olyan értelmezési réteget hoz be Weber a cselekvéselméletbe, amely segítségével el lehet különíteni az emberi cselekvéseket és állati (olykor emberi) viselkedéseket egymástól. A weberi szubjektívértelem-adás a viselkedéshez kapcsolható **intencionalitás** fogalmával ragadható meg igazán, aminek jelentését Weber nem igazán adja meg, csak néhány példát mutat be ezzel kapcsolatban. Az intencionalitás kategóriáját, valamint a belőle kibontható további fogalmakat – jelentőségük miatt – külön fejezetben tárgyalom.

A cselekvés fogalmához elegendő egyetlen ágens feltételezése, amennyiben egy cselekvő szubjektív értelmet kapcsol a viselkedéséhez, akkor már cselekvésről beszélhetünk. Ekkor azonban a cselekvésnek még nem tulajdoníthatunk társadalmi minőséget.[1] A tár-

[1] Azt a kérdést, vajon az intencionalitással "nem kerül-e be" valamilyen társadalmi mozzanat a cselekvés fogalmába, itt nem elemzem.

sas, társadalmi jelleg létezéséhez, létrejöttéhez szükség van legalább egy másik ember "jelenlétére". Ezt a mozzanatot Weber – a **társadalmi cselekvés** meghatározásákor – a következő módon ragadja meg.

> „ . . . »társadalmi« cselekvésnek . . . az olyan cselekvést nevezzük, amely a cselekvő vagy a cselekvők által szándékolt értelme szerint mások viselkedésére van vonatkoztatva, és menetében is mások viselkedéséhez igazodik." [575, 38]

Sajnos a társadalmi cselekvés fogalmának weberi definíciójában a 'szándékolt értelem' kifejezés szerepel, ami nem teljesen egyezik meg a cselekvés fogalmában szereplő 'szubjektív értelem' minőséggel. Ez az eltérés kicsit zavaró, mert az előbbi részben mást, részben kevesebbet az utóbbihoz képest. Ez a különbség lényegében megfeleltethető az intencionáltság és az intenció közti fogalmi külnbségnek, amit jobban kibontok az intencionalitásról szóló fejezetben. Ha a jelzett fogalmi pongyolaságtól eltekintünk, akkor a fenti definíció lényegi eleme az, hogy a társadalmi cselekvés fogalmába a "másikhoz való igazodás" minőségével már "belekerül" a társadalmiság mozzanata – még ha gyengébb formában is. Ekkor még mindig elképzelhetünk egyetlen cselekvőt, aki ugyan már társadalmi kötöttségben cselekszik, ha cselekvését valamilyen módon egy másik emberhez igazítja, de ezt még megteheti úgy is, hogy ez az igazodás csak fejben történik meg, valós kapcsolat nem jön létre a cselekvők között. A társadalmiságot igazán jól a társadalmi kapcsolat fogalmával tudja Weber megragadni, amikor már arról van szó, hogy a társadalmi cselekvők kölcsönösen egymáshoz igazodnak a cselekvéseik során, ezt a fogalmat azonban csak később tárgyalom (↬233.o.).

Ahhoz, hogy jobban megértsük, mi és miért kell társadalmi cselekvésnek tartanunk, jobban ki kell bontanunk az eddig jelzett fogalmi összetevők jelentését. Mielőtt ebbe belevágnék, érdemesnek tartok bemutatni pár példamondatot a cselekvésre, viselkedésre vonatkozóan, hogy a jelenség mélyén rejlő kérdéseket és problémákat jobban érzékelhessük. Nézzük meg tehát, miről is szólnak a következő mondatok.

i.)	Kiszáradt a fű, az eső elmosta a nyomokat.
ii.)	Kati elpirult, amikor Laci ránézett.
iii.)	Nagyot nőtt ez a Peti gyerek.
iv.)	A betörő ott hagyta az ujjlenyomatát az ajtókilincsen.
v.)	Profi volt az elkövető, nem hagyott nyomokat.
vi.)	A sofőr látta a balesetet, de nem állt meg segíteni.
vii.)	Dávid beteg volt, egész nap köhögött.
viii.)	Rita köhögött egyet, ezzel jelezte, hogy ő is ott van a szobában.
ix.)	Jani az izmait megfeszítve pózolt a tükör előtt, amikor görcsbe állt a vádlija.
x.)	Feri egész nap a kertben ásott.
xi.)	Pali bácsi megöregedett.

Vannak olyan mondatok, amelyek valamilyen történést írnak le (i, ii, iii, vii). Ezeket nem érezzük cselekvésnek, mert vagy nem emberi ágensekhez köthetjük őket, vagy ugyan emberekkel vannak kapcsolatban, de ezek az emberek ugyanúgy elszenvedői valaminek (például a köhögésnek vagy az elpirulásnak), mint ahogy a fű kiszáradásáról sem tehet semmit a fű maga. A példákból nyilvánvalóan kitűnik, hogy nem minden emberi viselkedést minősítünk cselekvésnek. Amikor valaki csuklik, köhög, (önkéntelenül) sóhajt, pislog, lélegzik, görcsbe rándul, elpirul, izzad, akkor valamilyen viselkedést ugyan tulajdoníthatunk neki, de ezekből hiányzik valamilyen értelemadás, valamilyen intencionáltság.

Ezek a dolgok megtörténnek velünk, nem mi cselekedjük őket. Az természetesen előfordulhat, hogy tudatosan tesszük meg valamelyiket a felsoroltak közül, például tudatosan köhögünk (viii) vagy megfeszítjük az izmunkat (ix), de ekkor már valóban cselekvésről (és intencionáltságról) beszélhetünk, amikor valamilyen szubjektív értelmet rendelünk a viselkedésünkhöz. Ilyenkor már úgy érezzük, valóban mi cselekszünk, nem pedig velünk történik valami.

A példamondatokban észrevehetünk további különbségeket is. Azt, hogy valaki ás a kertben (x), vitán felül cselekvésnek gondoljuk, de ha valaki nem áll meg segíteni egy közúti baleset áldozatainak (vi), azt már talán nem annyira nyilvánvalóan tartjuk cselekvésnek arra hivatkozva, hogy a sofőr még sokszor nem állt meg vezetés közben, és nem könnyű megmondani, hogy milyen alapon lehetne elkülöníteni ezeket attól a helyzettől, amikor a baleset helyszíne mellett haladt el segítségnyújtás nélkül. E példa azt szemlélteti, hogy a cselekvéseink között lehetnek olyanok, amelyeket tevőlegesen hajtunk végre, amikor megcsinálunk valamit, de cselekvésnek minősíthetjük a tartózkodásainkat is, amiket egy adott értelemben nem-cselekvésnek kell minősítenünk. A nagy kérdés az, hogy milyen módon tudunk különbséget tenni a tartózkodás és "sima" nem-cselekvések között. A külünbséget ebben az esetben is az intencionalitás fogalmával ragadhatjuk majd meg. További tipizálási szemponthoz adhat segítséget a bűncselekményekkel kapcsolatos két példamondat (iv, v). Az első mondat – a szokásos értelmezés szerint – nem is cselekvést, hanem inkább egy történést ír le, ha a mondatot úgy értjük, hogy a betörő nem is tudott róla, hogy az ujjlenyomata egyáltalán ott maradhat valahol. Ezzel szemben a másik mondat szereplője már tudta, hogy az ujjlenyomata lebuktathatja őt, ezért tudatosan arra törekedett, hogy ne hagyjon ilyen nyomokat. Ekkor a cselekvésének értelme az, hogy a cselekvő nem engedi, hogy megtörténjen valami a világban (ott maradjon az ujjlenyomata). Ez is cselekvés, mondhatjuk, hogy ez sem valamiféle tevőleges cselekvés csakúgy, mint a segítségnyújtás elmulasztása, de érezhetően különbözik egymástól a kétféle "tartózkodás".

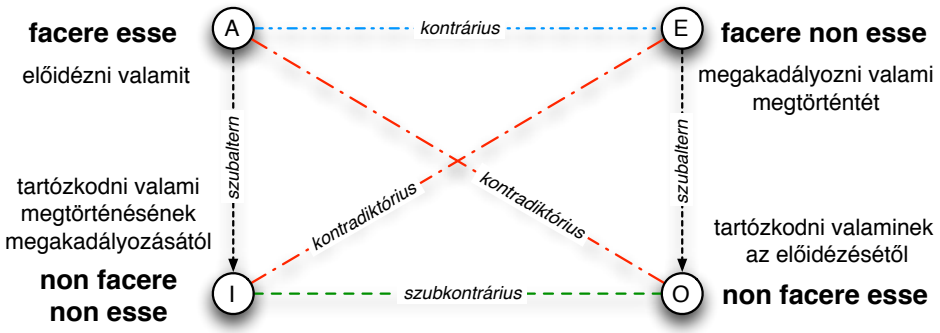
A fenti példák értelmezéséhez szükségünk van arra, hogy jól tipizálni tudjuk a cselekvéseinket. Kezdésként kiindulhatunk egy igencsak régi, ám nagyon is lényeglátó megoldásból, Szent Anzelm cselekvéstipológiájából.[2] Anzelm – a már akkor is jólismert logikai négszög mintájára – veti fel az ötletet, hogy a cselekvésnek négy típusát lehet elkülöníteni egymástól, és ezekre vonatkozóan fel lehet rajzolni egy logikai négyzetet. Szent Anzelm elméletének központjában a latin 'facere' ige áll, amit angolra a 'bring about ', magyarra a 'valamivé tesz, előidéz, létrehoz, teremt, csinál' értelemben fordíthatunk át. Ezzel a fogalommal ragadhatjuk meg a legszokásosabb értelemben vett cselekvés fogalmát, azt írhatjuk le vele, amikor valaki megcsinál, vagyis a viselkedésével "szándékosan" előidéz valamit. Erre használhatjuk a 'facere esse' kifejezést. Anzelm ezzel állította szembe a következő cselekvéstípusát, a 'tartózkodást valami megtételétől', amire latinul a 'non facere esse' kifejezést húzhatjuk. Az eddig említett két cselekvéstípusban közös az a vonás, hogy a cselekvés keretéül, feltételéül szolgáló világállapotokban nincsen valamilyen változást eredményező történés. Előfordulhatnak azonban olyan helyzetek is, amikor megtörténne valami, hacsak egy cselekvéssel valaki ezt meg nem akadályozná. Az ilyen helyzetekben megvalósuló cselekvésre húzta rá Anzelm a következő fogalmat, a 'facere non esse' kifejezést. Erre a fogalomra szolgálhat például a korábbi

[2] Anzelm cselekvésről szóló gondolatait alaposan bemutatta Douglas N. Walton [572], [570], [571], illetve Howard L. Dazeley és Wolfgang L. Gombocz [126].

bemutatott esetek közül az, amikor a betörő vigyázott arra, hogy ne hagyja ott az ujjlenyomatát. Végül Anzelm negyedik kategóriaként elkülönítette azt a cselekvéstípust, amelynek "végrehajtása" során a cselekvő tartózkodik attól, hogy egy történést megakadályozzon. Erre hozhatnánk a korábbi példamondatok közül azt az esetet, amikor valaki ott hagyja az ujjlenyomatát a kilincsen (iv), de ezt valamiért tudatosan teszi (mondjuk azért, mert "le akar bukni"). Összegezve az eddigieket Anzelm négy cselekvéstípusát az alábbi táblázatba rendezhetjük.

facere esse	to bring it about	előidézni valamit
non facere esse	to abstain from bringing it about	tartózkodni valaminek az előidézésétől
facere non esse	to prevent from having it come about	megakadályozni valami megtörténtét
non facere non esse	to abstain from preventing from having it come about	tartózkodni valami megtörténésének megakadályozásától

A táblázatban feltüntettem az anzelmi típusok angol fordítását is, mert a cselekvéselméleti szakirodalomban ezeket a kifejezéseket gyakran használják, és a későbbiekben is előkerülnek még. Ahogy említettem, Anzelm nemcsak a négy fenti cselekvéstípust különítette el egymástól, de érvényesnek tartotta a logikai négyszög alkalmazását is a négy elemre. A cselekvéstípusok logikai négyszöge abban az értelemben érvényes, hogy az egyes típusok között fennállnak a logikai négszög esetében elvárt relációk, de olyanfajta – kvantifikáció alapú – értelmezés, mint amilyen a logikai négyszög eredeti formájában rendelkezésre állt, itt nincs. És természetesen Anzelm nem ismerte a logikai hatszög fogalmát sem, így nem kereste azokat a cselekvésfogalmakat sem, amelyek a hatszög további két elemét alkothatták volna (ezt egyébként – tudomásom szerint – azóta sem kereste senki más).



Bár Szent Anzelm pontos és használható tipológiát állított fel, típustana gyakorlatilag feledésbe merült, vagy legalábbis a cselekvéselméleti diskurzus főáramába nem került be. Az anzelmi cselekvéstípusok azonban – persze más terminusokkal megjelölve – felbukkantak a későbbi elméletekben. A cselekvés fogalmának meghatározását keresve sokféle szempont szerint közelítettek a vizsgált jelenséghez. Egyes elméletek azt keresték, hogy lehet detektálni a cselekvéseinkben (vagy az azokat leíró mondatainkban) rejlő intencionáltságot, mások a cselekvés fogalmának logikai összetevőkre bontását végezték el, megint mások a tartózkodás fogalmának megragadására törekedtek.

Donald Davidson a cselekvést leíró mondatok specialitását keresve sokféle értelemben (irányban) hasznosítható javaslatot fogalmazott meg, de azzal nem igazán foglalkozott, mit is jelent igazából a cselekvés fogalma [118]. Tanulmánya végén annyit azért kimondott, hogy a cselekvésről akkor beszélhetünk, ha a viselkedésben intencionalitás van.

> „ . . . csak akkor beszélhetünk ágenciáról, ha értelmes az a kérdésfeltevés, vajon az ágens intencionálisan cselekedett-e. Bár, úgy tűnik számomra, vannak olyan esetek is, amikor csak akkor tulajdonítunk ágenciát, ha a kérdésre, vajon az ágens intencionálisa cselekedett-e, a válasz 'igen'. Ha egy ember elesik véletlenül vagy azért, mert egy teherautó elsodorta őt, nem beszélünk ágenciáról; de akkor igen, ha szándékosan esett el." [118, 94]

Egy másik tanulmányában ezt definíciószerűen ki is mondta, amikor azt állította, hogy „ . . . egy esemény akkor cselekvés, ha intencionálisként írható le." [122, 229] Azt azonban nem tisztázta Davidson, mit jelent az intencionalitás, csak a nyelvi intuíciónkra bízta annak eldöntését, mikor és miért érezzük azt, hogy egy mondattal intencionalitást írunk le. Erre a kérdésre az ún. **STIT-elmélet** sem ad(ott) választ, ennek ellenére az elméletet által javasolt tesztet nagyon széles körben elfogadták-elfogadják annak eldöntésére, hogy milyen mondatokra mondhatjuk azt, hogy azok ágentívek, vagyis valamilyen cselekvést írunk le velük. A STIT-elméletet Nuel Belnap és Michael Perloff alapozta meg (az első cikkük 1988-ban jelent meg [41], amit két évvel később követett egy javított változat [42]), Ming Xu volt az, aki axiomatizálta Belnapék elméletét, így hárman jelentettek meg egy közös könyvet [43]. Belnapék több lehetőséget is felsorolnak arra, hogy milyen teszt segítségével lehetne eldönteni a mondatok ágentivitását.

i.)	brings it about that	vi.)	allows it to be the case that
ii.)	makes it the case that	vii.)	takes steps in order that
iii.)	causes it to be the case that	viii.)	behaves so that in consequence
iv.)	is responsible for the fact that	ix.)	sees to it that
v.)	lets it be the case that		

Meggyőző érvelés végén Belnapék a táblázat utolsó sorában szereplő *sees to it that* operátort javasolják az ágentív mondatok kifejezésére (illetve detektálására) alkalmazni.[3] Ezt úgy lehet megtenni, hogy az operátorhoz mindig hozzáillesztünk egy-egy összetevőt, ami az ágensre, illetve a cselekvés tartalmára utal (utóbbit nevezik Belnapék leíró komponensnek). Egy ágentív mondatnak így kell kinéznie: $[\alpha \text{ STIT: } Q]$ A kifejezésben a STIT-operátor bal oldalán szereplő α helyére a cselekvő ágenst kell behelyettesíteni, míg a jobb oldali Q tag helyére pedig a cselekvés tartalmát leíró propozíciót. Belnapék tézise szerint amennyiben a Q propozíció, illetve az egész STIT-kifejezés mint propozíció ekvivalens egymással, akkor ágentív mondatról (cselekvésről) beszélhetünk.

(ZL.1) $Q \leftrightarrow [\alpha \text{ STIT: } Q]$

Belnap-Perloff ágentivitás tézise (HUN) – Belnap-Perloff's agentivity theorem (ENG)

Mielőtt példát mutatnék a STIT-operátoros teszt alkalmazására, meg kell még mutatnom azt is, hogy milyen magyar fordítást használhatnánk a STIT-operátorban levő 'sees to it

[3] A cselekvés fogalmában rejlő ágentivitás megragadására törekvő elméletek megismeréséhez két – történeti szempontot követő – cikket ajánlok: Michael Perlofftól [409] és Krister Segerbergtől [472].

that' angol kifejezés helyett. Nem könnyű feladat a fordítás itt. A legtöbbször alkalmazható megoldás a 'gondoskodik róla, hogy' fordulat lehetne, de bizonyos kontextusokban használható lenne a 'törekszik arra, hogy', 'tesz róla, hogy', a 'gondja van rá, hogy' vagy az 'ügyel arra, hogy' kifejezés is. Amennyiben nem okoz zavart, az egyszerűség kedvéért elhagyhatjuk az ágensre utaló paramétert a formulából, és csak a cselekvés tartalmát jelző paramétert hagyhatjuk meg benne. Ilyen esetben a Q paramétert tehetjük zárójelek közé: STIT:(Q).

A STIT-operátort tehát úgy használhatjuk ágencia-tesztelésre, hogy vesszük az eldöntendő mondatot, azt belehelyezzük a STIT-kifejezésbe, majd megvizsgáljuk, hogy fennálle az ekvivalencia a két propozíció között. Nézzünk meg egy példát!

példa, példa, példa, példa, példa, példa, példa, példa, példa, példa

A STIT-operátort megbízható tesztként használhatjuk az ágentivitás detektálására, de arra nem alkalmas, hogy fogalmilag pontosan leírhassuk vele az intencionalitás jelenségét. Nem segít sem abban, hogy tipizáljuk, sem abban, hogy logikai össszetevőkre bontsuk vele az intencionalitás fogalmát. Ez a "hiány" természetesen nem változtat a STIT-operátor hasznosságán a tesztelések során, de azt jelzi, hogy szükségünk lesz – nyilván más eszközökre támaszkodva – mondani valamit az intencionalitás fogalmáról. Ezt később, egy önálló fejezetben fogom megtenni (↪140.o.). Itt most a STIT-operátor egy további hasznos vonását emelem ki. Az operátor segítségével ugyanis könnyen és szemléletes módon kifejezhetjük a tartózkodást megvalósító cselekvéseinket. Mielőtt bemutatnám, hogyan, szükségesnek tartok egy rövidebb kitérőt tenni a tartózkodás fogalmával kapcsolatban.

A tartózkodás mint tétlen tett

Utaltam már rá, hogy a hétköznapi, reflektálatlan gondolkodás könnyen egybe mossa, azonosítja a cselekvés és a tevékenység, tevés kategóriáit, holott nyilvánvaló, hogy a cselekvés általános fogalma alá két másik alárendelt fogalmat kell besorolnunk: a **tevés**, megcsinálás, illetve a **tartózkodás**, tétlenség kategóriáját. Szokás ezeket **pozitív cselekvésnek**, illetve **negatív cselekvésnek** is nevezni. Ha valaki jelentéktelen, érdektelen kérdésnek tartaná a tartózkodás jelenségét, akkor annak érdemes belegondolni abba, hogy a sztrájk jelensége a nem-dolgozást (tehát a dolgozástól való tudatos tartózkodást) jelenti, amely igen sok, igen mély konfliktus forrása volt eddig a történelemben. Egy – Csuangce, taoista gondolkodótól származó – gyönyörű mondás szerint:

> „ ... az ember nem képes a sáros, zavaros vizet megtisztítani, de ha békén hagyja, az magától letisztul." [231, 27]

A taoizmus fontosnak tartja, hogy az emberi viselkedés a Wu-Wei elvhez igazodjon, aminek jelentése: cselekvés nem-cselekvés által. Ezzel voltaképp a tartózkodást írják le. De a történelemben sok helyen felbukkanó passzív ellenállás (Gandhi **szatjágraha** mozgalma, a erőszakmentes polgári engedetlenség) jelensége is azt példázza, hogy a tartózkodás értelmes, jelentéssel bíró fogalom lehet.

A tartózkodást természetes nyelvi kontextusban sokféle alakban kifejezhetjük. Először is bármikor használhatjuk a tevőleges igék tagadó formáját: 'nem csinál meg valamit', 'nem segített neki'. De a negatív cselekvést kifejezhetjük "önálló" módon is. Mondhatjuk, hogy 'tétlenül szemlél', 'tétlenkedik', 'elmulaszt', 'mellőz', 'hagy valamit megtörténni', 'visszafogja magát'. Ha valaki 'enged másnak valamit megcsinálni' vagy 'tűri, hogy más megcsináljon valamit', akkor is tartózkodásról beszélhetünk, bár ekkor már má-

sodrendű cselekvésről beszélhetünk. Olykor a ‘kiváltság’, a ‘privilégium’, a ‘mentesség’, az ‘immunitás’ is jelenthet valamilyen kötelezettség alóli mentességet, ami ebben az értelemben a tartózkodáshoz is kapcsolható.[4] Viszont ha valaki valamit elfelejt, és ezért nem csinálja meg, akkor azt nem sorolhatjuk a tartózkodás kategóriája alá, mert ebből hiányzik az intencionáltság. A negatív cselekvés vagy a tartózkodás mint “nem-cselekvés” az intenció fogalmával határozható meg a “nem cselekvéshez” vagyis a cselekvés tagadásához képest.

A tartózkodás jelenségét azért nehéz megragadni, mert egyszerre tartjuk cselekvésnek, illetve nem-cselekvésnek. Cselekvésként kezeljük, amikor a cselekvés egyik altípusaként értelmezzük, miközben a tartózkodás – egy másik értelemben – nem-cselekvést jelent. Utóbbi fogalom azonban sok mindent jelenthet, és el kell tudnunk választani egymástól a nem-cselekvés kétféle értelmezését. Amikor állok egy ajtó előtt és nem lépek ki rajta, mert nem akarom ezt valamiért, akkor ez az “nem-kilépés” nem-cselekvésnek és tartózkodásnak számít. De ugyanebben a pillanatban még milliónyi más dolgot nem csinálok: nem énekelek, nem beszélek, nem írok könyvet, nem eszem stb. Ezekre a nem-cselekvésekre nem mondhatjuk, hogy tartózkodom ezek végrehajtásától, mert ha így tennénk, akkor végtelen számú cselekvést lehetne (kéne) még felsorolnunk. Amikor tatózkodunk egy cselekvés végrehajtásától, akkor abban van valami többlet az “egyszerű” nem-cselekvéshez képest. Valahogy meg kell különböztetnünk a tartózkodást az ilyen “nem-cselekvésektől”. Ezt a többletet úgy foghatjuk meg, hogy azt mondjuk, a tartózkodás intencionális, szándékolt nem cselekvés. Ezt a STIT-elmélet segítségével könnyedén kifejezhetjük. Nem kell mást tennünk, mint hogy az intencionális cselekvés tagadását is intencionálisnak kell minősítenünk, vagyis kétszer kell a STIT-operátort alkalmaznunk (az egyszerűség kedvéért a formulából most elhagyhatjuk az ágensre utaló paramétert).

(ZL.2) $\text{REFRAIN}(Q) \stackrel{\text{def}}{=} \text{STIT}(\neg\text{STIT}(Q))$

tartózkodás (HUN) – refrain (ENG) – forbearance (ENG)

A képlet valamely cselekvés elvégzésére irányuló törekvés hiányára való törekvést vagyis a cselekvés nem elvégzésére való törekvést fejezi ki, de egyben mutatja azt is, hogy a tartózkodás – szemben a tevés kategóriájával – másodrendű fogalom. A tartózkodás fogalma arra is alkalmas, hogy segítségével könnyebben el tudjuk különíteni egymástól az intencionalitás cselekvéseket és a nem intencionális viselkedéseket, történéseket. Ha valamilyen viselkedés intencionális, akkor az attól való tartózkodás is az kell legyen. Ha tehát valamely viselkedéstől tartózkodni tudok, akkor az jele az intencionalitásnak, illetve fordítva: ha valamilyen viselkedéstől nem tudok tartózkodni, akkor az nem mondható intencionálisnak. Ha csuklom, ha görcsölök, ha pislogok, ha pirulok, akkor ezeket nem tudom saját akaratomra abbahagyni, tehát nem tudok tartózkodni tőlük. Ha viszont ások a kertben, iszom egy korsó sört, olvasok egy könyvet, akkor ezeket a cselekvéseket meg tudom szakítani, vagyis ilyen esetekben képes vagyok a tartózkodásra.

Mind a tevőleges cselekvés, mind a tétlen cselekvés (tartózkodás) jellemzése során használtam a viselkedés fogalmát, de még nem bontottam ki, mit is kell érteni alatta. A következőkben Georg Henrik von Wright könyve alapján azt mutatom be, hogy miként lehet a cselekvés fogalmát logikai összetevőkre bontva pontosabban leírni [564].

[4] Angolul is elég sok nyelvi olyan alak van, amely a tartózkodást fejezi ki, például ‘not-doing’, ‘omission’, ‘forbearance’, ‘refraining’, ‘inaction’, ‘abstention’, ‘abstaining’, ‘avoidance’, ‘nonaction’, ‘letting happen’, ‘neglecting’. És a sort még folytathatnánk.

A cselekvés fogalmi összetevői

Mielőtt a fogalmi összetevők kibontásába kezdenék, szükségesnek tartok néhány jelölést bevezetni. Egy formális cselekvéselméleti modellben nyilvánvalóan kulcsszerepe van a cselekvés fogalmának. Az alapkategóriák jelölésére többféle konvenció is létezik, ezek közül nekem is választanom kell, melyiket akarom használni. A választásom von Wright terminológiájára esett, amit a 'Norm and action' című könyvében fejtett ki először [564].

Az eddigiekből kiderült, hogy a **cselekvés** lehet **tevés**, illetve **tartózkodás**. Mivel ezeket a későbbiekben gyakran kell majd meghivatkozni, és konzekvens terminológia használatára van szükség, ezért a cselekvés általános fogalmát mostantól kezdve a $\pm$DO operátorral jelölöm. Megengedhetőnek tartom, hogy olykor – von Wright megoldását követve – az általában vett cselekvésre az ACT terminussal hivatkozzak. Amennyiben nem zavaró, vagyis nem okoz többértelműséget, akkor az egyszerűbb DO terminust is használhatónak gondolom. Ha a kontextus megkívánja, akkor a DO (esetleg a +DO) kifejezés a pozitív cselekvést vagyis a tevést, míg a $-$DO terminus a negatív cselekvést vagyis a tartózkodást jelöli.

A von wrigth-i DO pozitív cselekvésoperátor tartalmilag ugyanazt jelenti és ugyanazt a célt is szolgálja, mint a a Belnapék álta bevezetett STIT-operátor. A tartózkodás jelenségére ajánlott STIT-megoldás ismeretében viszont arra nagyon vigyázni kell, hogy a negatív cselekvésre vonatkozó $-$DO operátort nehogy úgy értelmezzük, mintha az csupán a pozitív cselekvésoperátor tagadása lenne. Láttuk, hogy a STIT-elmélet szerint a tartózkodás másodrendű intencionalitást jelent, ami szintaktikailag azt jelenti, hogy egymásba ágyazott cselekvésoperátorokkal lehet csak pontosan megragadni. A tartózkodást az új terminológia alapján az alábbi módon fejezhetjük ki. Az építkezés logikája a következő ábrával szemléltethető.

(ZL.3) $\text{REFRAIN} \equiv -\text{DO} \stackrel{\text{def}}{=} \text{DO}(\neg\text{DO}) \equiv \text{DO}\neg\text{DO}$

tartózkodás (HUN) – refrain (ENG) – forbearance (ENG)

Az egyszerűség kedvéért a fenti formulából elhagytuk a cselekvéstartalmára utaló paramétert, de természetesen a cselekvések leírásakor ezt mindig meg kell adni. Von Wright sem gondolja ezt másként, így a következő feladat megnézni, hogy von Wright milyen fogalmi összetevőkre támaszkodva véli felépíthetőnek a cselekvés fogalmát. Az már kiderült, hogy a cselekvés fogalma alatt von Wright is valamilyen ágens által végzett intencionális tevést vagy tartózkodást ért. A kérdés, hogy mire vonatkozik ez a kétféle cselekvési lehetőség. Von Wright válasza – röviden –erre az, hogy vannak a világállapotok, két világállapotok között valamilyen átmenet, változás (transition) valósul(hat) meg, és a cselekvéseinket úgy jellemezhetjük, hogy ezek eredményeként két világállapot közti átmenet vagy megvalósul vagy elmarad.

(világ)állapotok $\rightarrowtail$ állapotváltozások (átmenetek, tranzíciók) $\rightarrowtail$ cselekvések

Az állapotokat p, q, r betűkkel jelöli von Wright, amit az esetek egy részében én is követek, de a későbbiekben egy ponton eltérek ettől, amikor az állapotok jelzésére majd az F betűt fogom alkalmazni. Az állapotok fennállásának tagadására pedig a $\neg p, \neg q, \neg r$ terminusokat alkalmazom. Az átmenetek, tranzíciók jelölésére követem von Wrightet, és a T-operátorral hivatkozom rájuk, melynek két állapotváltozó jelzésére alkalmas paraméterét

kell feltüntetni a formulákban: $T(p, \neg p)$ vagy $T(p, q)$. Ha azt akarom kifejezni, hogy az állapotváltozás operátor valamelyik paraméterében előfordulhat egy állapot vagy annak tagadása, akkor ezt a $\pm p$ szimbólummal jelölöm. Ekkor így néz ki egy állapotváltozást leíró nyitott formula: $T(\pm p, \pm p)$. Ezzel a nyitott formulával négy lehetséges állapotváltozást ragadhatunk meg egyszerre: $T(p, p), T(\neg p, p), T(\neg p, \neg p), T(p, \neg p)$. Von Wright a cselekvés általános fogalmára az **act** (ACT) terminust használja, míg a tevés és tartózkodás fogalomkettősét a **do** és **forbear** igék segítségével jelöli, de képletekben a rövidebb d és f szimbólumokat is alkalmazza. Az általam épített formális modelbe nem ezt a jelölési konvenciót viszem tovább, hanem a $\pm$DO (illetve a DO és a $-$DO) szimbólumokkal jelölöm a cselekvésoperátorokat. Ezek alapján a cselekvésre vonatkozó legáltalánosabb nyitott formulát így fejezhetjük ki: $\pm\text{DO}(T(\pm p, \pm p))$. Ezt a következő négyféle módon jelölhetjük és értelmezhetjük (ahol a '|' szimbólum az **exklúzió** műveletét jelzi):

$$
\begin{array}{rl}
(i) & \pm\text{DO}(T(\pm p, \pm p)) \leftrightarrow +\text{DO}(T(\pm p, \pm p)) \mid -\text{DO}(T(\pm p, \pm p)) \\
(ii) & \pm\text{DO}(T(\pm p, \pm p)) \leftrightarrow \text{DO}(T(\pm p, \pm p)) \mid -\text{DO}(T(\pm p, \pm p)) \\
(iii) & \pm\text{ACT}(T(\pm p, \pm p)) \leftrightarrow do(T(\pm p, \pm p)) \mid forbear(T(\pm p, \pm p)) \\
(iv) & \pm\text{ACT}(T(\pm p, \pm p)) \leftrightarrow d(T(\pm p, \pm p)) \mid f(T(\pm p, \pm p))
\end{array}
$$

A cselekvéseket tehát – von Wright elmélete alapján– úgy lehet leírni, hogy legelőször magadjuk, ki a cselekvő, ezután pedig előbb vesszük a világállapotokat, majd felmérjük, hogy milyen átmenetek lehetségesek közöttük, végül rögzítjük azt, hogy milyen típusú cselekvéssel milyen eredményeket ért el a cselekvő. A cselekvés eredménye mindig valamilyen világállapot (ami egy átmenet végállapota). A cselekvés feltétele pedig valamilyen aktuális átmenet, ami két világállapot közti kapcsolatot ír le, és aminek végén mindig előáll az adott átmenethez "tartozó" végállapot. Jobban megérthetjük mindezt, ha felidézzük von Wright ajtós példáját.

Ha van egy ajtónk, akkor annak két világállapota lehet: nyitva (p) vagy csukva ($\neg p$) van. A két állapot között négy átmenetet képzelhetünk el: egy csukott ajtót ki lehet nyitni ($T(\neg p, p)$), egy nyitott ajtót be lehet zárni ($T(p, \neg p)$), illetve egy zárt ajtó zárva ($T(\neg p, \neg p)$), egy nyitott ajtó nyitva ($T(p, p)$) maradhat. A cselekvés feltételén azt értjük, hogy melyik folyamat jellemző arra a helyzetre, amelyben a cselekvés éppen zajlik. Ha egy "normál" ajtó zárva van, akkor a ($T(\neg p, \neg p)$) átmenet lesz a cselekvés feltétele. Ha a egy önműködően záródó ajtóról van szó, és az ajtó nyitva van, akkor a ($T(p, \neg p)$) átmenet szolgál a cselekvés feltételéül stb. Amikor a cselekvés lezajlik, akkor a cselekvő vagy tevőlegesen vagy tétlen módon cselekszik, és a négy lehetséges feltétel mentén a cselekvés más és más (pontosabban másként értelmezhető) eredményeket fog eredményezni.

A következő táblázatban mpost már általánosított módon mutatom be a cselekvések feltételei, a cselekvések, az átmenetek és a cselekvések eredményei közti kapcsolatokat [564, 49], azonban von Wright megoldásához képest három apróbb módosítást is végrehajtok. Egyrészt más jelölési rendszert alkalmazok az operátorokra vonatkozóan, átírom a posztfix jelölési rendszer szerint a kifejezéseket (például a '$pT\neg p$' helyett a '$T(p, \neg p)$' kifejezést használom). Másrészt a cselekvés paraméterei közé felveszek egy ágenst, és ezt '*Ego*'-val jelölöm. Harmadrészt a cselekvés operátort a négy lehetséges jelölési mód közül a *(ii)*-vel jelölöm (szemben a von Wright által alkalmazott *(iv)* jelöléssel).

cselevés feltétele	cselekvés	cselekvés eredménye
$T(p,\neg p)$ p állapot fennáll, de megszűnik, hacsak nem tartják fent	$\text{DO}(Ego, T(p,p))$ Ego a p állapotot megőrzi	$T(p,p)$ p megmarad
$T(p,\neg p)$ p állapot fennáll, de megszűnik, hacsak nem tartják fent	$-\text{DO}(Ego, T(p,p))$ Ego hagyja, hogy a p állapot megszűnjön	$T(p,\neg p)$ p megszűnik
$T(p,p)$ p állapot fennáll, és meg is marad, hacsak nem változtatják meg	$\text{DO}(Ego, T(p,\neg p)$ Ego a p állapotot megváltoztatja	$T(p,\neg p)$ p megszűnik
$T(p,p)$ p állapot fennáll, és meg is marad, hacsak nem változtatják meg	$-\text{DO}(Ego, T(p,\neg p))$ Ego hagyja, hogy a p állapot megmaradjon	$T(p,p)$ p megmarad
$T(\neg p,\neg p)$ p állapot nem áll fenn, és nem is következik be, hacsak nem állítják elő	$\text{DO}(Ego, T(\neg p,p))$ Ego a p-t állapotot előállítja	$T(\neg p,p)$ p megtörténik
$T(\neg p,\neg p)$ p állapot nem áll fenn, és nem is következik be, hacsak nem állítják elő	$-\text{DO}(Ego, T(\neg p,p))$ Ego hagyja, hogy a p állapot hiánya megmaradjon	$T(\neg p,\neg p)$ p hiánya megmarad
$T(\neg p,p)$ p állapot nem áll fenn, de bekövetkezik, hacsak nem akadályozzák meg	$\text{DO}(Ego, T(\neg p,\neg p))$ Ego megakadályozza, hogy a p állapot bekövetkezzen	$T(\neg p,\neg p)$ p hiánya megmarad
$T(\neg p,p)$ p állapot nem áll fenn, de bekövetkezik, hacsak nem akadályozzák meg	$-\text{DO}(Ego, T(\neg p,\neg p))$ Ego hagyja, hogy a p állapot bekövetkezzen	$T(\neg p,p)$ p megtörténik

Az ajtós példára visszatérve: ha az önzáró ajtó nyitva van, akkor az magától bezáródna. Ilyen feltétel ($T(p,\neg p)$) mentén ha valaki tartózkodik, akkor az ajtó bezáródik, tehát a cselekvés eredménye a zárt ajtó állapota lesz. Ha a cselekvő aktív megakadályozza, hogy az ajtó önmagától bezáródjék, akkor az eredeti állapotot megőrizné, tehát az ajtó nyitva maradna.

Cselekvő ágens, Ego és Alter

A társadalmi cselekvés elemzése során ágensekről beszéltem, de eddig nem volt fontos sem az, hogy több cselekvőről beszéljünk, sem az, hogy megkülönböztessük őket egymástól. A társadalmi cselekvések során – akaratlanul is – modellezzük magát a társadalmat is. A társadalom minimális modelljéhez két szereplőre, két ágensre van szükség. Sokáig elég ennyit feltételeznünk, hogy a társadalmi kapcsolatok legelemibb vonásait értelmezhessük. Ehhez viszont meg kell különböztetnünk két szereplőt egymástól, leginkább azért, mert a társadalmi cselekvések megértését vagyis a cselekvéshez kapcsolt szubjektív értelem felderítését úgy remélhetük, ha valamelyik cselekvő szempontrendszerét, optikáját magunkévá tesszük. Ezt a szereplőt **Egonak** fogom nevezni, és amíg csak lehet, kitartok e név mellett. Ego a cselekvéselméleti modell **fokális ágense**. A társadalmi kapcsolatot vagy rövidebben a társadalmiságot "reprezentáló" másik szereplő neve **Alter** lesz.[5] Mindig Ego cselekvését akarjuk megmagyarázni, tehát egy személyre fókuszálunk, de a társadalmi cselekvéshez szükséges másik szereplőt nem hagyhatjuk ki a modellből – mégha tudjuk

[5]Ezt az elnevezési konvenciót már a cselekvéselméleti könyvemben is alkalmaztam, és a vonatkozó referenciák közül akkor is, most is Talcott Parsons hasonló megoldását, illetve az ezt támogató érvelést emelném ki [405].

is, hogy ezzel a névvel a legtöbb esetben csak egy passzív szerepet/szereplőt jelölünk csak.

Amikor a későbbiekben játékelméleti apparátust fogok alkalmazni a társadalmi kapcsolatok modellezésére, akkor látható lesz, hogy a játékelmélet különböző ágaiban a cselekvőkre olykor más neveket aggatnak (donor, recipiens, actor stb.). Ilyen esetekben igazodni fogok az éppen bemutatott elméletek névhasználati konvencióihoz, de minden esetben jelzem majd, hogy az újonnan bevezetett szerepnevek melyik "főszerepnek" felelnek meg (Egonak vagy Alternek).

Bármennyire is lényeges kérdés, hogy mit tartunk ágensnek, pontosabban – mivel cselekvéselméleti modellről van szó – mit tartunk emberi ágensnek, és ehhez kapsolódóan mit értünk a tudat fogalmán, e kérdésekre nem próbálok meg választ adni. Sok más diszciplina illetékesebb a válaszok keresésében, az érdeklődők a kognitív tudomány, az elmefilozófia, az intencionalitás filozófiája és még több más diszciplina képviselőitől remélhetnek adekvát magyarázatokat erre.

2.2 Intencionalitás

A cselekvést a benne rejlő intencionalitás többletével különböztethetjük meg az egyszerű viselkedéstől. Az intencionalitás vizsgálatához a szociológia mellett sok más tudományterület tudására is szükség van, vagyis nem mondhatjuk azt, hogy a szociológia egyedül lenne illetékes e kérdés vizsgálatára, de ezen a ponton azt nem tehetem meg, hogy nem foglalkozom ezzel a problémával. A tudat és az ágencia fogalmának elemzését még átutalhattam más diszciplinák illetékességi körébe, az intencionalitásról mondanom kell valamit. Az intencionalitás mibenlétével kapcsolatban csak rövidebben fejtem ki, amit fontosnak tartok, és ennek során kicsit át kell merészkednem a filozófia területére is, ahol természetesen nagyon félszegen merek csak mozogni, de ezt nem nagyon lehet megúszni.

Az intencionalitás fogalmát sokak szerint Franz Brentano emelte be elsőként a tudományos diskurzusba, amikor a pszichikai és fizikai jelenségek közti különbséget keresve az intencionalitást jelölte meg a jelenségeket egymástól elválasztó minőségként. Ha Brentano felfogása a kályha, ahonnan érdemes elindulnuk, akkor nézzük meg először azt, hogy mit mondott Brentano az intencionalitás fogalmáról.

> „Minden pszichikai közös jellemvonása abban található, amit sajnálatos módon gyakran oly félreérthető kifejezéssel tudatnak neveznek, vagyis egyfajta szubjektív magatartásban, egyfajta, mint mondják, *intencionális* viszonyban, amely valamire irányul, ami talán valójában nem létezik, ám belsőleg tárgyként adott. Nincs hallás valami meghallott, hívés valami elhitt, remény valami reménylett nélkül, nincs törekvés a törekvés célja nélkül, és nincs öröm anélkül, hogy valaminek ne örülnénk, s még sorolhatnánk." [80, 55]

Bár ebben az idézetben még nincs egyértelmű fókusz rajta, de azért már szerepel az a gondolat, hogy az intencionalitás egyfajta **szubjektív valamire irányulás**. Hogy ezt hogy kell értelmezni, később bővebben kifejtem Husserl intencionalitásértelmezése kapcsán. Előtte azonban érdemes még felidézni azt, hogy Brentano milyen altípusokat rendelt az intencionlitás fogalma alá.

> „Az első alaposztály a legtágabb értelemben vett képzeteké ...
> A második alaposztály az ítéleteké ... az ítélet alapja, miként a vágyakozásé is, mindig valamilyen képzet; ... Mi tehát a megkülönböztető jegye annak, amikor nem egyszerűen képzetet alkotok, hanem ítélek is? – Ebben az esetben a képzetalkotáshoz egy második,

> a képzet tárgyára irányuló intencionális viszony is kapcsolódik, a képzet tárgyának elismerése vagy elvetése. Aki Istent említi, Isten képzetét fejezi ki, aki pedig azt mondja: van Isten, arról beszél, hogy hisz benne.
> A harmadik alaposztály a kifejezés legszélesebb értelmében vett érzelmi mozgásoké, kezdve attól, hogy valamely gondolat egyszerűen kellemesen vagy kellemetlenül hat rám, egészen a meggyőződésen alapuló örömig vagy szomorúságig, illetve a cél- és eszközválasztás legbonyolultabb jelenségéig. Arisztotelész mindezt az *órexisz* fogalmában foglalta össze." [80, 56-57]

Azért tartom különösen fontosnak a fenti idézetet, mert Brentano intencionalitásfelfogását – Husserl nyomán – majd pontosítani kell, de a brentanoi tipizálást elfogadhatónak tartom, és be fogom építeni a cselekvéselméleti modellbe, ezért minél pontosabban tisztáznunk kell, miről is van itt szó. Brentano szerint az intencionálisnak, a szubjektív valamire irányultságnak három típusa van. A **képzet**, az **ítélet** és az **érzelmi mozgás**. Az első kategóriára mondhatjuk azt, hogy a képzetek voltaképp a fogalmakat jelentik, az ítéletek kategóriájával nincs igazán értelmezési probléma, gond van azonban a harmadik kategóriával. Mit kell értenünk az 'érzelmi mozgások' kategóriája alatt? A fenti idézetben ezzel kapcsolatban Brentano utal Arisztotelész 'órexisz' kategóriája, ami a vágyakozás, törekvés fogalmával azonosítható. Sajnos, mikor Brentano más helyütt visszatér a típusainak bemutatására, akkor sem mond többet vagy pontosabbat az eddigieknél.

> „Minden pszichikai jelenség azáltal jellemzett, amit a középkor skolasztikusai intencionális (nyilvánvalóan mentális is) inegzisztenciának (valamire irányuló belső létezésnek) neveztek, s amit mi, habár nem is egészen kétértelműség nélküli kifejezéssel, egy tartalomra való vonatkozásnak, egy tárgyra való irányulásnak (ami alatt itt nem realitást kell érteni) vagy pedig az immanens tárgyiasultságnak nevezhetnénk. Mindegyik magába foglal valamit a tárgyakból, habár nem mindegyik egyenlő módon. A képzetben valamit megjelenítünk, az ítéletben valamit elismerünk vagy elvetünk, a szeretetben valamit szeretünk, a gyűlöletben valamit gyűlölünk, a vágyakozásban valamit kívánunk stb." [79, 161]

Ez az idézet is megerősíthet minket abban, hogy az első két irányultságtípusként a képzeteket (vagy fogalmakat), illetve az ítéleteket (vagy **propozíciókat**) különíthetjük el Brentano nyomán, amelyekhez ráadásul a modern logika eszköztárának igen nagy részét hozzárendelhetjük. Továbbra sem világos azonban az, hogy mit értsünk a harmadik kategória alatt. Brentano ide sorolja mindazt, amire a szerettünk, gyűlöletünk, vágyakozásunk, érzelmeink, örömünk, bánatunk irányul, de idetartozónak véli a cél- és eszközválasztás legbonyolultabb jelenségeit is. Utóbbi mozzanat zavarbaejtő, mert a (cél- és eszköz)választás már a mérlegelés területéhez tartozik, amit viszont nem szoktunk az érzelmi mozgások fogalma alá sorolni, sőt, ezeket sokkal inkább ellentétes lényegű jelenségeknek tartjuk. Mi lehet a megoldás? Brentanonak szerintem igaza van abban, hogy a felsorolt példák mind a harmadik kategória alá tartoznak, de nincs igaza akkor, amikor ezt érzelmi mozgásoknak nevezi. A racionális, mérlegelésen alapuló beállítódásokat valóban élesen szembe kell majd állítanunk az érzelmi cselekvésekkel, de ez nem jelenti azt is egyben, hogy ne lehetne mindezekben valami közös. Ez pedig nem más, mint az arisztotelészi órexisz kategóriája (amit Brentano egyébként áttételesen meg is nevez). Mivel ezt a régi terminust nem akarom továbbvinni, így helyette a továbiakban a törekvés, **késztetés** terminusát fogom használni, ami alatt a – képzeteinkre és ítéleteinkre vonatkozó – irányultságaink poláris minőséggel (vonzással vagy taszítással) és motivációs intenzitással ellátott értékviszonyulást érthetünk. A késztetéseink azért nagyon fontosak az életünkben, mert ezek aktuális vágyakat és averziókat képesek ébreszteni bennünk, amelyek közül

aztán vágyaink egy része a hiteinkkel kombinálódva aktuális szándékokat alakít ki, és ezek végül cselekvésekre sarkallnak minket. A késztetéseink irányultsága azonban egy fontos szempontból eltér az ítéleteink irányultságától. Searle munkássága nyomán vált széles körben ismertté és elfogadottá az a nézet, miszerint az intencionalitás leírásakor annak megfelelési irányát is ki kell fejeznünk [469, 92].[6] Ez az irány alapvetően kétféle lehet, világ-szó vagy szó-világ irányú, de Searle szerint elképzelhető olyan helyzet, amikor nincs az intencionálisnak nincs ilyen irányultsága, illetve lehetséges az is, amikor mindkét irány tételeznünk kell. De mit is ért megfelelési irányon Searle? A szó-világ megfelelés azt jelenti, hogy a kimondott, elgondolt ítéleteink tartalmilag a világra irányulnak. Amikor kijelentjük, hogy 'esik az eső', akkor a világ állapota felől érthetjük meg a kimondott mondatunkat. A világ itt adottnak tekinthető, és az intencionalitás funkciója itt a világ leírása. A világ-szó megfelelési irány esetében az intencionalitás funkciója más. Ekkor ugyanis az ágens szeretné valamilyen értelemben megváltoztatni a világot, és ezért a beszédaktusban megnyilvánuló cselekvése a szótól a világ felé irányul. Amikor ígéretet teszünk vagy utasítást adunk valaki másnak, akkor a megnyilatkozásunk arra irányul, hogy a világ valamilyen értelemben megváltozzék. A kétféle megfelelési irány segítségével tehát négyféle esetet képzelhetünk el. Amikor azonban Searle típuskra bontja a beszédaktus fogalmát, akkor ő öt altípust különít el.

beszédaktust	megfeleltetési irány	példák
asszertívum	szó-világ	kijelentés, leírás, osztályozás, magyarázat
direktívum	világ-szó	utasítás, parancs, kérés
komisszívum	világ-szó	ígéret, fogadalom, felajánlás, megállapodás, jótállás, fenyegetés
expresszívum	nincs	bocsánatkérés, köszönet, gratuláció, felköszöntés, részvét
deklaráció	szó-világ/világ-szó	hadüzenet, kinevezés, házastárssá nyilvánítás

Az öt searle-i beszédaktustípusból azonban kettő összevonható azon az alapon, hogy csak abban különböznek egymástól, hogy melyik ágensre (Egora vagy Alterre) vonatkozik a hatókörük. Mivel ezt a szempontot itt elhanyagolhatjuk, így a searle-i felosztás azt a négy esetet tartalmazza, amelyeket az előbb tételeztem. A intencionalitás fogalmának brentanoi hármas alábontása tehát abban az értelemben megalapozza a későbbi cselekvéselméleteket, hogy felkínálja az intencionalitás két fontos típusaként a propozíciót mint a szó-világ megfelelési irányt, illetve a késztetést mint a világ-szó megfelelési irányt megvalsító intencionalitást. Ezt a kettősséget Ruth Millikan – az emberi reprezentációk két típusaként – **kijelentő reprezentációnak** (vagy kijelentő doboznak), illetve **felszólító reprezentációnak** (vagy szándék doboznak) nevezte el [358], [359].

> „Az emberi hitek nincsenek közvetlenül a cselekvésekhez kötve. Ha nem kombinálják őket megfelelő vágyakkal, az emberi hitek tehetetlenek. És az emberi vágyak szintén tehetetlenek, ha nincsenek megfelelő hitekkel kombinálva. ...
> Mivel az emberi lények belső reprezentációs rendszerében a kijelentő és a felszólító funkciók szét vannak választva, szükség van újbóli integrálásukra. Ezért az emberek gyakorlati következtetéseket végeznek, a hiteket és a vágyakat újszerű módokon kombinálják, hogy azok először szándékokat, majd cselekvést eredményezzenek." [359, 267]

A képzetekkel az ontológiai elkötelezettségeinket fejezhetjük ki, rámutathatunk, szerintünk mit lehet kifejezni világban, mire, milyen dolgokra, jelenségekre irányulhat az inten-

[6] Ezt a megkülönböztetést már John L. Austin is felvetette [28], de G.E.M. Anscombe hozta arra a formára, amit a mai napig használunk [16], és amit John Searle is átemelt az elméletébe.

cionalitásunk, az ítéleteinkkel, propozícióinkkal kifejezhetjük azon hiteinket, vélekedéseinket, hogy mit gondolunk a képzeteink segítségével a világról, és a késztetéseink, vágyaink révén pedig valamilyen mozgásra, változtatásra, cselekvésre irányul az intencionalitásunk. A rendszer alján vannak a képzetek, és ezekre épül a millikani kijelentés- és szándékdoboz tartalma.

A pszichológiában régóta bevett nézet szerint három emberi képesség, fakultás létezik, a kognitív (vélekedés), az affektív (vágy, érzelem, érzés, öröm, fájdalom) és a konatív (intenció, akarat, elköteleződés). Ezek közül az első, a kognitív képesség a brentanoi értelemben vett fogalmak és ítéletek kezelésével kapcsolatos, a második, az affektív képesség a brentanoi értelemben vett késztetések rendszerével köthető össze, a harmadik, a konnatív képességet pedig később fogom tárgyalni, amikor az intenció fogalmát elemzem majd.

Az intencionalitás filozófiájából a fenomenalizmust megteremtő Edmund Husserl sokat átvett Brentano nézeteiből, de bizonyos pontokon alapjaiban újragondolta azokat. Husserl intencionalitásfelfogását csak olyan mértékben mutatom be, hogy az általa javasolt változtatások és kiegészítéseket értelmezni tudjam [248], [249], a husserli életmű, a fenomenalizmus filozófiájának részletes ismertetése nyilván nem lehet célom.[7] Husserl elsősorban a tudat, a tárgyra irányulás, illetve a belső reprezentáció fogalmának értelmezésével kapcsolatban támadja Brentano felfogását (illetve az ezzel rokon "természetes beállítódást", valamint az ugyan csak később kialakuló, ám nagyon széles körben elterjedt reprezentációelméleteket, illetve a Husserl által pszichologizmus címkéjével jelölt nézeteket). Szerinte ugyanis helytelen az az intencionalitásmodell, amely szerint egyrészt meg kell különböztetni mentális és a fizikai tárgyakat, másrészt az intencionalitást úgy kell elképzelni, mint a valamilyen módon elkülönült tudatnak a külvilág jelenségeiről "készült" belső reprezentációkra való irányulását, harmadrészt nem lehet az tudattal rendelkező én-t a tárgyra irányulás előtt, attól függetlenül, önmagában létező entitásnak tekinteni. Ezzel szemben – mondja Husserl – egyrészt az intencionális aktus nem függ sem a tárgy lététől, sem annak nem lététől, másrészt az intencionális aktus közvetlen tárgyra irányulás és az én ennek az élménynek egy mozzanata, harmadrészt az intencionalitás nem leképezés, hanem értelemadás [549, 141-178]. Az intencionalitás értelmezéséhez nem kell feltételeznünk az önmagában létező tárgyat, mert az intencionális élmény ennek hiányában is létezik.

> „... egy ilyen élmény, a maga intenciójával együtt, jelen lehet a tudatban, anélkül, hogy a tárgy egyáltalán létezne, vagy anélkül, hogy ez a tárgy egyáltalán lehetséges lenne. Ezt a tárgyat véljük, vagyis az ezt a tárgyat vélés az élmény; a tárgy azonban pusztán vélt és valójában semmi." [249, 29]

Ugyancsak fölösleges és félrevezető az elkülönülten, önmagában létező ént vagy tudatot feltételeznünk. Újracsak azt kell mondani ehelyett, hogy az én(tudat) az intencionális élmény egyik komponense, nem független létező.

> „Ám ha úgyszólván benne élünk a kérdéses aktusban, ha egyszerűen feloldódunk például a megjelenő esemény észlelő szemlélésében, vagy a fantázia játékában, egymese olvasásában, egy matematikai bizonyítás levezetésében stb., akkor a végrehajtott aktusokban egyáltalán nem vesszünk észre egy vonatkozási pont értelmében vett ént. Az én-képzet »készenlétben« állhat, hogy könnyűszerrel előtérbe nyomulhasson, vagy inkább hogy újra meg-

[7] Az intencionalitás husserli értelmezésének bemutatása során támaszkodom Ullman Tamás könyvére [549, 141-178] és Tőzsér János tanulmányára [544].

> valósuljon; azonban csak ha valóban megteremtődik, és a vonatkozó aktussal eggyé válik, akkor vonatkozunk *mi* úgy a tárgyra, hogy az én ilyen fajtájú vonatkozásának valami deskriptíve felmutatható felel meg." [249, 31]

Husserl frappáns kifejezésével élve: "az én semmi több, mint a »tudat egysége«". És mindazon túl, hogy nem kell feltételeznünk, pontosabban nem kell figyelembe vennünk sem az önmagában létező tárgyat, sem az elkülönült (én)tudatot, el kell vetnünk azt a nézetet is is, ami az intencionalitást a külvilág leképezéseként modellezi. Ha ugyanis azt feltételeznénk, hogy a külvilág dolgait egy belső reprezentációba képezzük le, és a tudatunk ezt "ismeri fel", akkor végtelen ciklusba kerülnénk, ugyanis fel kéne tenni a kérdést, hogy a belső reprezentációink hogyan állítjuk elő és hogyan észleljük.

> „Ha valaminek a képként való felfogása egy, a tudat számára intencionálisan adott objektumot előfeltételezne, akkor ez nyilvánvalóan végtelen regresszushoz vezetne, mivel maga ez az objektum ismét csak egy kép által konstituálódna;" Hua XIX/1.437. idézi: [549, 155]

A széles körben bevett, de épp ezért félrevezető, nem kívánatos konnotációval rendelkező 'intencionális tárgy' kifejezés helyett Husserl inkább az **(intencionális) matéria** vagy intencionális anyag terminust használja (ami mellett a továbbiakban a **(propozícionális) tartalom** kifejezést is használni fogom). Nem lehet elégszer hangsúlyozni, hogy ennek az intencionális tárgynak (vagy intencionális tartalomnak, matériának) a definiálása szempontjából teljesen indifferens annak létezése.

> „ ... az intencionális tárgy ugyanis *a szubjektum prespektívájából megállapított fogalom*. Intencionális tárgynak lenni nem más, mint annak a valaminek lenni, amit a szubjektum a 'mi az, amire gondolsz' kérdésre őszintén válaszol." [544, 172-173].

Fontos az a husserli kritika (és újraértelmezés) is, amely az értelemadással és az intencionális tárgy felfogásával kapcsolatos. Husserl azért tartja fontosnak bevezetni a **felfogás(értelem)** fogalmát az intencionalitás modelljébe, mert csak így lehet megragadni azt a tényt, hogy az intencionális tartalomra (matériára) irányulásunk mindig valamilyen szempontból, valamilyen apektusból valósul meg. Egy intencionális tartalomra sosem a maga teljességében tekintünk, hanem mindig csak valamilyen szempontból valósul meg az irányultság. És – ahogy ezt már jeleztem korábban – Husserl szerint ez a "részlegesség", ez a valamilyen aspektus szerinti irányultság nem leképzésként, hanem **felfogásként**, értelemadásként fogható fel inkább.[8] A nyilvánvaló példa erre az, hogy amikor egy tárgy van előttünk, mondjuk egy asztal, akkor sosem az egészet látjuk, csak valamilyen részét, aspektusát. Az intencionális tartalom mindig csak valamilyen szempont szerint, valamilyen aspektusból, valamilyen – szubjektív – értelemben van jelen az intencionális élményben. Ha egy adott nézőpontból látunk rá elsőként egy asztalra, és az egyik láb takarásban van, így háromlábú asztalként jelenik meg a tudatunkban, mindaddig nem változtatunk ezen az intencionális tartalmunkon, amíg az asztalt körbe járva ki nem derül, hogy annak van egy negyedik lába is.

Husserl intencionalitásértelmezésének harmadik fontos összetevője az **intencionális kvalitás** vagy intencionális minőség kategóriája, bár szerintem ezen a ponton nincs nagy különbség Husserl és Brentano felfogása között. Husserl is fontosnak tartja azt, hogy mindig számba vegyük az intencionális aktus minőségét, akárcsak Brentano. Az intencionális tartalom meghatározza azt, amire az intencionális aktus irányul, míg a kvalitás azt a módot, ahogy ez az irányulás megvalósul. Husserl többször használja érdemben

[8] John Searle a felfogás, az értelemadás fogalmára az aspektus terminust használja [544], de voltaképp' ugyanezt fejezi ki Max Weber 'szubjektív értelem tulajdonítás' kifejezése is.

Brentano tipizálását, és nem bírálja fölül azt: fogalmilag "legalul" vannak a képzetek, és ezekre épülnek az ítéleteink, illetve a vágyaink (késztetéseink).

Ebben a dimenzióban lehetne finomabb fogalmi felbontást is készíteni, ha a searle-i értelemben vett megfelelési irány fogalma által "kijelölt" felosztáson vagyis a kijelentések és késztetések kettősségén túl figyelembe vennénk további szempontokat. A cselekvések, beszédaktusok sokféle típusa (kérések, parancsok, ajánlatok, remények, aggodalmak stb.) beemelhető lenne a modellünkbe, ha így tennénk, akkor valóban érdemes lenne különböző cselekvési módokról, attitűdökről vagy – a beszédaktus-elmélet által bevezetett - – illokúciós erő ilyen-olyan típusairól beszélni, de ezt a részletesebb tipologizálást itt fölöslegesnek tartom, így egyelőre eltekintek ezek modellezésétől.

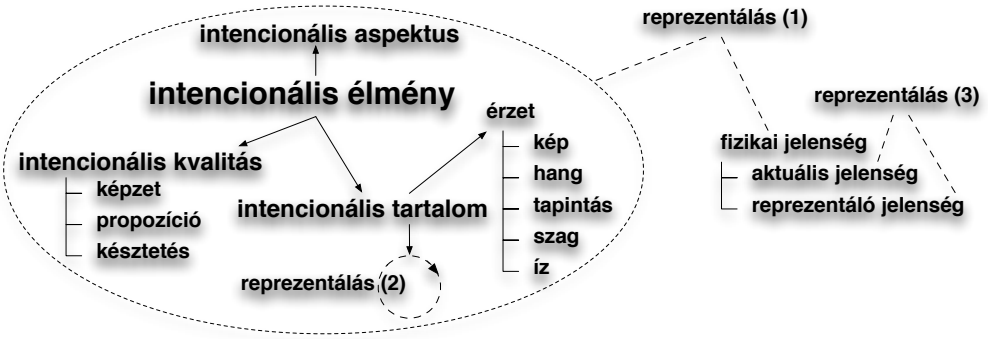
A hiteink, késztetéseink, propozícionális attitűdjeink összefüggő rendszert, sőt kauzális rendszert alkotnak [5]. Ha Ego úgy vélekedik, hogy kevés benzin van a kocsijában, és tudja, hogy a tervezett útra kevés lenne, akkor tudja azt is, hogy egy benzinkútnál tankolnia kell. A propozícionális attitűdökre (azok kialakulására, változására) hatnak a külvilág tárgyai, az intencionális állapotaink, vélekedéseink „kauzálisan kapcsolódnak a külvilág dolgaihoz, eseményeihez és sajátosságaihoz". Ez már egy második kauzális viszonyrendszert jelent, ami a belső és külső világ "tárgyai" közötti kapcsolatokat jellemzi. A kétfajta kauzális rendszer egy egységet (Altrichter kifejezésével élve: "kauzális teret") alkotva folyamatosan "kauzális befolyással" van az Egora. A kauzális hatásrendszer létezésén túl azonban az attitűdjeink további "kényszereknek" vannak alávetve. A propozícionális attitűdjeink világában ugyanis szemantikai, logikai kényszerek is érvényesülnek. A logikai-szemantikai kényszerek létezése és "befolyásolási képessége" azért különösen fontos, mert feltételeznünk kell azt is, hogy a racionalitásnak akár az ideális, akár a reális formáját is vesszük figyelembe, a racionalitásnak erős attitűdbefolyásolási képességet kell tulajdonítanunk. Érdemes is egyfajta zártság elvként megfogalmaznunk a racionális gondolkodásból fakadó logikai kényszereket.

> „A reális eseteknek sokkal jobban megfelel az az elv, amit a hitre vonatkozó zártsági elvnek nevezek. Ez az elv azt mondja ki, hogy a hit zárt a tudott vagy akárcsak elhitt logikai implikáció szempontjából, azaz ha valaki elhisz egy kijelentést, és tudja vagy csupán azt hiszi, hogy a kijelentés logikailag implikál egy másikat, akkor el kell fogadnia a másik kijelentést is. Ennek az elvnek azonban fontos tudatfilozófiai következményei vannak. Tegyük fel, azt hisszük, hogy *nem* q, és azt is hisszük, hogy p. Ha viszont rájövünk arra, hogy p implikálja q-t, akkor meg kell változtatnunk eredeti mentális állapontunkat q tagadását illetően. A zártság elve arra kényszerít mindannyiunkat, hogy többé ne higyjük el q tagadását, arra tehát, hogy eredeti propozicionális attitüdünket feladjuk. Az interszubjektív racionalitás által kikényszerített ilyen változás nem egyszerűen a mentális tartalom megváltozását jelenti, hanem magának a pszichikus állapotnak a megváltozását. A reális racionalitás regulatív szabálya a logika konstitutív szabályaival összhangban – hogy úgy mondjam – felülbírálja a kauzális tér elemeit, és bizonyos pszichikus állapotokat egyszerűen megsemmisít és kiküszöböl a kauzálisan releváns állapotok köréből." [5, 140]

A zártság elve megköveteli az önismeretet. Akkor változtathat valaki az attitűdjein, ha tud róla, hogy korábban milyen intencionális állapotban volt. Ez viszont már másodrendű intencionalitást jelent, hiszen az Egonak tudnia kell, hogy tudott valamit korábban. Olyan fogalmakkal kell tehát bővíteni mind az Ego egyes szám első személyű, mind a megfigyelő harmadik személyű ontológiáját, mint a hit, a vélekedés, vágy, a gondolat. Az önismeret tehát fogalmilag szükséges a modellünk számára, ez viszont megköveteli az öntudatnak, a szelfnek, a szubjektum fogalmának beemelését is, „hogy képesek legyünk

önmagunkat a propozícionális attitűdök szubjektumaként felfogni" [5, 140]. A szigorú fogalmi építkezés logikája csak ezen a ponton engedné meg nekünk, hogy a modellünkbe felvegyük a fokális ágens (Ego), illetve a társas ágens (Alter) fogalmát, és innen indulva lehet a társas preferencia kategóriája alá sorolható olyan fogalmakat is definiálni, mint egoizmus, altruizmus, önfeláldozás stb. (↬234.o.).

Husserl nyomán óvatosan kell kezelnünk a reprezentáció fogalmát, nehogy a természetes beállítódás által kínált naív elképzelések vagy a pszichologizmus csapdájába essünk, de ez nem jeelnti azt, hogy ne lehetne, sőt ne kellene használnunk ezt a kategóriát. Jelentőségét az a puszta tény is mutatja, hogy a terminus sokszorosan túlhasznált, több jelentésrétegeket is el lehet különíteni a fogalomhasználati gyakorlaton belül, ezért nagyon kell vigyázni arra, hogy ne keverjük össze ezeket. Az intencionális aktus értelmezésével kapcsolatban bemutatott eddigi fogalmakat egymáshoz való viszonyukban ábrázolva feltüntetem még a reprezentáció fogalmának fontosabb értelmezési lehetőségeit is.



Az intencionális élmény, intencionális aktus egészét jelzi a szaggatott vonallal határolt ellipszis. Az intencionális aktus az érzetadatokon keresztül valamilyen beállítódás mentén irányul az intencionális matériára, tartalomra valamilyen intencionális kvalitással, minőséggel jellemezhető módon. Az intencionális tartalom többféle lehet – az érzetadatoknak típusainak megfelelően. Eközben legalább háromféle reprezentációértelmezésről beszélhetünk. Egyrészt – a husserli intések ellenére – beszélnünk kell valamilyen reprezentáról a valós jelenségek és annak tudati feldolgozása, intencionális irányulása viszonylatában (1). Másrészt a tudat reflexív önmagára irányulása a reprezentáció más(od)ik szintjét teremti meg (2). Harmadrészt a valós jelenségek, fizikai objektumok világában is teremthetünk a valós jelenségek világán belül maradó reprezentációkat (3). A kulturális tárgyaink, jelenségeink, egy könyv, egy festmény, egy fénykép vagy egy karjelzés, egy visszavonulást jelző harcikürtjelzés lehetnek a példák erre.

Az intencionalitás modellje segít értelmezni azt a fontos jelenséget, hogy a reprezentációink lehetnek tévesek is. Intencionalitásunk irányulhat nem létező tárgyakra is (például egy szárnyas vizilóra), vagy a vélekedéseink lehet logikai értelemben tévesek (emiatt van szükség az episztemikus és doxasztikus logika elkülönítésére). A kérdés itt az, milyen funkciót tulajdoníthatunk a téves képzeteinknek. Érdemes komolyan vennünk Ruth Millikan megjegyzését ezzel kapcsolatban.

> „Ha az igaz hitek funkcionálisak, a hamis hitek pedig legtöbbször nem rosszabbak az üres elménél, akkor még a tévedésre rendkívül hajlamos hitrögzítő eszközök is jobbak lehetnek a hitrögzítő eszközök teljes hiányánál ... Mivel rendkívül pontos hitrögzítő eszközök tervezése nehéz, talán valójában előnyösebb túl sok hitet rögzíteni, megengedve,

hogy közülük egyesek hamisak lesznek. Ennek megfelelően hitfelhasználó mechanizmusainkat talán arra tervezték gondosan, hogy képesek legyenek tolerálni a hamis hitek magas arányát." [359, 259-260]

Teleológia

Az intencionalitás, a valamire irányulás fogalmával gyakran összekapcsolják a **teleológia** fogalmát, amely szűkebb terjedelmű kategória az intencionalitáshoz képest, de mivel sokan és széles körben hivatkoznak rá, ezért fontosnak tartom minél pontosabban tisztázni a jelentését. Annál is szükség van erre, mert világossá kell tenni azt is, hogy mi e fogalom viszonya egy másik kategóriához, a **teleonómia** fogalmához, amit a biológia használ. Amilyen gyakran hivatkozunk a társadalmi cselekvéseink célkövető, teleologikus jellegére, ugyanolyan gyakran alkalmazzuk a biológiai rendszerek, jelenségek, entitások jellemzésére az olyan fogalmakat, mint a funkcionalitás vagy célszerűség kategóriái. A későbbiekben nagyon fontos lesz, hogy el tudjuk precízen különíteni egymástól a társadalmi és biológiai rendszerekben egyaránt fellelhető "célra törekvés" eltérő minőségeit. Amíg a társadalmi cselekvést jellemző célra irányulást a **teleologikus**, addig a biológiai rendszerekhez rendelhető – az adaptivitásban megnyilvánuló – célszerűséget a **teleonomikus** jelzővel illethetjük.

A teleonómia fogalmát Colin Pittendrigh vezette be 1958-ban. Az ő meghatározását azonban túl tág értelműnek találta Ernst Mayr [328], aki a biológia által vizsgált kétfajta **célkövető** (end direction) jelenséget elkülönítve egymástól a teleonómia és teleológia fogalmainak párhuzamos használatát javasolta. Mayr azért kritizálta Pittendrigh értelmezését, mert az nem tett különbséget aközött, hogy egyfelő az individuumok (rendszerek) célkövető módon, szándékosan (purposefully) képesek viselkedni, másfelől a genetikai programok evolúciójában vélelmezett célra törekvés között [328, 1504]. Mayr a teleonómia fogalmát így valamely program, információs kód alapján viselkedő individuum, rendszer célkövető minőségére javasolta leszűkíteni. Ebben az értelmezésben a biológia mint tudomány csak a teleonómia jelenségére tud példával szolgálni [329], [330]. Hayne Reese a teleológia, illetve teleonómia kategóriakettőst két új fogalomra alapozva állította szembe egymással [436]. Reese megkülönböztette egymástól a **céltudatosság** (purposefulness) és a **célszerűség** (purposiveness) fogalmait egymástól. Céltudatosság fogalma alatt azt értette Reese, hogy az adott viselkedésnek, működésnek, entitásnak van valamilyen belső meghatározottsága, míg a célszerűség minőségét arra az esetre javasolta alkalmazni, amikor az adott viselkedés, működés, entitás valamilyen – hasznos – funkciót szolgál.

Abban a kérdésben, hogy miként lehet a "valamely program, információs kód" alapján történő viselkedést pontosabban leírni, a **funkció** fogalmát hívhatjuk segítségül [588], [357], amit Larry Wright vagy Ruth Millikan felfogásával szemben én a teleonómia fogalma alá gondolok besorolhatónak. Bár Wright és Millikam funkcióértelmezése nem egyezik meg, azért pár közös vonást találhatunk az elméletükben [172]. Az F funkció (működés, tevékenység) akkor és csak akkor egy X entitás funkciója (működése, tevékenysége), ha teljesül az alábbi két feltétel: *(i)* X egy korábbi Y entitás reprodukciója, amely korábban F-t csinálta, illetve *(ii)* X azért reprodukálódott, mert Y korában F-t csinálta. Ehhez szerintem érdemes még hozzávenni azt a feltételt is, hogy F egy olyan oksági láncon keresztül jött létre, amit valamilyen – a természet vagy az ember által létrehozott – kód váltott ki. Mindezt pedig az alábbi formulával fejezhetjük ki.

(ZL.4) $\text{TELEONOMIC_STATEMENT} \stackrel{\text{def}}{=} ((Y(t_1) \to F) \to (X(t_2) \to F)) \wedge (t_1 < t_2) \wedge X(t_1) \in R \wedge Y(t_2) \in R$

teleonomikus állítás (HUN) – teleonomic statement (ENG)

ahol R egy olyan reprodukciós osztály, amelynek tagjai mind megvalósítják az F funkciót, R-be tartozik mind az X, mind az Y entitás, és $X(t_1)$ időben megelőzi $Y(t_2)$-t. Nagyon fontos kiemelni, hogy a teleonómia funkció fogalmában van egyfajta meghatározottság, determinizmus, ami a teleológia fogalmából hiányzik.

Amikor Ronald de Sousa a teleologikus gondolkodás, beállítódás sajátosságát akarta kifejezni, a teleologikus állítás formuláját az oksági állítás formulájához hasonlította [128]. Amíg az oksági állítás formulája egy egyszerű kondicionálissal kifejezhető ($A \to B$), addig a teleologikus állítás leírásához bonyolultabb képletre van szükségünk. A teleologikus cselekvést leíró formula "középpontjában" a G **cél** (Goal) áll, az a jövőbeni állapotot, amit az ágens szeretne elérni a cselekvésén keresztül. A cél eléréshez szükséges valamilyen M **eszköz** (Mean), és – ezt Sousa nem vette figyelembe, én teszem hozzá a modellhez – számolni kell a cél megvalósulásakor jelentkező C **következménnyel** (Consequence). Ezekkel a paraméterekkel a teleologikus állítást a következőképpen írhatjuk le.

(ZL.5) $\text{TELEOLOGIC_STATEMENT} \stackrel{\text{def}}{=} (M \to (G \to C)) \to M$

teleologikus állítás (HUN) – teleologic statement (ENG)

A formula három egymásba ágyazott kondicionálisból áll.[9] A legfontosabb kondicionális azt írja le, hogy ha egy M cselekvést megvalósítok, akkor elérem a G célt. Ha ezt a kondicionálist mint egy egész állítást teszem meg egy második kondicionális előtagjává, akkor abból a cél elérését remélhetem, ezért kerül a második kondicionális utótagjába a G. Végül számolnom kell azzal is, hogy a céloknak lehetnek következményeik, vagyis ha a G cél megvalósul, akkor a C következmény is létrejön (ezt fejezi ki a harmadik, a formula közepébe beágyazott kondicionális).

2.3 Késztetés, vágy, érték

Az intencionalitás egyik kvalitása a késztetés. Ezt a fogalmat tartom a legáltalánosabbnak, amit fogalmi primitívként érdemes felvenni a modellbe, és amiből levezethetünk olyan további fontos fogalmakat, mint a vágy, az érték, szükséglet, cél vagy a szándék. A szükséglet jelentését itt azonosnak minősítem a késztetés fogalmának tartalmával, így a két terminust egymás szinonímáinak tartom. A továbbiakban a késztetés, vágy és érték fogalmai közti kapcsolatot szeretném tisztázni. Természetesen nagyon nehéz vagy talán nem is lehet tisztán elkülöníteni egymástól ezeket a fogalmakat. Régóta és gyakran használjuk őket, rengeteg jelentésréteg tapadt rájuk, a különböző kontextusokban részben változó és sokszor egymást átfedő jelentéseket adunk nekik. Azért érdemes a késztetés kategóriáját alapul venni, mert ez tűnik a legáltalánosabb jelentésűnek, ami még nem "követeli meg" feltétlenül a tudatosság jelenlétét. A vágyhoz vagy az értékhez – értelmezésem szerint – mindig érdemes valamilyen szintű tudatosságot rendelnünk, így ezeket speciálisabb fogalomnak kell tartanunk. A késztetéseknek (vágyaknak, értékeknek) mindig van polaritása,

[9] Csak a benne rejlő játékosság, ritmika miatt említem meg, hogy miként is kéne kimondani a fenti képletet: "ha, ha M, akkor, ha G, akkor C, akkor M".

amit a – fizikai világból ismert – vonzás és taszítás analógiájára hivatkozva ragadhatunk meg. Ezt az analógiát arra is használhatjuk, hogy jelezzük vele a késztetések (vágyak, értékek) energetizáló képességét, szerepét. A késztetés kifejezés maga szépen kifejezi ezen intencionális kvalitás meghatározó vonását, ti. azt, hogy a késztetés (vágy, érték) cselekvésre késztet. A késztetések polaritását magyar nyelvi terminuokkal nem tudjuk kifejezni, de a vágyak szintjén már megtalálhatjuk ezt a kettősséget a terminusok között, amikor – ellentétes polaritású – vágyakról és averziókról beszélünk. Ha a vágyat mint vonzást eredményező, pozitív késztetést, míg az averziót mint taszítást eredményező, negatív késztetést gondoljuk el, akkor már azt a megkülönböztetést származtathatjuk mind a késztetések, mind az értékek felé. Mondhatjuk, hogy a késztetéseink lehetnek pozitív, illetve negatív polaritásúak, vagyis vonzódás, ileltve taszítást eredményező intencionális kvalitások. Az energetizálási analógia alapján mondhatjuk még azt is, hogy a két pólushoz mindig kapcsolódik a kielégülés, illetve a kielégületlenség (frusztráció) jelensége, állapota. Azt a kérdést, hogy vannak-e és ha igen, milyen alapvető késztetései (szükségleti) vannak az embernek, itt nem tárgyalom, még akkor sem, ha egyébként a gyakorlati szempontokra figyelő cselekvéselméleti modell számára megkerülhetetlen kérdésnek tartom.

A modellépítés következő kérdése az, hogy mit is jelent a vágy fogalma. Értelmezésem szerint a **vágy** a tudatba került pozitív (polaritású) késztetés. És természetesen a vágy ellentéte, az averzió a tudatba került negatív (polaritású) késztetés. A vágy/averzió tehát úgy működik, hogy az elme részben a biológiailag adott, részben az egyén története során szerzett, tanult késztetéseket elraktározza, és ezek egy része a külső ingerek hatására valamilyen belső – a személyiségre jellemző – dinamika szerint eléri a tudatosság szintjét. Amíg a késztetésekről van csak szó, addig a tudattalanban levő **diszpozíciókról** beszélhetünk, és amint ezek közül valami tudatossá válik, akkor jelenik meg bennünk a vágyakozó irányultság valamire (vagy a valamitől való averzió).

A tudatosságnak egy következő szintjét is figyelembe kell vennünk ahhoz, hogy modellezni tudjuk az **érték** jelenségét. Az értéket én másodrendű vágynak tartom, ami voltaképp a tudatba került késztetések közti tudatos választással, elköteleződéssel azonosítható. A késztetéshez hasonlóan és a vágytól eltérően az érték diszpozicionális jellegű. A másodrendű vágyfogalmakkal kapcsolatos elképzeléseket röviden bemutatom a fejezet későbbi részében. A vágyak és a célok, akaratok, szándékok közti kapcsolatra viszont csak egy későbbi fejezetben fogok kitérni (↪155.o.), itt tehát nem foglalkozom ezzel a kérdéssel. Még ebben a bevezetőben kell azonban megválaszolnom azt a kérdést, hogy mi a vágy és a **preferencia** fogalma közt a kapcsolat. Egyrészt fontos hangsúlyoznom, hogy a prefrencia is a tudatos jelenségek közé tartozik. Ez a tétel azon az alapon védhető, ha arra hivatkozom, hogy ha a preferenciarendezés két (vagy több) alternaív összehasonlításának műveletét jelenti, akkor ebben látnunk kell a tudatos mérlegelés jelenségét. A preferenciákat így nem a késztetések, hanem legalább a vágyak szintjén kell elhelyeznünk. Mivel később beszélni fogok másodrendű preferenciákról is, ezért a preferenciákat mind a vágyak, mind az értékek fogalmával kapcsolatba hozhatónak gondolom. Mindezt figyelembe véve a preferenciát én a vágyak (vagy a preferenciák) közti választásnak vélem elgondolhatónak. Ezt megfordítva mondhatjuk azt is, hogy a vágy (averzió) valamelyik irányú reifikált preferencia (↪54.o.), vagyis a preferenciarendezésbe vont, és az összehasonlítás eredményeként preferált vagy diszpreferált elemek kiválasztása.

Diszkrepancia alapú vágyelmélet

Charles B. Cross a diszkrepancia fogalmára alapozott vágyelméletet írt le modális logika segítségével [110]. A diszkrepancia fogalma azért került az elmélet központjába, mert az ember vágyai, majd az ezekből fakadó szándékai, majd az ezek eredményeként megvalósuló cselekvései azokból az eltérésekből, **diszkrepanciákból** magyarázhatók meg, amelyek az ember által célul kitűzött, jövőbeli és a jelenbeli, ténylegesen létező állapotok közti léteznek. Ha a most is létező és a célul kitűzött állapotok nem különböznének egymástól, akkor nem beszélhetünk diszkrepanciáról köztük, de nem beszélhetünk vágyról, sőt, tulajdonképpen célról sem. Mielőtt a logika bemutatásába kezdenénk, nézzünk meg egy példát, amin keresztül be lehet vezetni az alapfogalmakat is.

Vegyük a lehetséges jelen- és jövőbeni világállapotokat, és az ezeket leíró propozíciókat jelöljük A, B, C, D betűkkel. Ha egy állapotok fennáll, akkor a rá vonatkozó A propozíció igaz, ha az állapot nem áll fent, akkor a propozíció hamis: $\neg A$. Az állapotok közül emeljünk ki két típust: G jelölje azt az állapotot, amit a cselekvő szeretne elérni, vagyis legyen G a cselekvő célja (goal), B pedig jelölje a cselekvő vélekedését (belief) a világ éppen adott helyzetéről.[10] Az ágens célja és hite is állhat több – igaz vagy hamis – propozícióból. Cross példája egy ágensről és annak négy macskájáról szól. Ha az ágens csak akkor megy el otthonról, ha tudja, hogy mind a négy macskája bent van a házában, akkor a célállapotként rögzíthetjük a következőt.

$$G = (E \wedge B \wedge S \wedge C)$$

ahol E, B, S, C rendre jelölje azt, hogy az E(lza), B(erci), S(aci), C(irmi) nevű macska bent van a házban. Induljunk ki abból a helyzetből, amikor az ágens annyit tud, hogy Elza és Berci nincs a házban, míg Cirmi ott van, viszont nem tud semmit arról, vajon Saci bent van-e vagy sem. Ezt jelöljük a kezdőállapotnak.

$$B_0 = (\neg E \wedge \neg B \wedge C)$$

Mit kell tennie emberünknek, hogy elérje célját vagyis a célállapotba kerüljön? Egyfelől meg kell keresnie Elzát és Bercit, és be kell vinnie őket a házba, másfelől meg kell győződnie arról, hogy Saci bent van-e a házban. Ha Sacit megtalálja odabent, akkor vele nem kell többet tennie, egy lépéssel közelebb került célja megvalósulásához. Ha viszont Saci nincs a házban, akkor az ágensünknek vele is ugyanazt kell csinálnia, mint a másik két kint kóborló macskával: megkeresni, behozni őket. Ezt a két opciót kétféleképpen kell kezelni, kétféleképpen is érdemes megnevezni őket mint a diszkrepancia két típusát. A célok és hitek közti diszkrepancia (goal/belief discrepancy) két típusa az inkompatibilitás és a nem-teljesség. Az **összeférhetetlenség**, inkompatibilitás (incompatibility) fogalmával azt a helyzetet jelölhetjük, amikor a G célállapot és az B_i aktuális állapotról tudja az ágens, hogy valamely pontban, valamilyen mértékig nem nem egyezik meg. Példánkban ágensünk azon hite (tudása), hogy Elza és Berci biztosan nincs a házban, a célhoz képest kifejezhető inkompatibilitásnak felel meg. Ezzel szemben a **hiányosság**, nem-teljesség (incompleteness) fogalmával írhatjuk le azt a helyzetet, hogy a negyedik macskáról, Cirmiről nem tud semmit az ágens. Inkompatibilitás esetén az ágens késztetése (vágya)

[10] A 'belief' fordításaként használhatnánk még a hiedelmi állapot, bizonyossági állapot, hit(állapot), vélemény(állapot) kifejezéseket is. Stikisztikai okokból adott esetben használni is fogom őket, sőt olykor még a tudás terminusát is alkalmazni fogom.

az lesz, hogy megváltoztassa a célhoz képest eltérő állapotokat, vagy másként: a célállapotba hozza őket. Ehhez képest a hiányosság esetén – részben – más a teendő, hiszen első lépésként információt kell gyűjteni, meg kell szüntetni a tudáshiányos állapotot, majd a beszerzett új információ alapján vagy cselekedni kell (amit megint a cél- és aktuális állapot közti eltérés megszüntetésének vágya hajt) vagy nem kell semmit sem tenni (ha az derül ki a beszerzett információból, hogy nincs diszkrepancia). De kövessük tovább Cross példáját. Mi történik akkor, ha a gazda megtalálja odakint Elzát és behozza a házba? Megváltozik az aktuális állapot, amivel kicsivel közelebb kerülünk a célhoz. Jelöljük ezt egy újabb formulával.

$$B_1 = (E \wedge \neg B \wedge C)$$

Tegyük fel azt, hogy ezután az ágens felfedezi Sacit az ágy alatt, tehát már tudja, hogy Saci a házban van. Ebben az esetben megszűnik a hiányosság értelemében vett diszkrepancia a cél- és aktuális állapotok között, de még megmarad az összeférhetetlenség. Az újabb állapot a következőképpen jelölhetünk.

$$B_2 = (E \wedge \neg B \wedge S \wedge C)$$

Ekkor már csak egyetlen cica van odakint, Berci. Ha a gazda a keresésére indulna, és közben észlelné, hogy Cirmi ott áll a főbejáratnál arra várva, hogy a gazda kinyissa az ajtó, és ő kiszaladhasson rajta, akkor az ágens dönthet úgy, hogy a hátsóajtón át indul Berci felkutatására, hogy megakadályozza Cirmi kiszökését. Ezt a "csavart" Cross azért tette bele a történetébe, hogy erre hivatkozva később el tudja különíteni a diszkrepanciák két fontos szerepét, funkcióját. De most még maradjunk a macskás sztorinál, fejezzük be, már csak egy lépés van hátra. Ha ugyanis a gazda a kertben megtalálja a negyedik macskát és behozza a házba, akkor ezzel olyan állapot áll elő, ami összehangba kerül az eredeti céllal. Ha ugyanis Bercit is behozta a gazda, akkor ezzel megint új állapot keletkezett.

$$B_3 = (E \wedge B \wedge S \wedge C)$$

Ez a B_3 állapot viszont minden összetevőjében megegyezik a G célállapottal, tehát kijelenthetjük, hogy az ágens elérte célját. És a bemutatott történet alapján már rámutathatunk arra is, hogy Cross hogyan javasolta elkülöníteni a cél/hit-diszkrepancia két szerepét, funkcióját egymástól. Szerinte a diszkrepancia lehet:

a) valami, amit egy racionális ágens megpróbál megszüntetni;
b) valami, amit egy racionális ágens megpróbál nem létrehozni.

A célkövető magatartás leírásakor mindkét funkcióra figyelemmel kell lennünk, mert csak így lehetünk képesek a cselekvések belső értelmét megragadni. A célkövetés folyamatában nemcsak arra kell figyelni, hogy lépésről lépésre megszüntessük az aktuális és a célállapot közti eltéréseket, de arra is ügyelni kell, hogy mindeközben ne hozzunk létre újabb diszkrepanciákat. Következő lépésként bevezetünk néhány operátort, amelyek segítségével majd megragadhatjuk a célkövetés folyamatának minden fontos mozzanatát. Előfordulhat, hogy a pontosság miatt olykor nyakatekertebben kell fogalmaznom, de törekszem a minél természetesebb nyelvi megfogalmazásra. Cross azt javasolja, hogy először rögzítsük azon propozíciók A együttesét, amellyel leírhatjuk az ágens célját. Ebben az esetben mondhatjuk azt, hogy egyrészt az ágens akkor éri el a célját, ha az A-ba tartozó propozíciók igazak, vagyis az általuk leírt (rész)állapotok mindegyike fennáll, másrészt

az ágens törekszik A megvalósulására, kívánja, akarja A-t. Az A-val kapcsolatban kifejezhetjük azt is, hogy az ágens hiszi-e, tudja-e, hogy A igaz, vagy A nem igaz, vagy az égens nem tudja sem azt, hogy A igaz, sem azt, hogy hamis. Végül, az elméleti teljesség kedvéért feltételeznünk kell, hogy előfordulhat az az eset is, amikor A közömbös az ágens számára. Ilyenkor az ágens nem vágyik sem A-ra, sem $\neg A$-ra. A hit- és célállapotok közti eltérő viszonyok más és más cselekvésre késztetik az ágenseket. Az egyezés esetén nincs szükség semmiféle cselekvésre, összeférhetetlenség esetén a cselekvő arra törekszik, hogy a vágyott A állapot bekövetkezzen, tehát megpróbálja előidézni azt valahogy. Amikor a hiányossági helyzet áll elő, akkor az ágens arra érez késztetést, hogy meggyőződjön A logikai értékéről vagyis megtudja, hogy A igaz-e vagy hamis. Dennis W. Stampe javaslatával [509] egyetértve Cross kijelenti, hogy a vágy fogalmához lényegileg hozzátartozik, hogy a vágy indokot szolgáltat a vágyott állapot létrehozására. Ha vágyunk valamire, akkor késztetést érzünk a vágy tárgyát képező állapot előidézésre. Márpedig – jelenti ki Cross – a hiányossági diszkrepancia esetében ez nem áll fent (az összeférhetetlenség esetében persze igen), így a hiányosságot mint diszkrepanciát nem kapcsolhatjuk összes a vágy fogalmával.

az ágens	az eset neve	jele	szükséges cselekvés
vágyja A-t és véli A-t	egyezés	$\oplus A$	nincs, mert beteljesült a vágy
vágyja A-t és véli $\neg A$-t	összeférhetetlenség	$\triangle A$	előidézni A-t
vágyja A-t, nem tudja, A vagy $\neg A$	hiányosság	$\triangledown A$	megtudni, hogy A vagy $\neg A$
nem vágyja sem A-t, sem $\neg A$-t	közömbösség	$\neg A$	nincs, mert nincs vágy

A négy modális fenti operátort alapul véve Cross kidolgozott egy teljes és helyes modális logikai nyelvet a vágy fogalmára, megadta a nyelv szemantikáját, a nyelv sémáját, következtetési szabályait. Ezek bemutatásától itt eltekintek. Azt is csak megemlítem, hogy Berislav Žarnić egy 2011-es tanulmányában javaslatot dolgozott ki a crossi vágylogika és az imperatív (deontikus) logikai közötti kapcsolat felépítésére vonatkozóan [595].

A vágyak (késztetések) egy további fontos felosztását végezte el Elliott Sober és David Sloan Wilson [503, 217-222], Soberék elkülönítették a **végső vágyakat** (ultimate desire), illetve az **instrumentális vágyakat** (instrumental desire) egymástól. Ez a szembeállítás elég nyilvánvalóan megfeleltethető a cél, illetve eszköz jellegű vágyak kettősségének, ezért a továbbiakban az **eszközvágy**, illetve **célvágy** kifejezéspárt is használni fogom. Soberék érdeme, hogy – relatíve – pontos definíciókat is adnak ezekre a fogalmakra. Pontosabban az instrumentális vágy fogalmát definiálták, a végső vágy fogalmát inkább primitív kategóriaként kezelték.

> „(1) S M-t kizárólag az E-re irányuló vágya eszközeként akarja akkor és csak akkor, ha (a) S akarja M-t, (b) S akarja E-t és (c) S csak azért akarja M-t, mert úgy hiszi, hogy M elérésével elősegíti (promotálja) E elérését.” [503, 217]

A meghatározásban szereplő rövidítések az S szubjektumra, az M eszközvágyakra, az E célvágyakra vonatkoznak. További erőssége még Sober és Wilson elméletének az, hogy a vágyfogalmakra támaszkodva két tézist is kimondanak, amelyek – bizonyos értelemben – a vágyakkal kapcsolatos előrejelzéseknek is tekinthetők. Íme a két tétel.

> „(2) ha S csak mint E elérésének eszközeként vágyik M-re, akkor ha S arra a belátásra jut, hogy M nem biztosítja E elérését, akkor S M-re irányuló vágya elmúlik, de továbbra is akarja E-t és megpróbál találni új eszközöket E eléréséhez.” [503, 218]

„(3) Ha S csak mint E elérésének eszközeként vágyik M-re, akkor ha S eléri E-t, de nem éri el M-t (pontosabban: ha S azt hiszi, hogy ez nem történt meg), akkor S nem vágyik többé M-re; azonban ha S eléri M-t, de nem éri el E-t, akkor S tovább vágyik E-re." [503, 219]

Soberék fontos észrevétele, hogy az eszközvágyainkat könnyebben feladjuk, míg a célvágyainkat már nehezebben. Azt is mondják még, hogy a fenti két tétel prediktív, vagyis segítségükkel következtetni tudunk jövőbeni viselkedésekre. Úgy lehet értelmezni ezeket, mint a diszpozíció fogalmát. Itt is, ott is tudunk mondani valamit a várható dolgokról. Viszont az is igaz, hogy egyik tétel sem retrodiktív, azaz nem tudunk semmit mondani arról, mi miért történt, miért lett olyan valaki, valami a múltban, amilyen lett [503, 221].

Másodrendű vágy: érték

Egyszer, amikor Diogenész Kraneioszban sütkérezett a napon, Nagy Sándor kíséretével épp arra haladva megállt előtte és azt mondta neki: – *Kérj tőlem, amit akarsz!* – mire Diogenész csak annyit felelt – *Ne álld el előlem a napot!* Évezredekkel később az amerikai Clay Intézet hét, régóta ismert matematikai probléma megoldásának ösztönzésére megalapította a Millennium-díjat. Bármelyik probléma megoldásáért egy-egymillió dollárt ajánlottak fel. Grigorij Perelman, orosz matematikus 2002-ben egy weboldalon közzétette a megoldást, amit a szakma idővel érvényesnek minősített, ezért 2010-ben a Clay Intézet felajánlotta Perelman számára a Millennnium-díjat és a vele járó egymillió dollárt. Perelman nem fogadta el egyiket sem. – *A díj teljesen irreleváns számomra.* – nyilatkozta – *Ha a bizonyítás jó, nincs szükség külön elismerésre.* [254]

Két történet, amelyekhez többféleképpen viszonyulhatunk. Többen lehetnek azok, akik értetlenkednének a történetek hallatán, de talán akadnak olyanok is, akik szeretnének a főszereplők helyében lenni. Akik szeretnének olyanok lenni, mint ők. Akik vágynak arra, hogy úgy ne vágyjanak a földi hívságokra, ahogy azt Diogenész vagy Perelman tette mintázta. Bármennyire is ritka, szinte életidegen a két példa, pont a szokatlansága miatt alkalmas arra, hogy segítségükkel könnyebben felismerhessük azt a nagyon fontos összetevőt, ami a két meséről szóló viszonyulásban közös, és amit megtalálhatunk sok más történetmesélésben is. Amikor erkölcsi tanulságokat fogalmazunk meg egy történet kapcsán, akkor a szereplők cselekvéseire, illetve az azok mögött meghúzódó vágyakra, késztetésekre vonatkozóan fejezzük ki az értékeléseinket. De mit is jelent ez az értékelés? Azt a vágyat (vagy averziót), amit a történetben szereplők cselekvéseivel, késztetéseivel (vágyaival vagy averzióival) kapcsolatban érzünk magunkban, és erre az érzésre a tudatunk reflektál valahogyan. Amikor erről van szó, akkor azt kell mondanunk, hogy a történetek értékelésekor másodrendű vágyak (averziók) dolgoznak bennünk, hiszen ilyenkor a történetekben levő késztetésekre irányuló késztetésekről kell beszélnünk.

A késztetésekkel kapcsolatos másodrendű fogalmakat Harry Frankfurt és Richard Jeffrey emelte be a tudományos diskurzusba a hetvenes években.[11] Frankfurt a **másodrendű vágy** és a **másodrendű akarat** fogalmára építette fel a személyiség fogalmának meghatározását, amit ő az ember legfőbb sajátosságának tekintett [175], míg Jeffrey bevezette a **másodlagos preferencia** fogalmát [259], amit a döntéselmélet keretein belül alkalmazott [259]. Frankfurthoz és Jeffreyhez képest David Lewis annyiban továbblépett,

[11] Tudomásom szerint Harry Frankfurt elméletéről magyar nyelven Balázs Zoltán írt először ismertetőt [31], és a 2013-ban megjelent válogatáskötetben több fontos tanulmány is megjelent a témával kapcsolatban [438]. Richard Jeffrey magyar nyelvű recepciójáról nem tudok.

hogy egy, az ember értelmezése szempontjából kulcsfontosságú kategóriát úgy definiált, hogy azt a másodrendű vágy fogalmával azonosította [501]. Lewis felfogása szerint az **érték** nem más, mint vágyra vonatkozó vágy, vagyis az értékeink a másodrendű vágyaink.

A frankfurti, jeffrey-i nézeteket szélesebb, a lewisi elképzelést szűkebb körben felkapták a vágyakkal, értékekkel foglalkozó diskurzuson belül. Charles Taylor a szelf, az identitás fogalmának meghatározásakor például Frankfurt másodrendű vágy és a különböző értékeléstípusok fogalmaira támaszkodik [540]. Amartya Sen is egyetértőleg utal a másodrendű preferencia (vagy ahogy ő nevezi: a metapreferencia) fogalmára miközben hivatkozik Frankfurt, Jeffrey elméleteire [482], ahogy Albert O. Hirschman is úgy ír kétféle preferenciaváltozásról egy cikkében, hogy közben utal Frankfurtékra [236]. Vannak persze olyanok is, akik kritizálják a másodrendű vágyfogalmak használatát. Gilbert Harman egy írásában áttekinti Harry Frankfurt, David Lewis és David Copp [106] javaslatait a másodrendű vágyak értelmezési lehetőségeiről, és mindegyik nézetet elutasítja [223].

A Frankfurt-recepció elsősorban a személyiség, a szabad akarat, a felelősség problémája köré épült fel, de ez a témakör a most tárgyalt szempontból nem igazán érdekes. Jeffrey másodrendű preferencia fogalma már sokkal inkább a mostani gondolatmenetbe illeszthető, bár csak akkor, ha pontosabban reflektálunk a másodrendű preferencia fogalmában meghúzódó lényegi mozzanatra. Hogy érthető legyen, mire gondolok, idézek egy rövid – más témakörben mozgó, és önmagában nagyon értékes – szövegrészt a politikai cselekvések logikájáról.

> „. . . a politikai cselekvés különbözni látszik a magánfogyasztók és termelők piaci viselkedésétől. Amellett is érvelhetünk, hogy a politikai cselekvés természeténél fogva publikus, abban az alapvető értelemben, hogy nemcsak egyéni nyereményeken alapul, hanem a világ lehetséges jövőbeli állapotaira értelmezett preferenciákon, amelyek magukban foglalhatják mások nyereményeit, valamint szabályokra és intézményekre vonatkozó preferenciákon, amelyek garantálhatják ezeket a nyereményeket.” [114, 567]

Mivel sem az idézet, sem az azt tartalmazó írás egésze nem a preferenciákról szól elsődlegesen, ezért nem kritikát fogalmazok meg vele szemben, csak mint szemléltetésre alkalmas szövegként használom arra, hogy rámutathassak, hogy a szövegrészben reflektálatlanul használt másdorendű preferencia fogalom létezésére. Amikor mások nyereményeiről van szó, akkor azok nyilvánvalóan kapcsolatba hozhatók mások preferenciáival, ezért a fenti idézetben leírt politikai cselekvéseket úgy lehet értelmezni, mint mások preferenciáival (nyereményeivel) kapcsolatos preferenciákat, melyek így más másodlagosak.

Ahhoz, hogy megértsük, mit jelentenek és hogyan működnek a másodrendű vágyak, az értékek bennünk, abból kell kiindulnunk, hogy a vágyaink a tudatba került belső és elsődleges késztetéseink (vagy kicsit pongyolábban elsődleges preferenciáink). Ameddig még csak arról van szó, hogy a vágyaink megjelennek a tudatunkban, addig még nyugodtan mondhatjuk, hogy ezek reflexió előtti (prereflektív) intencionalitások. Az elsődleges vágyaink, preferenciáink nem mérlegelések, nem tudatos döntések eredményei, hanem a bennünk levő – részben biológiailag örökölt, részben életünk során tanult, elraktározott – késztetések megnyilvánulásai a környezetünk ingerei adott válaszokként. Nagyon sok esetben kerülünk olyan esetbe, amikor nem egy, hanem több és többféle késztetés, vágy, preferencia jelenik meg egy időben a tudatosság szintjén. Ilyenkor ezek szinte mindig – kisebb-nagyobb – konfliktusba kerülhetnek egymással, és az ilyen ellentmondásokat a tudatos reflexió, a tudat mérlegelése oldhatja meg (próbálhatja feloldani). A cselekvést tehát nagyon sok esetben nem pusztán a vágyak vezérlik, hanem az egyszerre ható vá-

gyak (és averziók) együttese, valamint az ezekre reflektáló tudat kognitív megfontolásai. Ha pedig így értelmezzük a mérlegelés, a reflexió és a vágyak kapcsolatát, akkor lehet igazán megérteni (és elfogadni), miért javasolta David Lewis, hogy az érték fogalmát definiáljuk úgy, mint a vágyra vonatkozó vágyat. Amikor a vágyak (preferenciák) konfliktusba kerülnek a tudat szintjén, akkor maguk a vágyak válnak (pontosabban azok is) a mérlegelés tárgyává. Amikor a versengő vágyaink között mérlegelünk, amikor reflektálunk rájuk, akkor arról döntünk, hogy melyik vágyunkat soroljuk előbbre a többinél. Ez ugyanúgy egyfajta preferencia kifejezését, mégpedig az aktuális vágyak közti preferencia (vágy) vagy másként az aktuális vágyak közti prioritás kifejezését jelenti. A másodrendű vágy (preferencia) tehát mindig reflektált vágy (preferencia) is egyben. És ez – David Lewis-szal egyetértve – valóban megfelel az érték fogalom tartalmának.

2.4 Vélekedés, vágy, szándék

A társadalmi cselekvés meghatározó minősége az intencionalitás, ez utóbbi kifejezést azonban könnyű összetéveszteni az intenció fogalmával, sokan és sokszor egymással felcserélhető módon használják a két terminust. Pedig az intenció mint szándék az intencionáltságnak csak egy típusa, és vigyázni kell arra, hogy a két terminus alaki közelsége nehogy áthallásokat, fölösleges és szándékolatlan konnotációkat teremtsen.

> „Az *intencionalitás* meglehetősen balszerencsés szóalkotás, és mint annyi más szerencsétlen szót a filozófiában, ezt is a németül beszélő filozófusoknak köszönhetjük. A szó arra utal, hogy az intencionalitás az irányítottság értelmében valamiféle kapcsolatban áll a 'szándékkal', olyan értelemben például, hogy ma este moziba szándékozom menni. (A német nyelvben ez nem okoz problémát, mert az *Intentionalität* másképp hangzik, mint az *Absicht*, a mozilátogatás szándékának hétköznapi értelmében vett szándék.) Úgyhogy ne feledjük, az angolban a szándék csupán az intencionalitás számos formájának az egyike." [469, 89]

Searle figyelmeztetése természetesen nemcsak az angol esetében érvényes, a magyarban (és még nyilván sok más nyelvben) ugyanúgy megszívlelendő. Ennek a searle-i kritikának persze nem volt sem azonnali, sem széleskörű hatása. G.E.M. Anscombenak az intencionális emberi cselekvés elemzésében mérföldkőnek számító, *Intention* című könyvére nem is lehetett hatással, mivel ez a könyv 1957-ben jelent meg először [16]. Searle kritikája ettől még vonatkoztatható lenne Anscombe írására is, mert a könyv az intencionális cselekvés értelmezését célul tűzve az 'intenció' fogalmát az intencionalitás mint szándék jelentés mellett, tehát az intencionalitás egyik típusaként fogta fel.[12] Ez a "pongyolaság" azonban itt nem zavaró. Anscombe egy kérdéssel indítja értekezését.

> „Egy ember azt mondja, 'Sétálni fogok', és mi azt mondjuk erre, 'ez egy szándék kifejezése volt, nem egy előrejelzés'. De honnan tudjuk ezt?" [16, 6]

Megijedni, reszketni, összerezzenni, zihálni – ezek nem intencionális jelenségek vagyis nem cselekvések. Elmosogatni, olvasni, taxit hívni – ezek intencionális jelenségek vagyis cselekvések. Anscombe ezzel a szembeállítással kezdve rögtön a könyve elején olyan

[12] Az intenció és intencionáltság terminusait egyébként még Husserl is szinoním módon kezelte olykor, bár ő ilyen esetekben az intenciót az intencionáltság, intencionális állapot értelmében, tehát nem a szándék jelentésére leszűkítve értelmezte.

erős állításokat tesz, amelyeket azóta is sokan és sokat vitatnak a cselekvés filozófiájával kapcsolatos későbbi fejtegetésekben. Anscombe szerint az előrejelzésekhez, predikciókhoz tényeket, **bizonyítékokat** (evidence) rendelünk, **okokat** (causes) keresünk, az intenciókhoz, szándékokhoz cselekvésre vonatkozó **indokokat** (reason), és szerinte mindennek lehet oka, de csak az intenciónak lehet indoka. Ebből következően az intencionális cselekvések leírásához, megértéséhez mindig indokakat kell keresnünk és találnunk.

Anscombe szerint szándékolt cselekvések azok, amelyekre értelmesen feltehető a – bizonyos értelemben vett – 'Miért?' kérdés. Ha pedig ez a kérdés értelmesen feltehető, akkor a rá adott választ minősíthetjük a cselekvés szándékának. Nyilván tisztáznunk kell, hogy mit jelent ez itt a "bizonyos értelemben feltett »miért«" kérdés. Anscombe ezt több lépésben fejti ki. Először azt rögzíti, hogy milyen értelemben nem használható (mert nem nem tehető fel) a 'Miért?' kérdés. Ehhez két fogalmi dimenzióban elkülöníti az akaratlagos, illetve nem-akaratlagos tudást egyfelől, valamint a megfigyelés által ismert, illetve a megfigyelés nélkül tudott dolgokat másfelől. Az első szempont szerint vizsgálva a cselekvéseinket, Anscombe – egy nagyon termékenynek bizonyult gondolatként – felveti, hogy a cselekvéseinkre többféle leírást is adhatunk, és a cselekvés mindig **valamely leírás szerinti cselekvést** jelent (acting under a description), míg a többi leírás szerint nem számít cselekvésnek.[13] A cselekvések leírásaival kapcsolatos diskurzus egyik kedvenc példáját idézve: amikor a mutatóujjammal felfelé mozdítom a villanykapcsoló billenőlapját (1), akkor felkapcsolom a lámpát (2), amivel fényt csinálok a szobában, hogy olvasni tudjak (3). De ezeken túl elképzelhetők más értelmezések is: például az, hogy a lámpakapcsolással jelzést adok valakinek (4), vagy – villanyszerelőként – egyszerűen csak ellenőrzöm, hogy jól kötöttem-e be a kapcsolót (5). Mindegyik leírás egy-egy lehetséges és adekvát leírása, értelmezése lehet annak, amit tettem, a kérdés, hogyan tudhatunk meg többet erről.[14] Ez azért fontos, mert ha megkérdezik tőlem a negyedik leírás alapján, hogy miért adtam fényjelzést a lámpa felkapcsolásával, akkor erre felelhetem azt, hogy nem tudtam, hogy ezt teszem. Ilyen esetben nyilván kizárhatjuk azt, hogy az adott leírás mentén szándékolt cselekvésről lenne szó. A szándékom az volt, hogy fényt csináljak a szobában azért, hogy lássak, de nem volt szándékom, hogy valakinek jelzést adjak ezzel, és ha nem akartam ezt, akkor nem is tudhattam erről.[15] Ha valamiről nem tudok, akkor az nem lehet akaratlagos, akkor az arra vonatkozó 'miért' kérdésre nem válaszolhatok értelmesen, tehát nem minősíthetjük szándékoltnak azt a cselekvés(leírás)t. Megfordíthatjuk azonban a dolgot. Mondhatjuk egyszerűen azt is, hogy nem akartam valamit megtenni (például jelzést leadni a lámpával). Ehhez persze az szükséges, hogy meg tudjuk különböztetni az **akaratlagos** (voluntary), illetve **nem akaratlagos** (involuntary) cselekvéseket egymástól. Ezen a ponton azonban megfontoltnak kell lennünk, mert könnyen beleeshetünk az üres, önmagában forgó érvelés csapdájába (ha azt keressük, mi az, ami nem szándékolt, akkor nem felelhetünk azzal rá, ahogy mindaz, ami nem akarat-

[13] Ez az anscombe-i észrevétel nagyon fontos összetevője, vizsgálati eszköze lesz később kibontakozó narratív tudománynak [203], [204].

[14] A jelenség szemléletesen jelzi (és egyben vissza is igazolja) azt a tényt, hogy a cselekvés szubjektív értelme erősen függ a "választott" szemponttól vagyis az intencionális aspektustól (↳143.o.).

[15] Ritka, de létező példaként hivatkozhatunk arra az esetre, amikor egy tévéstáb számára a robbantómester bemutatta, hogy kell majd, egy későbbre tervezett időpontban a bontásra ítélt épületet felrobbantani. Abban a – téves – tudatban, hogy nincs áram alatt a rendszer, elfordította a robbantókart, mire az épület felrobbant, mivel minden élesítve volt (youtube.com/watch?v=z8HKkF2GsBQ). A robbantómester nyilván mondhatta, hogy ő el akarta forgatni a robbantókart szemléltetésként, de nem akarta idő előtt felrobbantani az épületet.

lagos). A megoldás kereséséhez Anscombe javasolta, hogy vizsgáljunk meg néhány példamondatot, amelyek cselekvések leírásaként tekinthetők [16, 13].

1)	A belünk perisztaltikus, összehúzódó mozgása.
2)	Az a furcsa rándulás, rángás, amit olykor az ember teste végez elalvás során.
3)	"Önkéntelenül hátrahőkölve visszahúzta a kezét."
4)	"Az a nem tervezett haszon, amelyhez egy ártó szándékú ütéssel jutottam."

A példákat arra használja Anscombe, hogy rámutathasson arra, merre kell továbblépnünk a szándékolt cselekvések leírásában. Az utolsó két példát például azért vethetjük el, mint amelyek a szándékos cselekvést szemléltetnék, mert azokban a szándékosság ugyan fellelhető, de nyilvánvalóan más akarati tartalomra vonatkozóan. Mivel ismerjük a cselekvésünk valódi indokát, tudjuk annak az okát is, ami nem-akaratlagosan végeztünk el. A negyedik mondatban például az ütés akartam (annak megadhatjuk az indokát is), de az ezáltal előált, nem várt hasznot nem akartam magamnak (amihez tehát nem tudunk indokot rendelni). A táblázat négy leírása közül a második mondatot más alapon lehet kizárnunk a szándékolt cselekvések lehetséges leírásai közül. Ezt azért tehetjük meg, mert ez a mondat egy egyszerű testi mozgás fizikai leírásaként értelmezhető, amiben nyilvánvalóan nincs a szándékoltságnak helye. Ez persze igaz az első leírásra is, de a két eset között mégis találhatunk egy fontos különbséget. Ennek megértéséhez azonban szükségünk van új fogalmak bevezetésére, mégpedig arra a kettősségre, miszerint vannak dolgok, események, amelyekről megfigyelés révén szerzünk tudomást, és vannak, amelyekről megfigyelés nélkül. A megfigyelés fogalma azért fontos itt, mert a megfigyelés során mindig okokat és ezekre szolgáló bizonyítékokat keresünk, és – ami még fontosabb – megfigyeléssel csak bizonyítékokat és okakat találhatunk, nem pedig a cselekvések indokait jelentő szándékokat. Ez azt is jelenti egyben, hogy amiről megfigyelés útján szerzünk tudomást, azt sosem sorolhatjuk a szándékolt cselekvés fogalma alá. Saját mozgásaink világából **megfigyeléssel szerzett tudás** lehet például az, amit a perisztaltikus bélmozgásunkról szerzünk be valahogyan. Ezzel szemben áll a másik eset, a **megfigyelés nélküli tudás**, amit az jellemez, hogy a jelenségek magyarázatához nem keresünk bizonyítékokat, okokat. Ilyenek a saját testi mozgásaink "cselekvései", történései, amikor nincsen semmiféle érzékelés (történés), amely alapján tudunk a dolgokról. Arról például, hogy a saját lábunk, kezünk, fejünk, testrészünk milyen helyzetben van, milyen az állapota, megfigyelés nélkül szerzünk tudomást. Ez utóbbi kategória azért fontos, mert Anscombe szerint ez alá tartozik a szándékolt cselekvések osztálya is (de nem azonos vele). Ezek alapján a példamondatok közül az elsőt azért zárhatjuk ki, mint ami nem szándékolt cselekvés leírását jelenti, mert egy megfigyelésen alapuló tudással kapcsolhatjuk össze.

A szándékolt cselekvések természetesen nem egy az egyben tartoznak a megfigyelés nélküli tudással jellemezhető jelenségek körébe, így további pontosításra van szükség. A megfigyelés nélküli tudás fogalmát egy lépésben még alábonthatjuk. Előfordulhat az, hogy megfigyelés nélkül tudunk a jelenségről, de a jelenség *okát* csak megfigyeléssel tudjuk meg. Példa rá az akaratlan izomrángás elalvás során vagy a reflexrúgás az orvosi vizsgálat során. A másik lehetőség pedig az, hogy megfigyelés nélkül tudunk mind a jelenségről, mind annak okáról. Anscombe azt a példát hozza erre, amikor egy krokodil hangjára hirtelen hátraugrunk. E kettősség bevezetése után már megfogalmazhatjuk, hogy milyen feltételek szükségesek a szándékolt cselekvésekhez. Ahhoz, hogy az ágens szándékoltan tegye meg az *a* cselekvést, szükséges, hogy:

- megfigyelés nélkül tudja, hogy *a*-t teszi, illetve
- megfigyelés nélkül tudja, hogy *a* megtételének mi az oka.

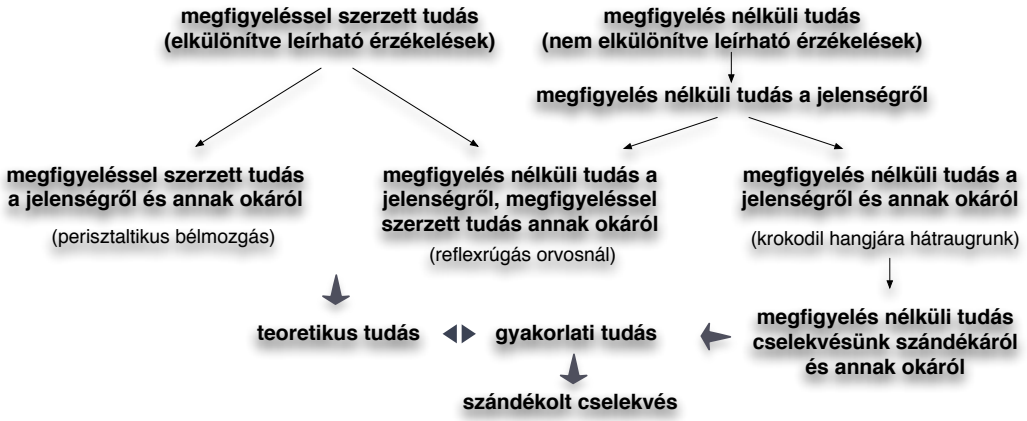
A tézis értelmezéséhez további pontosítások szükségesek. Először is tisztázni, pontosítani kell azt, hogy mit jelent a megfigyeléssel, illetve megfigyelés nélkül szerzett tudás fogalma. Anscombe válaszában abból indul ki, hogy a megfigyelések érzékszerveinken keresztül szerzett észlelések, tapasztalatok. Ez azt jelenti, hogy rendelkezünk egy **elkülönítve leírható érzékeléssel**, és ezt az érzéki benyomást függetleníteni tudjuk az azt kiváltó dologtól. Anscombe példája a liftben érzett "süllyedő" érzés, amit elkülöníthetünk a lifttől magától, hiszen ezt az érzést máskor is érezhetjük, mondjuk a vidámparki hullámvasúton ülve. Az elválasztható, elkülöníthető (separable) érzésekkel azonban szembeállíthatók a nem elválasztható, **nem elkülöníthető** (non separable) érzések. Ilyen például a reflexrúgás, aminek érzetét nem tudjuk elkülöníteni magától az azt kiváltó ütéstől. A kétféle érzés ismeretében Anscombe levonja azt a fontos következtetést, miszerint az elkülönítve leírható érzések a megfigyelés által tudott dolgokat jellemzik, míg a nem elkülönítve leírható érzések a megfigyelés nélkül tudott dolgokat. Ez azt jelenti, hogy az olyan eseteket kizárhatjuk a szándékos cselekvések köréből, amelyekhez ugyan nem elkülöníthető érzések kapcsolnak minket, viszont az okukat megfigyeléssel tudjuk csak beazonosítani.

A tisztánlátás érdekében le kell még választani pár értelmezési lehetőséget a szándékolt cselekvés fogalmáról. Amikor a szándékolt cselekvéseket akarjuk megragadni, amihez a megfelelő indokokat keressük, akkor meg kell különböztetni két dolgot egymástól: a mentális okot, a motivációt egyik oldalon, illetve a szándékot, a cselekvésre való indokot a másik oldalon. A **mentális ok**, mentális okozás (mental cause) más, mint a motiváció és más, mint az érzelmek tárgya. Anscombe elmesél egy történetet, amelyben egy kisgyereknek egy vörös ruhadarabra azt mondják, 'az csak egy darab szatén' (bit of satin), de a gyerek úgy érti, hogy 'egy darab sátán' (bit of Satan), ami miatt rémálmai lesznek. Anscombe szerint a félelem tárgya a ruhadarab volt, a félelem oka a félreértett mondat. Természetesen az érzelem oka és tárgya lehet ugyanaz. Ha éjszaka megpillantunk egy maszkos embert az ablakunk alatt, akkor az ijedtségünknek mind a tárgya, mint az oka az adott látvány (a félelmetes maszk) lesz. Ha egy cselekvés magyarázatában mentális okokat, motivációkat találunk csak, akkor azokkal sosem a cselekvési szándékokat tudjuk leírni, tehát az ilyen leírásokat nem tekinthetjük alkalmasnak. Mint ahogy a vágy fogalmát sem használhatjuk mindig a szándék leírásaként. Sokszor mondjuk ugyan, hogy azért csináltam meg valamit, mert egy vágyam erre késztetett, de az intenciót nem lehet mindig a vágyra visszavezetni. Anscombe példája erre az, amikor kopogtatnak az ajtón, és odamegyek, kinyitom, akkor ezt – legtöbbször – nem írhatjuk le a vágy kategóriájával.

A cselekvéseink szándékoltságában nem a – alapvetően a megfigyelésen keresztül megítélhető – sikeresség a döntő tényező, sőt, mondhatjuk azt is, hogy a sikeresség szempontja egyáltalán nem lényeges. Amikor ki akarunk tölteni egy online űrlapot, és a kitöltendő mezők között egymás után lépegetve egyszer véletlenül az egyik adatmező mellé kattintunk, vakon kezdjük el beírni az adatainkat, majd az írást befejezve felnézünk, és azt látjuk, hogy nem jelent meg semmi a képernyőn, vagyis a gép (vagy egy, a monitort figyelő másik ember) úgy érzékelte, hogy nem írtunk, nem csináltunk semmit, attól még mi ugyanazt gondoljuk magunkról, mintha sikeresek lettünk volna, ti. azt, hogy beírtunk egy sor adatot, vagyis az adatbeírás szándékolt cselekvését hajtottunk végre. Ezen a megfigyelés nélküli tudásunkon nem változtat semmit a "sikertelenség" ténye. Anscombe több

hasonló példát hoz erre [16, 50-51, 53-54, 82]. Arra, hogy nem a megfigyelés a megfelelő eszköz a szándékoltság "felderítésében", furcsa nézetből ugyan, de rávilágít Anscombe "bevásárlólistás" példája is [16, 56]. Ebben a feleség ír listát arról, hogy mit kell hozni a boltból, a férj elmegy a listával a boltba, miközben egy detektív is követi őt, és pontosan feljegyzi, hogy mit vásárolt a férj. A férj szándéka megfigyelés nélkül tudja, mit kell vennie (ami a példában kivetült a kezében tartott papírcetlire), a nyomozónak csak megfigyeléssel van módja a férj szándékát felderíteni. A különbség látványosan megmutatkozik akkor, amikor hibát követnek el. Ha a férj listája és a vásárolt termékek közt valami eltérés van, akkor a férj nem korrigálhat úgy, hogy kijavítja a listát a termékeknek megfelelően, akkor bizony vissza kell mennie a boltba, hogy pótolja a hiányzó terméket. Ezzel szemben a detektív nyugodtan megteheti ezt, vagyis kijavíthatja a megfigyelési listáját, és helyre áll a világ és a világ leírása közti megfelelés.[16]

Említettem már azt az anscombe-i tézist, hogy az intencionalitást nem minden leírás alatt rendelhetjük a cselekvéshez/jelenséghez, csak bizonyos leírások szerint beszélhetünk intencionális cselekvésről. Mivel a megfigyelés okok feltárására alkalmas bizonyítékot "eredményez", ezért kell kizárnunk abból a körből a megfigyeléseken alapuló értelmezéseket. Maradnak azok az esetek, amikor a cselekvő megfigyelés nélkül szerez tudomást saját szándékairól. Ezt nevezi el Anscombe **gyakorlati tudásnak**. A gyakorlati tudás az tudás, amikor a saját szándékomat nem megfigyelem, hanem arról megfigyelés nélkül tudok. A megfigyelés – bizonyos értelemben – passzív tevékenység, amire a kontemplatív, teoretikus, spekulatív tudás jellemző, a gyakorlati tudás viszont aktív. A kétféle tudást nem úgy kell elválasztani egymástól, mintha azok a cselekvéseinkre vonatkozó különböző leírások lennének. A gyakorlati és a spekulatív tudás nem ugyanannak a dolognak a tudása kétféle leírás szerint, hanem inkább a tudás két módja ugyanarra a dologra vonatkozóan. Ezzel Anscombe eljut a gondolatmenete végpontjáig, amikor kimondja a tézist, hogy a cselekvésünkre vonatkozó, megfigyelés nélküli tudás a **gyakorlati tudás** (practical knowledge) az, amivel a szándékolt cselekvést megragadhatjuk. Az ismeretés lezárásként egy ábrába sűrítve feltüntetem, hogy milyen logikai útvonalon jutott el Anscombe a gyakorlati tudás és a szándékolt cselekvés meghatározásáig.



[16]Ez a példa egyben rávilágít arra a mai napig használt megkülönböztetésre is, amit Anscombe dolgozott ki, és ami arról szól, hogy a megfelelési irány vagy **igazodás irányultsága** (direction of fit) szerint érdemes különbséget tenni a kognitív állapotok mint vélekedések és konnatív állapotok mint vágyak között. Ezt a kettősséget az intencionalitás szerkezetéről szóló részben már beemeltem a modellbe.

Anscombe intenciófelfogása az oksághoz való viszonyában nem volt egyedi, sőt, inkább azt kell mondanunk, hogy sokak, talán a kor nézetét visszhangozta.

> „Wittgenstein, Anscombe és Ryle érveinek hatására egészen a hatvanas évek végéig az volt az általánosan elismert álláspont, hogy, mivel a szándéktulajdonítás és a cselekvés leírása konceptuálisan nem független egymástól, az intenciók nem lehetnek a cselekvés (hume-i) okai. A konceptuális kapcsolat érvét azonban D. Davidson ... meggyőzően cáfolta." [246]

A cselekvést illetően az ok és indok megkülönböztetésével foglalkozva Donald Davidson 'Actions, Reasons, and Causes' című tanulmányában (angolul: [121], magyarul: [120]) az indokot az ok egyik fajtájaként értelmezte. Ebben szembe ment G. Ryle, G.E.M. Anscombe, G.H. von Wright nézetével, akik megkülönböztették a cselekvések gyakorlati indokait az események okaitól – nagyjából azon az alapon, hogy egyfelől a cselekvésekhez szükséges intencionális állapotok nem események, és az okságot események között állapíthatjuk meg, másfelől az indokokra nem találhatjuk meg az okokra jellemző általános törvényszerűséget, nomologikus minőséget. Ez fordulat persze nem kizárólag Donald Davidson nevéhez köthető, szerepet játszott ebben John Searle, Michael E. Bratman is.

> „Az az okság-értelmezés ugyanis, amely az intencionális események és állapotok kettős, logikai-fiziológiai természetét egyetlen kauzalitás-fogalommal képes megragadni, áthidalhatja a test-tudat probléma dualista és monista megközelítései között húzódó szakadékot, s ezáltal új tartalommal töltheti meg magát a problémát is. Ennek a sajátos, dualista-monista konszenzusnak a lényege pedig az, hogy bár minden egyes intencionális esemény és állapot azonos egy fizikaival, logikai jellemzőik meghatározása mégsem természettudományos perspektívából történik. Azaz a kauzalitás-problematikára vetítve ez azt jelenti, hogy jóllehet minden egyes intencionális oknak és okozatnak megfelel egy fizikai jelenség, a közöttük fönnálló viszonyok nem testesítenek meg egyetemes törvényszerűségeket. Az intencionális események és állapotok közötti oksági relációk logikai jellemzői ugyanis mindig valamilyen individuális kognitív háttérre vezethetők vissza. ... Az intencionális eseményeknek és állapotoknak ezt a kettős vonását, a kauzális függést és a nomologikus függetlenséget együttesen megvilágító és Davidson által »szabálytalan« monizmusnak nevezett fölfogás tehát egy sajátos, »intencionalizált okság-fogalomra épül." [532, 531]

Az okság problémakörét a filozófusokra hagyva nem akarok itt azzal a kérdéssel foglalkozni, hogyan kell a cselekvési indokokat érteleznünk. Anscombe elválasztatotta az okokat és az indokokat egymástól, Davidson nem. Szerinte „ ... a cselekvés elsődleges indoka nem más, mint a cselekvés oka." [120, 8]). Anscombe-ék az egyik oldalon, Davidsonék a másikon. Egy cselekvéselmélet számára fontosabb az, amiben egyetértettek.

> „ ... amikor indok által magyarázunk egy cselekvést, akkor valójában újraleírjuk azt; az újraleírás egy bizonyos mintába helyezi a cselekvést, és pontosan ezt nevezzük a cselekvés magyarázatának." [120, 13])

A szándék modellje

A cselekvések magyarázatára hosszú időn keresztül alkalmazott vélekedés-vágy (Belief-Desire – BD) modellek kritikáját Michael Bratman végezte el [76], akinek fellépése óta már inkább vélekedés-vágy-szándék (BDI) modellről beszélnek. A dolgokat kissé leegyszerűsítve valóban mondhatjuk, hogy Bratman nem tett egyebet, mint a vélekedés és a vágy kategóriája mellé felvette harmadik mentális tényezőnek az **intenció**, (szándék,

akarat) fogalmát.[17] Kortársaihoz hasonlóan Bratman is úgy vélte, hogy az intenciónak két fajtáját különíthetjük el, a **jövőre irányuló** és a **jelenre irányuló szándékot**, és az utóbbi kifejezés jelentése megegyezik azzal, amikor azt mondjuk, hogy szándékosan cselekszünk valamit. Mielőtt tisztáznám, hogy mi a kétféle intenciófogalom jelentése, miben térnek el ezek egymástól, előtte szükség van néhány bratmani gondolat bemutatására. Felmerül rögtön a kérdés, hogy mi a vágy és a szándék fogalma közt a különbség. Bratman válasza erre az, hogy az elköteleződés (commitment) mozzanata jelenti a különbséget köztük. Az ugyanis nincs benne a vágyban, de benne van a szándékban. A szándék megváltoztatható, módosítható, támadható, visszavonható. A szándék olyan viselkedésbefolyásoló (conduct controlling) proattitűd (pro-attitide), aminek tehetetlensége (intertia) van és önmagában érveket szolgáltat a további mérlegelések számára.

A szándék (és elköteleződés) szükséges a koordinációhoz, mégpedig a tervek megvalósításához szükséges cselekvések koordinációjához. A **terv** (plan) a szándékhoz hasonló fogalom, de a terv mindig a jövőre irányul. Bratman szerint az embernek mint tervező lénynek nagy szüksége van a részleges, hierarchikusan szervezett tervekre.

> „Komoly pragramikus előnye lehet annak a stratégiának, amely a jövőre nézve részleges, hierachikusan strukturált terveket dolgoz ki. Egyfelől szükségünk van arra, hogy koordináljuk saját tevékenységünket akár magunkban, akár egymás közt vagyunk. Mindezt úgy kell megtennünk, hogy közben figyelembe kell vennünk a rendelkezésre álló információ feldolgozásával és mérlegelésével kapcsolatos korlátos kapacitásainkat. E korlátos kapacitás birtokában szükségünk van arra is, hogy előzetesen mérlegelhessük mindazon szempontokat, amelyek a későbbi viselkedéseinket formálják. Mindezek erős érvek a tervezés mellett. A világ azonban változik, és mi messze nem vagyunk abban a helyzetben, hogy jól előre is tudjuk jelezni ezt; a távoli jövőre vonatkozó, nagyon kidolgozott tervnek könnyen haszontalanná válnak, nem érdemlik meg a vesződést. Ilyenkor a részleges, hierarchikusan strukturált tervek jelenthetik a kompromisszumos megoldást.” [76, 29]

Két fontos követelményt lehet megfogalmazni a tervvel kapcsolatban. Egyfelől a tervnek az ember vélekedéseihez képest **belsőleg konzisztensnek** kell lennie (internally consistent), másfelől a tervnek **cél-eszköz koherensnek** kell lennie (means-end coherence). Ez a két minőség adja a tervek, szándékok, illetve a vágyak, értékelések közötti megkülönböztetés lehetőségét. A vágyakra nem érvényesítjük egyik kritériumot sem. A vágyra nem áll a belső konzisztencia követelménye. Két vagy több vágy lehet inkonzisztens bennem, de egy szándék vagy egy terv nem. Vágyhatok egyszerre meccsre menni és vacsorázni egyet este, de nem tervezhetem őket egyszerre végrehajtani. De a vággyal szemben nem várjuk el a cél-eszköz koherenciát sem, vagyis a vágyakhoz (mint lehetséges célokhoz) nem kell eszközöket rendelnünk, szemben a szándékokkal és tervekkel, amelyekhez igen.

Azon a minősítésen, hogy a szándékaink alapján kidolgozott terveink részlegesek, azt kell értenünk, hogy nem minden részletükben kidolgozottak. Ez abban az értelemben hátrány, hogy ezeket a kidolgozatlanságokat ki kell a jövőben dolgozni a majdani körülményektől függően, de – és ez az igazán lényeges – pont e kidolgozatlanságok teszik lehetővé azt, hogy igazodhassunk a jövőben változó feltételekhez. A szándékok, tervek fontos szerepet visznek mindvégig a mérlegelés során, de fontos hangsúlyozni, hogy a szándékok nem szolgáltatnak érveket a cselekvés számára. A cselekvési indokokat a vá-

[17] Az ‘intenció’, a ‘szándék’, az ‘akarat’ termnusát mind ugyanazzal a jelentéssel használom itt, viszont ezek jelzős változatát már nem. Az ‘intencionális’ jelentése különbözik a ‘szándékos’ vagy az ‘akaratlagos’ jelentésétől.

gyainknak és vélekedéseinknek köszönhetjük. A súlyozást, az irányok és intenzitások kijelölését a hiteink által formált vágyaink, nem pedig a szándékaink végzik.

> „A gyakorlati következtetésnek két szintje van: az előzetes szándékok és tervek felvetik a problémákat és egyfajta szűrőként működnek a potenciális megoldási lehetőségeken; a vágyak és hitek megfontolásokként szolgálnak az érvek súlyozásához a releváns és elfogadható kimenetek közötti mérlegelés során. Ez a kétszintű szerkezet fontos minősége annak a működésmódnak, ahogy a szándékok és tervek betöltik saját koordináció-támogató szerepüket, illetve segítenek nekünk elkerülni azt, hogy olyan időszeletelő ágenssé váljunk, aki állandóan a semmiből kezdi újra a számításait a mérlegelései során." [76, 35]

Nagyon fontos eleme még az intenció jelenségének az a tény, hogy a jövőre irányuló vélekedéseink sok esetben lényegi módon a jelenbeli intencióinktól függenek. Az vélekedésem, hogy délben az egyetemen leszek, azon a tudásomon alapul, hogy oda szándékozom menni [76, 63]. Az intenció tehát elkötelezettséget jelent a versengő vágyak közepette, ami egyrészt azért fontos, mert – legalább időleges – állandóságot jelent, másrészt lehetséges további terveket ráépíteni.

Szándékaink természetesen könnyen ellentmondásba keveredhetnek vélekedéseinkkel, de ebben a "konfliktusban" mindig sokkal inkább kritizálható a racionalitás szempontjából a szándékok és hitek közti inkonzisztencia, mint a köztük keletkező hiányosság, nem-teljesség. Bratman ezt nevezi az aszimmetria tézisének [76, 38]. A szándékok persze belsőek, így csak az introspekció révén adhatunk számot róla, míg a viselkedés külsőleg is megfigyelhető. A szándékok figyelembe vételével tudjuk csak megmagyarázni, hogy miért térnek el gyakran a cselekvések belső és külső értékelései. A belső értékelés figyelembe veszi (hiszen "ismeri") az elköteleződéseket és szándékokat, míg a külső megítélés eltekint ezektől. Szándékainkat, terveinket a mérlegeléseink során mindig újragondolhatjuk, és fontos kérdés, hogy mi a tervezés, a tervezés újragondolása és a racionalitás viszonya, de Bratman nézetét erről itt nem mutatom be. Szándékaink tehát támadhatók, fölülbírálhatók, de mindenkor biztosítani tudnak egyfajta állandóságot, biztonságot nekünk.

> „...korlátos adottságaink, illetve terveink azon szerepe miatt, hogy a jelen megfontolásainak érvényességét megbízható módon kiterjesztik a jövőbeni cselekvéseinkre nézve, számíthatunk arra, hogy a nem-újragondolásnak ez a hasznos jelensége élteti bennünk azt a hajlamot, hogy ne vizsgáljuk felül újra a már kész tervet kivéve persze, ha annak végrehajtásakor valamilyen problémával kerülünk szembe." [76, 66-67]

A bratmani megközelítés nagy újdonsága a szándékok és tervek jelentőségének felismerése és ezek beemelése a cselekvés modelljébe. Ehhez képest másodlagosak azok a kritikák, amelyek a Bratman utáni korszak mérnöki megközelítéseiben merültek fel. Stuart J. Russel és Devika Subramanian például a racionalitás – túlzottan erős – elvárását kritizálták a bratmani elképzelésben, és a "teljes racionalitás" helyett a jóval "gyengébb" **aszimptotikus korlátozott optimalitás** (APO – Asymptotic Bounded Optimality) fogalmának alkalmazását javasolták helyette [453]. E kritikát lehet megalapozottnak tartani, de a cselekvéselméleti modell építése szempontjából ezt nem tartom kardinális kérdésnek. Másfajta racionalitáselképzelések mentén is meg lehet, sőt, szerintem meg kell tartani az intenció (szándék, terv) fogalmának Bratman által javasolt kiemelt szerepét a BDI-modellen belül.

Philip R. Cohen és Hector J. Levesque logikai formalizmust dolgozott ki Bratman intencióelméletére, ami azóta is gyakran hivatkozott forrás a mérnökök körében [101].

Már eleve beszédes a cikk címe: "A szándék elkötelezett választás" ("Intention is choice with commitment"), ami mutatja, hogy a szerzők azt a bratmani tételt tették a kidolgozott logikájuk központjába, amely szerint a szándék az a választás, amely mellett az egyén elköteleződik. Mivel szándék, az intenció, az akarat terminusait mind ugyanezen értelmezés mentén javaslom használni, ezért ezeket a terminusokat mind ehhez a fogalomhoz kapcsolom a cselekvéselmélet modelljében. Cohenék felsorolták a szándék hát legfontosabb jellemzőjét, amit érdemes itt felidéznünk.

- a szándék szokásos esetben valamilyen feladatot jelent az ágens számára, aminek megoldásához valamilyen – cselekvési – utat kell találnia (ha egy másik városban kell lennem másnap, akkor meg kell oldanom, hogy jutok el oda, kocsival, vonattal, biciklivel, valamit mindenképp tennem kell)
- a szándék valamiféle megvalósíthatósági keretet jelöl ki a későbbi szándékok számára (a vágyak lehetnek inkonzisztensek, a szándékok nem: ha kocsival akarok elmenni a szomszéd városba, akkor nem akarhatok biciklivel is menni)
- az ágens "nyomon követik" szándékaik megvalósítása érdekében végzett lépéseiket, és ha kell alakítanak az eredeti elképzeléseiken (ha elindulok biciklivel, de az első sarokban kidurran a kerekem, visszafordulok, és kocsira váltok)
- az ágensnek hinnie kell abban, hogy a kitűzött célja megvalósítható, nem irányulhat a szándékunk tudottan lehetetlen dolgok elérésére (ha fél óra múlva a másik városban kell lennem, amit biciklivel lehetetlen ennyi idő alatt elérni, akkor a kocsi használatára irányulhat a szándékom, a bicikliére nem)
- az ágens nem hiheti azt, hogy nem fogja előidézni a cél megvalósulását (nem hihetem azt, hogy vadászrepülővel fogok átmenni a szomszéd városba, amikor tudom, hogy nincs leszállópálya se itt, se ott, nincs repülőgép, nem tudok repülőt vezetni)
- az ágens hisz abban, hogy bizonyos feltételek mentén megvalósíthatja a célját (hinnem kell abban, hogy ha van benzin a kocsimban, nem durran ki menet közben a kereke, tudok vezetni, van jogosítványom, akkor odajuthatok a másik városba)
- az ágensnek nem kell akarnia a szándéka minden várható következményét is egyben (ha tudom, hogy kocsi használatának lesz benzinköltsége, ami a bicikli esetén elmaradna, de mégis kocsival akarok utazni, attól még nem akarom direkt kiadni a benzinre a pénzt)

Cohen és Levesque a dinamikus logika (↬76.o.) alapján felépített logikai nyelvébe négy modális operátort vett fel kiindulásként, a vélekedés (BELIEF), a cél (GOAL), az éppen előttünk álló, illetve a már megvalósult eseményt (HAPPENS, illetve DONE) leíró kategóriáit, illletve primitív fogalomként felvették még a cselekvő ágens (AGENT) fogalmát, valamint a cselekvések és történések fölötteseként az esemény (EVENT) katagóriáját is. A vágy helyett azért használták a cél kategóriáját, mert utóbbi már kiemelkedett és tudatossá vált az emberben dolgozó ezerféle vágy közegéből. Értelmezésükben tehát a cél kiválasztott vágyat jelent. Az alapkategóriák segítségével aztán további 40-50 állítás, definíció, hipotézis megfogalmazásával, valamint pár tétel bebizonyításával le tudták fedni a BDI-modell legfontosabb fogalmait. Az időbeli modalitásokat mint a később (LATER), korábban (BEFORE) a DONE saját operátor és a 'végül is' (eventually) jelentés mellett értelmezett LEHETSÉGESSÉG modális operátor segítségével már ki lehet fejezni. "Köztes" fogalmakként szükség van az ágens kompetenciája (COMPETENT), a tartós cél (PERSISTENT-GOAL), vagy az egyedi vagy egyszerű esemény definiálására, és Cohenék elkülönítettek és bevezettek kétféle szándék (INTENTION) kategóriát is, amelyek közül – csak a szemléltetés kedvéért –

bemutatom az egyik definícióját.

(ZL.6) $\text{INTEND}_1(x,a) \stackrel{\text{def}}{=} \text{PERSISTENT-GOAL}(x, (\text{DONE}(x, \text{BELIEF}(x, \text{HAPPENS}(a)))?; a))$

szándékozni (HUN) – intend (ENG)

A fogalmak paraméterei között az x az ágenst, a tetszőleges cselekvést jelöli. A formulát pedig úgy értelmezhetjük, hogy az ágensnek szándékához szükség van elköteleződésre, amit részben a perzisztens cél fogalmával fejezhetünk ki. Ez azonban még kevés lenne a szándék megragadáshoz, mert ez elvileg teljesülhet véletlenek sorozatán keresztül. Emellett szükség van arra is, hogy az ágens célra irányulását is leírjuk valahogy. Ezt fejezi ki a formula három egymásba ágyazott eleme: az ágensnek előzetesen hinnie kell ($\text{BELIEF}(x, a')$) a célul kitűzött cselekvés megvalósulásában ($\text{HAPPENS}(a)$), aminek aztán ténylegesen meg is kell valósulnia ($\text{DONE}(x, a'')$). A formula végén található műveleti jelek a dinamikus logika operátorai. A formula tehát azt a bratmani tézist fejezi ki, hogy a szándék (akarat) a választott vágy melletti elköteleződést jelenti.

Talán van, akinek meglepő, biztos van, akinek nem, de tény, hogy az akaratnak ezt a felfogását már a korábbi nagy elődök közül is többen magukénak vallották. Nem tartom valószínűnek, hogy Bratman, Cohen vagy Levesque sokat olvasta volna Thomas Hobbes, Arthur Schoppenhauer vagy Edmund Husserl műveit, mert ha így tettek volna, akkor rájöhettek volna, hogy előttük már jóval korábban is megfogalmazták az akarat elköteleződéssel kapcsolatos szoros viszonyát. Hobbes esetében talán még nem látszik annyira ez a fogalmi kapcsolat, amikor a vágy és akarat viszonyáról ír, de Schoppenhauer és Husserl esetében nyilvánvaló az összefüggés tételezése. Hobbes ezt írta:

> „A megfontolásban a cselekvésre vagy annak elmulasztására közvetlenül rátapadó legutolsó étvágy vagy ellenszenv, ez az, amit akaratnak hívunk; az akarás tette (nem a képessége)." [239, 44]

Furcsa kicsit, ahogy Hobbes étvágyról ír, de azért az nyilvánvaló, hogy a vágyra gondol, amit viszont nem nehéz összekapcsolni a preferencia vagy a választás fogalmaival, és 'legutolsó' vágy (vagy ellenszenv) minősége egyértelműen az elköteleződés mozzanatára utal. Schoppenhauer pedig az akarat szabadságáról szóló híres könyvében deklarálta a következőket:

> „Tehetsz, amit akarsz: de élted minden adott pillanatában csak egy határozott dolgot és szükségképen csakis ezt az egyet akarhatod, semmi mást. [465, 25]

> „Az öntudatnak egyetlen tárgy az akarati tény és annak feltétlen uralma a test tagjai felett, ezt értjük voltaképpen azon, hogy tehetek, »amit akarok«. Csupán ennek az uralomnak a gyakorlása, vagyis a *tett*, bélyegzi (az akaratot), még az öntudat számára is, akarati ténynek. Mert amíg keletkezőben van, addig *kívánság* a neve, ha kész, akkor *elhatározás*. De hogy valóban elhatározás, azt még az öntudatnak is csak a *tett* bizonyítja be, mert mindaddig változékony. ...
> *Kívánni* lehet ellentétest, de *akarni* csak az egyiket közülük: még pedig hogy melyiket, azt az öntudatnak is csak a tett teszi nyilvánvalóvá." [465, 19]

Itt még egyértelműbb a párhuzam aközött, ahogy Bratmanék, illetve Schopenhauer fogalmaz. A schopenhauri szövegben persze több is van, mint amennyi a bratmani tézisben, de ez most számunkra nem annyira lényeges.

Noha Edmund Husserl a teleologia és filozófia kapcsolatáról szóló tanulmányán belül, az alábbi idézetet is magába foglaló részben a személy fogalmáról, az akarat és a személy

viszonyáról fejti ki a gondolatait, ám eközben az akarat maradandóságának tézisét is "kimondja" egyben.

> „... az akarati tételezés a maga sajátos tartalmával (az akart céllal), minden további nélkül létrehozza azt a maradandó célra maradandóan irányuló akarást, amely a pillanatnyi, tovaáramló aktuson túlmutat. ... Az »Akarom« tehát sokkal inkább azt jelenti: kitűzöm vagy már korábban kitűztem magamnak a célt, és ettől kezdve folyamatosan – ameddig valami egyéb nem történik – én vagyok az, akinek ez a szándéka, aki ezzel az »akarattal« rendelkezik." [250, 175]

Bármennyire is másra fókuszál Husserl a cikkében, az idézett sorok magukban rejtik azt a mozzanatot, amiről eddig szó esett: az intencionális élmények folyamatosan változó áramában a célra mint saját tartalmára irányuló akaratot az állandósággal, a maradandósággal ragadhatjuk meg. A cikk egy későbbi pontján az akaratról mint "megmaradó elhatározásról" ír Husserl [250, 180], és ezzel a maradandó célra irányulás minőségével voltaképp az elköteleződés minőségét fejezi ki. Vagy mondhatnánk akár fodítva is: az elköteleződést magát definiálhatjuk a célok, vágyak megmaradóságának tulajdonságával. Ennek meglehetne az az előnye, hogy az itt használt 'elköteleződés' fogalmat kicsit távolíthatnánk a beszédaktusok, valamint a normativitás világában fontos szerepet betöltő 'elköteleződés'-értelmezésektől.

Máshol már utaltam rá, amit most újra megerősítek, hogy bármennyire is fontos kérdésnek tartom akaratszabadság filozófiailag releváns problémáját, én itt nem tudom, nem is akarom tárgyalni. Mindaddig, amíg az első- és másodrendű akaratokról, az akaratkonfliktusról mint egy személyen belül jelentkező problémáról beszélünk, addig a filozófiát érzem illetékesnek. Abban a pillanatban azonban megváltozik minden, ha két vagy több személy egymással konfliktusba kerülő akaratáról és az akaratok megtörésére szolgáló kényszerek különböző formáiról van szó, mert ez a jelenség már nagyon is a társadalmi cselekvések szociológiai tárgykörébe tartozónak vélem. Ezt azonban nem itt, hanem a hatalom és az uralom jelenségénél kell tárgyalni. Érdemes lenne az akaratszabadság kérdését a legújabb kognitív tudományi, neurológiai kísérletek fényében is körbe járni, de ezek tárgyalását – bár nagyon fontosnak és relevánsnak tartom őket – megintcsak mellőzhetőnek vélem. Benjamin Libet készenléti potenciállal kapcsolatos kísérlete óta ismert az a nézet, mely szerint valójában nincs igazán szabad akaratunk, mert már az akaratunk tudatosulása előtt dönt az agy, ami a készenléti potenciál állapotában kimutatható [576]. Vannak, akik támogatják, vannak, akik vitatják Libet eredményeit.[18]

A vélekedések és vágyak dinamikája

Cristiano Castelfranchi és Fabio Paglieri arra dolgoztak ki egy logikát, aminek középpontjában a vélekedések célkövetés dinamikájában betöltött szerepe áll [89]. Ha kicsit eltekintünk attól, hogy cél és szándékok kezelésének folyamat Castelfranchi és Paglieri kiindulópontja az, hogy a célok követését, a szándékok felülvizsgálatát legnagyobb részben a hiteink befolyásolják, és ezt az összefüggésrendszert érdemes kibontani, részleteiben

[18] Libet modelljének újraértelmezéséről, a szabad akarat lehetőségének neurológiai alapon folyó vitájáról lásd még: [506], [545].

is megvizsgálni, leírni. Ha így teszünk szakaszokra bonthatjuk ezt a folyamatot a célok aktiválásától kezdve a szándékok képzésén át, a célok és szándékok menetközbeni revízióján keresztül a célok, a cselekvések végrehajtásáig, különböző cél- és vélekedéstípusokat különíthetünk el egymástól, melyeket tipikus módon hozzárendelhetünk a folyamat egyes szakaszaihoz. Egy ilyen felbontású leírás sokkal pontosabb magyarázatot adhat a célkövető, mérlegelő cselekvésre.

A vélekedéseink és a céljaink egyaránt mentális reprenzentációknak tekinthetők, amelyek funkcióikban különböznek egymástól, nem pedig formájukban, tartalmukban vagy más belső jellemzőikben. Castelfranchiék itt használt cél fogalma közel van a mesterséges intelligencia világának vágy és szándék, a kognitív pszichológia motiváció, kívánság és cél vagy a társadalomtudományok preferencia fogalmaihoz.

> "...a cél úgy definiálható, mint egy, a jövőben feltételezett világállapotra vonatkozó belső reprezentáció, amelynek rendelkezik egyfelől azzal a potenciálal, másfelől azzal a funkcióval, hogy egy ágens viselkedését a cél megvalósításának irányába kényszerítse/vezesse. A célok meghatározó funkciója teleológiai értelemben formálni, irányítani a rendszer tényleges viselkedését" [89, 239]

A célok olyan, a jövőre nézve kialakított állapotok, amelyek ellentétesek a világnak azon jelen állapotaival, amelyek aktuálisan léteznek vagy inkább, amelyekről az érintett ágensek feltételezik (létezni vélik). A célkövető folyamat értelmezéséhez szükség van a vélekedések, hitek tipizálására. Kétféle hitet különíthetünk el a **célvonatkozású vélekedéseken** (goal-related belief) belül: céltámogató vélekedéseket, illetve vélekedéstámogató vélekedéseket. A **céltámogató vélekedéseknek** (goal-supporting belief) nélkülözhetetlen szerepe van a célok fenntartásában. A vélekedéseinknek ugyanis speciális struktúrába kell összeállniuk ahhoz, hogy fenntartsák, igazolják magukat a célokat. Az ilyen vélekedések hiányában az adott cél elválna az aktuális állapottól, megváltozna mind a természete, mind a funkcionális tulajdonságai. Mivel Castelfranchiék a *vélekedéstámogató vélekedések* (belief-supporting beliefs) fogalmát akarják elemezni a cikkükben, ezért ezt a fogalmat alaposabban kibontják. Az alábbi típusokat sorolják be alá.

- **motivációs hitek** (motivating beliefs)
  ezek a hitek késztetnek cselekvésre minket, bár ezek mellett léteznek nem-doxasztikus befolyások is, például az belső készenléti szintünk állapota vagy az érzelmi reakciók (physical arousal, emotional triggering)
  - kiváltó hitek (triggering beliefs)
  - feltétel hitek (conditional beliefs)
- **előszűrő/előértékelési hitek** (assessment beliefs)
  ezek a hitek előszűrik a vélekedéseket, és ezzel kialakítják a követhető célok körét
  - önmegvalósulási hitek (self-realization beliefs)
  - megvalósultsági hitek (satisfaction beliefs)
  - lehetetlenségi hitek (impossibility beliefs)
- **költségbecslő hitek** (cost beliefs)
  ezek a hitek a folyamat egyes összetevőihez rendelhető költségek (ráfordítások, nehézségek) nagyságára, típusára vonatkoznak
- **inkompatibilitási hitek** (incompatibility beliefs)
  ezek a hitek segítenek abban, hogy ki tudjuk szórni a mérlegelésbe vont célok, eszközök közül azokat, amelyek valamilyen szempontból diszkrepanciát, inkompatibilitást okoz(hat)nak

- végcél inkompatibilitási vélekedések (terminal incompatibility beliefs)
- instrumentális inkompatibilitási hitek (instrumental incompatibility beliefs)
- feleslegességi imkompatibilitási hitek (superfluity)

- **preferencia hitek** (preference beliefs)
  ezek a hitek "tartalmazzák" a preferenciáinkat, valamint azok – főleg időbeli – intenzitásait
  - érték hitek (value beliefs)
  - sürgősségi hitek (urgency beliefs)
- **előfeltétel hitek** (precondition beliefs)
  ezek a hitek azokra az előfeltételekre vonatkoznak, amelyek a folyamat sikeréhez elengedhetetlenek
  - inkompetencia hitek (incompetence beliefs)
  - feltételhiány hitek (lack of conditions beliefs)
- **cél-eszköz hitek** (means-end beliefs)
  ezek a hitek rendelik össze a potenciális célokat a megvalósításukhoz szükséges eszközökkel

Ez a vélekedéstipizálás Castelfranchiék szándéka szerint kimerítő, vagyis ezek és csak ezek a hiteink, mások nincsenek. Az egyes hittípusok elsősorban az általuk "közvetített" funkciók alapján határozhatók meg, de ezek általában nem magukban fejtik ki a hatásukat valamely részfolyamaton belül, hanem inkább több típus egyszerre gyakorol hatást szinte minden pillanatban méghozzá úgy, hogy ezek jól meghatározható viszonyban vannak egymáshoz képest képest, sok esetben már lehet struktúrákról beszélni. Ismerjük már a vélekedések tipológiáját, de még nem derült ki, milyen részfolyamatokról beszélhetünk egyáltalán. Nos, a szerzők az alábbi folyamatszakaszokat definálták.

aktiválás	↣	értékelés	↣	mérlegelés	↣	ellenőrzés	↣	cselekvés
(activation)		(evaluation)		(deliberation)		(checking)		(action)

Az egyes szakaszoknak megfelelően változhatnak a célok is, és ezt érdemes valahogyan megragadni, ezért Castelfranchi és Paglieri azt javasolja, hogy bontsuk alá a cél fogalmát is. Az alábbi négy célípust definiálták.

- aktív célok (active goals) = vágyak (desires)
- követhető célok (pursuable goals)
- választott célok (chosen goals)
- végleges/végehajtandó célok (executive goals)

A folyamaton belül használható vélekedések, a folyamat szakaszai és a szakaszok szerint változó célok együttese egy olyan keretrendszert biztosít, amivel már pontosan leírhatjuk a célkövető, mérlegelő beállítódással jellemezhető cselekvéseket. Ha a teljes folyamat vélekedés-, szakasz- és célkomponenseit egyetlen grafikonon tüntetjük fel, akkor az ábra szinte önmagát magyarázza. Ahogy a folyamatban haladunk előre, az egyes folyamatszakaszokban releváns vélekedéseink formálják a céljainkat, míg a végén ki nem alakul a végrehajtandó cél. Egy ilyen folyamat természetesen nem feltétlenül fut végig, pontosabban nem feltétlen ad végeredményt, vagyis nem biztos, hogy mindig találunk végrehajtható célokat. Az ábra jobb oldalán látható annak jelzése, hogy az egyes hitek hogyan járulnak hozzá a folyamat pozitív beteljesedéséhez. A pozitív minősítési hitek "támogatják" a célkiválasztási folyamatot, a negatív előjelő hitek a célok megvalósíthatóságának 'útjában állnak". Ami az ábrán nem igazán látszik, de a folyamathoz hozzátartozik,

céltípus	folyamatszakasz	támogató hitek	hit alosztályok	+/–
	AKTIVÁLÁS	motivációs hitek	kiváltó hitek	+
			feltétel hitek	+
aktív célok	***(= vágyak)***			
	ÉRTÉKELÉS	értékelő hitek	önmegvalósulási hitek	–
			elégedettségi hitek	–
			lehetetlenségi hitek	–
követhető célok				
	MÉRLEGELÉS	költségbecslő hitek		–
		inkompatibilitási hitek		–
		preferencia hitek	érték hitek	+
			sürgősségi hitek	+
választott célok				
	ELLENŐRZÉS	előfeltétel hitek	inkompetencia hitek	–
			feltételhiány hitek	–
		eszköz-cél hitek		+
végrehajtandó célok				
	CSELEKVÉS ⇨	*visszacsatolás és azt követő (1) véleményátalakítás és (2) tervdiagnózis*		

az a visszacsatolás és a cél- vagy tervfelülvizsgálat szakasza a cselekvési folamat egészén belül. Ezt úgy kell elképzelnünk, hogy a célkiválasztás után elindított cselekvési folyamat során a visszacsatolás eredményeként esetleg szükségessé váló újratervezés végigmehet az itt leírt célkiválasztási folyamat majdnem minden lépcsőfokán. A bemutatott folyamatábra a mérlegelő beállítódás, a racionális célkiválasztás folyamatát írja le. Nem foglalkozik azzal, hogy a választást magát hogy lehet értelmezni. A következő fejezetben áttekintem, hogy – a közgazdaság-tudományon belül – milyen értelmezéseket adtak a választás fogalmának, illetve a vele kapcsolatba hozható más, rokon kategóriáknak.

2.5 Választás

Ahhoz, hogy a választás fogalmát definiálni tudjuk, szükségünk van pár kisegítő fogalomra. A választás alternatívák, lehetőségek összehasonlítását, majd az összehasonlítási szempont szerinti legjobb elem(ek) kijelölését jelenti. A választás fogalmának meghatározása előtt így meg kell tudnunk mondani, mit jelent a legjobb elem fogalma. Ha adva van egy halmazunk, amin értelmezve van egy bináris reláció, akkor mindig megvizsgálhatjuk, hogy az adott reláció szerinti rendezés eredményeként lesz(nek)-e legnagyobb elem(ek). Ha igen, akkor ezek segíthetnek minket a választás fogalmának meghatározásához. Látni kell azonban, hogy ezek a tulajdonságok elsődlegesen nem magára a relációra, hanem a tartóhalmaz egy vagy több elemére vonatkoznak, de a legjobb, legnagyobb elem fogalma

megkívánja a rendezési reláció létét a definícióban, vagyis ezek az extremitást kifejező fogalmak a halmaz és reláció fogalmára egyaránt támaszkodik. Mégpedig a következő módon [514].

(ZL.7) $\text{GREATEST} \equiv G(S, \mathbf{R}) \stackrel{\text{def}}{=} (S \subset \mathbb{X} \wedge \mathbf{R} \subset \mathbb{X} \times \mathbb{X}) \rightarrow \{m \in S | \forall n \in S : (m, n) \in \mathbf{R}\}$

(R-)legnagyobb (elem) (HUN) – (**R**-)greatest (element) (ENG) – legjobb (elem) (HUN) – best (element) (ENG) – G-optimális (elem) (HUN) – G-optimal (element) (ENG) – (felülről) domináns elem (HUN)

(ZL.8) $\text{MAXIMAL} \equiv M(S, \mathbf{R}) \stackrel{\text{def}}{=} () \rightarrow \{m \in S | \forall n \in S : (n, m) \notin \mathbf{P}(\mathbf{R})\}$

(R-)maximális (elem) (HUN) – (**R**-)maximal (element) (ENG) – M-optimális (elem) (HUN) – M-optimal (element) (ENG) – (felülről) nem domináns elem (HUN)

Mind a legnagyobb, mind a maximális elem valamiféle extremitást jelent. Ha meg akarjuk érteni azt, hogy miben különböznek egymástól, akkor ebben leginkább az segíthet, ha felidézzük a preferenciastruktúrára vonatkozó tételeket, azok közül is leginkább a szigorú preferenciarendezés, valamint annak konverze, az indifferencia és az összehasonlíthatatlansági reláció mint négyes partíció létezésére vonatkozó tételt (↷90.o.). Utóbbi szerint az említett négy reláció egymást kölcsönösen kizárja, és együttesen pedig kiteszik az univerzális relációt. Ebből viszont az következik, hogy amíg a legnagyobb elemek halmazába azok az elemek tartoznak, amelyek kielégítik a szigorú preferenciarendezés, illetve az indifferencia reláció feltételeit, addig a maximális elemek halmaza azokból az elemekből áll, amelyek nem felelnek meg a konverz szigorú preferenciarendezésnek, vagyis amelyek kielégítik vagy a szigorú preferenciarendezés vagy az indifferencia vagy az összehasonlíthatósági reláció feltételeit. Ebből perdig nyilvánvalóan az következik, hogy a legnagyobb, illetve a maximális eleme halmazok között az összehasonlíthatatlansági reláció jelenti a különbséget. Ezt úgy is megfogalmazhatjuk, hogy a legnagyobb eleme halmazban azok az elemek vannak, amelyek minden más elemnél "nagyobbak" és egymás között indifferencia reláció áll fent, a maximális eleme halmaz pedig azokat az elemeket tartalmazza, amelyek elemei között vagy az indifferencia vagy az összehasonlíthatatlansági reláció érvényesül. A legnagyobb elemek tehát mind indifferensek egymáshoz képest, a maximális elemek vagy indifferensek vagy összehasonlíthatatlanok. Az igazi különbség tehát az összehasonlíthatatlansági relációban van.

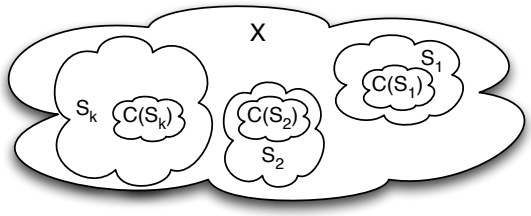
A két extremitás között két további fontos összefüggést is megállapíthatunk [65]. Egyfelől igaz az, hogy $G(S, \mathbf{R}) \subset M(S, \mathbf{R})$, vagyis a legnagyobb elemek halmaza mindig része a maximális elemek halmazának, másfelől bizonyítható az a tétel is, hogy ha egy reláció teljes és reflexív, akkor a maximális és legnagyobb elemek halmaza megegyezik.

Az extremitások, azaz legnagyobb (legjobb) és maximális elem fogalmaira támaszkodva már meg lehet határozni a **választási függvény** fogalmát, de előtte még tisztázni kell pár dolgot. Első lépésként be kell vezetni a $(\mathbb{X}, \mathscr{S})$ **választási tér** fogalmát, ahol $S \subset \mathbb{X}$ $(S \in \mathscr{S})$ az alternatívák teljes $\mathbb{X}$ halmazán képzett részhalmazokat, vagyis az ágens számára a választás során elérhető állapotok halmazait jelentik, $\mathscr{S}$ pedig az S részhalmazok hatványhalmazát. A **választási halmaz** a választási tér valamelyik konkrét S halmazának egy nem-üres részhalmaza, és a **választási függvény** a választási tér választási halmazra való leképezését végzi el.

(ZL.9) $\text{CHOICE_FUNCTION} \stackrel{\text{def}}{=} \forall S \in \mathscr{S} (C : \mathscr{S} \mapsto C(S) \mid C(S) \neq \varnothing)$

választási függvény (HUN) – choice function (ENG) – döntési függvény (HUN) – decision function (ENG)

A választási függvény alapján előállt választási halmazok elemei még nincsenek rendezve. Vannak a kiválaszottak, azok az elemek, amelyek $C(S)$-ekbe tartoznak, és vannak a választhatók, azok az elemek, amelyek S-ekbe tartoznak.



Itt még nem feltétel a választási függvény egyeleműsége, lehet benne több kiválaszott elem is. Ha ez fordul elő, akkor döntetlentörő szabályok szerint lehet eldönteni a végleges (egyetlen) választást, de ez nem jelent igazi probléma, nem is nagyon fogom elemezni. Érdemes viszont pár példát bemutatni a választási függvényre, mégpedig olyanokat, amelyek majd szembeállíthatók a később bevezetendő racionális választási függvény fogalmával. A lottósorsolás esetében az alternatívák halmazát a leadott szelvények jelentik, és a választás a véletlen alapján történik meg. A résztvevők megtippelik a számokat, és a gép véletlenszerűen választ a számok közül, de ezzel voltaképp a leadott szelvények (és áttételesen az azokat birtokló személyek) közül választ ki egyet vagy többet. Ez a példa mutatja azt is, hogy a választási halmaz fogalma még megengedi, hogy a procedúra több kimenetet adjon. A döntetlentörő szabály itt annyi, hogy a nyereményt annyi részre kell osztani, ahányan bekerültek a választási halmazba. A lottóhúzáshoz képest csak egy ponton van különbség a tombolajáték esetében, amikor a leadott szelvények közül – megintcsak véletlenszerűen – egyetlen győztest húznak ki a "jelöltek" közül.

A véletlen választás elve mellett létezik más nem-racionális módszer is. Például az általános iskolai osztályokban a hetes szerepébe a névsor szerinti kerülnek be a tanulók. Ilyenkor is választásról beszélünk, hiszen egy alternatívahalmazból egy (vagy több) elemet kiválasztunk. Meg lehetne szavaztatni mindenkit arról, hogy ki legyen a hetes, lehetne pénzfeldobással dönteni, és lehet a nevek kezdőbetűi alapján. Ezt a mechanikus elvet egyébként alkalmazzák, alkalmazták a politika világában is. Amikor még létezett Jugoszlávia, akkor a tagországok vezetőiből álló államelnökség tagjai közti úgy döntötték el, hogy ki legyen az államelnök egy adott időszakra, hogy az államok neve szerinti sorrendben váltották egymást a tagországok vezetői. De a görög városállamokban is előfordult, hogy bizonyos tisztségek betöltésére a névsor szerint jelölték ki a polgárokat egymás után.

Ezek a példák szépen mutatják azt is egyben, hogy a választási függvény létezése önmagában még nem "garantálja" annak racionális jellegét. Pedig a racionalitás feltételének érvényesülését jó lenne elvárni. Ezt a minőséget valamilyen gyenge rendezési preferenciarendezési reláció, illetve az ennek alapján képzett legnagyobb elem halmaz választási függvényhez rendelésével biztosíthatjuk.

(ZL.10) $\text{G-RATIONAL_CHOICE_FUNCTION} \stackrel{\text{def}}{=} \exists \mathbf{R} \subset \mathbb{X} \times \mathbb{X}\ \forall S \in \mathscr{S}(C(S) = G(S, \mathbf{R}))$

(G-)racionális választási függvény (HUN) – (G-)rational choice function (ENG)

Akkor racionális egy választási függvény, ha az általa kiválasztott választási halmaz megegyezik egy adott gyenge preferenciarendezés szerinti legnagyobb elemek halmazával. A racionalizálást tehát a preferenciarendezés "végzi el" (mondhatjuk is azt, hogy az adott preferenciareláció a választás/döntés racionalizálása). Ez azért is tűnik szerencsés megoldásnak, mert így jobban kezelni tudjuk azt a vitatott kérdést, hogy ti. milyen kritériumai is vannak a racionális választásnak. Ha megengedjük, hogy a preferenciareláció többféle lehessen, akkor nem kell kizárólagos módon állást foglalnunk ebben a kérdésben, és a fenti meghatározáshoz képest elfogadhatunk másfajta racionalitásértelmezéseket is. Például ha a másik extremitásfogalmat, a maximális elem kategóriáját alkalmazzuk itt, akkor máris más racionális választási függvényhez jutunk.

(ZL.11) M-RATIONAL_CHOICE_FUNCTION $\stackrel{\text{def}}{=} \exists \mathbf{R} \subset \mathbb{X} \times \mathbb{X}\ \forall S \in \mathscr{S}(C(S) = M(S, \mathbf{R}))$

(M-)racionális választási függvény (HUN) – (M-)rational choice function (ENG)

Van tehát egy választási függvényünk, van egy hozzárendelhető preferenciarelációnk, amire vonatkozóan különböző minőségű és erősségű tulajdonságokat tételezhetünk abban a reményben, hogy ezáltal jellemezni tudjuk a választói döntéseket, beállítódásokat.

> „Akkor tekintünk egy aggregált gazdasági szereplőt – tehát a döntési függvényét – racionálisnak, ha bevezethető olyan reláció az alternatívák halmazán, amely szerinti optimalizáló választás ugyanazt eredményezi, mint a döntési függvény szerinti választás. Világos, hogy ebben a keretben annyiféle racionlitás vethető fel, amennyi tulajdonsága lehet a racionalitást biztosító relációnak, így a tranzitivitás ennek csak egy fontos speciális esete." [317]

Ez a hozzáállás azt is lehetővé teszi, hogy a preferenciarelációra vonatkozó feltételek mentén különböző erősségű racionalitáskritériumokat állítsunk fel a választási függvényre vonatkozóan. Hogy ezt megtehessük, szükségünk van még egy fontos fogalomra, amivel kapcsolatot teremthetünk a választás és racionalitáskritériumot "szállító" reláció között. Ehhez fel kell idéznünk a közgazdaságtan egyik központi kategóriáját, a **kereslet** fogalmát. A közgazdászok ezt a fogalmat nagyon mélyen beágyazták a mikroökonómia szövetébe, ami – sok minden mellett – azért is különösen fontos volt a közgazdaságtan számára, mert "külsőleg" lehetett mérni – szemben a preferencia (vagy a hasznosság) fogalmával, amit meg nem. A kereslet változását a jószágok fogyasztásán keresztül "objektíven" követhetjük, mérhetjük, a jószágokra vonatkozó "belső" hasznosság szubjektíven érezhető, a megfigyelőnek nincs direkt hozzáférése. A közgazdászok körében sokáig vitatott volt, hogy a hasznosságot (preferenciát) és keresletet hogyan lehet összekapcsolni egymással. Az áttörést a – Paul A. Samuelson és Hendrik Hoathakker nevéhez köthető – **kinyilvánított preferencia** fogalmának megjelenése hozta el [289]. Samuelsonék újítása az volt, hogy a – keresletet generáló – fogyasztók vásárlásait, választásait értelmezhetjük úgy, mint saját belső preferenciáik kinyilvánításait, és az egyének választásainak teljes halmaza voltaképp az egyének preferenciáit adja meg. Erre már fel lehet írni egy összefüggést, és akkor a preferencia fogalma "kezelhetővé" válik. A választás és a kinyilvánított preferencia közti kapcsolat persze többféleképpen is értelmezhető, adhatunk egy erős és egy gyenge meghatározást, az alábbiak szerint.

(ZL.12) WEAKLY_REVELALED_PREFERENCE $\equiv \mathbf{R}_c \stackrel{\text{def}}{=} \bigcup_{S \in \mathscr{S}} [C(S) \times S]$

gyengén kinyilvánított preferencia(reláció) (HUN) – weakly revealed preference (relation)(ENG) – Richter-reláció (HUN) – Richter-relation (ENG)

(ZL.13) $\text{STRONGLY_REVELALED_PREFERENCE} \equiv \mathbf{R}_c^* \stackrel{\text{def}}{=} \bigcup_{S \in \mathscr{S}} [C(S) \times \{S \setminus C(S)\}]$

erősen kinyilvánított preferencia(reláció) (HUN) – strongly revealed preference relation (ENG) – Samuelson-reláció (HUN) – Samuelson-relation (ENG)

Mindkét relációnak ugyanaz a lényege: létezik az alternatívák teljes készletének egy olyan halmaza, amelynek m is, n is eleme, és ezek közül az egyén az előbbit preferálja az utóbbival szemben, vagy másként: az egyén mindkét elemet választhatta volna, de ténylegesen az m-t választotta. A kétféle értelmezés csak abban tér el egymástól, hogy az egyikben megszorítás van a preferenciareláció második argumentumára vonatkozóan (a formulában $S \setminus C(S)$ szerepel a S helyett), míg a másikban nincs. Ezt úgy interpretálhatjuk, hogy az egyén n-t választhatta volna, de nemcsak, hogy nem választotta, de aktuálisan még vissza is utasította (ezt fejezhetjük ki azzal a szűkítéssel, hogy n mindig csak az S-nek azon részhalmazában szerepelhet, amely az S alaphalmaz és a $C(S)$ választási részhalmaz különbsége).

A továbbiakban azonban a választási függvénnyel, a kinyilvánított preferenciarelációval kapcsolatos racionalitásértelmezéseket Kotaro Suzumura megközelítését elfogadva mutatom be, aki könyvében precízen és kimerítően tárgyalja ezt a témakört [514]. Történelmileg Samuelsonék axiómái voltak az elsők, amelyek kényszereket fogalmaztak meg a kinyilvánított preferencia fogalmára vonatkozóan. Samuelson feltételét el is nevezték gyenge axiómának, míg Houthakkerét erősnek. Ezek mellé Suzumura definiált egy harmadik kritériumot is, amit Houthakkerről nevezett el, és a három fogalmat az alábbi módon definiálta.

(ZL.14) $(\text{WARP}) \stackrel{\text{def}}{=} \forall m \forall n \in \mathbb{X}(\mathbf{R}_c^*(m,n) \to \neg \mathbf{R}_c(m,n))$

kinyilvánított preferencia gyenge axiómája (HUN) – Weak Axiom of Revealed Preference (ENG) – Weak Axiom of Consumer Behaviour (ENG)

(ZL.15) $(\text{SARP}) \stackrel{\text{def}}{=} \forall m \forall n \in \mathbb{X}([\mathbf{R}_c^*]^+(m,n) \to \neg \mathbf{R}_c(m,n))$

kinyilvánított preferencia erős axiómája (HUN) – Strong Axiom of Revealed Preference (ENG) – Strong Axiom of Consumer Behaviour (ENG)

(ZL.16) $(\text{HARP}) \stackrel{\text{def}}{=} \forall m \forall n \in \mathbb{X}([\mathbf{R}_c]^+(m,n) \to \neg \mathbf{R}_c^*(m,n))$

kinyilvánított preferencia Houthakker axiómája (HUN) – Houthakker's Axiom of Revelaed Preference (ENG)

Mindhárom tulajdonság esetén szerepel a gyenge és erős értelemben kinyilvánított preferencia a definícióban, és kétszer alkalmazni kell az adott relációra az $\mathbf{R}^+$ tranzitív lezárás műveletét is (↰41.o.).

Amartya K. Sen a 1970-es években írt cikkeiben több racionalitáskritériumot is megfogalmazott, és ezeket α-, β, γ-tulajdonságoknak nevezte el. Maga Sen is megjegyezte, hogy az α-tulajdonság megegyezik Herman Chernoff által definiált tulajdonsággal, amit a tétel megjelenése óta Chernoff-axiómának neveznek [93]. Elsőre bemutatom, ahogy Sen megadta az α-tulajdonság definícióját, majd másodjára melléteszem azt a formulát, ahogy Suzumura fejezte ki Chernoff axiómáját.

(ZL.17) $(\text{SEN-}\alpha) \stackrel{\text{def}}{=} \forall m((m \in S_1 \wedge S_1 \subset S_2) \to (m \in C(S_2) \to m \in C(S_1)))$

Sen α-tulajdonsága (HUN) – Sen's α-property (ENG) – irreleváns alternatívák függetlensége (HUN) – independence of irrelevant alternatives (ENG) – (basic) contraction consistency (ENG)

$$(\text{CA}) \stackrel{\text{def}}{=} \forall S_1 \forall S_2 \in \mathbb{S}(\, S_1 \subset S_2 \rightarrow (S_1 \cap C(S_2) = \varnothing \vee S_1 \cap C(S_2) \subset C(S_1)))$$

Chernoff axiómája (HUN) – Chernoff's axiom (ENG) – örökítő (HUN) – öröklődő (HUN) – megőrző (HUN) – preserving (ENG)

Sen azt javasolta, értelmezzük úgy ezt a tulajdonságot, hogy ha a világbajnok egy pakisztáni, akkor ennek a világbajnoknak Pakisztán bajnokának is kell lennie [477]. Ez a kritérium az örökítésről, örökletességről szól. Ha egy nagyobb halmazból kiválasztunk valamit "győztesnek", akkor annak bármely olyan részhalmazából is ki kell választanunk ugyanazt a győztes, amely részhalmazok tartalmazzák azt a legjobb elemet. Sen megjegyezte még, hogy erre feltételre használják az irreleváns alternatívák függetlensége kifejezést is, de fontos tudni, hogy ilyenkor ezt más értelemben teszik, mint ahogy az a társadalmi választások elméletében szokás (↰433.o.). Itt az egyén racionalitásának jellemzésére, ott a kollektív preferencia kialakítását végző eljárás egyik kritériumaként használják ezt a terminust. Kenneth J. Arrow egy 1959-es cikkében definiált egy tulajdonságot [24], ami egyetlen apró mozzanatban tért csak el a Chernoff axiómától. Arrow tulajdonságát később Suzumura elnevezte Arrow-axiómának, és a következő formában mutatta be.

(ZL.18) $$(\text{AA}) \stackrel{\text{def}}{=} \forall S_1 \forall S_2 \in \mathbb{S}(S_1 \subset S_2 \rightarrow (S_1 \cap C(S_2) = \varnothing \vee S_1 \cap C(S_2) = C(S_1)))$$

(választás konzisztenciájának) Arrow axiómája (HUN) – Arrow's axiom (of choice consistency) (ENG) – Arrow's independence condition (ENG) – erősen örökítő (HUN) – erősen megőrző (HUN) – strongly preserving (ENG)

Az a pici kis eltérés, ami megkülönbözteti a Chernoff- és az Arrow-axiómákat egymástól, a formulák utolsó tagjában látható: eltérő műveleteket írnak elő a választási függvények között. A különbség miatt nyilván nem vonhatjuk össze a két tulajdonságot, de a hasonlóság miatt az Arrow-feltételt is megőrzési tulajdonságnak tarthatjuk, olyannak, ami az örökletességre erősebb feltételt támaszt a Chernoff-axiómához képest. Más nézőpontból tekintve a dolgokra kicsit másfajta interpretációt is adhatunk az örökítés problémájára. Az egyszerűség kedvéért egyelemű halmazokra váltva: ha van egy m alternatívánk, amit előnyben részesítünk n-nel és o-val szemben, akkor ez a beállítódás nem szabad, hogy megváltozzon attól, hogy még egy elemet "beveszünk" az összehasonlításba. Vagyis nem lehetséges az, hogy egy p elem felvétele után már nem m-t preferáljuk n-nel vagy o-val szemben. Mindezt egy gyakran idézett gyakorlati példával szemléltetve: ha a csoki- és vaníliafagyik közül a csokit választjuk, majd ezután közlik velünk, hogy van még málna is, akkor – ha konzisztensek akarunk maradni – nem mondhatjuk másodjára, hogy inkább mégis vaníliát kérünk.[19]

Sen az általa definiált β-tulajdonságot a következő formában határozta meg.

(ZL.19) $$(\text{SEN-}\beta) \stackrel{\text{def}}{=} \forall m \forall n((m, n \in C(S_1) \wedge S_1 \subset S_2) \rightarrow (m \in C(S_2) \leftrightarrow n \in C(S_2)))$$

Sen β-tulajdonsága (HUN) – Sen's β-property (ENG)

[19] Természetesen az új alternatívát (példánkban a málnát) preferálhatjuk a régi opciókkal szemben.

Sen javaslata értelmében ezt a minőséget úgy interpretálhatjuk, hogy ha a pakisztáni bajnokok közül legalább egy világbajnok is egyben, akkor az összes pakisztáni bajnoknak világbajnoknak kell lennie [477]. Ha bővítjük az alternatívák halmazát, és az eredeti – szűkebb terjedelmű – halmaz legjobb (maximális) elemei közül egy a bővítés után is bekerül a legjobbak (maximálisak) közé, akkor a "korábbi" társaival is ennek kell megtörténnie. Sen definiált egy harmadik fontos tulajdonságot is, amit először γ-tulajdonságnak [479], majd később a **kiterjesztés konzisztenciájának** nevezett el [473]. Sen az új kritériummal azt akarta biztosítani, hogy ha az alternatívák tetszőleges halmazaiból ugyanazt az elemet mindig kiválasztják, akkor a halmazok uniójából is ki kell választani.

(ZL.20) $(\text{EC}) \stackrel{\text{def}}{=} \forall i(S_i \in \mathscr{S} \rightarrow (\bigcap_i C(S_i) \subseteq C(\bigcup_i C(S_i))))$

kiterjesztési konzisztencia (HUN) – (basic) expansion consistency (ENG) – Sen γ-tulajdonsága (HUN) – kiterjedő (HUN) – Sen's γ-propertiy (ENG) – összehangoltság (HUN)

Igazodva a Sen által korábban megadott interpretációk "műfajához", azt adhatjuk meg a γ-tulajdonság értelmezéseként, hogy ha valaki minden országban megnyeri az országos versenyt, akkor a végső összesítésben is nyernie kell, tehát a világbajnoknak is neki kell lennie. Sen megjegyezte azt is [479], hogy ha csak a választási részhalmazok véges halmazaira fókuszálunk, akkor érdekes komplementaritást vehetünk észre az α- és γ-tulajdonságok között. Mindez akkor válik különösen szembetűnővé, ha az α-tulajdonság formuláját kicsit átalakítjuk.

(ZL.21) $(\text{SEN-}\alpha) \stackrel{\text{def}}{=} \forall i(S_i \in \mathscr{S} \rightarrow (\bigcup_i C(S_i) \subseteq C(\bigcap_i C(S_i))))$

A Sen által definiált tulajdonságok segíthetnek a racionális választással, választási függvénnyel szemben támasztott elvárások értelmezésében. A racionalitáskritériumok értelmezésében és meghatározásában azonban Marcel Richter hozzájárulása fontosabbnak mondható [439]. Ő vezette be a **kongruencia** fogalmát, ami gyorsan az egyik legelfogadottabb racionalitáskritériummá vált. A kongruencia fogalmának kétféle (erős és gyenge) értelmezést is lehet adni – az alábbi módon:

(ZL.22) $(\text{WCA}) \stackrel{\text{def}}{=} \forall S \in \mathscr{S}(\forall m \in S \wedge (\exists n \in C(S) : \mathbf{R}_c)) \rightarrow m \in C(S)$

gyenge kongruencia axiómája (HUN) – Weak Congruence Axiom (ENG)

(ZL.23) $(\text{SCA}) \stackrel{\text{def}}{=} \forall S \in \mathscr{S}(\forall m \in S \wedge (\exists n \in C(S) : [\mathbf{R}_c]^+)) \rightarrow m \in C(S)$

erős kongruencia axiómája (HUN) – Strong Congruence Axiom (ENG)

A kétféle tulajdonság meghatározása csak abban tér el egymástól, hogy amíg a gyenge kongruencia a Richter-reláció, addig az erős kongruencia a Richter-reláció tranzitív lezártjának való megfelelést írja elő a formulában.

A racionalitáskritériumok között érvényes összefüggéseket tételszerűen nem mutatom itt be, Suzumura társadalmi választásokról szóló könyvében minden tételt kimond és be is

bizonyít [514]. A könyvből viszont átemelek egy grafikus ábrázolást, amely szemléletesen mutatja a kritériumok közti kapcsolatrendszert.[20]

$$
\begin{array}{ccccccccc}
\textbf{HARP} & \leftrightarrow & \textbf{FR} & \leftrightarrow & \textbf{QTR} & \leftrightarrow & \textbf{AR} & \leftrightarrow & \textbf{R} \\
\updownarrow & & & & & & & & \\
\textbf{SARP} & & \updownarrow & & \updownarrow & & \updownarrow & & \downarrow \\
\updownarrow & & & & & & & & \\
\textbf{WARP} & \leftarrow & \textbf{AA} & \leftrightarrow & \textbf{(CA + SUA + GC)} & \leftrightarrow & \textbf{(CA + GC)} & \leftrightarrow & \textbf{CA} \\
\updownarrow & & & & & & & & \\
\textbf{WCA} & & \updownarrow & & \updownarrow & & \updownarrow & & \uparrow \\
\updownarrow & & & & & & & & \\
\textbf{SCA} & \leftarrow & \mathbf{C = C_{cu}^{R_C}} & \rightarrow & \mathbf{C = C_{so}^{R_C}} & \rightarrow & \mathbf{C = C_{co}^{R_C}} & \rightarrow &
\end{array}
$$

A közösségi döntésekről szóló fejezeten belül még egyszer szóba kerül majd a racionalitás problémaköre, amikor – egy egészen eltérő megközelítés alapján – a racionalitásfogalom picit más értelmezést kap. Mind az egyéni, mind a Φ csoportos véleménykészletre (collective set of judgement) vonatkozóan az lesz a racionális, ha a véleményekre áll majd a teljes, konzisztens és deduktíve zárt minősítés (↪477.o.).[21]

Választás és kontroll

Ugyan eddig nem reflektáltam rá, de ettől még tény, hogy a választásról eddig úgy beszéltünk, hogy az teljesen szabad volt, vagyis semmiféle kontrollt nem gyakoroltak fölötte. Nem volt sem külső, sem belső akadálya annak, hogy a választás előtt álló egyén saját késztetéseinek (és csak azoknak) megfelelően dönthessen. Pedig sokszor és sokféle módon elképzelhető, hogy van valamilyen kontroll a választásaink fölött. A kontroll egyik – belső – formája az **önkontroll**, míg a másik – belső vagy külső – formája a **normatív kontroll**. Már Thomas Schelling is használta, később Jan Elster is, és persze sokan mások az **előzetes elköteleződés** (precommitment) fogalmát, ami alatt valamilyen másodlagos preferencia kifejeződését, valamilyen szabadon mérlegelő tevékenység eredményét kell érteni, és semmiképpen sem szabad összekeverni a normatív elköteleződéssel, amelyben mindig van egy deontikus összetevő. Odüsszeusz és a szirének története szolgáltatja a jelenséggel foglalkozók kedvenc példáját, amikoris a hős vágyva arra, hogy hallja a szirének csodás énekét, de egyben vágyva arra is, hogy ne menjen utánuk a halálba, lekötöztette magát azért, hogy ne tudja megcselekedni azt, amiről a jelenben nem akarta, hogy a jövőben megtörténjék. Az előzetesen elköteleződés lényege tehát az, hogy a jövőbeni cselekvési lehetőségeket (vagy azok egy részét) kizárja valaki a jelenbeli döntésévelm, vagyis egy előzetes, ex ante döntéssel csökkenteni a majdani választási alternatívák számát [375]. Ezt a jövőre irányuló, jelenbeli elköteleződést úgy ragadhatnánk meg, ha a választás modelljébe beemelnénk az időt, de ezzel most nem próbálkoznék meg.

Fontosabbnak tartom, hogy megpróbáljam azt modellezni, hogy miként lehetne a választás és normatív elköteleződés vagy csak röviden a választás és norma viszonyát leírni, jellemezni. Amartya K. Sen az 1990-as években fenntartásokat fogalmazott meg a

[20] Néhány racionalitásfogalmat nem mutattam be azok közül, amelyek szerepelnek Suzumura ábrájában, például a teljes racionalitás kritériumát, és nem beszéltem az útfüggetlenség axiómáiról sem.

[21] A magyar nyelvű szakirodalomban Mészáros József és Málik József Zoltán tágabb nézőpontból foglalkozik a racionalitás kérdésével [342], [322], míg Magyarkúti Gyula kiváló axiomatikus megközelítését adja a racionalitás fogalmának [317].

racionális választás addigra egyeduralkodóvá vált paradigmájával szemben [473], [483], és a társadalmi normák figyelembevételének fontossága mellett érvelt. Szemléletes példája a következő volt. A racionális választás paradigmája alapján – akármelyik racionalitáskritériumot is fogadjuk el – az alábbi két választást leíró választási függvényt inkonzisztensnek kell minősíteni:

$$C(S_1) \equiv C(\{m,n\}) = \{m\}$$
$$C(S_2) \equiv C(\{m,n,o\}) = \{n\}$$

Pedig könnyen kerülhetünk olyan helyzetbe, amikor teljesen racionális magyarázatot adhatunk arra, hogy miért választunk látszólag nem konzisztens módon, ha bővítjük a választható alternatívák halmazát. Gondoljunk csak arra – javasolta Sen –, amikor egy gyümölcsöstálban egyetlen alma van. Ekkor a választást értelmezhetjük úgy, hogy m jelenti azt, hogy nem vesszük el az utolsó almát a tálból, míg n-nel jelölhetjük azt az opciót, amikor elvesszük az almát. A tényleges gyakorlatban sokszor azt tapasztalhatjuk, hogy az emberek nem veszik el az almát, tehát m-t választják. És ha nem veszik el az almát, akkor azt azért teszik, mert érvényesnek tartják azt a társadalmi normát, amely tiltja az utolsó gyümölcs elvételét a tálból. Amennyiben viszont van egy másik alma is a tálban, amit a harmadik o alternatíva "felvételével" fejezhetünk ki, akkor ez a norma már nem tiltja valamelyik alma elvételét a kettőből. Elvehetjük akár azt is, ami előzőleg még tiltva volt, tehát ekkor választhatjuk az n alternatívát.

Sen 1977-es cikkében már arról írt, hogy az eredeti konzisztenciaelvárásokat kétféleképpen is megsérthetjük, noha ilyenkor is racionálisnak érezhetjük a cselekvéseinket [483].[22] Az egyik lehetőséget **választófüggőségnek** (chooser dependence), a másikat **menüfüggésnek** (menu dependence). Kotaro Suzumura és Walter Bossert a normativitás jelenségének megragadására, a **normafüggő választás** (norm-conditional choice) modellezésére azt javasolta,[23] hogy az eredeti választási modellt ki kell egészíteni a normák tiltó hatását kifejező kivételhalmazokkal [64]. Ezt úgy tehetjük meg, hogy a normák által tiltott m_k alternatívákat összekapcsoljuk adott választási halmazokkal (S_k, m_k), és ezek uniójából létrehozzuk a norma alapú kivételhalmazunkat. Ha ezek után ezt a kivételhalmazt kivonjuk az eredeti választási halmazból, akkor azzal megkapjuk a normafüggő választási halmazunkat. Mindezt a következő formulákat írhatjuk fel.

(ZL.24) $N = \{\bigcup_k (S_k, m_k)\}$

tiltott alternatívák halmaza (HUN) – kivételhalmaz (HUN) – set of forbidden alternatives (ENG)

(ZL.25) $C^N(S,N) = C(S) \setminus N$

normafüggő választási halmaz (HUN) – norm-conditional choice set (ENG)

Miután előállítottuk a normafüggő választási halmazunkat, azt nyugodtan behelyettesíthetjük a "sima" választási halmaz helyébe, és azt használhatjuk tovább a már megismert formulákban, képletekben.

[22] Sen cikkét lefordították magyarra is [485].

[23] Lehetne a 'normakondicionált választás' kifejezését is használni itt, ami szószerinti fordítása lenne az angol terminusnak, de én jobbnak tartom a 'normafüggő' minősítést. Jobban kifejezi a norma választást befolyásoló "jelenlétét".

Chapter 3

A társadalmi cselekvés típusai

„Az érzelmek kölcsönöznek irányultságot és értelmet az életnek, de egyúttal meg is gátolnak bennünket abban, hogy kitartóan ebben az irányban haladjunk.” [145, 76]

„Fölöttébb téves az a nézet, ... miszerint fejlesztenünk kéne magunkban azt a képességet, hogy gondoljunk arra, amit éppen teszünk. Ennek pont az ellenkezője az igaz. A civilizáció azáltal halad előre, hogy gyarapítja ama fontos műveletek számát, amelyek végrehajthatók a róluk való gondolkodás nélkül. A gondolkodási műveletek a csaták lovasrohamaihoz hasonlítanak – számuk szigorúan korlátozott, megkívánják a friss lovakat, és csak a döntő pillanatokban vethetők be.” [581, 61]

Amikor a cselekvés, illetve a társadalmi cselekvés fogalmait definiáltam (↰131.o.), már akkor is jeleztem, hogy a (társadalmi) cselekvéshez hozzátartozik a cselekvő értelemtulajdonítása vagy másként: az intencionalitás. Természetesen intencionalitás nélkül is lehet szabályszerű az ember viselkedése, de ezt a szabályszerűséget biológiai vagy fizikai tényezőkkel magyarázhatjuk. Bármennyire is fontos azonban számunkra a cselekvés fizikai, illetve biológiai szintje, a cselekvéselmélet számára mindez nem igazán érdekes. Arra persze jó, hogy a társadalmi szinten található szabályszerűségeket olykor viszonyíthassuk a fizikai vagy biológiai meghatározottságainkhoz.

Ha szabályszerűséget keresünk valamilyen jelenségben, olyan fogalmakat használunk, mint alakzat, minta, mintaszerűség, mintázat, séma, szabály, szabályosság, szabályszerűség, törvényszerűség. Ezek a kategóriák valamilyen felismerhető, ismétlődő minta, séma létezésére utalnak, semmi másra. A társadalmi cselekvések szabályszerűségeinek, a cselekvések típusainak megragadására használt fogalmainkban azonban mindig van valami többlet. Gondoljunk csak ezekre: divat, erkölcs, etikett, habitus, illem, jog, konvenció, morál, ösztön, reflex, rend, rutin, szabályzat, szokás, norma, szokásnorma, szokáserkölcs, törvény, tradíció. Ezekkel a fogalmakkal nemcsak a szabályszerűség létezését

fejezzük ki, de szinten minden esetben utalunk a szabályszerűség kiváltó okára is. A kérdés csak az, hogy milyen módon tipizálhatjuk a társadalmi cselekvések szabályszerűségeit kiváltó indokokat. Erre nyilván sokféle lehetőség adódik. Ha a cselekvés differencia specifikáját az intencionalitásban jelöltük meg, akkor kézenfekvő lehetőségnek tűnik, hogy igazodjunk az intencionalitás fogalmának a felosztásához. Az effajta tipizálási kísérletek azonban túlzottan messze vannak azoktól a fogalmaktól, amelyekkel a cselekvés szabályszerűségeit, típusait szoktuk megragadni (lásd erre például: [367]). Azt, hogy milyen a viszony a cselekvő (Ego) és a világ között, milyen az értelemadás minősége, olyan fogalmakkal fejezhetjük ki leginkább, mint beállítódás, irányultság, motiváció, attitűd.

Úgy gondolom, hogy tartalmilag nincs miért meghaladni Max Weber híres cselekvéstipológiáját, de terminológiailag egy ponton érdemes azt korrigálni. Bár Weber – többékevésbé – pontosan definiálta az egyes cselekvéstípusokat, de arra nem reflektált, hogy lehet-e, és ha igen, akkor milyen dimenziók alapján lehet a négy cselekvéstípust elrendezni egymáshoz képest egy fogalmi térben. Erre többen is kísérletet tettek, amelyek közül kiemelném Bertalan László javaslatát [49], amire a későbbiekben – legalább részben – én is támaszkodni fogok. A weberi tipizálás mögöttes logikájának feltárása előtt azonban röviden elemeznem kell az egyes cselekvéstípusokat. Weber az alábbi négy cselekvéstípust különítette el egymástól [575]:

- értékracionális cselekvés
- célracionális cselekvés
- indulati-érzelmi cselekvés
- habituális (tradicionális) cselekvés

Weber negyedik típusként a tradicionális cselekvést sorolta fel, ami mind a jelentését, mind a taxonómiai helyét tekintve helyénvaló és elfogadható, de az általa adott megnevezés – nézetem szerint – nem igazán szerencsés. A probléma itt az, hogy a 'tradicionális' jelző nagyon szoros, "eltéphetetlen" kapcsolatban áll a 'tradíció' fogalmával, miközben a cselekvéselméleti kontextusban a két kategória jelentése nagyon messze áll egymástól. A legfontosabb szempontok szerint gyakorlatilag nincs vagy csak nagyon kevés, áttételes kapcsolat van a weberi tradicionális cselekvés és a tradíció fogalma között. Ezt a nézetemet itt természetesen nem tudom megvédeni, de az egyes cselekvéstípusok fogalmának kibontása után remélhetőleg jobban elfogadható lesz az álláspontom. Azt is csak később tudom megmagyarázni, miért a habituális minőséget emeltem be a weberi kategória helyett. Dióhéjban megelőlegezve a későbbi gondolatmenetet: az osztályozási logika kulcsa a bizonyos értelemben a racionalitás fogalma, de csak akkor érthetjük meg igazán ezt a logikát, ha a racionalitás emellett figyelembe vesszük még a reflektivitás fogalmát is.

Az, hogy Weber két racionális beállítódást (a cél- és az értékracionálist), illetve két aracionális motivációt (az indulati-érzelmit és a habituálist) különít el, sokan leírták már. Hogy mit is jelent pontosan a racionalitás fogalma, a későbbiekben alaposabban körbejárom majd. Fontos azonban látnunk azt is, hogy a cselekvő beállítódásának jellemzésekor a racionalitás mellett van még egy másik fontos dimenzió is, mégpedig az, hogy a cselekvő milyen szempontokra figyel egyáltalán: van-e, amire reflektál, ha igen, mik ezek, és mit tesz akkor, ha több szempontot is figyelembe kell, figyelembe akar venni. A racionális, mérlegelő beállítódás fogalma szükségszerűen feltételezi, hogy a cselekvő több szempontra figyel. A mérlegelésnek enélkül nincs értelme. Eléggé nyilvánvaló-

nak látszik, hogy Weber is ezért nevezte racionálisnak az értékracionálist, mert a mérlegelés mozzanata hozzátartozik. A habituális cselekvés során azonban nincs mérlegelés, mert a cselekvő nem is igazán reflektál saját magára, a cselekvés rutinszerű, automatikus. Az érzelmi cselekvésben már tetten érhető valamiféle tudatosság, reflektáltság, de ekkor sincs még "szükség" mérlegelésre, mivel a cselekvő irányultságában az egyetlen érzelmi motívum van a fókuszban, semmi más. Mondhatjuk ezt úgy is, hogy az aracionális cselekvések azért nem-mérlegelő cselekvések, mert a semmit és az egyet nem lehet mérlegelni. A mérlegelés és a reflektáltság fogalmait tehát egyszerre kell alkalmaznunk, amikor az osztályozási logikát meg akarjuk érteni.

A négy jólismert weberi cselekvéstípust nem abban a szokásos sorrendben mutatom be, ahogy maga Weber tette ezt, és ahogy azóta még sokan mások követték őt ebben. A fogalmi építkezés belső hivatkozásainak rendszere, a szükséges ontológiai elköteleződések egymásra épülésének "kényszerei", valamint a – legalább részben elfogadhatónak tartott – evolúciós logika követése mind azt a sorrendet "kívánják meg", hogy elsőként az indulati-érzelmi, majd a habituális, aztán a célracionális, végül az értékracionális cselekvés jeleségét elemezzem.

3.1 Indulati-érzelmi cselekvés

Amikor Max Weber kategóriatana második paragrafusában a cselekvések harmadik típusaként megadta az indulati-érzelmi cselekvést [575, 2.§], akkor nem adott sok túl sok támpontot a fogalom értelmezéséhez. A paragrafus definíciós részében semmi használható elem nincs, hiszen ott a megnevezésbe emelt fogalmakat ismételte meg újra, amikor azt állította, hogy a cselekvés lehet "*indulati*, különösen *emocionális*: ha tényleges indulatok és érzelmi állapotok ... határozzák meg a cselekvést". A paragrafushoz tartozó magyarázó blokk harmadik bekezdésében írt ugyan egy tartalmas mondatot az indulati cselekvésről, de ezen kívül mást nem mondott.

> „Indulati-érzelmi alapon cselekszik az, aki ténylegesen meglévő vágyait elégíti ki: bosszút áll, élvezetet szerez magának, odadást tanúsít, belefeledkezik a szemlélődésbe, vagy (mindegy, hogy egy az egyben vagy szublimált módon) levezeti az indulatait." [575, 53:2.§.3.]

A kissé felszínes kezelésmód ellenére az persze nyilvánvaló, hogy Weber fontosnak tartotta ezt a cselekvéstípust. Ezt az a tény önmagában is igazolja, hogy Weber önálló típusként nevezte meg az indulat-érzelmi cselekvést. Ahhoz azonban, hogy ezt a cselekvéstípust egyértelműen el lehessen helyezni egy cselekvéselméleti modellen belül, szükséges néhol pontosítani, néhol kiegészíteni a weberi definíciót. Fontos ezt megtenni, mert egyfelől a pontosítások és kiegészítések után lehet egyértelművé tenni az indulati-érzelmi cselekvés viszonyát a többi motivációtípushoz képest, másfelől el kell tudnunk választani az érzelmi jelenségek igen széles körét az indulati-érzelmi cselekvések szűkebb tartományától.

Amikor valaki beszámol egy indulati-érzelmi cselekvéséről, akkor gyakran ilyen kifejezésekkel érzékelteti a történteket: "elborult az agyam", "pánikrohamot kaptam", "semmit sem mérlegeltem", "elveszettem az ítélőképességemet", "önkívületi állapotba kerültem", "elragadott az indulat", "elveszettem a fejem", "megfeledkeztem magamról", "túláradó érzelem", "érzelemkitörés". Ez nem véletlen. Az indulati-érzelmi cselekvés lényege a "mindent elsöprő szenvedély". De mit is jelent ez a szenvedély, ez az önkívületi állapot, ez az elragadottság, ez az elvesztés, ez a megfeledkezés, ez a kitörés, ez a robbanás?

Azt, hogy a cselekvő tudata egyetlen dologra irányul, a cselekvés egyetlen tényezőssé válik, minden más eltűnik a cselekvő tudatának horizontjáról. Intencionalitás van, valamilyen szintű tudatosság van, de nincs mérlegelés, és azért nincs, mert – a tudatot teljesen kitöltő – egyetlen dolgon nincs mit mérlegelni. Amikor a tudat kizárólag egyetlen tárgyra irányul, akkor beszélünk indulatról.

Az érzelem, az affektivitás persze ennél sokkal többet jelent. A cselekvéstípusok közé besorolandó indulati-érzelmi cselekvéseket úgy kell értelmeznunk, hogy azok az érzelmi világunknak egy kis részét fedik csak le. Az érzelmi jelenségek egészét (affekció, affektivitás, affektus, emóció) három nagyobb részre bonthatjuk. Egyrészt vannak az indulati cselekvések, amikor felindulásról, indulatról, gerjedelemről, szenvedélyről beszélünk. Ezeket illesztem be itt a cselekvéselméleti modellbe. Ezek relatíve rövid ideig tartanak, és lényegük az indulat tárgyára történő kizárólagos irányulás. Pont a kizárólagosság miatt nem lehetnek ezek tartósak. Másrészt vannak a hangulatok. Ide olyan fogalmak sorolhatók, mint a kedély, kedélyállapot, kedv, mód, hajlandóság, de ide tartozik az empátia, szimpátia, antipátia, apátia, ellenszenv is. Ezek az érzelmi jelenségek már hosszabb ideig is képesek fennmaradni, viszont ezek már nem uralják kizárólagos módon a tudatunkat, bár erős befolyással lehetnek rá. Harmadrészt az érzelmek világába sorolhatjuk a (lelki)alkat, a lelkület, a mentalitás jelenségét is, és a kapcsolódó társfogalmakat: passziót, pátoszt, hajlamot, személyiséget, temperamentumot. Ezek a fogalmak persze nem rendelkeznek pontos határvonalakkal, és természetesen az affektív mező iménti felosztása sem mondható igazán egzaktnak. Az affektusmező három alosztálya elsősorban az időhöz való viszonyában rendezi el az affektív jelenségeket: az alapérzelmek nagyon rövid ideig (percekig, órákig) létező jelenségek, a hangulatok már hosszabb időszakon (napokon, heteken, akár hónapokon) át érvényeslő beállítódások, az affektív személyiségéjegyek pedig korszakokon át tartanak (akár egy életen keresztül). Amennyiben az érzelmi jelenségeket felbontjuk, vagyis az időbeliség mentén affektív alosztályokat képzünk, akkor már kezelni tudjuk azt a többek által ellentmondásosnak tartott helyzetet, hogy ti. egyszerre tarthatjuk igaznak azt a két tézist, hogy az indulati-érzelmi cselekvések intencionálisak, a hangulatok és temperamentumok viszont nem. Amikor dühösek vagyunk, amikor félünk, mindig meg tudjuk mondani a haragunk, félelmünk tárgyát, de a tartós szorongás, a rosszkedv nem irányul semmire.

Természetesen az affektivitás mindhárom területe ugyanolyan fontos részét képezi a mindennapi cselekvéseinknek. A cselekvéelméleti modellezés számára azonban a legszűkebb értelemben vett indulati-érzelmi cselekvés az igazán fontos. Amikor a cselekvéseinket ideáltipikus módon akarjuk jellemezni, akkor az affektív-érzelmi mező legbelsejében meghúzódó réteget kell feltárnunk. A hangulatok, illetve a affektív személyiségjegyek ettől még az érzelmi világunk ugyanolyan fontos részét képzik, mint az indulati-érzelmi cselekvések, csak amíg az utóbbiak elméleti modell egyik primitív (vagyis ontológiailag magasabb szintű) típusfogalmának kell tekintenünk, addig az előbbieket – fogalmi összetettségük miatt – más fogalmakból levezethető kategóriaként érdemes kezelnünk.

Ha az érzelmi cselekvések eddig említett példáira gondolunk, akkor szembetűnő lehet két fontos vonásuk: egyfelől a tudatos ént bizonyos értelemben mindig passzív szerepben tapasztaljuk ilyenkor, másfelől az érzelmi megnyilvánulások együtt járnak testi reakciókkal, érzésekkel. Utóbbit önmagában is szépen jelzi az ‘érzés’ és ‘érzelem’ terminusok közti szoros etimológiai kapcsolat, de a magyar ‘indulat’, az angol ‘emotion’ terminusoknak a ‘mozgás’ fogalmához való kapcsolódása is ugyanezt fejezi ki.

Noha az indulati-érzelmi cselekvés pont arról szól, hogy ilyenkor a tudatunkat egyetlen indulat, érzelem határozza meg, ez közel sem jelenti azt, hogy az érzelmekből hiányozna az intencionalitás. Nem beszélhetünk arról, hogy az indulat, az érzelem valamire irányul, hogy az érzelemnek tárgya van, ha nem feltételezzük, hogy az indulati-érzelmi cselekvésben inherens módon benne rejlik az intencionalitás. Ez persze nem jelenti azt, hogy az érzelem akaratlagos, szándékos lenne. A szenvedély nincs az akarat közvetlen irányítása alatt. Nem lehet akarni, hogy szeressünk, csodáljunk, tiszteljünk valakit vagy féljünk tőle, haragudjunk, irigykedjünk rá. "Lehet, hogy valaki kontrollálni tudja az érzelmei kifejezését, de az élményt magát nem ..." [593]. Az is igaz, hogy elő lehet állítani tudatosan olyan helyzetet, amelyben jelentkezhetnek érzelmek, de ilyenkor a helyzet maga váltja ki az érzelmeinket, nem pedig a saját magunkra irányuló akarat. Ebben az értelemben az indulati-érzelmi és a racionális cselekvés valóban élesen szemben áll egymással, és az érzelmeket tényleg irracionálisnak kell tartanunk. Arra, hogy mire képesek az irracionális érzelmek (például a félelem), vagyis az indulati-érzelmi állapot mennyire figyelmen kívül képes hagyni minden racionális érvet, azt az esetet hozhatom fel kedvenc példámként, ami egy repülőgépen történt 1991-ben az USA-ban. A gép útját ugyanis meg kellett szakítani, és kényszerleszállást kellett végrehajtani vele, amikor olyan sok utas kezdett el összevissza rohangálni az ülések között egy egér miatt, hogy az már veszélyeztette a gép stabilitását [147, 711].

A teljes érzelmi mezőből nemcsak az időben tartósabb affektusokat kell kizárni a cselekvéselméleti modellezés jelen szintjéről, szükség van egy további szűkítésre is. Az érzelmeket abból a szempontból nézve is rétegekre bonthatjuk, hogy milyen mértékben tekintjük őket biológiai-neurológiai értelemben meghatározottnak, illetve az érzelmek működésén belül milyen szerepet tulajdonítunk a kognitív folyamatoknak, valamint a reflexió jelenségének. Ezzel kapcsolatban is többféle tipizálási javaslattal találkozhatunk a szakirodalomban. Az egyik felosztás szerint léteznek alapérzelmek és összetett érzelmek. Az előbbiek érzelmi primitíveknek tekinthetőek, amelyeket nem lehet visszavazetni más érzelmekre, nem lehet őket felépíteni fogalmi értelemben vett alacsonyabbrendű érzelmek segítségével. Az összetett érzelmek pedig több alapérzelem keverékeként tekinthetők. Ezt az osztályozási logikát a későbbiekben még előveszem, amikor az alapérzelmekkel kapcsolatos elméleteket mutatom be röviden. Vannak, akik pozitív, illetve negatív érzelmekről (büszkeségről, jóindulatról, együttérzésről, illetve rosszindulatról, irigységről, haragról) beszélnek aszerint, hogy – valamilyen morális mérce alapján – helyesnek vagy helytelennek tartjuk-e őket [238]. Egy másik tipizálási javaslat szerint elkülöníthetünk alapérzelmeket és kognitív érzelmeket egymástól. Justin D'Arms és Daniel Jacobson a 'természetes érzelmek' és a 'kognitív hangolások' kifejezéspárt húzza rá erre a kettőségre [117]. Az alapérzelmek (természetes érzelmek) evolúcióbiológiai mechanizmusok termékei, amelyek célja, hogy az életben maradás szempontjából kulcsfontosságú külső ingerekre reagáljanak, funkciójuk a környezeti ingerek monitorozása, figyelemkeltés veszély vagy rivális megjelenése esetén. A kognitív érzelmek (kognitív hangolások) az alapérzelmek alá besorolható – származtatott – érzelmek, amelyek már valamilyen jellemző és meghatározó kognitív összetevővel is rendelkeznek. Például a honvágyat mint egyfajta szomorúságot akkor érezzük, amikor távol vagyunk az otthonunktól, a gyászt mint szomorúságot egy hozzánk közeli ember halála váltja ki, a szerelem egy konkrét ember iránti érzelmek bonyolult keveréke, amiben van várakozás, öröm, bizalom (és még sok minden más). Az alapérzelmekből kevés van, a kognitív érzelmekből sokkal több áll rendelkezésünkre,

hiszen bármilyen gondolatot összekapcsolhatunk egy természetes érzelemfajtával és ezzel egy új kognitív érzelemhez juthatunk.[1] A kognitív érzelmek reflexiót kívánnak, és elsődleges testi ingerek nélkül is létrejöhetnek, vagyis az elméből is származhatnak [340]. Amikor Antonio R. Damasio definiálja az elsődleges és másodlagos érzelmeket, akkor a tudatos reflektivitás minőségével különíti el az utóbbiakat az előzőektől [116]. A másodlagos érzelmeket a minket ért élmények alapján alakítjuk ki, tehát tanulással jönnek létre, de ehhez nyilvánvalóan tudatra, megismerésre, relflexióra van szükség. Mérő László – az alapérzelmek és kognitív érzelmek mellett – az érzelmek harmadik nagyobb csoportjaként hivatkozik az ún. emberi érzelmekre, amelyek sajátossága, hogy ezek már a szelf jelenlétét is megkívánják [340]. Amíg az alapérzelmek és a kognitív érzelmek (legalábbis egy részük) az állatokban is megjelenhetnek, addig az emberi érzelmek nyilván nem.

Az érzelmi mező rétegekre bontásának indokoltságához lehet újabb adalék az a tény, hogy bizonyos érzelmek igenis tanultak, nem pedig velünk született reakciók. A New Hampshire-i Egyetem kutatói Gina Mireault és John Sparrow vezetésével fél- és egyéves kor alatti csecsemőkkel végeztek kísérletet arra vonatkozóan, hogy a humort hogyan tanulják meg a gyerekek. A csecsemőket és szüleiket különböző helyzetekbe hozva, a videóra rögzített eseményeket elemezve arra az eredménye jutottak, hogy a csecsemők a szüleik reakcióiból tanulják meg, hogy mi a mulatságos, min hogyan lehet nevetni [360].

Bármennyire is érdekes és fontos az affektivitás többféle értelemben is sokrétegű jelensége, nem tudok és nem is akarok az érzelmi jelenségek mindegyik típusával foglalkozni. Elegendőnek tartom azt, hogy értelmezzem kicsit az indulati-érzelmi cselekvés jelenségét, valamint – abból a célból, hogy lehessen látni az utat, amin majd tovább lehet menni – kicsit foglalkozom még az alapérzelmek osztályozási és formalizálási kérdéseivel. Innentől kezdve amikor érzelemről, indulatról, affekcióról beszélek, akkor mindig ebben a – fent jelzett – leszűkített értelemben használom ezeket a terminusokat.

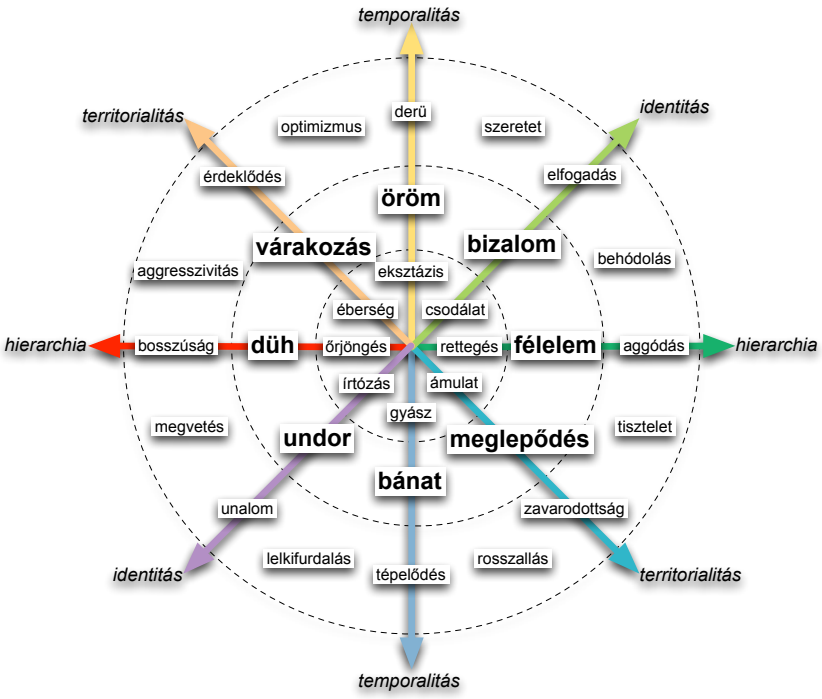
Bár a pszichológusok elég régóta vizsgálják az érzelem jelenségét, egyelőre nem alakult ki konszenzus sem abban a kérdésben, hogy vannak-e alapérzelmeink, sem abban, hogy ha vannak, vajon melyek azok. A cselekvéselmélet számára talán nem is annyira lényeges ez a kérdés, még akkor sem, ha az nyilvánvaló, hogy az érzelmek beemelése a racionális döntések elmélete által dominált diskurzusba komoly előrelépésnek tekinthető. Az alapérzelmekkel kapcsolatos konszenzus hiánya ellenére érdemesnek tartok pár tipizálási kísérletet bemutatni, legfőkképpen olyanokat, amelyeket esélyesnek látok arra, hogy képesek lehetnek megkönnyíteni az érzelmi jelenségek logikai formalizálására törekvő kísérleteket.

Az első ilyen osztályozási javaslat Robert Plutchiktól származik, aki átvette Darwin nézetét az érzelmekre vonatkozóan, miszerint az érzelmeknek adaptív funkciójuk van [417]. Amikor Plutchik az alapérzelmeinket megpróbálta tipizálni, akkor feltette, hogy négyféle életprobléma van, amelyekkel szembe kerülve minden állat, így az ember is, megfelelő érzelmekkel reagál. Plutchik szerint az alábbi négy alaphelyzettel szembesülünk életünk során:

- identitás
- temporalitás (reprodukció)
- hierarchia
- territorialitás

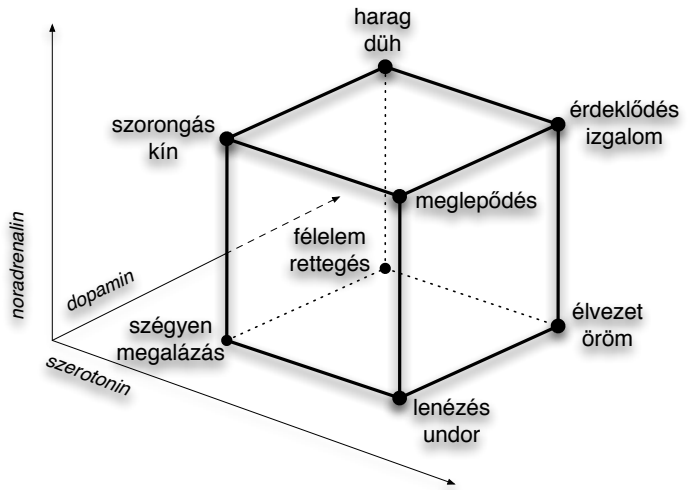
[1] A kognitív érzelmek létezése azért is fontos, mert az ilyen érzelmek feltételezésére támaszkodva már elemezni lehet a felelősség és az érzelmek közti kapcsolatot [437].

Plutchik szerint amikor a négy alaphelyzet valamelyikébe kerülünk, akkor a helyzet pozitív vagy negatív értékelésétől függően az adott helyzetre jellemző érzelempár valamelyik pólusán levő érzelem aktiválódik bennünk. Plutchik azt a nyolc alapérzelmet rendelte a fenti négy alaphelyzethez, amelyek az alábbi ábrán a tengelyek közepén láthatóak (nagyobb méretű betűkkel szedve).



Plutchik mind a nyolc alapérzelemhez hozzárendelt még két-két olyan érzelmet is, amelyek közel vannak az adott alapérzelemhez csak éppen erősebb vagy gyengébb intenzitással rendelkeznek. Az ábra középpontjához közelebb feltüntett érzelmek nagyobb intenzitásúak, a középponttól távolodva pedig mindnen irányban gyöngül az érzelmek hevessége. A legkülső körön már olyan érzelmek is látszanak, amelyek két alapérzelem keverékeként minősíthetőek Plutchik szerint. A négy alaphelyezet és az alapérzelmek közti viszonyrendszer érzékeltetésére pedig álljon itt egy rövid felsorolás, miként kapcsolja össze adolgokat Plutchik: a szomorúság a szeretett személy elvesztését, a félelem fenyegetettséget, a harag akadályt, az öröm potenciális társat, a bizalom csoporttagot, az undor förtelmes tárgyat, a várakozás új territóriumot, a meglepődés hirtelen tárgyat jelent. Warren D. TenHouten tiszteletben tartva és elfogadva Plutchik alaprendszerét még tovább lépett, és az alapérzelmekből további másod- és harmadrendű érzelemfogalmat vezetett le az affektív birodalomról szóló könyvében [542].

Izgalmas cikket publikált 2012-ben Hugo Lövheim az alapérzelmek biológiai meghatározottságával kapcsolatban [312]. Lövheim az általa megadott nyolc alapérzelmet aszerint tartja elrendezhetőnek egy háromdimenziós térben, hogy hogy viszonyul egymáshoz három neurotranszmitter aktuális szintje. Aszerint, hogy három neurotranszmitter, a **szerotonin**, a **dopamin**, valamint a **noradrenalin** szintje aktuálisan milyen értéken áll valakiben, Lövheim megmagyarázhatónak véli a nyolc alapérzelem keletkezését. Az alábbi ábra mutatja a neurotranszmitterek szintje és a lövheimi alapérzelmek közti összefüggést.



Lövheim tanulmánya még nagyon friss, elgondolásai épphogy csak bekerültek a tudományos diskurzusba, ezért nem árt némi óvatossággal kezelni az eredményeit, érdemes megvárni a szakma fogadtatását. Az azért elég nyilvánvaló, hogy amennyiben igazolódik Lövheim tézise, akkor az rendkívül erős alapot biztosít az érzelemelmélet számára.

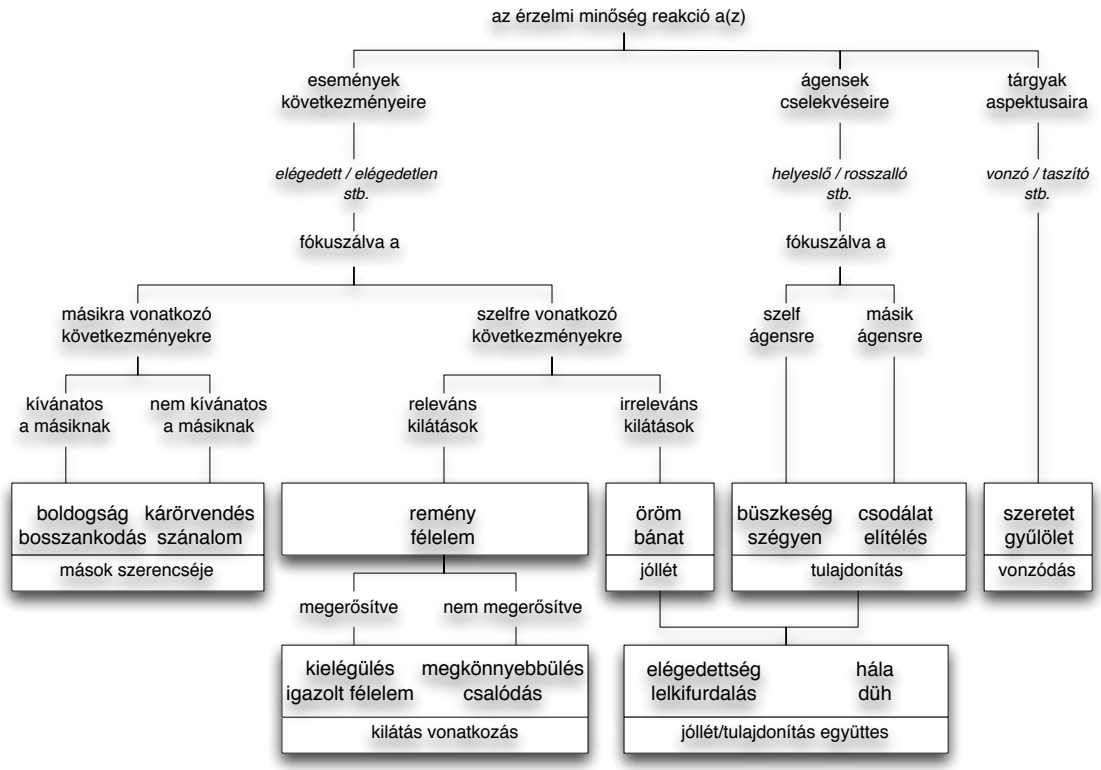
Mielőtt ismertetnék egy harmadik érzelemosztályozási javaslatot, előtte bemutatok egy összegzést arra vonatkozóan, hogy az érzelmekkel foglalkozó legjelentősebb szerzők közül kik milyen alapérzelmeket emeltek be saját elméletükbe. Warren D. TenHouten gyűjtötte össze és rendezte egybe ezeket az adatokat [542, 4], tőle veszem át az alábbi táblázatot.

Descartes (1647)			öröm	bánat				csodálat, szeretet, gyűlölet, vágy
Tomkins (1963)	félelem-rettegés	düh-őrjöngés	élvezet-öröm	szorongás-kín	érdeklődés-izgalom	meglepődés-megijedés	undor	szégyen-megalázottság
Plutchik (1980)	félelem	düh	öröm	bánat	várakozás	meglepődés	undor	bizalom
Osgood et al. (1975)	félelem	düh	(csendes) öröm	szorongás-bú	érdeklődés-várakozás	ámulat	unalom, undor	
Arieti (1970)	félelem	őrjöngés	kielégülés		feszültség, (ét)vágy			
Izard (1972)	félelem	düh	öröm	szomorúság	érdeklődés	meglepődés	undor	szégyen (szégyellősség, bűntudat), lenézés
Ekman (1980)	félelem	düh	boldogság	bánat		meglepődés	undor	
Emde (1980)	félelem	düh	öröm	bánat, szorongás	érdeklődés	meglepődés	undor	szégyen, szégyellősség, bűntudat
Scott (1980)	félelem	düh	gyönyörűség	magányosság, aggodalom				szeretet
Panksepp (1982)	félelem	őrjöngés			remény			pánik
Epstein (1984)	félelem	düh	öröm	bánat				szeretet
Trevarthen (1984)	félelem	düh	boldogság	bánat				
Johnson-Laird, Oatley (1992)	félelem	düh	boldogság	bánat			undor	
Turner (2002)	irtózás-félelem	követelés-düh	kielégülés-boldogság	csalódottság-bánat				
kínai tradíció	félelem	düh	boldogság	bánat				
Darwin (1872)	félelem	düh	boldogság	bánat		meglepődés	undor	
Lövheim (2012)	félelem-rettegés	harag, düh	élvezet, öröm		érdeklődés, izgalom	meglepődés	lenézés, undor	szégyen, megalázás

TenHouten gyűjtését kiegészítettem három további sorral, a kínai tradíció, Darwin és Lövheim típusaival. A táblázatból kiderül, hogy valóban nincs teljes konszenzus az alapérzelmeket illetően, de azért látszik az is, hogy valójában nincs igazán nagy eltérés a legfontosabb szerzők elképzelései között. A táblázat első nyolc oszlopa azokat az érzelmeket mutatja, amit Plutchik definiált, így azoknál a szerzőknél, akik valamelyik plutchiki alapérzelmet nem emelték be saját rendszerükbe, az adott oszlopban nem szerepel érték. Az utolsó oszlopba kerültek azok az érzelmek, amelyek nem illeszkedtek a plutchiki rendszerhez. Nem tudom, és nem is akarom itt az alapérzelmekkel kapcsolatos elméletek átfogó kritikáját elvégezni, de annyit azért megjegyeznék, hogy amennyiben az érzelmi jelenségek egyik alapminőségének tekintjük a poláris jelleget, akkor minden érzelemhez tartoznia kell valamilyen az ellentétes pólusú párnak, így azok a tipizálások, amelyekben ez nem valósul meg teljeskörűen, hiányosnak mondhatóak. Ebből a szempontból a plutchiki vagy a lövheimi megoldás jónak mondható, de a táblázatban jelzett érzelemelméletek többsége nem. A másik megjegyzésem csak annyi, hogy az alapérzelmek egy része (például a meglepődés vagy az érdeklődés) már meglehetősen távol áll az indulati-érzelmi cselekvésektől.

Az a harmadik érzelemelmélet, amit vázlatosan elemezni szeretnék, nem annyira az érzelemosztályzási javaslata miatt érdekes (bár ebből a szempontból is mutat érdekességeket), hanem inkább azért foglalkozom vele, mert ennek az elméletnek a logikai reprezentálásával jutottak el a legmesszebbre eddig az érzelmek formalizálásával próbálkozó mérnökök. R.W. Picard 1995-ben írt először arról, hogy megérett az idő arra, hogy elkezdjük az érzelmek mérnöki, számítógéptudományi vizsgálatát és a megfelelő logikai formalizmusok kidolgozását [412]. A kezdeti szórvány kezdeményezések után a 2000-es évek második felében jelentek azok a kísérletek, amelyek már komolyabb eredményeket értek el, bár ezek széleskörű kritikájához és befogadásához még nem telt el kellő idő. Andrew Ortony, Gerald L. Clore és Allan Collins az érzelmek kognitív struktúrájáról szóló könyve [400] alapján több elképzelés is megjelent az érzelmek logikai formalizálására vonatkozóan [2], [39]. Ezek közül itt a Bas R. Steunebrink, Mehdi Dastani és John-Jules Ch. Meyer által javasolt megoldást mutatom be vázlatosan. Átvéve Steunebrinkék megoldását én is OCC-vel hivatkozom a logikai formalizmus tartalmi forrását adó érzelemelméletre (a könyv szerzői nevének kezdőbetűi alapján). Az OCC-ben 22 érzelemre próbálnak kognitív magyarázó modellt adni.

Az OCC-modell szerint az érzelmek háromféle intencionális tárgyra irányulhatnak: események következményeire, ágensek cselekvéseire, illetve tárgyak aspektusaira. Ha ezt a három fő szempontot kombináljuk olyan további szempontokkal, mint hogy melyik ágensről a fokális személyről vagy a partnerről van szó, relevánsak-e, kívánatosak-e vagy sem a következmények, akkor eljutunk az OCC-elméleten belül tárgyalt érzelmekig. A kérdés az, hogy miként lehet az ábrán látható struktúra elemeit egy logikai nyelv segítségével reprezentálni. A Steunebrinkék által kifejlesztett logika az ún. KARO ágens-specifikációs keretrendszerre épül [355], [354]. Ezt azért érdemes megtenni, mert az érzelmek leírásához szükség van egyrészt ágensekre, másrészt a cselekvések értelmezéséhez kellenek a vélekedés, vágy, cél, terv, valamint a szándék fogalmai. Márpedig ezeket korábban mások már ellátták megfelelő logikai formalizmussal, amit érdemes újrahasznosítani, vagyis az érzelemlogika alapjául elfogadva egy alsó logikai rétegbe “betenni”.



Steunebrinkék első menetben "pontosítják", némileg átrendezik az OCC-modellt (ami tartalmilag nem jelent változást, csak a szintaktikai szempontokhoz könnyebben alkalmazható struktúrát biztosít a formalizálandó érzelemfogalmaknak). Ezután egy félig formális nyelven leírják az OCC-modell fogalmat. Azzal indokolják ezt a lépést, hogy ezen a szinten csak érzékeltetni szeretnék az OCC-modell fogalmainak szemantikáját, aminek alapján szabadon választható logika segítségével lehet formalizálni a modellt. Ezen a félformális tárgyalási szinten meg kell alapozniuk azokat a "segédfogalmakat", amelyek segítségével az érzelemfogalmakat definiálni lehet. Például fel kell venni a nyelvbe azokat a tulajdonságfogalmakat, amelyekkel értékeljük az események tárgyait (eseményeket, ágenseket/cselekvéseket, illetve tárgyakat). Amikor a jóság-rosszaság dimenziójában az események következményeit értékeljük, akkor azokat kívánatosnak (desirable) vagy nem-kívánatosnak (undesirable) tartjuk, amikor mások cselekvéseivel kapcsolatban ítélkezünk, akkor dicséretesnek (praiseworthy) vagy elítélendőnek (blameworthy) tartjuk azokat, amikor valamely tárgyat ítélünk meg, akkor vonzónak (appealing) vagy taszítónak (unappealing) véljük azt. A jóság (Good), illetve a rosszaság (Bad) fogalma alá tehát aszerint sorolható be három-három altípus (kívánatos, dicséretes, vonzó, illetve nem-kívánatos, elítélendő, taszító), hogy milyen típusú dologgal (eseménnyel, cselekvéssel, tárggyal) kapcsolatos. Ahhoz, hogy kifejezhessék az OCC-modellben leírt huszonkét érzelemet, Steunebrinkéknek – az itt érzékeltetett megoldások mintájára – tizenhét tulajdonságfogalmat kellett definálniuk.

Az OCC-modell fogalmainak félig formalizált jellemzése után következik a szerzők két logikai nyelv segítségével is bemutatják, hogy lehet szabatosan formalizálni az OCC-modell érzelmeit. Először dinamikus episztemikus logikát használnak, majd a KARO-

keretrendszert, ami a dinamikus logika, az episztemikus logika egyfajta keveréke, amit kiegészítettek még néhány, a cselekvések leírásához szükséges motivációs fogalmakat reprezentáló, modális logikai operátorral.

Paul Ekman szerint az (alap)érzelmek elsődleges funkciója "felkészíteni" az egyént az interperszonális helyzetek gyors kezelésére a múltban sikeresnek, adaptívnak bizonyult megoldások, tevékenységek "felidézésével, újrahasznosításával" [142, 46]. Ekman tehát – sok más pszichológussal együtt – osztja Darwin nézetét, miszerint az érzelmek, az affektív reakciók evolúciós eredetűek, és megtalálhatók az állatvilágban. Rovert B. Zajonc, egy másik híres érzelemkutató is hasonló módon vélekedett erről.

> "Az affektív válaszkészség – eltérően a nyelvtől vagy a gondolkodástól – általánosan megtalálható valamennyi állatfajnál. Egy nyúl egy kígyóval szembekerülve nem vizsgálhat meg minden szóbajöhető lehetőséget a kígyóval kapcsolatban, hogy azokból sikeresen következtessen a kígyó támadására, annak időpontjára és irányára. A nyúl nem állhat meg a kígyó hosszáról, méregfogáról vagy bőrének geometriai mintázatáról elmélkedni. Ha a nyúl menekül, ezt jóval az emlékezeti jegyek mérlegelése előtt kell megtennie; valójában az előtt, mielőtt megállapította és tisztázta volna az elkövetkező mozgást, amire egy összetekert kígyó képes lehet. A futási elhatározásnak minimális kognitív emlékek alapján kell születnie." [593, 16]

Az érzelmek funkcióját más irányból is, más megközelítés alapján is lehetne elemezgetni. Jack Hirshleifer például a közgazdaságtan felől mutatott rá az érzelmek egy, a racionális, önérdekkövető cselekvő számára hasznos funkciójára [238]. Hirshleifer szerint mind a pozitív, mind a negatív érzelmek szolgálhatnak arra, hogy segítségükkel a cselekvők kellő támogatást biztosítsanak fenyegetéseik és ígéreteik beteljesüléséhez. A közgazdasági szemléletmód persze meglehetősen ritkának mondható az érzelemkutatásban, így, bár érdekesek Hirshleifer gondolatai, nem fejtem ki őket alaposabban (azért sem tehetném meg itt, mert az értelmezésükhöz játékelméleti apparátus ismeretére is szükség lenne). Ahogy viszont azt már említettem, az érzelemkutatásban az (alap)érzelmek evolúciós felfogása széles körben elterjedt. Bizonyságul erre a következő idézet, amelyben Zajonc nyolvanas évek elejéről származó gondolatait harminc évvel később szinte visszhangozza Alan H. Goldman.

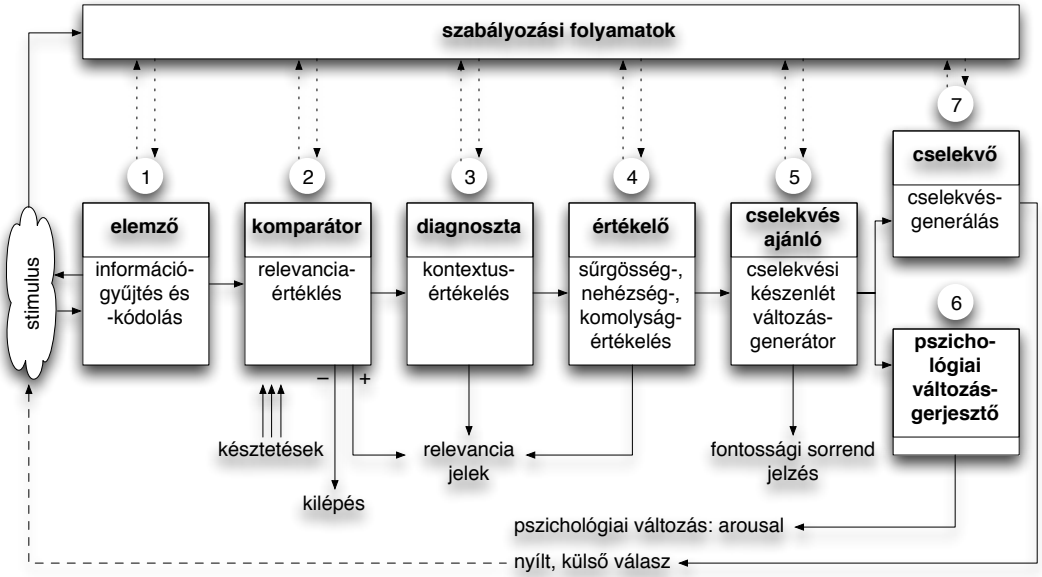
> „A paradigmatikus vagy alapvető érzelmek funkciója az, hogy (átmenetileg) felváltsák a cselekvési lehetőségek mellett, illetve azok ellen szóló érvek, szokásos, racionális mérlegelését. A paradigmatikus érzelmek robbanásszerű motivációs állapotok. Adaptív erejük abban az eseményekre való reakciókészségükben van, amikor a racionális kalkuláció, annak akár a leginkább automatikus fajtája is, túl lassú. Durva, de nagyon gyors kognitív értékelésekre támaszkodva olyan korlátos viselkedéskészletet aktiválnak, amelyek evolúciós értelemben sikeresek voltak a múltban. ...a félelem például a szokásos racionális mérlegelést kikerülve azonnali reakcióra késztet a veszély érzékelésekor. Ennek az adaptív ereje az azonnaliságban van. Több esetben sarkall hirtelen futásra, mint ahányszor valóban valós ok lenne a menekülésre, de adaptív szempontból nézve ez egy jó biztonsági stratégia. Jobb félni, mint megijedni."[193, 87-88]

A Goldman-idézet azért is érdekes, mert az érzelmi cselekvés több fontos összetevőjére találunk utalást benne. Itt az idő összeszedni és egy modellben bemutatni ezeket. Először is az érzelmeknek mindig van **tárgya**, mindig találnunk kell valamilyen intencionalitást, valamire való irányulást az indulati-érzelmi cselekvésekben. Az érzelemnek mindig van **valenciája**, érzelmi minősége, vagyis pozitív vagy negatív vonatkozása. Thomas Hobbes

szerint a cselekvést megelőző gondolatokat "rendszerint törekvésnek nevezzük", és ebből kétféle irányultságot származtathatunk:

> „Ez a törekvés, amikor valamire irányul, étvágynak vagy vágynak nevezik, amikor valamitől távolodni igyekszik, általában ellenszenvnek hívják." [239, 38]

Voltaképp ez a bennünk levő törekvés, vagy másként: a dolgokkal való törődés adja a valencia, az érzelmi minőség alapját. Ha van valaminek valenciája, akkor, mivel valamilyen rendezési művelet áll a dolgok hátterében, ahhoz a jelenséghez könnyen lehet mértéket rendelni, amikoris mód nyílik intenzitást mérni. Így van ez az érzelmek esetében is: az érzelmeknek van **intenzitása**. Ebből fakad az érzelmek **motiváló ereje**. Az érzelmek egy fő funkciója valamely helyzetre, stimulusra adott megfelelő és gyors **reakciókészség** kialakítása. Ez talán legfőképpen abban nyilvánul meg, hogy az egyén úgy reagál a helyzetre, hogy egy **korlátos viselkedési készletből** automatikusan, mérlegelés nélkül kiválaszt egy cselekvési mintát és azt végrehajtja. Az érzelmi folyamatok talán legalaposabb leírását Nico H. Frijda adta meg [178]. Érdemes itt felidéznünk azt, ahogy Frijda egyetlen folyamatábrába sűrítetve jelenítette meg az érzelmi folyamatok legfontosabb összetevőit.



Minden érzelemnek van egy külső és egy belső előfeltétele. Egyfelől kell egy külső inger, másfelől kell egy belső jelzésrendszer, ami a késztetéseket, a fontosságokat jelöli ki az egyén számára. A külső inger, a stimulus az, ami elindítja az érzelmi folyamatot. A stimulust az elemző (1) részleg észleli és dolgozza fel, és erről jeleket ad tovább a komparátornak (2), ami összehasonlítja a stimulust mint új ingert az egyén számára belsőleg adott értékkészlettel, a késztetések (érintettségek, törődések) rendszerével. Az összehasonlítás eredményeként kijöhet az, hogy a stimulus érdektelen. Ebben az esetben a folyamat leáll. Amennyiben a komparátor azt jelzi, hogy a stimulus valamilyen törődést érdemel, akkor a folyamat vezérlése átkerül a diagnosztához (3), ahol a kontextus is felmérve egyfajta relevanciaérték rendelődik a stimulushoz (helyzethez). A értékelő egység (4) ezután a sürgősség, a bonyolultság és a komolyság dimenzióira figyelve meghatározza azt, hogy mennyire kell komolyan venni az éppen adott helyzetet. Ez alapján egyfelől megjelenik egy fontossági sorrend jelzés, másfelől a cselekvésajánló egység felkészíti az új – és azon-

nali – cselekvésre az egyént, ami megváltoztatja az egyén belső állapotát, arousal szintjét (6), illetve egy korlátos cselekvési készletből merítve egy viselkedési választ generál (7). Ezt a folyamatot természetesen rengeteg szabályozási mechanizmus kontrollálja, és ezek a részmechanizmusok az egyes belső egységekkel kétirányú kapcsolatrendszerrel vannak összekötve.

A pszichológia területén a tudományos ismeretek közül a James-Lange elmélet biztos azok között van, amelyek szemben állnak a közgondolkodást uraló nézetekkel. A hétköznapi tudat számára egyáltalán nem tűnik kézenfekvőnek, hogy előbb vannak bennünk a testi változások, és ezek hatására érzünk majd érzelmeket magunkban – amiről a James-Lange elmélet legfőbb tézise szól. Ezt a tézist persze a tudományon belül is vitatják egyesek. Stanley Schacter és Jerome Singer 1962-es kísérletei szerint nem teljesen igaz a James-Lange tézis [386]. Schacter és Singer kísérleti eredményei arról szóltak, hogy a kísérleti alanyoknak befecskendezett stimuláló anyag attól függően váltott ki eufórát vagy dühöt a páciensekből, hogy a milyen mintákat tapasztaltak a környezetükben más embereken, vagyis az érzelmek kialakulására kognitív folyamatok is hatnak.

Az érzelmek testi vonatkozásaival kapcsolatban érdekes tanulmányt jelentetett meg 2013-ban Lauri Nummenmaa, Enrico Glerean, Riitta Hari és Jari K. Hietanen [384]. A kutatócsoport a kísérleti alanyok közreműködésével elkészítette az érzelmek kognitív hőtérképét. A kutatók különböző stimulusok (szavak, történetek, arcképek, filmek) segítségével érzelmeket keltettek a kísérleti alanyokban, majd arra kérték őket, hogy emberi alakot ábrázoló sziluetteken színekkel jelöljék be, hogy a testükön hol, melyik testrészükben észleltek erősebb érzeteket, érzéseket, illetve hol volt gyengébb szintje valamely érzetnek az átlagoshoz képest. Magasnak mondható korrelációs értékek mellett jöttek ki azok az eredmények, amelyek szerint a kísérleti alanyok a különböző típusú érzelmekhez nagyon hasonló, karakteres mintázatokat rendeltek. Mivel az eredmények a kísérleti alanyok saját jelentésein alapultak, ezért nehéz megmondani, hogy ezek a kognitív testi mintázatok, kognitív érzelmi hőtérképek mennyire a valós testi folyamatok modelljei, illetve mennyire a kognitív folyamatok kivetülései. Mindenesetre amíg – fizológiai mérések alapján – el nem készül a testünk valós érzelmi hőtérképe, ezek az adatok addig is érdekes mintát szolgáltatnak az érzelmek testi vonatkozásaival, a test és elme viszonyával kapcsolatos kutatások számára.

Akármit is gondolunk a test és elme viszonyáról az érzelmek kapcsán, állásfoglalásunk nem kardinális a cselekvéselmélet modellje számára. A cselekvéselmélet számára csak a fontos, hogy létezik egy olyan cselekvéstípus, az indulati-érzelmi cselekvés, amely jól elkülöníthető a többi típustól, és ebből a szempontból tekintve másodlagos kérdés az, hogy miként is zajlanak az érzelmi folyamatok az emberen belül. Nagyon oda kell azonban figyelnünk arra, amit Frijda az érzelmi folyamatok talán legfontosabb összetevőjének tartott. Nem véletlen, hogy amikor Frijda az érzelemet többször is, többféle hosszúságban definiálta, akkor a legrövidebb meghatározásaként csak annyit adott meg, hogy az érzelem nem más, mint a "késztetések kielégítése" [178, 473]. Ahhoz, hogy egyértelművé tehessem, mit is értett ezalatt Frijda, segíthet, ha – mintegy kontraszként – felidézzük, hogyan védte meg Jon Elster azt a tézisét, miszerint "érzelmekből áll az élet".

> „Az érzelmileg semleges tapasztalatnak, ha egyáltalán elképzelhető, nem volna semmi értelme. Érzelmek nélküli teremtményeknek nem volna miért élniük, de életüknek véget vetni sem." [145, 67]

Elster szerint az érzelmek adnak értelmet az életünknek, mégpedig azáltal, hogy irányultságot adnak neki (ahogy ezt a fejezet első mottójaként szereplő idézetben Elster szószerint ki is mondja). A kérdés itt csak az, hogy mindez hogyan valósul meg. A kulcsmozzanat – ahogy azt már említettem – Frijda **késztetés** fogalma. Mivel a Frijda által használt angol 'concern' terminus sokjelentésű, így nehéz jó magyar fordítást találni rá. A 'késztetések' mellett lehetne az 'érintettség', a 'törődés', a 'törekvés', az 'értékirányultság', az 'érdeklődés', 'érdekeltség', sőt, akár még az 'érdek' terminusát is alkalmazni itt. Végül – a másik lehetséges terminus, az 'érintettség'-gel szemben – azért döntöttem az 'késztetés' terminusa mellett, mert az 'érintettség'-et idegenebbnek éreztem ebben a kontextusban, a 'törődés, törekvés' terminuspárban erősebb a kelleténél a szándékoltság minősége, az 'érdek' terminuscsalád pedig annyira erősen kötődik az önérdekkövető magatartással kapcsolatos diskurzushoz (egyébként alaptalanul!), hogy nagyon zavarónak éreztem volna ezt az erős konnotációt. Amikor Frijda a késztetés fogalmát bevezette, jelezte, hogy tudatosan választotta a 'motiváció' vagy a 'cél' fogalma helyett ezt az új fogalmat elmélete középpontjába, majd röviden definiálta is, mit ért ezen a kategórián.

> „A késztetést olyan diszpozícióként definiálhatjuk, mint valamely helyzet megtörténte vagy be nem következése iránti vágyakozást." [178, 335]

Frijda a késztetést tehát **vágyakozási diszpozíciónak** tartja, amihez nyilvánvaló tisztáznia kéne, mit ért a vágy fogalma, de ezt nem teszi meg. Én korábban megpróbáltam tisztázni a fogalom jelentését (↪148.o.), itt abban az értelemben használom ezt a fogalmat, ahogy ott bemutattam.[2] Ha valamiben érintett vagy érdekelt vagyok, ha valamivel törődöm, akkor valamilyen vonzást (vagy taszítást) érzek a szóbanforgó dologgal kapcsolatban. A vágy és taszítás fogalomkettősében ez a kétpólusosság, a valencia, a valamire való pozitív vagy negatív jellegű irányulás jelenléte a meghatározó mozzanat. Mindez persze nem tűnhet annyira meglepőnek. Az viszont már sokkal inkább, hogy ez az értelmezés mennyire egybevág azzal, amit Pierre Bourdieu gondol az érdek fogalmáról.

> „Ahhoz, hogy megértsük az érdek fogalmát, látnunk kell, hogy az nemcsak az önzetlenség vagy az ingyenesség, de a *közömbösség* fogalmával is szemben áll. Közömbösnek lenni azt jelenti, hogy nem érint meg a játék, ... érdektelen számomra. A közömbösség olyan axiológiai állapot, amely egyfelől a nem-preferálás etikai állapotának felel meg, másfelől olyan tudásállapot, amelyben képtelen vagyok különbséget tenni az felmerülő lehetőségek között. Ilyesmi volt a sztoikusok célja is: az *ataraxia*, az aggódásmentes állapot elérése. Az *illusio* az ataraxia tökéletes ellentéte: amikor a helyzetben és a helyzet által figyelmet, érdeklődést mutatunk a lehetőségek iránt." [68, 116]

Ez a bourdieu-i érdekfelfogás éles ellentétben áll azzal a ténnyel, ahogy a racionalitást, a racionális cselekvéseket annyian és annyiszor akarták (és akarják) az **önérdek** (self-interest), önérdekkövetés fogalmával megmagyarázni. Sokak szemében az érdekvezérelt cselekvés a tökéletes ellentéte az indulati-érzelmi cselekvésnek, így az érdek kategóriáját az érzelmi cselekvések középpontjába tenni értelmetlennek és ellentmondásosnak tűnhet. Pedig nem az. Ezt rövidesen bővebben kifejtem. Nyugodtan használhatnánk tehát az érdek fogalmát is itt, de praktikus okok miatt nem érdemes. Azért nem (de csak azért nem), mert az önérdek és az önérdekvezérelt cselekvés fogalmának gyakori használata

[2] A vágy fogalmáról, a vágy és az érzelmek viszonyáról természetesen rengeteget írtak már eddig a történelemben. Nagyon tanulságos történeti áttekintést ad e fogalmak "múltjáról", történelmileg változó felfogásaik "eszmetörténetéről" Susan James [256], [257], Orthmayr Imre és Steiger Kornél pedig a szenvedélymentesség fogalmának szoikus értelmezését elemzi [396], [510].

során egyszerűen lekopott az önmagára való irányulást kifejező 'ön' szóösszetevő, és az érdek fogalmának értelmezése így valóban eltávolodott attól a felfogástól, amit fentebb Bourdieu kifejtett. Ez az átértelmez(ődés)és annyira széleskörű, hogy nem érdemes szembe menni vele.

A továbbiakban tehát a késztetés terminusát használva itt az idő bővebben kifejteni, mit is jelent pontosan az érzelmi cselekvés fogalma. Robert B. Zajonc fontos tanulmányát érdemes először megidéznünk.[3] Bármennyire is korszakos állításokat fogalmazott meg Zajonc, az alapfogalmakat nem különítette el pontosan egymástól egymástól. A mondanivalója szempontjából ennek nem volt akkora jelentősége, de azért érdemes megemlíteni, hogy Zajonc sajnos összemosta az értékítélet, az attitűd, a preferencia és az érzelem fogalmait egymással. Ezt jól szemlélteti az a mód, ahogy az alábbi idézetben felcserélhetően használja ezeket a fogalmakat (egy, a preferenciákról szóló tanulmányban).

> „Az egyszer már kialakult értékítélet nem hajlamos a megváltozásra. Az érzelem fennmaradásával és az impressziók alakulását jelentősen meghatározó érzelemmel kapcsolatos vizsgálatok, és az a tény, hogy az attitűdök látszólag megközelíthetetlenek a kommunikatív meggyőzés számára, mind az érzelem hatalmas erejét és tartósságát tanúsítják." [593]

Ez a terminológiai pongyolaság nem csökkenti annak a megkülönböztetésnek az értékét, amit Zajonc a **kognitív** és **affektív ítéletek** fogalomkettős bevezetésével teremtett meg. Miután alaposan bemutatta, hogy az érzelem (preferencia) alapú affektív ítéleteknek milyen sajátosságai vannak (elkerülhetetlenek, visszavonhatatlanok, a szelf által sugalmazottak, nehezen verbalizálhatóak, nem kell a gondolkodástól függeniük és külön válhatnak a tartalomtól, áthatolhatatlanok), Zajonc rámutatott a kétféle ítélkezés eltérő minőségére.

> „A kognitív ítéleteket az ingerek jelen levő minőségei vezetik: 'Ez a macska fekete', 'A camambert és a brie puha érlelt sajt'. Ezek az ítéletek az ingerekhez rendelt I-skálákon születnek ... Az affektív ítéletek azonban J-skálákon jönnek létre, ezek azok a skálák, amelyeken a különböző ingerek és a személy elképzelt preferenciapontjai együttesen helyezkednek el. 'Utálom ezt a fekete macskát' vagy 'A briénél jobban szeretem a camambert' – ezek az ítéletek a J-skálán vannak, így az affektív ítéletek mindig a szelfet érintik. Ezek határozzák meg az ítélkező álláspontját az ítélet tárgyához való kapcsolatában." [593, 18]

Zajonc itt a Clyde H. Coombs által bevezetett skálafogalmakra utal [105]. Ha viszont különböznek a kognitív és affektív ítéleteink, akkor meg kellene mondanunk, hogy miben is áll ez a különbség. Zajonc erre egy újabb fogalomkettőst vezet be, aminek értelmét akkor érthetjük meg igazán, ha gondolunk arra, hogy Zajonc azonos minőségeknek tekintette az érzelem és a preferencia fogalmait.[4] Abból indult ki, hogy a kognitív ítéleteink arra támaszkodnak, hogy a világból érkező ingereket megkülönböztjük (képesek vagyunk a diszkriminálásra), amikoris minél pontosabb diszkriminációra vagyunk képesek, annál finomabb reakciókra van esélyünk. Azokat a tulajdonságokat, amelyek lehetővé teszik ezt a külvilág ingereinek feldolgozása során, Zajonc **diszkriminandáknak** nevezte el, és ezekkel állította szembe az affektív téleteink során használatos ingertulajdonsá-

[3] Sokatmondó a tanulmány címe: 'Érzelem és gondolkodás (A preferenciák nem igényelnek következtetéseket' [593].

[4] Ezt szépen mutatja a következő idézet: „ ... nem alaptalan az a feltételezés, hogy sok döntésben fontosabb szerepe van az érzelmeknek, mint amit el akarunk ismerni. Gyakran áltatjuk magunkat, hogy egy racionális utat követünk és mérlegeljük az alternatívák pro és kontra érveit. De ez a konkrét esetben valószínűleg ritkán történik így. A leggyakrabban, amikor 'X mellett foglaltam állást', az nem jelentett egyebet annál, hogy 'kedveltem X-et'." [593].

gokkal, amikre ráhúzta az – általa alkotott – **preferenda** kifejezést. Utóbbiak azonban más minőséggel rendelkeznek az előbbiekhez képest.

> „Azok az ingertulajdonságok, amelyek segítenek minket a tárgyak és események elkülönítése, felismerése, kategorizálása során, nem feltétlen használhatók ezen tárgyak értékelésekor. Ha pedig a dolgok így állnak, akkor léteznie kell egy olyan tulajdonságosztálynak, amely könnyen kombinálódik az érzelmekkel, és ezáltal lehetővé teszi számunkra, hogy elvégezhessük ezeket az értékeléseket, hogy megtapasztalhassuk a vonzódást, a taszítást, a kellemességet, a konfliktust és az érzelem további formáit, valamint biztosítsák számunkra, hogy megjelenhessenek bennünk ezek az érzelmi reakciók szinte azonnal a szenzoros élmények felbukkanása után. Ezek a tulajdonságok lehetnek egészen durvák, elmosódottak, globálisak, így könnyen előfordulhat, hogy nem szolgáltatnak elégséges alapot a legtöbb kognitív ítélet számára – akár még a legegyszerűbb megismerés-felidézés számára sem. Abból a célból, hogy megkülönböztethessük ezt a tulajdonságosztályt a diszkriminandumoktól, a továbbiakban *preferendumnak* fogom nevezni őket.” [593, 159]

A Zajonc nézetét idővel egyre többen fogadták el. A követők közül Seymout Epstein elméletét emelném ki, aki ugyan új kifejezéseket vezetett be, viszont pontosabban definiálta voltaképp ugyanazt a kettősséget, amit már Zajonc is felállított. Epstein szerint kétféle megismerési, gondolkozási mód létezik, az intuitív-tapasztalati és a racionális-analitikai.

> „Nem szűkölködünk a bizonyítékokban arról, hogy a mindennapi életben az emberek két, lényegileg különböző módon értelmezik a valóságot. Az egyiket nevezhetjük intuitív, automatikus, természetes, nem-verbális, narratív vagy tapasztalati módszernek, a másikat analitikainak, mérlegelőnek, verbálisnak vagy racionálisnak.” [147, 710]

Epstein sokféle címkét megengedhetőnek tart a két típusra való hivatkozásban, az elsőre legtöbbször az intuitív és a tapasztalati jelzőket alkalmazza, míg a másodikra elsősorban analitikai címkét ragasztotta, de gyakran hivatkozik rá racionálisként is. Ezzel némileg nehezíti annak a – maga által is vallott – tézisnek az elfogadtatását, miszerint mindkét megismerési, gondolkozási mód tartalmaz racionális elemeket – igaz, az analitikus többet, mint a másik. Az intuitív-tapasztalai rendszerünk ősi, az evolúció terméke, ami lehetővé tette elődeinek a gyors reakciókészséget. Korábban is, most is nagyobb adaptivitást biztosít nekünk, hogy rendelkezünk intuícióval, ösztönös reakciókkat, zsigeri érzésekkel – például arról, hogy egy állatot szabad-e megközelíteni vagy a víz iható-e [496]. Idővel pedig kiépült a racionális-analitikai megismerési rendszerünk is, ami nagyon más elvek szerint működött a tapasztalai-intuitív rendszerhez képest. Epstein egy táblázatba összefoglalva mutatta be a kétféle megismerési, gondolkozási mód jellegzetességeit [147].
Az epsteini tulajdonságok részben tovább erősítik, részben kiegészítik az eddig elmondottakat a kétféle emberi működési mód különbségeit illetően. Egy-egy fontos mozzanatot már csak azért is érdemes kiemelni közülük, mert azok nagyon egybecsengenek más, ekkortájt híressé váló kortárs elméletek eredményeivel (Epstein egyébként hivatkozik is mindkettőre). Az egyik Allan Paivio duálkódolási (vagy kettőskódolási) elmélete a mentális reprezentációkról [402], a másik Jerome S.Bruner elképzelése a narratív gondolkodási módról.

Allan Paivio szerint a fogalmi és képi reprezentációk párhuzamosan és egyenrangú módon léteznek egymás mellett az emberben úgy, hogy természetesen a kétféle rendszer között léteznek egymásra mutató hivatkozások, azaz a fogalmi és képi reprezentációk össze vannak kapcsolva egymással. Mindkét alrendszer saját folyamatai vannak, a fogalmi alrendszert verbális, a képi alrendszert nem-verbális mechanizmusok működtetik.

holisztikus	1	analitikus
affektív: öröm-fájdalomorientált (mi jó)	2	logikus: értelemorientált (mi ésszerű)
asszociációs kapcsolódások	3	logikai kapcsolódások
a viselkedést a múltbeli tapasztalatokban gyökerező megérzések vezérlik	4	a viselkedést a történtek tudatos értékelése vezérli
a valóság konkrét képekben, metaforákban, narratívákban képeződik le	5	a valóság absztrakt szimbólumokban, szavakban, számokban képződik le
a feldolgozás gyorsabb, közvetlen cselekvésre irányul	6	a feldolgozás lassabb, késleltett cselekvésre irányul
a változások lassabbak, az ismétlődő, intenzív tapasztalatokhoz igazodnak	7	a változások gyorsabbak, a gondolkodás sebességéhez igazodnak
durvább differenciáltság, túlzó általánosítási hajlam, sztereotipikus gondolkodás	8	magasabbfokú differenciáltság
durvábban integrált, disszociatív, érzelmi komplexek, kontextusfüggő feldolgozás	9	magasfokon integrált, kontextusközi feldolgozás
passzív és tudatelőttes módon megtapasztalt: az érzelmeink irányítanak	10	aktív és tudatosan megtapasztalt: a gondolataink vezérelnek
magától értetődő érvényesség, a tapasztalatban való hiten alapul	11	logikán és bizonyítékon alapuló igazolást követel

A kettőskódolás elmélete a képek jelentőségét értékeli fel a más reprezentációs elméletekben domináns fogalmi reprezentációkhoz képest, és ennyiben azt az emberképet erősíti, amely az emberi megismerés, gondolkodás kettős természetén alapul. Paivionak a képi feldolgozás létezését, jelentőségét hangsúlyozó elmélet könnyen összeegyeztethető az affektív ítélkezés, megismerés zajonci vagy a tapasztalati-intuitív gondolkodás epsteini elképzeléseivel.

Amikor Epstein a tapasztalati-holisztikus "üzemmód" másik fontos jellemzőjeként említ a valóság narratívákban történő leképzését, akkor – többek között – Jerome S. Bruner elképzeléseit emeli be saját elméletébe. Ekkor már ismert volt Brunner elmélete a **paradigmatikus** (vagy propozícionális) és **narratív gondolkodásmód** kettősségéről [82]. A paradigmatikus gondolkodáshoz Bruner olyan jellegzetességeket kapcsol, mint az elvont összefüggések, a fogalmak, képletek, formulák keresése, kezelése, a matematikai, logikai, kísérleti tudáshoz kapcsolható igazságra vonatkozó szabályokhoz való igazodás. Emiatt az általánosításra törekvés miatt válik mindig egyre hangsúlytalanabbá a kontextus figyelembe vétele, sőt a paradigmatikus gondolkodás számára a kontextusfüggetlenség az igazi érték. Ezzel szemben a narratív gondolkodást mindig összekapcsolhatjuk a történetek, narratívák létezésével, illetve a narratívákhoz köthető intenció, szándék, indok, vélekedés, cselekvés fogalmakkal. Ebben az esetben nem igazságkritériumokhoz igazodunk, hanem a jóság, az életszerűség megragadása, az elképzelhetőség, a szubjektív tapasztalat hitelessége mentén tartunk valamit megfelelőnek. A történetekben mindig vannak szereplők, akik cselekszenek, fontos a cselekvők karaktere, a rendelkezésükre álló készletek, a cselekvések részletei, a bennük foglalt, bennük felismerhető vagy vélelmezett intenciók, hiszen ezek teremthetik meg azt a kontextust, ami alapján a történet értelmét "felfejthetjük". Ezért a narrativitás számára mindig fontos a kontextus.

A harmadik momentum, amit ki akarok emelni Epstein jellemzéséből, az az affektivitás, illetve az ehhez kapcsolódó ismérveknek a tapasztalati üzemmódhoz rendelése. A

passzivitás, a viselkedésnek a tudatelőttesből érkező, a múltbeli tapasztalatokban gyökerező megérzések alapján történő vezérlése olyan minőségek, amelyek létezését, működését akkor érthetjük meg, ha felidézzük Frijda érzelemmodelljét (és folyamatábráját). A késztetés, az érdek(eltség) fogalmának jelentőségét ezen a ponton érthetjük meg. Az érzelemmodell középpontjában ugyanis lennie kell egy olyan összetevőnek, ami az indulati-érzelmi működésmód, az indulati dinamikáját biztosító energiaforrásaként minősíthetünk. Nos, ezt ragadhatjuk meg a frijdai késztetés vagy a bourdieui érdek fogalmával. És, hogy ez hogyan működik, azt írja le – mégha nem is erre a célra jött létre – Antonio R. Damasio szociális marker elmélete [116].[5] Damasio nagyjelentőségű felismerése az volt, hogy a racionális gondolkodást úgynevezett szomatikus markerek segítik a döntéshozatalban úgy, hogy a korábbi tapasztalatainkat érzelmeken, zsigereken keresztül közvetítik számunkra, és ezeken keresztül el tudjuk kerülni az egyébként könnyen végtelen ciklusba futó döntésképtelenség állapotát. Először nézzük meg, hogyan definiálta Damasio a szomatikus markerek fogalmát. Szerinte ezek:

> „ *...a másodlagos érzelmek által generált érzések sajátos esetei.* Ezeket az érzelmeket és érzéseket *tanulás révén összekapcsoljuk bizonyos szcenáriók előrejelezhető jövőbeni kimenetével.* Amikor egy negatív szomatikus jelölő együtt jár egy bizonyos jövőbeni kimenettel, a kombináció vészcsengőként működik. Ha ehelyett egy pozitív szomatikus markert kapcsolunk össze a dologgal, az vonzerő jelzés lesz.” [116, 174]

A szomatikus markerek tehát úgy működnek bennünk, hogy erős élmények hatására elraktároztunk egy értékmozzanatot a valóság valamely összefüggéséről, és bizonyos feltételek teljesülésekor ezt az élményt testünk felidézi, azonnali testi jelzést ad, amely erősen befolyásolja, gyakran teljesen meghatározza döntéseinket. Ez funkcióját tekintve megegyezik azzal, amit Frijda mondott a késztetés (érdek) mint működési komponens értelméről. Az a damasioi gondolat, hogy a szomatikus markerek mint az affektív rendszerünk fontos összetevői elemi erővel hatnak a racionális gondolkodára is (előbbiek nélkül nem is létezhetne az utóbbi), itt most másodlagos (majd visszatérek rá a racionalitás tárgyalásakor). A szomatikus markerek azt építik be a késztetéseink, érdekeltségeink “affektív raktárkészletébe”, hogy milyen irányú, polaritású, illetve milyen erősségű élmények értek a korábbi cselekvéseink során, és a későbbiek során az új ingerek felbukkanásakor az érzelmi folyamat mechanizmusai már ebből a folyamatosan változó jelzéskészletet használva működnek (ahogy azt Frijda leírta). Újra hangsúlyozom, hogy Damasio szomatikus marker elmélete nincs teljes fedésben egyik eddig bemutatott elmélettel sem, jelentősége is inkább a racionális gondolkodásunkra vonatkozó magyarázóerejében van, itt csak a szomatikus markerek és a késztetés/érdekeltség fogalmai közti hasonlóságot próbáltam meg kihasználni.

Az a belső késztetés- és gátlásrendszer, amit Frijda a késztetés, Bourdieu az érdek, Damasio a szomatikus markerek fogalmával próbált megragadni, nem az indulati-érzelmi cselekvéstípushoz tartozik, és még csak nem is a teljes affektív mezőhöz vagyis az érzelmi jelenségek jóval tágabb köréhez, hanem az összes cselekvéstípushoz, vagy méginkább, a cselekvéshez általában. Nincs cselekvés vonzások és taszítások, késztetések és gátlások nélkül. Mérő László szerint Damasio szomatikus markerei másodlagos jelenségként épülnek rá a természet által bennünk kiépített alapérzelmekre úgy, hogy a megoldás egyáltalán nem értékelhető elegánsnak, meglepően barkácsolt jellegűnek mondható [340]. Abban

[5] Epstein elméletének első publikálásakor még nem ismerhette Damasio elméletét, hiszen ugyanabban az évben jelent meg mindkettejük munkája.

igazat adok neki, hogy a szomatikus markerek valóban rátelepülnek a már létező, biológiailag adott belső mechanizmusainkra, és ennyiben másodlagos jelenségek. A barkácsolt jelleget sem vitatnám. Csak azt látom másként, hogy mi is az az elsődleges rendszer, amire ráépülnek ezek második szinten. Szerintem nem az (alap)érzelemekre általában, hanem a biológiailag adott késztetéseinkhez, késztetésrendszerünkhöz képest másodlagosak. Az érzelmeinket is lehet szintekre bontani, de az másfajta felosztást jelent, mint a késztetések, érdekeltségek egész rendszerének két szintre osztása. Késztetéseinknek biztos van egy olyan rétege, ami biológiailag adott számunkra, és erre épülhetnek rá az életünk során a szomatikus markerek.

Érdemes még megemlíteni azt, hogy a bennünk zajló érzelmi folyamatok eddig bemutatott jellegzetességeiből további következtetéseket vonhatunk le. Paul Slovic például rámutatott arra, hogy a tapasztalati-intuitív alapú megismerési és viselkedési üzemmódunk értelmezhető úgy is, mintha **érzelmi heurisztikákat** dolgoznánk és működtetnénk magunknak [496]. Az affektív mechanizmusok ismertetett vonásai, durvább, de előre kidolgozott, gyors válaszadás képessége könnyebbé és sok esetben hatékonyabbá teheti a viselkedést, mint az előnyöket, hátrányokat mérlegelő, összehasonlítgató, a memóriában emléknyomok után keresgélő racionális döntéshozatali mechannizmus. Különösen akkor van ez így, ha komplex döntéseket kell hozni, miközben a rendelkezésre álló mentális erőforrások valamiért korlátosak. Ilyenkor jöhetnek jól az slovici érzelmi heurisztikák.

Az is igaz, hogy ha ezek a heurisztikák (affektív mechanizmusok, érzelemautomaták) a mindennapi életünk részei, akkor könnyen elképzelhető, hogy mások ezt kihasználhatják. Tudjuk jól, könnyen becsaphatóak, manipulálhatóak vagyunk. Sokszor (talán legtöbbször) azért, mert döntéseinkben a tapasztalati-érzelmi üzemmódra hagyatkozunk. Ezt ismerték fel és ezt használják ki a reklám- és marketinggépezetek. A viselkedésgazdaságtan területéről sok olyan példa hozható, amelyek azt szemléltetik, hogy hogyan képesek félrevezetni minket saját affektív-heurisztikus döntéseink (↪224.o.).

Persze, látni kell azt is, hogy az érzelmi heurisztikáinknak vannak komoly, lényegükből adódó korlátai. A villámgyors reakció evolúciós eredetű követelményének azon az áron lehet(ett) megfelelni, hogy az affektív gépezetbe a cél teljesítéséhez szükséges elfogultságok, egyoldalúságok épültek be. Úgy tűnik, hogy az affektív rendszerünk arra lett tervezve, hogy a környezetünk kis mértékű változásaira tudjunk érzékenyen reagálni, a nagyobb léptékű változásokat már nem tudjuk jól kezelni. Slovic – kissé morbid – példája erre az, hogy nagyon éles különbséget tudunk tenni 0 és 1 ember halála között, de annál inkább érzéketlenné válunk erre, minél nagyobb számú haláleset közti különbséget kéne "értékelnünk". Nem tudjuk sehogy sem kezelni az 500 és 600 haláleset közti különbséget. Ugyanezt fejti ki Szent-Györgyi Albert is – csak kicsit érzékletesebb megfogalmazásban.

> „A közös emberi szenvedés ugyancsak könnyen válik elvonttá. Engem személy szerint nagyon mélyen érint a szenvedés, a halál pedig felzaklat. Gyakran megkerülök egy háztömböt, hogy ne találkozzam egy temetési menettel; százezer ember halála azonban nincs rám semmilyen hatással. Csak mosolygok, mert képtelen vagyok megszorozni a halált vagy a szenvedést százezerrel. Ez nekem már csak puszta szám, absztrakció. Egy halál tragédia – százezer halál statisztika . . . " [533]

Tulajdonképpen ehhez hasonló jelenség az is, amikor a zsigeri érzéseket (éhség, szomjasság, addiktív sóvárgás) a jelenben erősnek érezzük, de a jövőbeni hatásukat, a jövőbeni cselekvéseinkre irányuló befolyásolási képességüket alábecsüljük. Egyáltalán: az időbeni folyamatokat, azok hatásait nem tudjuk igazán jól becsülni, előrejelezni. Ennek oka vél-

hetőleg az, hogy a tapasztalati gondolkodásmódunk a mostban, a jelenben zajló cselekvéseink megkönnyítésére jött létre. E korlátunknak egyértelmű példája az, hogy az egynapos differencia a ma és a holnap között sokkal nagyobbnak tűnik számunkra, mint ugyanaz az időtartamnyi különbség az 'egy év múlva' és az 'egy év plusz egy nap múlva' között.

Ugyancsak és ugyanazért tudjuk nehezebben kezelni az időben lassan zajló változásokat az érzelmi rendszerünkön keresztül. Ezzel kapcsolatban Slovic meggyőző magyarázatot mutat be arra vonatkozóan, hogy a dohányosok miért kezdenek el a dohányzásba, miközben tisztában vannak annak káros hatásaival. Vitatható módon ugyan, de a dohányzást meg lehet kísérelni racionális alapon is megmagyarázni [247], de a tapasztalati döntések figyelembe vételével más – és jobb – értelmezést is adhatunk. Amikor a dohányos a cigi után nyúl, akkor rövidtávú előnyöket, az azonnali élvezet lehetősége áll szemben a racionális rendszer által mérlegelhető hosszútávú kockázatokkal. Ebben az összehasonlításban az érzelmi heurisztika könnyen és gyakran felülkerekedhet a racionális mérlegelésen.

Az érzelmekről szóló gondolatmenetem elején jeleztem, hogy az indulati-érzelmi cselekvések az affektív mezőnek csak egy részét képezik, és vannak olyan affektív jelenségek, amelyeket már nem lehet az indulat cselekvés terjedelmébe sorolni. A hangulatok, a temperamentumok is érzelmi jelenségek körébe tartoznak, de már nem jellemzi őket az az egydimenziós, egyfókuszú irányultság, és ebből adódóan csak korlátozott szándékosság van jelen vagy esetleg nem is érhető tetten az akaratlagosság. Az indulati-érzelmi cselekvések intencionálisak, de szándékoltak. Viszont minden érzelmi jelenség mélyén ott van az érdekeltségek rendszere, amelynek tartalmát, összetételét éppen az affektív mechanizmusok frissítik. Ez a mentális energiaközpont azonban nemcsak az affektív jelenségek számára, de minden más ugyanazt a szerepet tölti be.

3.2 Habituális cselekvés

Nem véletlen, hogy Max Weber cselekvéstipológiájában utolsóként említi meg a **tradicionális cselekvés** típusát. Nem véletlen, mert Weber nem igazán szentel teret ennek a fogalomnak saját elméletén belül. Pedig sokkal több figyelmet érdemelne ez a cselekvéstípus, mert a hétköznapok világában nagyon sok esetben ez jellemzi a cselekvéseinket. Hogy ez így van, azt – igaz, csak egy mondat erejéig – még maga Weber is elismerte. Bizonyságul erre, meg egyébként is, érdemes teljes terjedelmében felidézni azt, amit Weber erről a cselekvéstípusról írt. Nem sok.

> A társadalmi cselekvés „. . . *tradicionális:* ha meggyökeresedett szokások határozzák meg a cselekvést." [575, 53]
>
> „A szigorúan tradicionális cselekvés – akárcsak a puszta reakciószerű utánzás . . . – teljesen a határán van annak – sőt, gyakran túl van azon –, amit egyáltalán »értelemhez« igazodó cselekvésnek lehet nevezni. Nagyon sokszor ugyanis csupán valamilyen homályos reagálást jelent a megszokott ingerekre, s a reakció irányát az egyszer meggyökeresedett beállítottság szabja meg. A megszokott mindennapi cselekvések többsége közel áll ehhez a típushoz, amelyet nemcsak azért kell fölvenni a rendszertanba, mert határesetet jelent, hanem azért is, mert . . . a megszokotthoz való kötődés különböző mértékben és többbféle értelemben tudatosan is fenntartható. Ha ez a helyzet, a tradicionális cselekvés a 2. pontban tárgyalandó [*értékracionális – Syi*] típushoz kerül közel." [575, 53:2.§.1.]

Azt már korábban jeleztem (↪178.o.), hogy bár ezt a kategóriát fontosnak tartom, és úgy gondolom, hogy egyértelműen ott a helye a cselekvéstípusok között, de az elnevezését szerencsétlennek tartom, és más javaslok használni helyette. Sajnos a **habituális cselekvés** kevésbé terhelt a fogalomhoz kapcsolódó, fölösleges konnotációkkal, bár ezt a kifejezést is "veszélyezteti" az etimológiai közelsége a **habitus** terminusához. Ezt azonban kisebb és kezelhetőbb problémának érzem, mint a tradicionális cselekvés és a tradíció fogalmai közti kapcsolatból fakadó nehézségeket. Ennél a fogalomnál is nagyon kell vigyáznunk arra, hogy a lehető legpontosabb értelmezést adjuk rá, és minél egyértelműbben elkülönítsük azoktól a jelenségektől, amelyek bizonyos szempontokból hasonlítanak rá, de mégsem egyek vele. A cselekvéstípus kucsminősége ugyanis az automatizmus, de ez olyan **cselekvésautomatizmus** (vagy automatikus cselekvés), amit el kell tudunk különíteni azoktól az automatizmusoktól, amelyek egyfelől az affektív cselekvéstartományt, másfelől a biológiailag meghatározott, reflexszerű viselkedéseinket jellemzik.

Ott tartunk tehát, hogy terminológiai szinten lecseréltem Weber 'tradicionális cselekvés' kifejezését a habituális cselekvés terminusára. Amikor a későbbiekben kitérek arra (↪222.o.), Weber hogyan értelmezte a **szokás** fogalmát, akkor majd kiderül, hogy a szokást a habituális cselekvés felettes fogalmának tartotta. Ebből az következik, hogy ha a habituális cselekvésekre ráhúznánk a szokás minőségét is, nagy hibát nem követnénk el vele. Ha úgy adódik, akkor ezt meg fogom tenni én is – leginkább stilisztikai kényszerek miatt. Mielőtt belekezdenék a habituális cselekvés fogalmának tárgyalásába, célszerűnek látom, ha előtte pontosítom, mivel is nem szabad összekevernünk ezt a jelenséget. A habituális cselekvés nem keverhető ösze Pierre Bourdieu habitus fogalmával, bár nagyon sok szempontból átfedésben van a két fogalom. Mindkettőre mondhatjuk, hogy társadalmilag kondicionált, **tanult automatizmusok** rendszere, de a bourdieui habitus fogalom sokkal több mindent magába foglal, mint amennyit egy koherens cselekvéstipologizáláshoz én szerencsésnek tartanék, ezért gondolom jobbnak a – sokkal szűkebb jelentéstartományú – habituális cselekvés fogalmát. Ezzel egy pillanatig sem támadom Bourdieu habituselméletét, sőt, nagyon relevánsnak, értékesnek tartom azt, csak modellépítési szempontokból nem tartom jónak annyi mindent egyetlen fogalomba sűríteni, ahogy ezt Bourdieu teszi a habitus kategóriájával. A habituális cselekvés nem azonos a Gilbert Ryle által tárgyalásba vont **diszpozíció** fogalmával sem. Bizonyos mértékig persze megintcsak átfedésben van a két fogalom, és mondhatjuk, hogy a habituális cselekvés egyfajta diszpozíció kialakulását, létezését feltételezi, illetve fordítva: a habituális cselekvés diszpozíciókat alakít ki és azokat tart életben. De a két fogalom többféle értelemben is különbözik egymástól. Bármennyire is termékenynek és relevánsnak gondolom Polányi Mihály **személyes tudásról** szóló elméletét [419], nem fogom tárgyalni itt, pedig a készségek és jártasságok, a hallgatólagos tudás jelensége sok szempontból nagyon egybevág azzal, amit a habituális cselekvésről mondani lehet, de annyira eltér a Polányi által alkalmazott fogalom- és szempontrendszer az általam preferálttól, hogy túlzottan nagy munka lenne összehangolni a kétfajta megköközelítést. El kell még választanunk azokat az értékelési, érzelmi automatizmusokként értelmezhető **affektív mechanizmusokat** is a habituális cselekvésektől, amelyeket korábban már tárgyaltam az indulati-érzelmi cselekvések elemzése során. Végül különbséget kell tudnunk tenni a habituális cselekvések és a reflexszerű testi viselkedések között, a testi reflexeket ugyanis biológiailag determinált, a gének által vezérelt automatizmusoknak kell tartanunk, míg a habituális cselekvéseket szerzett, tanult jelenségnek. Az emberi viselkedések világában is tetten

érhetjük azt a különbséget, amit a teleonómia és teleológia kategóriáival tudunk megragadni és elkülöníteni egymástól (↪147.o.), és itt is az intencionalitás minőségével lehet a reflexszerű (teleonómiai) és a rutinszerű (teleológiai) viselkedéseket elkülöníteni egymástól. Azt persze még meg kell mutatnom, hogy a habituális cselekvésekhez hogyan és miért lehet/kell hozzárendelnünk az intencionalitás jelenségét. Amikor a habituális cselekvés fogalmát pozitív módon akarom meghatározni, akkor olyan fogalmakkal kerülhetünk részleges átfedésbe, mint diszpozíció, észjárás, ethosz (szokás), éthosz (lelki alkat), jártasság, hajlam, implicit hit (hiedelem, előfeltevés, tudás), hallgatólagos tudás, belefoglalt tudás, implicit tudás, jellem, készség, képesség, habitus, hészix, rejtett tudáselem, személyes tudás, szokás, tudatalatti, rutin. Miután tisztáztam, hogy mi mindent nem tartozik a fogalom terjedelmébe, most már itt az idő megadni azt, hogy akkor mit is jelent valójában a fogalom.

Nehéz észrevenni, még nehezebb a súlyának megfelelő módon kezelni az embernek vagy talán minden élőlénynek azt a hajlamát, hogy az idő elteltével pozitív viszonyt (preferenciát) alakít ki a környezete tárgyaival, entitásaival szemben. A pszichológia ezt a jelenséget a **kitettség** (explosure) fogalmával ragadja meg [594]. Ezt a jelenséget még nem sorolnám a habituális cselekvés fogalma alá, mégis fontosnak tartom ezzel kezdeni utóbbi tárgyalását. A kitettség jelensége ugyanis azt jelzi (szerintem elementáris erővel), hogy hajlamosak vagyunk **megszokni** a dolgokat. Az ilyen beállítódásnak elég nyilvánvalónak látszanak az adaptív előnyei: ha a tartós együttlét során kiderül, vagy legalábbis úgy látszik, hogy a környezetünk tárgyai nem veszélyes, károsak ránk nézve, akkor ezeket érdemes pozitív jelzésekkel ellátni. A habituális cselekvések mögött is mindig érdemes lesz ilyen és ehhez hasonló előnyöket keresnünk, és a tartósan fennmaradni képes habituális cselekvések mögött vélhetőleg mindig, de legalábbis sokszor találni is fogunk ilyeneket. A habituális cselekvések másik – és egyébként legfontosabb – vonása az, hogy automatizmusként, mégpedig tanult automatizmusként működik. A habituális cselekvés mindig rutinná vált cselekvés. Észre sem vesszük magunkon, milyen sok viselkedési séma, "tudásmorzsa" válik rutinná bennünk. Testtartásunk, gesztusaink, járásunk (sőt, maga a járni tudás), az öltözködés, a tisztálkodás vagy bármilyen kézügyesség, az írástudás, fejszámolás, egyáltalán a számolás maga, a beszédtechnikák, az akcentusok vagy általában a nyelvtudás maga, a jobbkezesség-balkezesség technikái, a cipőfűzés, a tisztálkodás, öltözködés módja, az autóvezetés technikái, a katonai, sportbeli, művészeti, foglalkozási, szakmai fogások, készségek. Mindezek olyan elemi fogásokból, rutinokból épülnek föl, amelyeket – miután megtanultuk őket – nem tudatosan irányítunk. Egyszer meg kell tanulni őket, tehát a kezdetekben nyilvánvaló a cselekvés intencionális jellege, de ha már készséggé vált bennünk, akkortól kezdve automatizmusként működik. A legnyilvánvalóbb példa erre az autóvezetés begyakoroltatása. Az elején mindenre figyelnünk kell, aztán egyszercsak már minden megy magától úgy, hogy oda sem figyelünk rá. Nem hiszem, hogy ez lenne itt a lényeges kérdés, mennyire fontos is ez a jelenség az életünkben. hanem sokkal inkább az, hogy vajon cselekvéseink mekkora hányadát teszik ki az ilyen tanult automatizmusok. Nem keveset. Mégha nem is könnyű erre mérési metodikát kialakítani, azért legalább nagyságrendjét tekintve elfogadhatjuk Bas Verplanken és Wendy Wood becslését, miszerint a mindennapi cselekvéseink negyvenöt százaléka a szokás, habituális cselekvés fogalma alá sorolható [560].

A jelenséget előbb az antropológia, szociológia, majd kicsit később a pszichológia felől kezdték vizsgálni. Elsőként Marcel Mauss foglalkozott vele, aki testtechnikáknak

(body techniques) nevezte el a testbe épült rutinokat [325], majd őt követve Norbert Elias [143] és Pierre Bourdieu [67] a **habitus** fogalmának alaposabb kidolgozásával. Ezen a szemléleti vonalon a testtechnikák kulturális-társadalmi meghatározottságára fókuszáltak elsősorban. Mauss hivatkozott például az úszástechnikák különbségeire a nemzetek és időszakok között, aminek kulturális meghatározottságai egyértelműen kimutathatóak voltak. Ha azonban egy másik sportágra figyelünk, mondjuk a magasugrás technikáinak változásaira, akkor beazonosíthatunk egy kultúrafüggetlen komponenst is a jelenségen belül. A magasugrást a kezdetekben az ollózás, majd a hasmánt, végül a flop technika uralta, és azekben a betanult rutinszerű mozgásokban már egyáltalán nem látszik lényegesnek társadalmi, kulturális meghatározottság ténye. Talán már nem is létezik. A két példa szembeállítása csak azért fontos itt, mert segítségével rámutathatok arra, hogy érdemes elválasztani két kérdést egymástól. Az egyik oldalon ott a kérdés, hogy hogyan alakulnak ki és hogyan működnek az ilyen rutinok a cselekvő emberben, és természetesen a másik oldalon ott vár minket a feladat, hogy megmagyarázzuk, mindez miért, hogyan alakult ki, illetve miért maradt fent.

Mausshoz visszatérve érdemes idézni tőle azt – a jelenséget nagyon jól szemléltető – példát, miszerint az első világháborúban minden egyes alkalommal, amikor az angol és francia csapatok egymást váltották, hadosztályonként nyolcezer ásót kellett kicserélni, mert másfajta formát szoktak meg (és tudtak használni) az angol, illetve francia katonák [325, 427]. Aztán további példaként Mauss említést tett arról, hogy "az emberiség elég jól kettéválasztható bölcsős és bölcsőtlen népekre" (437), az alvástechnikákról szólva megemlítette, hogy vannak "gyékényes vagy gyékény nélküli népek, illetve kispárnás vagy kispárna nélküli népek" (438), a pihenéstechnikák kapcsán elkülönített "guggoló és ülő népeket" (439), és – őt követve – nyilván még sokáig lehetne folytani a sort az étkezési technikák, evőeszközök, alapanyagok, fűszerek használata, a főzési szokások stb. felemlegetésével. Mauss megpróbált magyarázatot is adni minderre, és bár az – szerintem – még az általa felsorolt példák mindegyikére sem érvényes, azért tanulságos itt megidézni.

> „. . . az önmagukat meghatározott eredmény elérése érdekében szelektáló emberfajták nevelésének a fogalma a történelemnek is alapvető mozzanata: a látás tanítása, a járás, földmászás, lemenés, futás tanítása. Ez kiváltképp a higgadtság elsajátításában áll. Ez pedig elsősorban késleltetési mechanizmus, az összehangolatlan mozgások megakadályozása; a késleltetés pedig lehetővé teszi az összehangolt mozdulatok immár összehangolt válaszát, amelyek így az akkor megválasztott cél irányába hatnak. Az embert megrohanó nyugtalansággal szembeni ellenállás a társadalmi és mentális életben alapvető dolog." [325, 445]

Van szépség és biztos, hogy van egy nagy adag igazság is abban, amit itt Mauss ír, de azt nem hiszem, hogy minden habituális cselekvés fogalma alá sorolható szabályszerűséget így lehetne és kellene megmagyaráznunk. Még kevésbé fogadható el vagy talán pontosabb úgy fogalmaznom, hogy még kevésbé lényeges számomra az, ahogy Mauss összekapcsolja a társadalmat és egyént ebben a jelenségben.

> „A társadalomnak köszönhető, hogy a tudat beavatkozik. Nem a tudattalannak köszönhető, hogy beavatkozik a társadalom. A társadalomnak köszönhető, hogy létezik a kész mozdulatok biztonsága, a tudat uralkodik az érzelmen és a tudattalanon." [325, 445-446]

Cselekvéselméleti nézőpontból csak első szinten azt kell tudnunk elkülöníteni egymástól, hogy milyen típusai vannak a cselekvésnek, és mivel a cselekvés fogalmát a hozzá kapcsolat szubjektív értelemmel, az intencionalitással magyarázhatjuk meg, ezért ezen a szinten

nincs még szükségünk a válaszolni arra, hogy miért is alakulnak ki, miért is léteznek egymás mellett az egyes cselekvéstípusok. E programhoz képest Pierre Bourdieu habitusértelmezése sokkal bővebb, sokkal ambíciózusabb, de ezt a kiterjesztést itt én szükségtelennek – és magam szempontjai szerint, sajnos, kicsit zavarónak – tartom. Újra hangsúlyozom: mindezt csak azért említem, mert szeretném indokolni, hogy az itt tárgyalt cselekvéselméleti modell számára miért "sok" Bourdieu habitusfogalma. Természetesen Bourdieu elemzése zseniális, mint ahogy frappánsak az olyan kifejezései is, mint "a habitus nem más, mint tudattalanított rutin" [418] vagy "a habitus társadalmasított szubjektivitás", [68, 126]. Bourdieu magára a fogalomra a következő meghatározást adta.

> "A habituson olyan tartós és átvihető hajlamok rendszerét értjük, amely, valamennyi múltbeli tapasztalatot beépítve, minden pillanatban mint *észlelési, értékelési és cselekvési minta* működik. [67, 213]

Ebbe a definícióba tartalmilag nem tudok belekötni, bár nem tartom elég pontosnak. Mielőtt megadnám azt, hogy mit hiányolok belőle és hogyan javasolom kiegészíteni, szeretném megmutatni egy újabb idézet segítségével, hogy mi is az igazi kifogásom a buerdieu-i felfogással szemben. Bourdieu szerint a habitus egyszerre a társadalmi gyakorlatok eredménye (opus operandi), illetve a társadalmi gyakorlatok működésmódja (modus operandi), ebben az értelemben azok fenntartója. Ebben igaza van, mint ahogy a következő szövegrésszel sem tudnék, de nem is akarnék vitatkozni:

> „ ... a habitus nem más, mint emberi természetté vált történelem, vagyis második természetté vált és így letagadott történelem. A „tudatalatti" valójában nem más, mint a történelem elfelejtése, ami abból fakad, hogy a történelem a habitusokba, e kvázi természetekbe ágyazza bele az általa létrehozott struktúrákat." [67, 215]

Bármennyire is elfogadom az idézet minden pontját, alkalmazhatóságát vitatom. Vitatom azon az alapon, hogy ez az értelmezés észrevétlenül átvisz minket a módszertani individualista megközelítésből egy olyan holisztikus szemléletmódba, ami egy cselekvéselmélet számára nem jelenthet adekvát nézőpontot. Igaza van (lehet) tehát Bourdieu-nek mindabban, amit mond, de minden gondolata inadekvát, ami nem a módszertani individualizmus talaján mozog. Márpedig amikor letagadott történelemről, osztályokról, természetté vált történelemről, szimbolikus hatalom érvényesítéséről az osztályhabitusokon keresztül, és ezekhez hasonló fogalmakról beszél Bourdieu, akkor már messze eltávolodik a módszertani individualizmus szintjétől.

A bourdieu-i habitusfogalom egy másik értelemben is szélesebb terjedelmű ahhoz képest, mint ami elfogadható számomra. Az ő habitusfelfogása ugyanis nem érzékeny arra kérdésre, hogy milyen mechanizmusok révén alakul ki és marad fent a habitusba rögzült szabályszerűség. Számára oda tartozik az is, amikor az adott viselkedési mintázat valóban a puszta ismétlődés eredményeként jön létre, és az is, amikor a mintázat létét az attól való eltérést követő társadalmi szankció biztosítja. Weber fogalmaival leírva ugyanezt: a bourdieu-i habitusfogalom alá besorolható a weberi szokáshoz igazodó cselekvés, illetve a weberi rendhez igazodó cselekvés kategóriája egyaránt. Sőt, még a bourdieu-i értelmezésbe belefér az érzelmi cselekvések során beégetett értékelő jelzések, a szomatikus markerek teljes jelensége is. Ezt az értelmezést – saját modellem számára – túl tágnak tartom.

És akkor most lássuk, hogy szerintem hogyan lehet értelmezni a tanult automatizmusok jelenségét, hogyan működnek valójában ezek a folyamatok, és hogyan lehet őket jellemezni a többi, részben hasonló jelenséghez való viszonyukban. Az egyik nagyon

fontos vonása a habituális cselekvéseknek az, hogy van bennük egy olyan "gazdaságossági" mozzanat, ami elementáris erővel meghatározza a létüket és működésüket. Ezek az automatizmusok az agyunkat sokszor komoly mértékben tehermentesíteni tudják, és ez a képességük az, amivel a leginkább magyarázhatjuk az ilyen rutinok kialakulását és fennmaradását. Patkányokon, majmokon végzett mérések szerint a szokások kialakulása előtt a később szokásba rögzülő tevékenységsor még teljes agyi igénybevétel mellett zajlik, majd a szokás kialakulása után a rutinszerűen végrehajtható viselkedési elemek már nem dolgoztatják az agyat. Ez az agy számára rendkívül gazdaságos, mert egy ilyen viselkedési séma levezénylése sokkal kevesebb munkát igényel tőle. A rutineljárásokat az agy a legősibb területeken tárolja el, ahonnan már tudatos kontroll nélkül is irányíthat minket. Charles Duhigg megfogalmazásában: cselekvéseink jelentős részét "ciklikus agyi folyamatok", szokáshurkok (habit loop) vezérlik [136]. E kifejezéssel Duhigg a viselkedést kiváltó jel, a végrehajtandó rutin és a tevékenység révén megszerezhető jutalom ciklikusságában megnyilvánuló automatizmust jelöli.[6] A lényeg az, hogy a testünket gyakran trenírozzuk arra, hogy gépként hajtson végre valamilyen – többé-kevésbé bonyolult – mozgásfolyamatot.

Vélhetnénk úgy, hogy egyáltalán nem véletlen az, hogy Mauss még testtechnikáknak nevezte az általa elemzett tanult automatizmusokat. Ha valahol van valamilyen értelme valamilyen tevékenység ismételt megvalósításának úgy, hogy minél kisebb mértékben kelljen terhelni a tudatos agy feldolgozókapacitását, a figyelmet, akkor érdemes kialakítani ezeket az automatizmusokat. Ezek az automatizmusok a testi folyamatok esetén látványosak és egyértelműen működőképesek. Azonban koránt sincs arról szó, hogy csak testi mozgásokra lehetne ezt alkalmazni. Az előszóban már idéztem Neumann Jánost, aki azt mondta, hogy „ ... a matematikában az ember nem megérti a dolgokat, hanem megszokja", aztán ennek a fejezetnek az elején olvasható az a Whitehead idézet, ami szerint a mentális folyamatainkban is rengetegszer rögzített, tehát automatikus gondolatmeneteken, hivatkozásokon keresztül haladunk előre, végül álljon itt még egy idézet e tárgyban.

> „A szabályok létét az indokolja, hogy az emberek nem képesek erőfeszítések nélkül és hibátlanul alkalmazni az utilitariánus elvet; szükség van rá, hogy időt takarítsanak meg, és útjelzőket tűzzenek ki." J. Rawls [433, 387]

A tanult automatizmusok nemcsak testi, de mentális folyamatok esetében is előfordulhatnak. Az eddig említett példákon túl gondoljunk csak az anyanyelven, majd az idegennyelven való megszólalás vagy az olvasás, az írás, a számolás, a kottaolvasás, a szakszövegolvasás és még ezernyi mentális készségre mint a tanult mentális automatizmusok példáira. Ezeket is rengeteg gyakorlással, trenírozással, belső szokáshurkok kialakításával építjük be magunkba, és miután már készségeinkké váltak, nem kell figyelmet szánnunk rájuk – vagyis minden ugyanúgy történik, mint a "hagyományos" testtechnikáink esetében.

Bizonyos értelemben mondhatnánk, hogy nemcsak a habituális cselekvések tekinthetők automatizmusoknak, de az indulati-érzelmi cselekvések is – abban az értelemben legalábbis, hogy egyik cselekvéstípusban sincs mérlegelési, döntési lehetőség, s így nincs alternativitás sem a cselekvésben. Ehhez meg kell vizsgálnunk, mit is jelent az automatizmus fogalma. Mielőtt ezt megtenném, rá kell mutatnom egy nagyon lényeges különb-

[6] Duhigg és mások könyvéről bővebben írtam az aracionális jelenségek racionalitása témakörén belül [521].

ségre a habituális és az indulati-érzelmi cselekvések között. Ha felidézzük azt a Mauss-idézetet, amelyben a testtechnikák értelméről volt szó (↫199.o.), akkor emlékezhetünk rá, hogy ott olyan kifejezések bukkantak fel, mint 'higgadtság elsajátítása', 'késleltetési mechanizmus az összehangolatlan mozgások megakadályozására' vagy 'nyugtalansággal szembeni ellenállás'. Mauss (és mindenki más, aki utána a habitusról értekezett) érték- és érzelemmentes automatizálásról beszélt, és ezáltal egy olyan vonást emelt ki és tartott jellemzőnek, ami pont "ellentétes" az érzelmek egész világával. Ez fontos különbség. A habituális cselekvések, a tanult automatizmusok – a jólismert ryle-i fogalmat használva – **tudni hogyan** jellegű készségek kialakítását és működtetését jelentik. Az indulati-érzelmi, illetve a habituális cselekvés abban tehát megegyezik, hogy nem a tudatos, szándékos irányítás, tehát nem a kontroll jellemzi őket, hanem inkább az automata jelleg, az viszont jelentős eltérés köztük, hogy az előbbiek az értékirányultságaink, utóbbiak a – tudni hogyan típusú – készségeink kezelésével kapcsolatosak.

És akkor következzék az automatizmus fogalmának elemzése. Automatizmuson itt nyilván emberi cselekvések automatikus jellegét kell értenünk, amit érdemes szembeállítani egy másik cselekvéstípussal. Ha az automatikus cselekvés fontos minőségének érezzük a tudatosság hiányát, akkor ezzel szemben definiálhatjuk azokat a cselekvéseket, amelyek tudatosan vezéreltek. Daniel M. Wegner és John A. Bargh a **kontrollált** vagyis tudatosan irányított és az **automatikus** vagyis nem-tudatosan vezérelt folyamatokat elemezte egy tanulmányában [577], amiből sokat tanulhatunk. Mielőtt bemutatnám, milyen tulajdonságokkal jellemezték Wegnerék a kontrollált folyamatokat, meg kell adnom azt, hogyan különítették el a folyamatok két típusát egymástól. A mérnökök világából jólismert fogalmakat átvéve azt mondhatjuk, hogy az automatikus folyamatokat a **nyitott hurok**, az **előretáplálás** (open-loop/feed-forward),[7] a kontrollált folyamatokat **zárt hurok**, a **visszacsatolás** (close-loop/feedback) jellemzi. A zárt-hurkú, előretáplált folyamatokat Wegnerék a következő tulajdonságokkal jellemezték:

- tudatos szándék irányul a folyamat révén elérendő célra
- a kontroll érzése, észlelése
- a várt eredmény eléréséhez szükséges "költséges" erőfeszítés
- visszacsatoláson alapuló ellenőrzés a kimenet fellett

A kontrollált folyamatban mindig tudatos szándék érhető tetten, és erre a szándékra folyamatosan reflektálni tudunk. A folyamatnak van valamilyen tervezett, szándékolt kimenete, aminek elérése érdekében folyamatos figyelmet kell szentelni a folyamat összetevőinek. Ez abban az értelemben költséges, hogy a tudatos figyelem fenntartásához erőfeszítéseket kell tenni, ami azt is jelenti egyben, hogy más folyamatoktól kell elvonni a korlátos kapacitásokat. A folyamat irányításához hozzátartozik a folyamat monitorozása, ellenőrzése, és, ha szükséges, az eredeti szándék részleges vagy teljes átértelmezése. Wegnerék nem emelik ki külön tulajdonságként, de könnyen belátható, hogy az irányított folyamatok nem feltétlen gyorsak, sőt, minél összetettebb szándéktulajdonítást fedezhetünk fel egy folyamatban, annál biztosabban számíthatunk arra, hogy a folyamat lelassul, mert a tudatos kontroll igénye egyre nagyobb terheket ró a monitorozásra, mérlegelésre, döntésre kényszerülő cselekvőre. A kontrollfolyamatok központjában a visszacsatolás (zárt hurok) elve áll. A kontrollált rendszer megváltoztat valamit (előállít egy kimentetet) valamely cselekvésen (kontrollműveleten) keresztül minden olyan esetben, amikor felfedezi (a mo-

[7] A 'feed-forward' kifejezést magyarra leggyakrabban az 'előrecsatolt' terminussal fordítják, de jobbnak tartom az 'előretáplált' kifejezést, mert – érzésem szerint – ez jobban kifejezi a folyamat "egyirányúságát".

nitoring folyamat eredményeként), hogy eltérés (diszkrepancia) van a folyamat aktuális kimenete és a kívánatos állapot (a kontrollkritérium) közöttm miközben a kontrollkritériumait a rendszer bemeneti feltételei szabják meg, amelyek sokféle forrásból származhatnak (a rendszer környezetétől kezdve bármely más kontrollrendszerig bezárólag).

Nyilván nem meglepő, ha a kontrollált folyamatok legfőbb vonásainak pont az ellenkezője húzható rá az automatikus folyamatokra. Wegner és Bargh szerint az automatikus folyamatokat három tulajdonság jellemzi, amihez – az összehasonlíthatóség kedvéért – hozzátennék egy negyediket.

- autonómia (miután a folyamat elindult, nincs szüksége tudatos irányításra)
- gyorsaság
- hatékonyság (kevés figyelemkapacitásra van szüksége)
- előretáplált működésmód

Az automatikus folyamatok az előretápláltságuk (vagyis a nyitott hurok elve) felől érthetők meg. A folyamat elején szükség van egy tudatos aktusra, hiszen el kell indítani a folyamatot, de miután ez megtörtént, a folyamat nem igényli a tudatos figyelem fenntartását. Ez a "szabadjára engedett működésmód", ez az autonómia a legfőbb értelme (és minősége) az automatizmusnak. Ez magyarázza meg, miért mondják azt Wegnerék, hogy az automata folyamat hatékony. Azért, mert nem igényel különösebb agyi kapacitást, figyelmet. Ugyanennek köszönheti a gyorsaságát is, hiszen nincs szükség monitorozásra, összehasonlításra, mérlegelésre, döntésre, amelyek mindegyikének kisebb-nagyobb időszükséglete lenne, ha megjelennének. A figyelem mint erőfeszítés tehát döntő minősége a tudatosságnak.

> „. . . a kontrollt nemcsak pszichológiai folyamatotként kell érteni, de úgy is, mint egy érzést: annak tapasztalatát, hogy az ember valamit kontrollál vagy épp kontroll alatt van. A kontroll erőfeszítést követel meg, és pont ezen erőfeszítés elvárása teremt egy olyan érzést, hogy az ember csinál valamit, nem pedig csak hagyja történni a dolgokat. A tudatos kontroll során ugyanúgy maximálisnak tűnik a kognitív erőfeszítések tényleges ráfordítása, mint a mindent kitöltő erőfeszítés érzékelhető tapasztalata." [577, 458]

Az a tény, hogy a figyelem nagyon korlátos kapacitás, az információfeldolgozásban is "megköveteli és kitermeli" az automatikus folyamatokat [487], [463]. W. Schneider és R.M. Shiffrin kísérleteiben kiderült, hogy a kísérleti alanyok egy begyakorlási folyamat után tizenhat stimulus feldolgozásával is ugyanúgy megbiróztak, mint amikor a kísérletek elején csak hat vagy kilenc stimulust kaptak, amiből az következik, hogy az információfeldolgozásban is vannak automatikus folyamatok. Feltehető a kérdés, vajon a két alapfolyamat jellemezhető-e, elkülöníthető-e aszerint, hogy miként viszonyulnak a többszörös folyamatok vezérléséhez, vagyis több folyamat létezése esetén a soros vagy párhuzamos vezérlés jellemzi-e őket. Tendenciaszerűen annyi mondható, hogy a nemtudatos folyamat sokszor párhuzamos szálakból áll, míg a soros feldolgozás inkább tudatos folyamatokra jellemző, vagy fordítva: a tudatos folyamatokat inkább soros módon vezéreljük. Ez megintcsak a tudatos figyelem szűkös jellegéből kiindulva érthető meg. Amikor monitorozunk, mérlegelünk, döntünk, választunk, vagyis a tudatos folyamatot esetén mindig maximális figyelmi erőfeszítés állapotában vagyunk, és ilyenkor nyilván nem tudunk egy másik tudatos folyamatot is ugyanilyen módon figyelemmel ellátni.

Az automatizmus azonban nem jelenti sem a tudatosság, sem az intencionalitás teljes hiányát. David M. Armstrong a tudatosság alapvonásait, az elme, a szellem modelljét kereste, azt, hogy hogyan lehet megmagyarázni az intencionalitást, az elmét, a menális

jelenségeket [22]. Armstrong szerint a behaviorizmus: rossz választadott akkor, amikor kijelentette, hogy – karteziánus nézetekkel szemben – nincs szellem a dobozban, a mentális jelleg magában a cselekvésben fejeződik ki. Ezzel a magyarázósémával az a probléma ugyanis, hogy vannak olyan helyzetek, amikor a mentális folyamatok léteznek, de semmilyen viselkedésben nem reprezentálódnak: például magában (belül) mérges valaki. Armstrong kritizálta Gilbert Ryle – a diszpozíció fogalmára támaszkodó – magyarázatát is, aminek az a lényege, hogy az egyén nem viselkedik az adott módon, de diszpozicionált (beállított, előhangolt) úgy viselkedni (Armstrong ezt a minőséget a törvényekhez, normatív szabályokhoz hasonlítja). Ryle így írt a diszpozíció fogalmáról:

> „Ha valami egy diszpozicionális tulajdonsággal rendelkezik, akkor ez nem azt jelenti, hogy az illető dolog valamilyen különleges állapotban van, vagy hogy valamilyen speciális változáson megy át, hanem azt, hogy amikor valamilyen sajátos feltétel megvalósul, akkor szükségképpen vagy valószínűleg valamilyen sajátos állapotban lesz, vagy valamilyen speciális változáson megy át.” [455, 59]

Armstrong szerint ez jó előrelépés, de kevés. Amikor mentális állapotban vagyunk minden külső viselkedési nyom nélkül, akkor valami csak történik ott belül, nem pusztán potencialitásról beszélhetünk.

> „Amikor én gondolkozom, de a gondolataim nem valamely cselekvés végrehajtására irányulnak, akkor teljesen nyilvánvaló, hogy valami zajlik bennem, ami a gondolataimat eredményezi. Nem egyszerűen arról van itt szó, hogy beszélnék, cselekednék, ha bizonyos feltételek fennállnának, amelyek jelenleg nem állnak fent. Valami éppen akkor – a szó legerősebb, szószerinti értelmében véve – zajlik, végbemegy bennem, és ez a valami az én gondolatom. A ryle-i behaviorizmus tagadja ezt, ezért nem is kilégítő, nem is elfogadható az elme elmélete számára.” [22, 298]

Armstrong szerint a behaviorizmus akkor téved, amikor a mentális folyamatokat, az elmét egybemossa (azonosítja) a viselkedéssel. Ennél gyengébb kapcsolat elképzelhető köztük. Armstrong javaslata: az elme, az egyéni mentális állapotok és a viselkedés között logikai kapcsolat létezik. A gondolat nem megfelelő feltételekre “váró” beszéd, hanem inkább valami az emberen belül, ami megfelelő körülmények között előidézi a beszédet. Ehhez viszont nem kell feltétlenül teljes tudatában lennek annak, amit éppen végrehajtunk. Armstrong az autóvezetést hozza fel például. Amikor az ember hosszú időn keresztül vezet, automatikussá válik szinte minden tevékenysége: gázt adott, fékezett, forgatta a kormány, indexelt stb. de mindezt nem tudatosan tette. Pontosabban: a vezetése nagyobb részben nem volt tudatos, mert persze bizonyos értelemben azért használta a tudatát. A tudatosság teljes hiányában nem észlelte volna, hogy az út kanyarodik, nem forgatta volna el megfelelő mértékben a kormányt, és a kocsija az első kanyarban leszaladt volna az útról, vagy beleszeladt volna az előtte haladó autóba, ha nem észlelte volna, hogy az fékezett, és ne lépett volna rá ő is a fékre (a tudatos figyelem hiányában). Mégha a teljes tudatosság nincs is jelen, az intencionalitás akkor is megkérdőjelezhetetlenül benne van a folyamatban. És nemcsak az intencionalitást érhetjük tetten ilyenkor, de mindig kell találnunk valamilyen tudatosságot a folyamat valamely pontján, tipikus módon legalább a folyamat legelején. A vezetés, az írás, olvasás és minden más automatikus készségünk gyakorlás után válik automatikussá, de minden esetben valamilyen tudatos akarat van jelen az effajta folyamatok beindításakor. Amikor elindulok autóval, akkor tudatosan teszem ezt. És amikor váratlan, szokatlan esemény történik (például jégre futok a kocsival), akkor szükség lesz figyelemre, és egyből a tudat homlokterébe kerül az addig automatikusan végzett mozdulatok vezérlése.

Az értelmezés kedvéért eddig élesen szembeállítottam egymással a kétféle folyamatot, ami sugallhatta azt, hogy ezek egymástól élesen elkülönülve működnek, működhetnek csak. Ez nem így van. Mégha egy ideáltipikus megközelítésben egyértelműen elválaszthatók is egymástól a kontrollált és automata folyamatok, a valóságban nagyon sok esetben egymásba fonódnak, összekapcsolódnak egymással. Wegnerék felsoroltak pár olyan kapcsolódási lehetőséget, ami a kontrollált és automatikus folyamatok között valósulhat meg, és amelyek valamilyen átmenetet, átalakulást eredményeznek az addigi folyamatokban.

többszálúság	párhuzamosan futnak kontrollált és automata folyamatok
feladatátruházás	egy kontroll folyamat elindít egy automata folyamatot
orientálás	egy automata folyamat elindít egy kontroll folyamatot
rátelepülés	egy automata folyamat felülír egy kontroll folyamatot
felülvizsgálat	egy kontroll folyamat felülír egy automata folyamatot
automatizálódás	egy kontroll folyamat átalakul automata folyamattá
megszakítás	egy automata folyamat átalakul kontroll folyamattá

Azt már korábban említettem, hogy a tudatos folyamatok általában soros-vezérlésűek, vagyis egy időben csak egy tudatos folyamatot tudunk/szeretünk/szoktunk kezelni, míg az automata folyamatokból többet is. Az viszont könnyen előfordulhat, hogy egy tudatos folyamat mellett/alatt több automata folyamat zajlik párhuzamosan. Miközben telefonon beszélünk az autónkat vezetve, tudatos kontroll nélkül is képesek vagyunk vezetés műveletére (többszálúság). De eközben képesek vagyunk még arra is, hogy ha egy nagy csomagot viszünk az autónkban, és fogni kell, hogy ne boruljon ránk, akkor ezt megtegyük anélkül, hogy állandóan figyeljünk erre, vagy ha a telefonbeszélgetés és autóvezetés közben a rádióban meghalljuk a kedvenc számunkat, akkor esetleg odanyúlunk, és felhangosítjuk (többszálúság és orientálás). Ha közben jégbordára futunk, akkor a vezetés automatizmusa megszakad, és teljes figyelmünk a vezetési manőverre irányul: fékezünk, kormányozunk, tesszük, amit a váratlan helyzet megkövetel tőlünk (felülvizsgálat és/vagy megszakítás). Ha pedig túljutottunk a veszélyes szakaszon, egyszercsak újra elvonjuk a figyelmünket a vezetéstől (automatizálódás vagy rátelepülés).

Összegezve az eddigieket: a habituális cselekvések, a bennünk/velünk zajló, nagyrészt tudatlan, automatikus, de azért intencionális folyamatok fontosak is, gyakoriak is az életünkben. Hogy ennek pontos értelmezéséhez le kell-e számolnunk a tudatos akarat illúziójával, ahogy azt Daniel M. Wegner javasolja könyvében [576], nem tudom. A cselekvéselmélet felépítése szempontjából azonban ezt nem érzem lényeges kérdésnek. Cselekvéseink tipizálása, az intencionalitásuk, szubjektív értelmük kibontásának igénye felől nézve mindegy, hogy a testünkben, az agyunkban mikor jelenik meg a tudatos akarat (vagy annak illúziója), mindegy, hogy mi van előbb: a cselekvés vagy az akarat. Sok más tudományág számára ez megkerülhetetlen kérdés, de a cselekéselmélet számára nem az. Ami fontos, az csak annyi, hogy el tudjuk különíteni a habituális cselekvéseket egyfelől az indulati-érzelmi, másfelől a mérlegelésen alapuló racionális cselekvésektől. Eddig azt mutattam meg, hogy hogyan írhatjuk le a habituális cselekvéseket, illetve miven különböznek ezek az indulati-érzelmi cselekvésektől. A következőkben a racionális cselekvés jellemzését végzem el.

3.3 Célracionális cselekvés

Kimondva-kimondatlanul, hivatkozva-hivatkozatlanul, de akik cselekvéselmélettel foglalkoztak az elmúlt, mondjuk, egy évszázadban, azok mindig kiemelt figyelmet szenteltek a racionális cselekvés, és általában a racionalitás fogalmának. Persze nem is nagyon tehettek mást. Annyira a cselekvésünkhöz, sőt, ember minőségünkhöz tartozik a racionalitás fogalma, hogy megkerülhetetlen. Nem lehet a cselekvésről beszélni nélküle. Ez még akkor is így van, ha a (cél)racionális cselekvés evolúciós értelemben megelőzi mind az indulati-érzelmi, mind a habituális cselekvés. Mondhatjuk azt is, hogy a racionalitás mechanizmusa ezekre "ül rá". A célracionális cselekvés esetében is Max Weber a kályha, tőle érdemes elindulnunk. Weber a következő meghatározást adta.

> „ ... a cselekvést az határozza meg, hogy a cselekvő milyen viselkedést vár a külvilág tárgyaitól és más emberektől, és mennyiben képes e várakozásokat mint »feltételeket«, vagy mint »eszközöket« fölhasználni arra, hogy saját, racionálisan kiválaszott és mérlegelt céljait sikeresen elérje." [575, 53:2.§]

Mivel Weber kétféle racionális cselekvést különít el, ezért nyilván meg kell majd mutatni azt, hogy miben egyezik meg és miben különbözik a kétféle racionális cselekvéstípus. Ehhez elég lesz egyfajta racionalitásfogalmat felvenni a modellbe, és a különböző fogalmi összetevők, paraméterek eltérő értékeire rámutatva lehet majd elkülöníteni a racionális cselekvés kétféle altípusát egymástól. A racionalitás fogalmának meghatározása előtt azonban érdemes azt megvizsgálni, hogy miben tér el a kétféle racionális cselekvés az eddig tárgyalt másik két cselekvéstípustól, az indulati-érzelmi és a habituális cselekvéstől. A válasz egy szóval megadható, ami persze csak annyit jelent, hogy egy másik kifejezést használunk a racionalitás terminusa helyett. Az indulati-érzelmi és a habituális cselekvés során a cselekvő tudatában vagy nem volt ott a cselekvés intencionalitását adó értelem, mint a habituálsi cselekvés esetében, vagy egyetlen tárgyra irányult a cselekvő tudata. Mindkét esetben igaz az, hogy a cselekvőnek nem volt mit **mérlegelnie**. Ez a feltétel adja a lényegi közös vonásukat. Amikor ez a helyzet megváltozik, vagyis amikor a cselekvő tudatában több lehetséges szubjektív értelem is bekerül, melyek közül választania kell, amikor tehát alternativitás kerül a tudat elé, akkor beszélhetünk arról, hogy a cselekvő mérlegel. Amikor mérlegelünk, akkor alternatív cselekvési lehetőségeket vetünk össze egymással, melyek közül választanunk kell, és ez a **választás**, vagyis az általunk hozott **döntés**t határozza meg, hogy végülis mit cselekszünk. Meg kell tehát vizsgálnunk a választás (döntés) fogalmát. Amikor mérlegelünk, választunk, a tudatunk saját tartalmaira irányul, ebben az értelemben a mérlegelés mindig **magasabbrendű tudattevékenységet** jelent, hiszen a mérlegelés mentális művelete más tudattartalmak közti választásként (tehát magasabbrendű fogalomként) értelmezhető.

Ha egy cselekvésre azt mondjuk, hogy racionális, akkor ésszerűnek gondoljuk. A kérdés az, hogy mit is tartunk ilyenkor ésszerűnek. Sok mindenre vonatkozhat ugyanis az 'ésszerű' minősítés, a célok, az eszközök, a feltételek közti választás techiikájára, a döntési folyamat egészére vagy annak valamely részére, a döntési mechanizmus "logikájára" vagy akár a választási folyamat végeredményére. Jack Hirshleifer azt hangsúlyozza, hogy a racionalitás elsősorban módszert, nem pedig eredményt jelent.

> "A racionalitás instrumentális fogalom. Ha valaki saját céljait (preferenciáit) figyelembe véve megfelelő eszközöket (cselekvéseket) választ, akkor racionális; ha nem, akkor irracionális. A 'megfelelő' itt inkább a módszerre, mint az eredményre vonatkozik. A

> racionális viselkedés olyan cselekvés, amit a logika és az érvényesség más normái szerint mérlegelve hajtanak végre. A változásoknak köszönhetően pedig egy jó megoldás nem mindig vezet jó eredményhez." [237]

Hirshleifer (el)választásával egyetértve a továbbiakban a fenti racionalitásértelmezést próbálom meg alaposabban kibontani. Amennyire fontos fogalom a racionalitás, annyira sokféleképpen próbálják meg a fogalom jelentését megragadni, a definícióját megadni. Ez nyilván nem véletlen. Az eltérő szakmai nézőpontok máshová teszik a hangsúlyt. Az sem meglepő, hogy a racionalitás fogalmával a közgazdaságtan foglalkozott messze a legtöbbet. A választásban megnyilvánuló racionalitás fogalmának kidolgozását nagyon komoly logikai-matematikai apparátussal támogatták meg az idők során. Mégha egyoldalúnak, elfogultnak is tartom és ezért kiegészítendőnek vélem ezt a közgazdasági értelmezést, ez természetesen nem jelenti azt, hogy ne tartanám szükségesnek bemutatni és a cselekvéselméleti modellbe beemelni ezt a fogalmat (pontosabban fogalomcsaládot). Szükség van azonban arra is, hogy kiegészítsük a közgazdasági értelmezést azokkal a szempontokkal, amelyeket más tudományágak, a filozófia, a szociológia képviselői tartanak fontosnak.[8]

Racionalitásértelmezések

John Broome filozófusként tágabb horizonton vizsgálta a racionalitás jelenségét [81]. Broome három racionalitásfeltételt kötött ki, melyek közül az egyikbe "belevetíthető" a közgazdászok által preferált értelmezés is (de erről majd később). Broome első feltételét az alábbi formulával fejezhetjük ki.

(ZL.26) $(\mathbf{B}_i(\phi) \wedge \mathbf{B}_i(\phi \to \psi)) \to \mathbf{B}_i(\psi)$

modus ponens alkalmazása (HUN) – applying Modus Ponens (ENG)

A feltételt úgy értelmezhetjük, hogy ha az i személy hisz a ϕ állításban, és azt is hiszi, hogy a ϕ állításból "következik" egy másik, a ψ állítás, akkor i-nek hinnie kell ψ-ben is. Ez nem jelent mást, mint hogy alkalmazza a modus ponens következtetési szabályt, vagyis képes következetések levonására, azaz képes a logikus gondolkozásra. Vannak, akik szerint ez kevés, a racionalitáshoz a logikai helyesség önmagában még nem elég. Evans-Pritchard a zandék törzsénél végzett vizsgálatai alapján azért különböztette meg a tudományos és a logikus gondolkozást, hogy rámutathasson, a törzsi társadalmak tagjai is logikusan (de még nem tudományosan) gondolkodnak.

> „A tudományos képzetek azok, amelyek összhangban vannak az objektív valósággal mind a premisszáik érvényességét, mind az állításaikból levont következtetéseket tekintve ... A logikai képzetek azok, amelyekben a gondolkodás szabályainak megfelelően az igaz premisszákból levont következtetések is igazak, a premisszák igazsága lényegtelen. ... Egy fazék eltört égetés közben, valószínűleg a homoknak köszönhetően. Vizsgáljuk meg a fazekat, nézzük meg, tényleg ez volt-e az ok. Ez logikus és tudományos gondolkodás. A betegséget a boszorkányok okozzák. Egy ember megbetegedett. Kérdezzük meg a jósdát, melyik boszorkány a felelős érte. Ez logikus és tudománytalan gondolkodás." [149]

[8] Érdeklődőknek ajánlom Dagfinn Follesdal rövid, de alapos elemzését a racionalitás fogalmáról és értelmezési lehetőségeiről [171].

Evans-Pritchard megkülönböztetésével egyet lehet érteni, de ez az egyetértés még nem mond semmit arról, mit tartsunk racionálisnak, a logikus vagy a tudományos gondolkozást. Ha a tudományos gondolkodást követelnénk meg, azzal túl erős feltételt teremtenénk. Ráadásul a tudománytörténetből rengeteg olyan példát hozhatnánk, amikor egy tudományosan megalapozott premissza igazságáról kiderült annak hamissága, és ilyen esetekben egy darabig racionálisnak, majd utána nem-racionálisnak kellene tartanunk ugyanazt a hitet. Ezért én azt az értelmezést fogadom el, amely szerint – a most vizsgált kontextusban – a racionális feltételéül a logikus gondolkodás képességét kell megszabnunk. Evans-Pritcharddal szemben (és persze így Broome-mal egyetértve) Peter Winch és Steven Lukes is ezt az értelmezést fogadja el [586], [314].

Broome második racionalitáskritériuma a cselekvés céljaira és eszközeire vonatkozó vélekedéseket és szándékokat kapcsolta össze. A feltételt így fejezhetjük ki formálisan.

(ZL.27) $(\mathbf{B}_i(M \to G) \wedge \mathbf{I}_i(G) \wedge \mathbf{B}_i(\mathbf{I}_i(M) \to M)) \to \mathbf{I}_i(M)$

szükséges eszközök akarása (HUN) – wanting necessary means (ENG)

A második broome-i feltétel annyit jelent, hogy ha az i személy úgy véli, hogy az általa megvalósulni kívánt G cél akkor teljesül a jövőben, ha az M eszköz is megvalósul, és i úgy gondolja, hogy M teljesülése az ő szándéktól függ, akkor i akarni fogja M megvalósulását. Magyarán: ha valamilyen célt el akarunk érni, és tudjuk, hogy a cél elérésének eszközeként valamit meg kell tenni, ami a mi szándékunkon múlik, akkor az utóbbit akarnunk kell megtenni. Ez a feltétel összeköti a célra irányuló cselekvések megvalósításának szándékát és a célok eléréséhez szükséges eszköz jellegű cselekvésekre irányuló szándékokkal. Broome harmadik kritériuma is a szándékainkkal kapcsolatos.

(ZL.28) $(\mathbf{B}_i(\mathbf{O}F) \wedge \mathbf{B}_i(\mathbf{I}_i(F) \to F)) \to \mathbf{I}_i(F)$

krasia (HUN) – krasia (ENG)

A harmadik Broome-féle kritérium szerint ha i úgy véli, hogy meg kell tennie F-t, miközben úgy gondolja, hogy F-t csak akkor teheti meg, ha akarja megtenni, akkor i-nek akarnia kell megtenni F-t. Ez a feltétel a cselekvés szándékoltságát ragadja meg.

A közgazdászok közül Harsányi János volt az, aki a szokásos, leszűkítő jellegű megközelítéshez képest jóval általánosabb szempontokra figyelembe véve foglalkozott a racionalitás problémakörével [224], [225]. Egy 1976-os cikkében Harsányi abból indult ki, hogy a racionalitás szabályokhoz való következetes, koherens, konzisztens igazodást jelent. Nem fogadta el azonban a racionalitás **cél-eszköz leírását** vagyis a célok és eszközök viszonyrendszerén belüli, leszűkítő jellegű értelmezését mondván, hogy az nem foglalja magába a kritériumhoz igazodó viselkedést: amikor a cél eléréséhez – például felmászni a legmagasabb hegyre – nem eszköz(ök)re, hanem kritériumra van csak szükség. A másik probléma a cél-eszköz racionalitás modellel az, hogy csak az adott cél eléréséhez szükség alternatív eszközök mérlegelését teszi lehetővé, és ezáltal nem fedi le az alternatív célok közti racionális választás jelenségét, tehát nem lehet megmagyarázni, ha valaki egy adott cél helyett valami mást választ. Ezért van az, hogy ebből a modellből hiányzik a változás, a dinamika megragadásának lehetősége. Harsányi szerint ezen a ponton még előnyös a közgazdaságtani megközelítés hozadéka, mert a cél-eszköz alapú leírás helyett jobb a **preferencia-lehetőség leírás** modell, amely szerint a cselekvőnek vannak preferenciái, lehetőségei, és az ágens utóbbiakból választ az előbbiek alapján. Amennyiben

az alternatív lehetőségek halmazából valaki kiválaszt egyet, akkor ez egyben azt is jelenti, hogy a többi lehetőséget nem választja, és a nem választott lehetőségekben rejlő kívánatosságpotenciált értékelhetjük a választási egyfajta költségeként is. A dolgokat így értékelve egyszerűen megmagyarázhatjuk azokat a jelenségeket, amikor valaki akkor is célt vált, ha az alternatív célokra vonatkozó preferenciái változatlanok maradtak, egyszerűen csak azért, mert az alternatívák költségei (még pontosabban: az effajta költségekről való ismeretei) megváltoztak. A preferencia-lehetőség modell magába foglalja mind a cél-eszköz, mind a kritérium-teljesítő modellt. Ha az ágens preferenciái megfelelnek bizonyos konzisztencia- és folytonossági feltételeknek, akkor ez a modell lehetővé teszi, hogy a preferenciáinkat **hasznosságfüggvénnyel** reprezentáljuk, és ebben az esetben a racionális cselekvés **haszonmaximalizáló** viselkedést jelent.

Ezzel azonban nem tudunk mindenfajta racionalitás leírni. A döntéselmélet hozadéka az a fontos felismerés, hogy nem mindig, sőt, inkább ritkán cselekszünk biztos tudás birtokában vagyis a teljes **bizonyosság** (certainty) állapotában. Az életünket sokkal jobban jellemzi az a tény, hogy **bizonytalanság** (uncertainty) és **kockázat** (risk) mellett kell döntéseket hoznunk. Olyan racionalitásértelmezésre tehát van szükség, amely a bizonytalanság feltételei között is használható marad. Mielőtt megnéznénk, hogy tudunk ennek megfelelni, tegyünk egy rövid kitérőt a bizonytalanság fogalomkörével kapcsolatban. Először rögzítsük a három alapfogalom jelentését.

(ZL.29) $\text{CERTAINTY} \stackrel{\text{def}}{=} \forall j(\boldsymbol{B}_i(p(A_j) = 1))$

bizonyosság (HUN) – certainty (ENG)

(ZL.30) $\text{UNCERTAINTY} \stackrel{\text{def}}{=} \exists j(\boldsymbol{B}_i(p(A_j) \neq 1))$

bizonytalanság (HUN) – uncertainty (ENG)

(ZL.31) $\text{RISK} \stackrel{\text{def}}{=} \exists j\ \forall x(\neg \boldsymbol{B}_i(p(A_j) = x) \wedge (0 \leq x \leq 1))$

kockázat (HUN) – risk (ENG) – bizonytalanság (HUN) – uncertainty (ENG)

Akkor beszélünk bizonyosságról, ha a cselekvés során figyelembe venni kívánt, lehetséges kimenetre vonatkozó információink biztosak. Azt a helyzetet tarthatjuk bizonytalannak, amikor egyes kimenetek bekövetkezésének valószínűségét nem ismerjük. Ha létezik akár csak egyetlen kimenet, aminek megvalósulására nem tudunk még valószínűségi becslést sem adni, akkor azt bizonytalannak kell tartanunk. Kockázatról viszont akkor van szó, ha az egyes kimenetek bekövetkezésének valószínűsége (jól) ismert, kiszámítható. A lottózás során az egyes számok kihúzásának esélye pontosan ismert, tehát ezt kockázatnak kell minősítenünk. A kockázat és bizonytalanság itt megadott értelmezése – logikai szempontból tekintve – nem problémamentes, mert nem igaz az a sokszor hangoztatott nézet, miszerint a kockázat a bizonytalanság alárendelt fogalma lenne. Mivel azonban itt nincs szükségünk erre a tézisre, ezért nem is kell a modellbe felvenni. Azt viszont nem én állítom, hanem hazai közgazdászok (amit készséggel elfogadok), hogy a kockázat és bizonytalanság közti határvonalat tartalmilag is nehéz (vagy talán nem is lehet) pontosan meghúzni, és e fogalmak kezelését is sok esetben a pongyolaság, túlhasználtság jellemzi [44], [45], [338], [531], [288].

> „A modern közgazdaságtan egyik fundamentális észrevétele a kockázat és a bizonytalanság különbsége. A bizonytalanság melletti döntés kérdése alapvetően korlátozza a közgazdasági, így többek között a pénzügyi döntések lehetőségét. A kockázat feltárására, eliminálására számos lehetőséget tartalmaz az irodalom. Ugyanakkor, függetlenül attól, hogy

ezeket milyen hatékonysággal hajtják végre a piaci szereplők, az inherens bizonytalansággal nem tudnak mit tenni. Bizonytalanság melletti döntés esetén az egyetlen lehetséges megoldásnak a 'több szem többet lát' módszere tűnik. Az optimális, igaz döntés lehetetlensége nem teszi szubjektívvá a rossz vagy előkészítetlen döntés fogalmát. A statisztikai módszerek által nyújtott, objektív kritériumok azonban nagyrészt illúziók, amelyek – minden kifinomultságuk ellenére – csak a múlt alapján következtetnek a bizonytalan, és így inherens módon ismeretlen jövőre.
Ugyanakkor a bizonytalanság és a kockázat megkülönböztetése is csak egy elmélet, amelynek az igaz vagy hamis volta nem vethető fel. Egyetlen döntési szituációban sem lehetünk biztosak abban, hogy kockázattal vagy bizonytalansággal állunk-e szemben, vagyis soha sem tudhatjuk előre, hogy az előrejelzések nem bizonyulnak-e végül igazaknak." [338]

Az értelmezési nehézségeket tudomásul véve abban az irányban léphetünk tovább, hogy megnézzük, miként lehet a bizonytalanság problémakörét összekapcsolni a racionális cselekvésekkel. Nos, a racionalitás elvét a bizonytalan, kockázatos helyzetekre úgy lehet érvényesíteni, hogy – a várható hasznosság hipotézise szerint – a lehetséges kimenetek hasznosságát a hozzájuk rendelhető valószínűségekkel súlyozzuk [225]. Ekkor a mérlegelési folyamatba bevont tényezőket nem önmaguk értéke, preferáltsága szerint vesszük figyelembe, hanem annak figyelembe vételével, hogy az adott tényezőnek mekkora a megvalósulási esélye. A döntéselmélet lényege – Harsányi szerint – az, hogy elfogadunk pár konzisztencia és folytonossági axiómát, és ezek alapján azt feltételezzük, hogy a racionálisan cselekvő ágens a saját szubjektív várható hasznossága maximalizására törekszik.

Ez a modell azonban még mindig nem elég az összes racionális cselekvés leírásához, hiszen a játékelméleti helyzetekre nem alkalmazható. A játékokban ugyanis olyan bizonytalanság van, amivel kapcsolatban nem tudunk – direkt módon – esélyeket latolgatni, hiszen nem tudhatjuk, hogy a másik játékos hogyan lép a játék során. Ráadásul a racionalitás teljes modelljének le kell még fednie az etikai cselekvések világát is, amit Harsányi úgy definált, hogy az a racionális morális értékítéletek elmélete. Ez alatt pedig azt értette, hogy az ágensek úgy ítélkeznek, hogy **pártatlan** (impartial) és **személytelen** (impersonal) kritériumokra alapozott preferenciáik közül választanak – racionális. Amíg a játékelmélet a potenciálisan konfliktusos egyéni érdekek elmélete (ahol az egyéni érdekek lehetnek egoista és altruista irányultságúak egyaránt), addig az etika a társadalom, a közösség mint egész közös érdekeinek elmélete [225, 98].

A játékokat mint a játékelmélet által kezelt **interdependens helyzeteket** többféleképpen is lehet tipizálni. Az egyik lehetőség a játékosok rendelkezésére álló információk szerint megkülönböztetni a **komplett** vagy **inkomplett információval**, illetve a **perfekt** vagy **imperfect információval** rendelkező játékokat. A játékok másik felosztása a **kooperatív** és **nem-kooperatív játékok** megkülönböztetése, amiről a játékelmélet egyik klasszikusa, John F. Nash írt először [225]. A kooperatív játékokban egyrészt a játékosok tehetnek kikényszeríthető elköteleződéseket, egyezségeket, másrészt szabad a kommunikáció köztük, míg a nem-kooperatív játékokban mindez nem lehetséges. Harsányi szerint az előbbi szempont az erősebb, bár az utóbbit szokták inkábbb hangsúlyozni. Harsányinak igaza van: a fogolydilemma csak akkor lesz igazából kooperatív játék, ha kikényszeríthető egyezséget lehet kötni a játékosok között, a kommunikáció, a megegyezés lehetősége ehhez még kevés. A nem-kooperatív játékokban ezért csak az **önérvényesítő megegyezéseknek** (self-enforcing aggreements) lehet esélye a sikerre. Ezt nevezik **egyensúlyi pontnak** (equilib-

rium point) megintcsak John F. Nash alapján.[9] A **legjobb válasz** (best reply) az a lépés (stratégia), amelyik a legjobb kifizetést eredményezi a játékosnak akkor, ha feltételezzük, hogy a másik játékos stratégiája konstans (nem változik). Ezek alapján az egyensúlyi pont a játékosok legjobb válaszaiból álló lépéskombináció.

Nash egyensúlyi pont fogalmának bevezetése után azt gondolták, hogy ez elégséges és megfelelő feltétel lehet a racionalitás meghatározásához, ám 1965-ben Reinhard Selten olyan példát mutatott, ami ugyan egyensúlyi pont, de irracionális, így ez az elképzelés megdőlt [225].

	$Alter_1$	$Alter_2$
Ego_1	1,3	1,3
Ego_2	0,0	2,2

Két egyensúlyi pont van: $EqP_1 = (Ego_1, Alter_1)$ és $EqP_2 = (Ego_2, Alter_2)$, de csak az utóbbi **perfekt egyensúlyi pont**, és csak is ez reprezentálhatja a racionális cselekvést (ezt igazából csak a játék extenzív formájú reprezentációja képes megmutatni). A magyarázat viszont egyszerű: a második játékosnak az egyik döntési ágon nincs valódi döntési lehetősége (döntetlen számára a helyzet, hiszen a saját döntésétől független a saját és a társa nyereménye.

A kooperatív játékokat nevezik **alkujátékoknak** (bargaining game) is. Arra utalnak ezzel, hogy a játékosok közti kommunikáció arra irányul, hogy megállapodjanak egy olyan kimenetben, ami valamennyiük számára elfogadható lehet [225]. Az alkufolyamat eredménye azonban nagyban függ attól, hogy milyen szabályok érvényesek a tárgyalási folyamatra: ki kinek beszélhet, ki kezdheti a tárgyalást, milyen a tárgyalások nyilvánossági foka (nyilvános vagy privát), visszavonhatók-e a korábban megkötött egyezségek (van-e mód újratárgyalásra vagy végleges, visszavonhatatlan ajánlatokat engednek csak meg), lehet-e egyoldalúan kötelező ígéreteket vagy fenyegetéseket tenni vagy sem. A szemléltetés végett képzeljünk el egy játékot, amelyben van három játékos és három kifizetési feltétel: *(i)* egyedül játszó játékos nem nyer semmi, *(ii)* a kétszemélyes koalíció 100 dollárt nyer, *(iii)* háromszemélyes koalíció 100 dollárt nyer. Milyen szabályok lehetnek érvényesek ezen felül, és milyen végeredmények várhatók ezek alapján?

1) ha a játékosok elkülönülten páronként is találkozhatnak, akkor az első – mindegy, hogy kik között – létrejött találkozó szereplői vélhetőleg azonnal koalíciót formálnak és egyenlően elosztják maguk között a nyereményt;
2) ha nincs mód szeparált találkozókra, és nincs elég idő a hármas találkozón arra, hogy kettős koalíció alakulhasson ki, akkor a három fél megegyezhet a hármas koalícióban, és harmadolhatják az összeget;
3) ha a fentiekhez képest más feltételek adottak, akkor ilyen esetben bármilyen koalíció kialakulhat, de ezt nem lehet előre megjósolni.

Amikor a játékosok maguk között alkukat, megegyezéseket, elköteleződéseket formálhatnak ki, akkor érezhetően – legalább részben – megváltoznak a racionalitásfeltételek. Ezeket figyelembe kell vennünk akkor, amikor racionális cselekvésekről gondolkozunk.

Harsányi egy másik cikkében foglalkozott azzal a kérdéssel, hogy – a racionalitás fogalmának kidolgozása mellett – mi volt a közgazdaságtan sikerességének másik fontos

[9] Nash legfontosabb idevágó cikkei: [373], [374], [372].

magyarázótényezője [224]. Harsányi szerint a siker másik oka az volt, hogy a közgazdászoknak sikerült megfelelő motivációs elmélet is kidolgozniuk. Ennek a motivációs elméletnek azonban, bár valóban sok esetben komoly magyarázóerővel bírt a gazdasági cselekvések leírásában, volt egy komoly gyengesége. Mégpedig az, hogy egyszálú motivációs magyarázat volt, mert az gazdasági önérdek mint egyetlen létező motiváció fogalmára támaszkodott. Ez ugyan az élet sok területén működőképesnek bizonyult, de még a gazdaság területén sem lehetett ezzel a túlegyszerűsített modellel mindent megmagyarázni, és persze a gazdaságon túli, gazdaságon kívüli világban, például a politikában, még inkább igaz ez, hogy figyelembe kell venni azt a tényt, hogy vannak nem-gazdasági és nem-egoista motívumok is. Ha bővíteni akarjuk a modellt azzal, hogy más motívumokat is figyelembe veszünk, akkor felmerül a kérdés, hogy honnan tudhatjuk meg az emberekről, hogy milyen relatív súlyokat, hasznosságokat rendelnek az egyes célokhoz, eszközökhöz, gazdasági és nem-gazdasági érdekeikhez, normáikhoz stb. Harsányi válasza az, hogy meg kell figyelni az emberek tényleges viselkedését, és az így szerzett empirikus tényekből fel lehet állítani hasznosságfüggvényeket, majd ezek alapján már meg lehet magyarázni az emberek további cselekvéseit. Fontos azonban, hogy ez csak akkor nem jelent körbenforgó érvelést, ha az empirikus megfigyelésből szerzett paraméterek száma hangsúlyosan kevesebb, mint a megmagyarázni kívánt jelenségek, paraméterek száma. Olyan motivációs elméletre van szükség, amely nem egydimenziós, tehát nem csak a gazdasági önérdek motivációjára épít, de nem akar túl sok új motivációra sem támaszkodni. Harsányi javaslatot is tett egy ilyen motivációs rendszerre, amely csak négy posztulátumra épül (az első három az önző és önzetlen motivációk kérdésével foglalkozik, a negyedik a két legfontosabb alapmotívumot adja meg). Ezek a következők.

1) *az alacsony költségigény esetén követjük a pártatlanság és közjó elvét*
   Ha olyan helyzetekben vagyunk, amelyben nincs közvetlenül érdekeltségünk, akkor hajlandók vagyunk pártatlanok lenni, hajlandók vagyunk a köz érdekében, altruista módon fellépni, mindezt azonban csak akkor, ha nem túl megerőltető számunkra. A bírót kivonják az ügyintézésből, ha érdekeltsége van abban, de ha nincs, akkor képes részrehajlás nélkül működni – enélkül nem lenne működőképes közigazgatásunk. Elmegyünk szavazni annak ellenére, hogy racionális módon gondolkodva nem érné meg nekünk, és csak azért tesszük meg, mert kevés fáradságba kerül.

2) *konfliktus esetén sikeres lehet a pártatlan közvetítőkön keresztüli érdekaggregálás*
   A posztulátum értelme az, hogy ha ennek alapján egy harmadik – pártatlan – felet is bevonunk az érdekellentétes helyzetekbe, akkor a konfliktusban állók közötti kompromisszumkötést segítheti elérni az új szereplő. Ez magyarázza meg a csoportérdek, osztályérdek, nemzetérdek fogalmát, jelenségét, de ez adja morál alapját is, hiszen olyan megoldásokat lehet ezáltal találni a konfliktusok feloldására, amelyek minden ember érdekére tekintettel vannak.

3) *mások iránti elköteleződéseink korábbi racionális választásaink eredménye*
   Az ember személyes elköteleződései (a családtagok, a munkatársak, a barátok iránt) gyakran altruizmust eredményeznek. Ezek az elköteleződések azonban valamilyen racionális választás eredményeiként jöttek létre korábban. Miután az ember így vagy úgy elköteleződött, utána ezek az elköteleződések gyakran "altruista irányban" befolyásolják a mindenkori racionális döntéseket, de ez nem változtat azon a tényen, hogy az elköteleződések valamilyen korábbi racionális döntési helyzetben születtek.

4) *két fő motívumunk a gazdasági előnyök követése és a társas elfogadás igénye*
Cselekvéseinket alapvetően két domináns érdekkel magyarázhatjuk. Az egyik alapmotiváció a gazdasági előnyök maximálása és a gazdasági veszteségek minimalizálása, a másik pedig a társas elfogadás iránti alapkésztetés. Cselekvéseink során mindig e két fő motívumból fakadó célok megvalósítására törekszünk.

Harsányi szerint a fenti posztulátumokra támaszkodva a legtöbb esetben már magyarázatot adhatunk a cselekvéseinkben tapasztalható irracionalitásokra, lehetőségünk nyílik a társadalom statikus és dinamikus magyarázatára, amiből persze az utóbbi az igazán fontos, ez az izgalmasabb (és persze nehezebb is egyben). Talán a legfontosabb feladat megmagyarázni a társadalmi értékek aktuális eloszlását, amit csak a múlt ismerete alapján tehetünk meg. Ha azt tapasztaljuk, hogy változnak az értékek, akkor az a legtöbbször a feltételek változása miatt következik be. Például a nagycsalád adta előnyökért sokáig megéri vállalni a sok gyereket, de idővel már többet jelent az embereknek a kényelmetlenségektől megszabadulni, tehát kevés gyereket vállalni. Az értékváltás nehézsége sokszor puszta tudatlanságból fakad, mert nem tudunk eleget az új alternatíváról, és a régi ismertségéből fakadó biztonságérzet erősebb. Az értékváltások legnagyobb akadályai azonban inkább mások szankciói (leginkább persze azoké, akik érdekeltek a változatlanságban).

Az eddigiekben a racionalitás feltételeit a mérlegelés folyamatában, annak minőségében kerestük, miközben adottnak vettük, hogy vannak döntési lehetőségek. Nem foglalkoztunk azzal, hogy a mérlegelt alternatívák hogyan "álltak elő", miért pont azokat ítéljük meg, bevonhatók lennének-e más alternatívák is a döntésbe. A továbbiakban erre a szempontra figyelve egy újabb racionalitáskritériumot ajánlok felvenni a modellünkbe. Robin Horton az afrikai tradicionális gondolkodás és a nyugati tudomány kapcsolatáról írt tanulmányában használta a nyitott, illetve zárt társadalom fogalmát [242]. Horton felfogásában ez a nyitottság-zártság kettősség kevesebbet, viszont pontosabbat mit jelent a Karl Popper által ismerté tett (nyitott kontra zárt társdalom) fogalompárhoz viszonyítva. Horton kizárólag az alternatívák létezése vagy nem létezése mentén állítja szembe a nyitott és zárt társadalom fogalmát. Definíciós javaslata értelmében ha az emberek vitáiban léteznek alternatívák, akkor nyitott, ha nem léteznek, akkor zárt társadalomról beszélhetünk. De miért is lehet ez érdekes a racionalitás fogalmának tárgyalásakor? Jürgen Habermas szerint egy "megnyilatkozás racionalitása a bírálhatóságára és a megindoklásképességére vezethető vissza" [202, 16]. Ez a definíció azt a szempontot hozza be a racionalitásértelmezések világába, amit már Robin Horton is fontosnak tartott. Legalábbis akkor, ha kicsit továbblépünk egy egyénhez kötött racionalitás fogalmán. A habermasi javaslatot (ami persze visszavezethető lenne legalább Weberig) vagyis a racionális diskurzus feltételeként megadott bírálhatóságot és megindoklásképességet egy új dimenzióban is értelmezhetjük. A kommunikációban megnyilvánuló racionalitás, a nyilvánosság racionalitása azt (is) jelenti, hogy nincs semmilyen tilalom a megnyilatkozásokra vonatkozóan. Amennyiben ugyanis a közösség bizonyos normái nem teszik lehetővé bizonyos alternatívák választását, akkor az nyilván befolyással lehet a cselekvés racionalitására. Ez a szempont átvezet minket az értékracionális cselekvések világába, de ezek elemzését külön fejezetben végzem el (↪215.o.). Amit viszont a célracionális cselekvéshez kapcsolódóan kell még megvizsgálnom, az egy olyan másodrendű preferenciafogalom, amelynek segítségével meg lehet ragadni a célracionális orientáció egyik – ha nem a – legfontosabb vonását.

Arkhimédeszi preferencia

A mérlegelés, a célracionális cselekvés fogalmának tárgyalásakor korábban már jeleztem, hogy ezek – bizonyos értelemben – másodrendű fogalomként értelmezhetőek, hiszen a mérlegelés folyamata azt jelenti, hogy célokat, eszközöket, következményeket veszünk figyelembe, vagyis több különböző szempontokat követve valamilyen döntést hozunk. A másodrendű minőség abban a mozzanatban érhető tetten, hogy amikor a mérlegelési folyamatban a különböző szempontokat veszünk figyelembe és azok mentén hozunk részdöntéseket, akkor voltaképp belső késztetéseink, illetve ezek alapján a tudatossá váló preferenciarendezéseink határozzák meg a részdöntéseinket. A mérlegelés folyamatának egésze viszont nem jelent mást, mint ezen részleges belső szempontok (preferenciarendezések) egymáshoz viszonyítását, vagyis valamiféle másodrendű preferenciarendezés kialakítását az elsőrendű preferenciák felett. Mondhatjuk ezt a műveletett preferenciaaggregálásnak is, hiszen arról van szó, hogy több preferenciarendezésből kell egyetlen végső rendezést kialakítani. Ez az "elvárás" ugyanaz, mint a közösségi döntések elméletében leírt közösségi preferenciaaggregálás feladata (↬411.o.). A másodrendű preferenciarendezés persze többféle lehet. Én két fajtáját fogom bemutatni. A **arkhimédeszi preferencia(rendezés)** fogalmával kezdem azért, mert ezt a célracionális cselekvéshez kapcsolhatjuk.

Vegyük át még egyszer, miről is van szó. Amikor a preferenciánkat fejezzük ki, akkor valamilyen szempont szerint valamely alternatívákon végrehajtunk egy preferenciarendezést. A szempontok tartalmával nem kell foglalkoznuk, viszont figyelnünk kell arra a mozzanatra, hogy könnyen előfordulhat (és gyakran elő is fordul), hogy ugyanazt az alternatívahalmazt több szempont szerint is értékelnünk (rendeznünk) kell. Ilyenkor merül fel a kérdés, hogy miként lehet egyszerre több szempontot is figyelembe venni. Azért fontos ez a kérdés, mert feltételezhető, hogy a különböző szempontok szerint más és más végeredményhez lehet jutni a rendezések során, miközben a mérlegelési folyamattól az egyértelmű döntést "várjuk el", hogy ti. a folyamat végén egyetlen kiválasztott cselekvési alternatíva álljon rendelkezésre a cselekvő számára. A következő fontos "technikai" kérdés, hogy milyen módon lehet a preferenciarendezést végrehajtani több elem (és/vagy több szempont) esetén. Ahogy már említettem, az egyik ilyen mód az arkhimédeszi preferenciarendezés alkalmazása.

Az arkhimédeszi preferencia matematikai értelemben vett leglényegesebb tulajdonsága az, hogy nincs sem végtelen nagy, sem végtelen kicsi eleme, vagy másként: nincs ugrás az őt leíró függvényben. Ez persze már azt feltételezi (jelenti), hogy az arkhimédeszi preferencia folytonos függvénnyel reprezentálható. Amikor a preferenciát így tudjuk jellemezni, akkor alkalmazhatjuk rá a közgazdaságtan egyik központi kategóriáját, a hasznosság jenningsi fogalmát. Minden olyan esetben tehát, amikor a hasznosság fogalma kerül elő a cselekvések leírásában, feltételezhetjük/kijelenthetjük, hogy a cselekvők a döntési szempontjaikat az arkhimédeszi preferencia jellegzeteségének megfelelően veszik figyelembe. Az arkhimédeszi jelzővel tehát a cselekvők (döntéshozók) másodrendű preferenciarendezéseit minősíthetjük. Nem mindenkire és nem minden helyzetben jellemző ez az arkhimédeszi beállítódás, bár sokak szerint a modernizáció egyet jelent az archimédeszi preferencia elterjedtségének folyamatos növekedésével. Az arkhimédeszi rendezés dominánssá válásának és ezzel szoros összefüggésben a – következő fejezetben tárgyalandó – lexikografikus rendezés háttérbe szorulásának folyamatát Csontos László és társai "uralmi" tézisnek nevezték el [115]. Ez a dominancia azt jelenti, hogy egyre több

emberben, egyre több szempontra vonatkozóan kiterjed(t) az arkhimédeszi tulajdonság érvényessége. Amikor megfogalmazzák az "eladó az egész világ" szlogenjét, vagy amikor a reklámozásban való részvétellel szembeni ellenállás feladását fejezik ki a "van az a pénz, amiért korpás lesz a hajam" bonmot-tal, akkor az arkhimédeszi tulajdonság terjedését, új területekre benyomulását példázzák.

A kérdés már csak az, hogy miként tudjuk formalizálni ezt a fogalmat. A relációelméleti részben bevezettem a sorrendezés vagy archimédeszi rendezés fogalmát (↪36.o.), az arkhimédeszi preferencia voltaképp ez alá a rendezési fogalom alá sorolható be. Az arkhimédeszi preferencia meghatározásakor a preferenciarelációt fogyasztói kosarakra (matematikailag vektorokra) alkalmazzák, tehát az összehasonlítást, értékkifejeződést nem egyedi entitásokra, hanem entitások (mégpedig különböző típusú entitások) halmazaira vonatkoztatva fejezik ki. Erre a „módosított" preferenciarelációra írják elő az arkhimédeszi axióma érvényesülését, amely a következőképpen formalizálható (az arkhimédeszi preferenciarelációt az $\mathbf{R}$, a termékek különböző mennyiségeit x_i és y_j szimbólumokkal jelölve).[10]

(ZL.32) $\text{ARCHIMEDEAN_PREFERENCE} \stackrel{\text{def}}{=} \forall i \forall j \exists k (\mathbf{R}((x_i, y_i), (x_j, y_j)) \rightarrow \mathbf{R}((x_j, y_k), (x_i, y_i)))$

arkhimédeszi preferencia (HUN) – archimedean preference (ENG)

A képlet azt fejezi ki, hogy bárhogy is állítunk össze két csomagot az X és Y termékekből, amelyek közül az egyik csomag (x_i, y_i) többet ér a másiknál (x_j, y_j), ha az Y jószág mennyiségét növeljük (y_j-t kicseréljük y_k-ra), akkor megfordulhat a preferenciareláció iránya, azaz (x_j, y_k) csomag értékesebbé válhat az (x_i, y_i) csomagnál.

Az arkhimédeszi preferencia lényege az, hogy a szempontok összehasonlításakor, tehát másodrendben úgy járunk el, hogy nincsenek a szempontok között olyanaok, amelyeket külön kellene kezelnünk a többiektől. Az egyes szempontokat természetesen mindig rendezi a döntéshozó, vagyis egyes szempontokat fontosabbnak, olykor sokkal fontosabbnak tart, mint a többit, de mindig előállhat olyan konstelláció, amelyben a legerősebb szempont is hátrább kerülhet egy másik szemponttal szemben. Mindez jobban érthetővé válhat azután, hogy a másik fontos másodrendű preferenciarendezést, a lexikografikus preferencia fogalmát is bemutattam. Előtte azonban tárgyalnom kell az utolsó cselekvéstípust, az értékracionális cselekvés fogalmát.

3.4 Értékracionális cselekvés

A weberi cselekvéstipológia negyedik eleme az **értékracionális cselekvés** kategóriája. Ez is nyilván racionális, mérlegelő cselekvés, és a legfontosabb kérdés az, hogy miben tér el a célracionális cselekvéstől ez a típus. Weber a következő meghatározást adta.

> „ . . . a cselekvés . . . lehet . . . *értékracionális*: ekkor a cselekvést egy meghatározott magatartásnak – pusztán mint olyannak, függetlenül attól, hogy sikeres-e – a feltétlen etikai, esztétikai, vallási vagy bármilyen más néven nevezendő *ön*értékébe vetett tudatos hit határozza meg." [575, 53:2.§]

A definíció egyik fontos eleme, hogy Weber önértékbe vetett tudatos hitről beszél. Ebből számunkra az következik, hogy egyfelől ezt a cselekvéstípust a mérlegelés világába kell

[10] A formula forrása [115].

besorolnunk, másfelől a mérlegelést, a választást itt valamilyen értékelköteleződés határozza meg. Az érték jelenléte lényegi módon befolyásolja a mérlegelést. A fogalom meghatározásának másik lényeges összetevője az, hogy a mérlegelésnek nem a sikeresség a szempontja, hanem az értékelköteleződésnek történő megfelelés. Weber a definícióhoz fűzött magyarázó megjegyzései között ezt tovább egyértelműsíti, amikor összehasonlítja az indulati-érzelmi és értékracionális cselekvéseket.

> „A cselekvés indulati-érzelmi és értékracionális orientációja abban különbözik egymástól, hogy az utóbbi tudatosan megszabja a cselekvés végső célját, és tervszerű *következetességgel* igazodik hozzájuk. Egyébként közös bennük, hogy a cselekvés értelmét nem a rákövetkező sikerben, hanem magában a meghatározott fajtájú cselekvésben látják." [575, 53:2.§.3.]

Tudatos céltételezés és a célhoz való tervszerű igazodás – a mérlegelő cselekvés két legfontosabb eleme érhető tetten ebben az idézetben is. Mindez azonban nem jelenti azt, hogy az értékracionális cselekvést jellemző értékelköteleződésnek csak a célválasztásra lehet hatása. A mérlegelés során az eszközök kiválasztását is meghatározhatják értékválasztásaink. A mérlegelés egy újabb, különös fajtájáról van itt szó, és ezt a különös jelleget fogom feltárni a továbbiakban. Az értékelköteleződés fogalma alatt itt azt értem, amit a késztetésekről, vágyakról szóló fejezetben fejtettem ki (↪153.o.). Az érték az egyszerű késztetésekhez vagy vágyakhoz képest a vonzódás tudatosságában különbözik. Az értékeink úgy működnek, hogy amikor a mérlegelés során a lehetséges célok vagy eszközök között olyan alternatívák tűnnek fel, amelyeket értékelköteleződés köt minket, akkor azt az értékkötött alternatívát nem szabad vagy épp ellenkezőleg azt kell mindenáron figyelembe vennünk. Ha egy férfi tervezi a hétvégi szórakozását, és mérlegeli, hogy a pénzéből mire mennyit költsön, akkor – ha számára a nőválasztás kérdése érték – eldönti, hogy kajára, piára mennyit adjon ki, de nem kalkulál azzal, hogy a szexuális szolgáltatás tervezett igénybevételének mennyi lesz az ára. A 'szerelmet nem lehet pénzért venni' értéke nem engedi meg számára azt, hogy prostituálthoz menjen. Az 'igazi háziasszony maga gyúrja a levesbe való tésztát' értékelkötelezettsége nem engedi meg, hogy a családi ebéd során bolti tésztát egyék a család. A cselekvési célt az értékelköteleződés jelöli ki, és nem számít, hogy mibe kerül. Nem lehet szempont, hogy vasárnap korábban kell felkelni és többet kell dolgozni, sok esetben még az sem számít, hogy a gyerekek már a bolti tésztát szeretik jobban. A kérdés már csak az, hogy hogyan lehet ezt a kivételes szerepet megmagyarázni. Ebben segít minket a lexikografikus preferencia fogalma.

Lexikografikus preferencia

Natura non facit saltum. A latin mondás szerint a természet nem szereti az ugrást. Az emberi természet más. A világhoz való viszonyaink között van olyanok, amelyeket nem tudunk folytonos függvénnyel reprezentálni. A **lexikografikus preferenciarendezés** művelete is ilyen.

A döntéselméletben úgy mondják, hogy az arkhimédeszi preferencia kompenzatórikus jellegű, a lexikografikus preferencia pedig nem-kompenzatórikus. Kompenzatórikus egy döntési eljárás, ha bármelyik értékelési kritérium szerinti eredmény kompenzálható a többi szempont szerinti értékelésekkel, és nyilván a nem-kompenzatórikus eljárásban valamelyik kritériumra ez nem igaz [132]. Szemléletes példa erre – bár a preferencia jelenségéhez való kapcsolata nem annyira egyértelmű – az az oktatási helyzet, amikoris egy diák nem mehet tovább a következő félévre, ha nem teljesítette egy tantárgy

minimálkövetelményét, mégha az összes többi tantárgyból jelest is szerzett. Ez úgy általánosíthatjuk, hogy előfordulhatnak helyzetek, amelyekben egy szempontot kiemelten kezelünk, és minden máshoz képest erősebbnek tartunk. Amikor a helyzetben releváns módon figyelembe lehet venni azt a szempontot, akkor az adott szempont kell, hogy meghatározza a döntésünket, és minden más szempontot figyelmen kívül kell hagyni.

A közgazdasági szakirodalomba Nicholas Georgescu-Roegen vezette be a lexikografikus preferencia fogalmát [184], [185]. A fogalom formális elemzésével Peter C. Fischburn foglalkozott a legtöbbet [165], [166], [167], [164], [169], [299], [298], [75], a fogalom elméleti alkalmazásáról, a vele kapcsolatos kísérleti eredményekről lásd: [300], [103], [144]. Említettem az arkhimédeszi rendezés dominánssá válásának tézisét [115]. Itt nyilván e folyamatnak azt az ezzel párhuzamosan érvényesülő összetevőjét érdemes kiemelni, hogy az arkhimédeszi preferenciák terjedésével együtt a lexikografikus preferenciák fokozatosan háttérbe szorulnak. Ezt sokan értékvesztési folyamatként, az értékelvű magatartás eltűnéseként minősítik. Ebben a könyvben ennek a kérdésnek a tárgyalását nem tartom relevánsnak, de annyit azért jeleznék – a tendenciák létezét elismerve –, hogy az arkhimédeszi preferencia térnyerése sosem lehet kizárólagos. Bizonyítani itt nem tudom ezt a tézisemet, nem is akarom, de szemléltetni azért szeretném, mégpedig azzal a reklámmal, amely egyébként az – arkhimédeszi preferenciát reprezentáló – bankkártyát próbálja népszerűsíteni. A Mastercard bankkártya 'Felbecsülhetetlen' szlogennel bevezetett reklámjában a kártyát magát azzal próbálják meg magasra értékelni, hogy a reklámban rámutatnak egy értékre (barátságra, családra, szerelemre, meghitt pillanatra stb.), ami "megfizethetetlen", majd a kártyát rögtön emögé pozícionálják. A reklámokat lezáró szlogen is ezt a kettősséget fejezi ki: „Van, amit nem lehet pénzért megvenni. Minden másra ott a MasterCard." Bármennyire is dominánssá válik az arkhimédeszi preferencia beállítódása, az értékekre való hivatkozást nem tudja nélkülözni.

Nézzük meg tehát, hogy milyen módon tudjuk leírni a lexikografikus prefeerncia fogalmát. Ha van két termékünk (szolgáltatásunk vagy bármink, amire a preferenciareláció alkalmazható), akkor azok különböző mennyiségeit egy-egy vektorral reprezentálhatjuk: $(x_1, x_2, \ldots, x_i)$, $(y_1, y_2, \ldots, y_j)$, ahol x_i az X jószág csökkenő nagyság szerinti sorrendbe állított i. egyede, míg y_j az Y jószág j. példánya. Ebben az esetben a lexikografikus preferencia fogalmának formális meghatározását a következőképpen adhatjuk meg (a preferenciarelációt $\mathbf{R}$-fel jelölve).[11]

(ZL.33) $\text{LEXICOGRAPHIC_PREFERENCE} \stackrel{\text{def}}{=} \forall j \exists i (j < i \rightarrow (\mathbf{R}(x_j, y_j) \wedge \mathbf{R}(y_j, x_j) \wedge \mathbf{R}(x_i, y_i)))$

lexikografikus preferencia (HUN) – lexicographic preference (ENG)

A formula értelmezésekor ki kell használnunk azt a feltételt, hogy a két jószágból vett különböző mennyiségeket – nagyság szerint csökkenő – sorrendbe állítva jelöljük x_i-vel és y_j-vel. Így a fenti képlet azt fejezi ki, hogy akármekkora mennyiséget veszünk is a két jószágból, azaz vehetjük akárhányadik (j.) elempárt a sorrendbe rakott termékekből, a két terméksor j. elemeit megelőző párjaira még lehet döntetlen az állás köztük (ezt fejezi ki az $\mathbf{R}(x_j, y_j) \wedge \mathbf{R}(y_j, x_j$ részformula), de az i. elemtől kezdve már az X termék a preferáltabb az Y-nal szemben ($\mathbf{R}(x_i, y_i)$. Ezt azt is jelenti egyben, hogy az X termék „összességében" kívánatosabb a másikhoz viszonyítva.

[11] A formula forrása [115].

A közgazdász értelmezés után megmutatom a lexikografikus rendezés matematikusok által kidolgozott modelljét is, mert egyrészt nagyobb kifejezőerőt, másrészt az általánosabb megközelítésmód miatt további hasznosítási lehetőségeket nyerhetünk vele. A modellt Andréka Hajnal, M.D. Ryan és P.Y. Schobbens dolgozta ki [14]. Andrékáék abból indulnak ki, hogy a preferenciarendezés fogalma nem elég a lexikografikus rendezés fogalmának definiálásához, szükség van a *prioritás* operátor bevezetésére. Azért, mert többféle szempont szerint is rendezhetünk dolgokat (autóvásárláskor figyelembe vehetjük az autó árát, gyorsulását, fogyasztását, vezethetőségét, színét, méretét, alakját stb.). Minden egyes szempont szerinti rendezés egy-egy preferenciarendezésnek felel meg, tehát sok rendezési relációnk is lehetséges. Ha pedig több szempont szerint kell rendezni ugyanazt az elemkészletet, akkor más és más sorrendek alakulhatnak ki a különböző szempontok (rendezések) szerint, amikor pedig szükség van arra, hogy a szempontok (rendezések) is valamilyen prioritást alakítsunk ki (vagyis valamilyen sorrendbe állítsuk őket. A rendezések rendezésének problémáját általánosítva túl lehet lépni az egyéni rendezések problémakörén, és felállítandó modell hatókörét ki lehet terjeszteni a közösségi döntések területére is. Amikor a többszerelős döntési folyamatokban egyetlen közösségi döntést kell hozni, akkor az egyéni preferenciákat kell egy közös preferenciába aggregálni. Matematikailag ezt ugyanúgy lehet kezelni, mint azt a helyzetet, amikor egy személy többféle szempontját kell egyetlen döntés során figyelembe venni. Andrékáék mindkét helyzetre (sőt, még továbbiakra is) érvényes modellt állítottak fel.

Kiinduláshoz egy $\mathbf{R}$ tetszőleges reláció, valamint az $\underset{\perp}{\mathbf{F}}$ teljes reláció és az $\underset{\perp}{\mathbf{E}}$ üres reláció fogalmára van szükség (↰27.o.). Bár az $\mathbf{R}$-re preferenciarelációként hivatkozunk, a korábbi erős feltételeket (↰81.o.) itt nem követeljük meg. Nem feltételezzük sem azt, hogy $\mathbf{R}$ reflexív, sem azt, hogy tranzitív lenne, semmilyen tulajdonságot nem követelünk meg tőle. A "legalább olyan jó" (jobb vagy ugyanolyan jó) jelentés mentén értelmezett $\mathbf{R}$ relációra támaszkodva a következő további relációkat vezethetjük be:

(BA.34) $\mathbf{R}(m,n)$

$\mathbf{R}$ a **legalább olyan jó** (jobb vagy ugyanolyan jó; jobb vagy közömbös) reláció; $\mathbf{R}$ a $P = \{m,n,\ldots,o\}$ (világ)állapotok halmazán van értelmezve.

(BA.35) $\overline{\mathbf{R}}(m,n) \stackrel{\text{def}}{=} \neg\mathbf{R}(m,n)$

$\overline{\mathbf{R}}$ a **rosszabb** (se nem jobb, se nem közömbös) reláció; $\overline{\mathbf{R}}$ az $\mathbf{R}$ negáltja.

(BA.36) $\textsc{better}(m,n)\mathbf{R}^{<}(m,n) \stackrel{\text{def}}{=} \mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m)$

$\mathbf{R}^{<}$ a **jobb** (szigorúan jobb; erősen preferált; szigorúan preferált) reláció; $\mathbf{R}^{<}$ aszimmetrikus.

(BA.37) $\textsc{indifferent}(m,n)\mathbf{R}^{\equiv}(m,n) \stackrel{\text{def}}{=} \mathbf{R}(m,n) \wedge \mathbf{R}(n,m)$

$\mathbf{R}^{\equiv}$ a **közömbös** (indifferens) reláció; $\mathbf{R}^{<}$ szimmetrikus.

(BA.38) $\textsc{incomparable}(m,n) \equiv \mathbf{R}^{\#}(m,n) \stackrel{\text{def}}{=} \neg\mathbf{R}(m,n) \wedge \neg\mathbf{R}(n,m)$

$\mathbf{R}^{\#}$ az **összehasonlíthatatlan** (inkompatibilis) reláció.

Az univerzális és üres reláció, valamint a fenti preferencia-vonatkozású relációk között megállapíthatóak a következő összefüggések:

(BA.39) $\overline{\underset{\perp}{\mathbf{F}}} = \underset{\perp}{\mathbf{F}}^{<} = \underset{\perp}{\mathbf{F}}^{\#} = \underset{\perp}{\mathbf{E}}^{<} = \underset{\perp}{\mathbf{E}}^{\equiv} = \underset{\perp}{\mathbf{E}}$

(BA.40) $\underset{\perp}{\mathbf{F}}^{\equiv} = \underset{\perp}{\mathbf{E}}^{\#} = \overline{\underset{\perp}{\mathbf{E}}} = \underset{\perp}{\mathbf{F}}$

Ezekután az $\mathbf{R}(m,n)$ preferenciarendezés mintájára vegyük az $\{\mathbf{R}_v(m,n)\}_{v\in V}$ preferenciarelációk valamely halmazát (ahol $m,n \in M$). A preferenciarelációinkra legyen adott még egy $|V|$-edrendű $<$ (szigorúan részben rendezett) **prioritás** operátor is.

(BA.41) $< (\mathbf{R}_x(m,n), \mathbf{R}_y(m,n)) \equiv < (x,y) \equiv x < y \overset{\text{def}}{=} \text{STRICT_PARTIAL_ORDER}(x,y)$

prioritás reláció (HUN) – priority relation (ENG)

A formula azt fejezi ki, hogy a prioritás reláció szigorú részben rendezés, azaz a reláció reflexív (↳28.o.), tranzitív (↳30.o.) és antiszimmetrikus (↳29.o.). Ekkor a $\{\mathbf{R}_v(m,n)\}_{v\in V}$ preferenciarelációk halmazán, azok kombinációjaként értelmezett $\langle \mathbf{R}_v(m,n)\rangle_{v\in V}$ rendezett $|V|$-es, amit innentől $\text{LPREF}(m,n)$ **lexikografikus preferenciarendezésnek** nevezünk, így definiálható:

(BA.42) $\langle \mathbf{R}_v(m,n)\rangle_{v\in V}(m,n) \equiv \text{LPREF}(m,n) \overset{\text{def}}{=} \forall i \in N(\mathbf{R}_{v(i)}(m,n) \vee \exists j \in N(<(j,i) \wedge \mathbf{R}^{<}_{v(j)}(m,n)))$

lexikografikus rendezés (HUN) – lexicographic order (ENG)

A formula egy diszjunkció. Az első komponens akkor lesz igaz, ha az aggregátum összes eleme ($\mathbf{R}_x(m,n)$) ugyanúgy rendezi el a két elem (m és n) viszonyát. Ennek kiderítéséhez végig kell menni az aggregátumba bevont összes preferenciarendezésen, hiszen univerzális kvantor zárja le ezt a részformulát. A képlet második összetevője megadja a kiértékelési folyamat gyorsabb lezárásának lehetőségét (ebben az értelemben rendkívül praktikus szabálykomponenesnek minősíthetjük). A diszjunkció második eleme ugyanis azt rögzíti, hogy ha elindulunk a prioritási sorba állított preferenciarendezések kiértékelésében, és ezek között van olyan reláció (a prioritási sor y. pontján), amely az összehasonlított két elem (m és n) közül az egyiket szigorúan preferálja a másikkal szemben ($\mathbf{R}^{<}_y(m,n)$), akkor ez a tény eldönti az teljes lexikografikus rendezést is (az éppen adott prioritási ponton szigorúan preferált alternatíva vagyis az m javára).

Értelmezéséhez segíthet, ha a szótári rendezés példájára gondolunk. A szótárakban (lexikonokban) szavakat kell sorba állítanunk, ezeket jelöljük itt az m,n változókkal. A $<$ prioritás operátor itt a betűk szavakon belüli pozícióinak sorrendjét határozza meg (európai írásmód esetén balról jobbra halad a sorrend). A $\mathbf{R}_x(m,n)$ preferenciarelációink mind megegyeznek, és az ábécé szerinti rendezést jelentik. Annyi preferenciarelációnk van ($|V|$), ahány betűje van a szótár legtöbb betűből álló szavának. Minden szó tetszőleges x. pozíciójában a $\mathbf{R}_x(m,n)$ preferenciareláció alapján meg tudjuk mondani, hogy az ott található betű alapján az adott szó mely szavakat előz meg, mely szavak vannak előtte, illetve mely szavakkal van azonos helyzetben (amikor ugyanaz a betű szepel az x. pozíción). Ez azonban kevés a lexikon szavainak sorrendbe állításához. A végleges rendezést a $<$ prioritás operátor segítségével lehet csak kialakítani. Nézzük meg ezt a következő példa segítségével:

> E példában a pék a példa arra, hogy ki pótolható, de ő példázza azt is, hogy kit nem lehet pótolni.

A későbbi azonosíthatóság miatt érdemes megjelölni a szavak helyét (sorrendjét a mondaton belül). Erre azért lesz szükség, hogy kezelni tudjuk a szavak többszöri előfordulásait is. Mivel a szótárba nem kerülnek be a központozási jelek, csak a szavak, így a fenti mondatot húsz szóalakra bonthatjuk fel:

E(1) példában(2) a(3) pék(4) a(5) példa(6) arra(7), hogy(8) ki(9) pótolható(10), de(11) ő(12) példázza(13) azt(14) is(15), hogy(16) kit(17) nem(18) lehet(19) pótolni(20).

Ezt a húsz szót a lexikografikus elv alapján rendezve a szabály két összetevője közül két esetben alkalmazhatjuk az első komponenst. Az 'a' és a 'hogy' szavak esetében a formula első komponesét kell mindvégig alkalmazni, vagyis e két-két szóelőfordulásra a szabály azt hozza ki, hogy legalább olyan jók, mint a másik, hiszen a szavak minden pozíciójában megegyeznek a betűk. A többi esetben a lexikografikus szabály mindkét összetevőjét figyelembe kell venni, és mindig találhatunk egy olyan pozíciót az egyes szavakon belül, amelyekben az éppen összehasonlított szavak valamelyikének betűje megelőzi a másik szó betűjét. Az 'azt' és az 'arra' szavak esetében például a második pozícióban levő betűk esetében az 'r' előbb van a 'z' betűhöz képest az ábécé szerinti rendezés szerint, amiatt az 'arra' szavunkat előbbre soroljuk az 'azt' szavunkhoz képest. Az egybetűs szavakat már az első pozíció alapján rendezni lehet (ezért kerül előbbre az 'a' az 'e' vagy 'ő' szavakhoz képest), a több betűből álló szavaknál a későbbi pozíciók alapján lehet csak sorrendet kialakítani a szavak között. "Példaszótárunk" összeállításában a hatodik pozícióig el kellett menni ahhoz, minden szóról el tudjuk dönteni, hol a helye (a 'példában' és 'példázza', illetve a 'pótolni' és 'pótolható' szavak esetében).

#	szó	(#)	PR_1	PR_2	PR_3	PR_4	PR_5	PR_6	PR_7	PR_8	PR_9
1-2	a	(3)	a								
	a	(5)	a								
3	arra	(7)	a	r	r	a					
4	azt	(14)	a	z	t						
5	de	(11)	d	e							
6	e	(1)	e								
7-8	hogy	(8)	h	o	g	y					
	hogy	(16)	h	o	g	y					
9	is	(15)	i	s							
10	ki	(9)	k	i							
11	kit	(17)	k	i							
12	lehet	(19)	l	e	h	e	t				
13	nem	(18)	n	e	m						
14	ő	(12)	ő								
15	pék	(4)	p	é	k						
16	példa	(6)	p	é	l	d	a				
17	példában	(2)	p	é	l	d	á	b	a	n	
18	példázza	(13)	p	é	l	d	á	z	z	a	
19	pótolható	(10)	p	ó	t	o	l	h	a	t	ó
20	pótolni	(20)	p	ó	t	o	l	n	i		

A lexikografikus és arkhimédeszi rendezés különbsége egyszerűen szemléltethető. Elég csak azt bemutatni, hogy a kétféle módszer hogyan teszi sorba egész számok adott halmazát. Nézzünk meg egy példát! Vegyük az alábbi számok halmazát: {2, 100, 3, 10, 101, 20, 11, 12, 1, 13, 21, 22}. A kétféle rendezés eredményét mutatja a következő táblázat.

arkhimédeszi rendezés	1, 2, 3, 10, 11, 12, 13, 20, 21, 22, 100, 101
lexikografikus rendezés	1, 10, 100, 101, 11, 12, 13, 2, 20, 21, 22, 3

A lexikografikus rendezés fogalmának definálásával azonban még nincs vége. Következő lépésként Andrékáék bemutatják, hogy a prioritás operátor alapján felépíthető gráf mindig kifejezhető két bináris operátor segítségével, amelyekre aztán egy egész algebrát lehet építeni. A két operátort az 'x/y' és az '$x\|y$' szimbólumokkal, illetve a **de** (but) és a **másfelől** vagy másrészt (on the other hand) nevekkel jelölhetjük, és a következő módon határozhatjuk meg őket:

(BA.43) $x/y \equiv \text{BUT}(x,y) \stackrel{\text{def}}{=} (x \cap y) \cup y^{<}$

de (operátor) (HUN) – but (operator) (ENG)

(BA.44) $x\|y \equiv \text{ON_THE_OTHER_HAND}(x,y) \stackrel{\text{def}}{=} x \cap y$

másfelől (operátor) (HUN) – on the other hand (operator) (ENG)

Az operátorok jelentésének értelmezéséhez segítséget jelenthet, ha arra gondolunk, hogy miként vesszük figyelembe a hierarchiába szervezett emberek véleményét. Amikor emberek egyetértenek, akkor bárhol is legyenek a szervezeten belül, a "közös véleményük" is az lesz, amit egyenként helyeselnek. A kérdés az, hogy miként összesítsük az eltérő véleményeket. Ezen a ponton lehet megérteni a két operátor szemantikáját. Amikor alá-fölérendeltségi viszonyban levő emberek véleménye eltér, akkor a magasabb ponton levő (a szervezeten belül nagyobb prioritással rendelkező) ember véleményét vesszük figyelembe. Ha i és j a két személy, és i van az alárendelt szerepben j-hez képest ($i < j$), akkor azt mondjuk, hogy "i véleménye az, hogy x, *de* j azt gondolja, hogy y", és ez utóbbi számít. Alá-fölérendeltségi viszony esetén tehát a **de** (BUT) operátort alkalmazzuk. Amikor viszont két ember azonos hierarchikus szinten van, akkor az eltérő véleményeik között nem tételezünk prioritási különbséget, azonos mértékben vesszük figyelembe őket. Ekkor azt mondjuk, hogy "(egyfelől) i azt gondolja, hogy x, *másfelől* j úgy véli, hogy y". Ilyen esetekben a **másfelől** (ON_THE_OTHER_HAND) operátort használjuk.

Érdemes megjegyezni, hogy bár matematikai értelemben elégséges a fenti két operátort definiálni, létezik egy harmadik is, amit leginkább akkor lehet értelmesen igénybe venni, amikor az elmélet alkalmazni akarjuk a társadalmi cselekvések leírásakor. A BUT operátorban ugyanis megcserélhetjük az argumentumok változóit. Ez a reláció (operátor) **konverzének** felel meg, amit logikailag kifejezhetünk új kategória felvétele nélkül is. Azért érdemes mégis említeni (és felvenni a modellbe) ezt a "másik irányú" relációt (operátort), mert közel sem mindegy, hogy melyik végén áll az ember egy alá-fölérendeltségi viszonyban. Az új operátor definíciója a következő:

(BA.45) $y/x \stackrel{\text{def}}{=} (y \cap x) \cup x^{<} \equiv x/^{-1}y$

Mivel az új operátor matematikai értelemben redundáns, az elmélet további kifejtésekor nem számolunk vele, noha később még hivatkozhatunk rá. A két alapoperátor viszont két ok miatt is nagyon fontos. Egyrészt bizonyítható, hogy bizonyos feltételeknek[12] megfelelő, végesrendű operátor meghatározható a két alapoperátor és a változók segítségével

[12] Andrékáék cikkükkel a közösségi döntések területén centrális Arrow-féle lehetetlenségi tételhez is hozzá akartak szólni, ezért az ott használt feltételekhez hasonló saját kritériumokat defináltak, amiket IBUT-feltételeknek neveztek el. Ezek azonban az itt tárgyalt témakör szempontjából nem lényegesek, ezért nem térek ki rájuk.

([14, Corollary 46]), másrészt a két operátorra támaszkodva definiálhatunk egy **preferenciális algebrát** (preferential algebra) is, ami az elméletet nagyon erőssé teszi. Ehhez az alábbi axiómákat kell rögzítenünk:

(BA.46) $x \| x = x$

(BA.47) $x \| (y \| z) = (x \| y) \| z$

(BA.48) $x \| y = y \| x$

(BA.49) $(x/x) = x$

(BA.50) $x/(y/z) = (x/y)/z$

(BA.51) $(x \| y)/z = (x/z) \| (y/z)$

(BA.52) $(x/y) \| x = x \| y$

Ez az algebra azért is különösen fontos, mert ez a modellt értelmezhetjük a hierarchia jelenségének leírásaként is. Márpedig a hierarchia fontosságát nem kell különösebben ecsetelni. A biológiában, kognitív tudományban a hierarchia két fontos típusaként a **beágyazott** és **nem-beágyazott hierarchia** típusait elkülönítve egymástól fontos tételként kezelik azt, hogy az élő rendszerek (így az emberi agy is) beágyazott hierarchiaként jellemezhetőek [161]. A szervezetszociológia a munkaszervezés világának vagy még általánosabban a szervezetek szervezésének egyik kiemelten fontos módjaként elemzi a hierarchiát [323]. Az általában vett szervezés, irányítás területén a hierarchia elvének előnyeiről szól Herbert Simon híres órás példája [489]. A mérnökök az általuk tervezett rendszerek irányításának eltérő logikáját próbálják megragadni a **hierarchikus** és **heterarchikus** irányítás kettősségével [553]. Az archívumok, könyvtárak világában a tartalmi eligazodást segíteni hivatott tudásszervezési rendszerek szervezési módjai között elkülönítenek egymástól kétféle hierarchiafogalmat, a **monohierarchia**, illetve a **polihierarchia** típusait [515].[13]

Nagy lépést tettek meg Andrékáék, amikor egyetlen modellbe vonták össze a személyen belül, illetve személyek között egymás mellett létező, versengő preferenciarelációk leírását mondván, hogy formális szempontból a személyen belüli preferenciakonfliktusok azonos lényegűek a személyek közti preferenciaütközésekkel. Ez azt is jelenti egyben, hogy a mindkét önállóan kutatott szakterületnek lehet mondanivalója a másik számára, azaz a társadalmi választások elméletének relevanciája lehet az egyéni preferenciakezelést leíró elmélek számára, és fordítva. Persze arra azért vigyázni kell, hogy ne kövessünk el túláltalánosításokat. Mindenesetre nagyon izgalmasak azok az új kísérletek, amelyek ezen a nyomvonalon haladva próbálják meg minél pontosabban leírni a preferenciák kezelésével, dinamikájával, változásaival kapcsolatos jelenségeket [127], [311], [191]

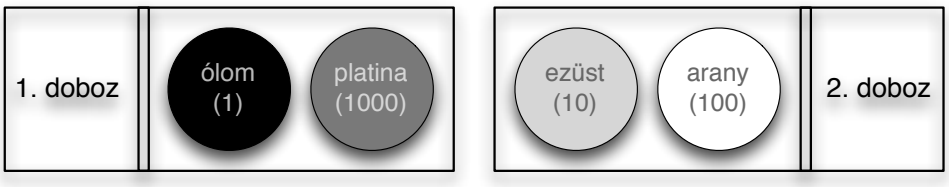
3.5 A racionális aracionális és az aracionális racionális

Akármelyik értelmezését is fogadjuk el a racionalitás fogalmának, ha elkötelezzük magunkat valamelyik definíció mellett, és ennek szellemében vizsgálni kezdjük az emberek

[13] A hierarchia fogalmának használatát a természet- és társadalomtudományok területén belül David Lane tekintette át egy tanulmányában[297], és a hierarchia szerepéről az élet különböző területein én is írtam korábban [525].

viselkedését, akkor biztosak lehetünk benne, hogy a tényleges társadalmi gyakorlatban rengetegszer megsértik a racionalitás elvét. Ennek sok oka van. Szemléltetésként vegyük a többszempontú döntési helyzetek példáját. Amikor valaki egyetlen szempont szerint akar választani az alternatívák halmazából, akkor ezt megteheti a szemponthoz "illeszkedő" preferenciareláció segítségével. Amikor azonban több szempont is felmerül a választás során, akkor a választás döntési helyzetté alakul, ami annyiban lesz több és más, hogy az eltérő szempontok más és más rendezéseket eredményezhetnek. Ezt nyilván kezelni kell valahogy, amit azonban többféleképpen tehetünk meg. A döntéselmélet – legalább részben – ennek a problémának a tárgyalására jött létre. A sokféle szempont egyeztetése során lehetséges, hogy a szempontokat egyenrangúaknak, nem rendezetteknek tekintik, és lehetséges, hogy különböző szintekre sorolják, vagyis rendezik őket. Nem célom itt az egyéni döntési eljárások, módszerek, szabályok ismertetése.[14] A probléma érzékeltetése végett röviden bemutatok három jól ismert, asokat elemzett döntési stratégiát, a maximin, a maximax és a minimax szabályt.

A **maximin szabály**, amit neveznek még pesszimista, biztonságra törekvő, kockázatkerülő módszernek, Wald-féle Wald kritériumnak, a legrosszabbak legjobbikát választja ki az alternatívák közül. A szabály lényege, hogy első körben kiválasztja az alternatívák egyes szempontok szerinti leggyengébb értékelését, majd második körben ezen leggyengébb elemek közül kiválasztja a legjobbat. A **maximax szabály** vagy másként az optimista módszer első körben kiválasztja az alternatívák egyes szempontok szerinti legerősebb értékelését, majd második körben ezek közül veszi a legjobbat. Ez módszer tehát a legjobbak legjobbikát keresi meg. A **minimax szabály** megkeresi az egyes szempontok szerint várható legrosszabb eredményeket, és ezek közül kiválasztja a legkisebb veszteséget ígérő alternatívát. Ez a szabály kockázatsemleges, az elmaradó haszon minimalizálására törekszik, arra, hogy a döntéshozónak a végén a legkisebb legyen a megbánása. A módszert nevezik még Savage vagy Savage-Niehans-kritériumnak is. A maximin és a maximax szabályok közti különbség szemléltetésére megidézem itt Carl Hempel példáját [232].



A maximin szabály szerint első lépésben megkeressük a két doboz minimumát, ami az ólom (1) és az ezüst (10), második lépésben pedig vesszük a minimumok közül a legnagyobbat, ami az ezüst. A maximax szabály először megkeresi a két doboz maximumát, ami itt a platina (1000) és az arany (100), majd másodjára kiválasztja utóbbiak maximumát, vagyis a platinát. Láthatjuk tehát, hogy két – egyaránt elfogadhatónak és racionálisnak tűnő – módszertan alapján más eredményekhez jutunk. Ez mindenképpen magyarázatra szorul. Amikor értelmezni akarjuk a stratégiákat, megakarjuk magyarázni a különbségeket, akkor már nem segítenek minket az eddig bemutatott racionalitásdefiníciók. Itt már más tényező(ke)t kell beemelnünk a magyarázó modellbe. Úgy tudunk magyarázatot adni, ha hivatkozunk a döntést hozó személyek személyiségjegyeire,

[14] A témakört jól bemutatja ez a magyar nyelvű tankönyv: [541].

temperamentumára, beállítódásaira, hangulataira. Ezek ismerős fogalmak – az effektív jelenségek világából (↪179.o.). A racionális elvek helyett az érzelmi heurisztikákra, érzelmi-tapasztalati mechanizmusokra támaszkodó cselekvések elemzése a viselkedés-gazdaságtan területére tartozik.

A viselkedés-gazdaságtan a magatartástudományok és a gazdaságtan metszetében helyezkedik el. Bár alapvetően gyakorlati irányultságú, vagyis az emberek tényleges viselkedését megfigyeléseken és kísérleteken keresztül próbálja meg leírni és értelmezni, azért két fontos előzményét vagy inkább elméleti alapját érdemes említenem. Az egyik az Amos Tversky és Daniel Kahneman munkássága alapján kibontakozó **kilátáselmélet** (prospect theory), a másik a Herbert Simon nevéhez köthető **korlátozott racionalitás** (bounded rationality) elmélete. Az új megközelítések közös vonása az a felismerés, hogy kockázatos vagy bizonytalan döntési helyzetekben, amikor úgy kell döntenünk, hogy nem áll rendelkezésre minden információ, akkor eltérünk a racionalitáselméletek által elvárt elvektől, megoldásoktól. A kilátáselmélet arra fókuszál, hogy milyen hüvelykujjszabályok, döntési heurisztikák alapján hozzuk meg tényegesen a döntéseinket, a korlátozott racionalitás elmélete pedig azt hangsúlyozza, hogy a tökéletes döntések helyett legtöbbször megelégszünk az elégséges, még elfogadható döntésekkel, vagyis korlátozott racionalitás vezérel minket.

Herbert Simon a **korlátozott racionalitásról** szóló tézisében arra hivatkozik, hogy az ember saját korlátozott információfeldolgozó képessége miatt nem tud teljesen racionális lenni [488], [490], [491]. A racionális döntéshez minden szempontot össze kellene gyűjteni, minden lehetséges kimenetet, minden célt, a hozzájuk tartozó eszközökkel, azok következményeivel, mellékhatásaival együtt figyelembe kellene venni. Ezzel szemben sosem áll rendelkezésünkre az elvileg szükséges információ. Ha viszont elképzelnénk azt az esetet, hogy valamiért, valahogyan mégiscsak szert tennénk minden információra, azt a hatalmas mennyiségű és bonyolultságú információhalmazt nem vagyunk képesek feldolgozni. Gyakorlatilag mindig beszerzési és feldolgozási gondjaink adód(ná)nak.

> „Egy olyan világban, amelyben viszonylag szűkösek az információk, és amelyben csupán néhány, egyszerű döntési problémát kell megoldani, az információ majdnem mindig pozitív jószágnak számít. Egy olyan világban, amelyben a figyelem az egyik legszűkösebb erőforrás, az információ költséges luxuscikk is lehet, mivel figyelmünket a fontos dolgokról kevésbé fontosakra terelheti.” [490, 81]

Az ilyen helyzetben az embernek az lehet az egyetlen értelmes válasza, hogy “feladja” az teljes racionalitás iránti elvárását, pontosabban az ilyen helyzetben a cselekvést elemző, értelmező kutatónak kell feladnia az ember teljes racionalitására vonatkozó hipotézisét. Simon is ezt lépte meg, amikor bevezette a korlátozott racionalitás tézisét, ami szerint a hétköznapi gyakorlatok során megelégszünk a (még éppen) elfogadható döntésekkel, nem törekszünk teljességre. A racionalitás legnagyobb “ellensége” tehát sajátmaga, pontosabban a **racionalitás illúziója**, vagyis annak fel nem ismerése, hogy az ember (a döntéshozó) szükségszerűen korlátos kognitív képeségekkel rendelkezik. Ennek számtalan jele, bizonyítéka van, ami leginkább abban nyilvánul meg, hogy az előttünk álló lehetőségekre vonatkozóan mindig kockázatokkal, bizonytalanságokkal kell számolnunk, és ezek kezelésében nagyon sokszor gyengék vagyunk, tévedünk, hibázunk, félremagyarázunk [339]. Balhiteink egyik látványos példája a véletlen jelenségének félreértelmezése, amire a “szerencsejátékos tévedéseként” szoktak hivatkozni. A hétköznapi ember számára nem evidens, hogy minden lottóhúzás alkalmával ugyanannyi esélye van vala-

mennyi számkombinációnaknak a nyerésre (például az 1-2-3-4-5 sornak). Ezt a tény, hogy ti. a véletlenen alapuló rendszereknek nincs emlékezete, az emlékezettel rendelkező emberek nagy része nem tudja megfelelően kezelni [339].

Amos Tversky és Daniel Kahneman kilátáselmélete, sőt a viselkedés-gazdaságtan egésze felfogható egy olyan heurisztikagyűjteményként is, amelynek segítségével értelmezhetjük, hogy mikor, milyen módokon és miért döntünk a tényleges társadalmi gyakorlatban.[15] A kilátáselmélet egy legismertebb fogalma a **keretezési hatás** (framing effect) jelenségét írja le. A keretezés lényege, hogy kockázatos helyzetekben az emberek attól függően (is) döntenek, hogy milyen módon, pozitív vagy negatív formában tálalják a döntési lehetőségekre vonatkozó információt [269]. Tverskyék híres kísérletében a következő lehetőségekből kellett választani. Ha egy 600 ember életét veszélyeztető járvány leküzdésére két programot dolgoznak ki, és *(i)* az első program szerint 200 ember életét lehet biztosan megmenteni, míg *(ii)* a második programban egyharmadnyi eséllyel megmenthetik mind a 600 embert, de kétharmadnyi valószínűséggel mindenki meghal, akkor a kísérleti alanyok 72%-a az első opciót választotta – miközben a két opció a várható hasznosságok szempontjából megegyezik. Ha viszont úgy tették fel a kérdéseket, hogy *(iii)* a harmadik program alapján 400 ember biztos meghal, viszont *(iv)* a negyedikben egyharmad valószínűséggel senki sem hal meg, míg kétharmadnyi esélye van annak, hogy mindenki meghal, akkor megfordult a szavazatarány, és a negyedik alternatívát választották 78%-nyian. A keretezésnek tehát iszonyú ereje tud lenni. A reklám, a marketing igen jelentős részben erre a jelenségre épül rá.

Tverskyék másik "újítása" a **tükrözési hatás** (reflection effect) fogalmának bevezetése volt, amivel azt írták le, hogy az emberek döntő többsége kockázatvállalóként viselkedik, ha rizikós döntésében lehetséges veszteségekkel szembesül, de kockázatkerülővé válik, amikor lehetséges nyereségek között választ [270].A tükrözési hatás fogalmával azt a jelenséget írhatjuk le, hogy ha a kísérleti alanyokat pozitív, illetve negatív keretbe helyezzük, akkor választásaik egymás tükörképei lesznek. Az előző járványos példasorban a pozitív keretben előadott bizonytalanabb *(ii)* alternatívát jóval kevesebben választották, tehát kockázatkerülők voltak, míg a negatív előadásmód prezentálásakor kockázatvállalóak lettek, miközben a bizonytalanabb *(iv)* opcióra szavaztak inkább. Ezt jelenti az, hogy a két választás egymás tükérképe.

A **bizonyossági hatás** (certainty effect) fogalma azt a jelenséget írja le, amikor az emberek túlértékelik a biztos kimeneteket a bizonytalanokkal szemben. Eszerint amennyiben valaki választhat a teljesen biztos 80 forint vagy a 80%-os bizonyosságú 100 forint között, akkor az előbbit választják (noha a két kimenet várható hasznossága teljesen egyforma). E hatás következtében egyébként várható nyeremények esetében a kockázatkerülő beállítódás válik dominánssá.

Tversky és Kahneman elmélete természetesen követőkre talált, akik további anomáliák, irracionalitások létezésére mutattak rá. A **felidézési előítélet** (availability bias) szerint az ember nem is annyira aszerint tudja előhívni az információt a memóriájából, hogy az milyen gyakran vagy ritkán történik meg, tehát nem az ismertség, megszokottság kontra ismeretlenség, szokatlanság a felidézési képesség és gyorsaság fő meghatározója, hanem inkább az eseményekhez kapcsolt szubjektív fontosság. Könnyebben felidézzük az érzelmi töltéssel elraktározott eseményeket, kimeneteket. Ezért értékeljük felül a gyilkosságokat, a súlyosabb bűncselekményeket.

[15] A kilátáselmélet felfutását, eredményeit, kritikáját mutatja be magyarul [210].

A **birtoklási hatás** (endowment effect) szerint a már birtokolt tárgyak iránti elkötelezettségünk indokolatlanul magasabb. Az emberek azokat a javakat, amelyek már birtokában vannak, sokkal magasabbra értékelik, mint amelyeket még csak meg kellene szerezniük. Dan Ariely és munkatársai végeztek kísérleteket a tulajdonlás érzelmi töltetének vizsgálatára [21]. Az egyik rangos egyetemen az éves kosaras szuperrangadóra az egyetemisták összlétszámához képest csak nagyon kevesek juthattak be, ezért sorsolással döntötték el az érdeklődők tömegéből ki juthat jegyhez a meccsre. A kutatók megtudakolták, mennyiért vennének meg egy jegyet azok, akik elestek a vásárlási lehetőségtől azoktól, akiknek lett jegyük, óriási különbség alakult ki a kereslet és kínálat között: a lemaradók legfeljebb 175 dollárt voltak hajlandók áldozni egy jegyre, míg legalább 2400 dollárt kértek volna érte a szerencsés jegytulajdonosok. „Érzelmi szakadék támadt azok között, akik lélekben már a mérkőzés nagyszerűségére készültek, és azok között, akik arra gondoltak, mi mást is vehetnének a jegy áráért." A magyarázat: a tulajdonlás átalakít bennünket, mert érzelmi kapcsolatot teremt a tulajdonunk és saját magunk között. Tulajdonosi tudat esetén már működni kezd a tulajdon elvesztésétől való félelem. Sőt ez már a részleges tulajdonlás esetén is így van! Ezért hajszolják bele magukat olyan gyakran a magasabb árajánlatokba a licitáló felek az aukciók során, de – részben – ezért működnek a virtuális tulajdonlás reklámtechnikái (elképzeljük magunkat a reklámozott termékkel együtt és azonnal érzelmi viszonyba kerülünk vele) és az elköteleződést kialakító ingyenes kipróbálási időszakok.

A birtoklási hatás mélyén elég nyilvánvalóan a **status quóhoz** való ragaszkoda?s húzódik meg. A status quótól való eltérést ugyanis az emberek mindenképpen kockázatosnak ítélik. A **status quo előítélet** (status quo bias) hasonló ehhez, csak ekkor nem az eredetileg birtokolt tárgyhoz, hanem az eredetileg kialakított döntéshez ragaszkodnak az emberek. Erre mindenki sok példát hozhatna saját életéből.

A **reprezentativitási előítélet** (representativeness bias) az a potenciális irracionalitás bennünk, hogy nehezen tudjuk jól kezelni a statisztikai reprezentativitás fogalmát a gyakorlatban. Erre Hámori Balázs példáját idézem [210].

> „Tételezzük fel, hogy Péter halálmegvetően bátor ember. Mi a valószínűbb, hogy alpinista, vagy az, hogy könyvtáros? Tegyük fel továbbá, hogy az alpinisták 90 százaléka bátor, míg a könyvtárosoknak csak 5 százaléka. Az országban 100 alpinista található és 5400 könyvtáros, azaz ekkora a minták nagysága. Ebből az következik, hogy 90 bátor alpinistára 270 bátor könyvtáros jut. Háromszoros a valószínűsége tehát annak, hogy Péter könyvtáros, mint annak, hogy alpinista. Mégis 100 megkérdezett ember közül 99 az előbbi kérdésre azt feleli, hogy Péter alpinista, noha nyilvánvalóan tudatában van annak, hogy sokkal kisebb valószínűséggel hozza össze a sors egy alpinistával, mint egy könyvtárossal."

Amit Tverskyék a **beakaszkodás és igazodás** (anchoring and adjustment) kifejezéssel ragadtak meg, azt Ariely az etológiából ismert **bevésődés** (imprinting) fogalmával írta le. A bevésődés vagyis az első élmények meghatározó jelentősége megfigyelhető az emberi döntésekben is. Egy új termékkel, szolgáltatással kapcsolatban az első élmények alapján kialakítunk referenciapontokat, amelyektől aztán a későbbiekben sem nagyon vagy csak nehezen mozdulunk el. Sok esetben nem (csak) a kereslet-kínálat törvénye szabja meg a piaci árakat, hanem a bevésődés is hathat az árak alakulására. Egy érdekes kísérletben a kísérleti alanyoknak különböző termékek árát kell megbecsülniük, de mielőtt belekezdtek volna, le kellett írniuk a saját társadalombiztosítási számukat a papírra. Akiknek nagyobb száma volt, azok szisztematikusan magasabb árakat írtak be a megadott terméklistára, mint azok, akiknek kisebb volt a száma.

Dan Ariely azt állítja, hogy a – hagyományos racionalitásfelfogás szerint – irracionálisnak minősíthető döntéseink sokszor nagyon is kiszámíthatóak [21]. A döntéseinkben sokszor felismerhető heurisztikák belső logikája néha nagyon könnyen megmagyarázható.[16] Sokszor egyszerűen csak arról van szó, hogy a nehéz döntések helyett a könnyebbet választjuk, ami viszont befolyásolja, olykor meghatározza az alternatívákra vonatkozó tartalmi döntéseinket is. Ariely egyik példája az Economist előfizetési opcióival volt kapcsolatos. Először három lehetőséget kínáltak fel a leendő olvasók számára: *(i)* a nyomtatott változatot 85, *(ii)* az online hozzáférést 125, végül *(iii)* mindkettő lehetőséget egyszerre 125 dollárért. Ilyen feltételek mellett az előfizetők nagy arányban a kettős hozzáférési lehetőséget kérték. Amikor azonban – egy második kísérletben – a középső opció elmaradt, akkor sokkal inkább az első, csak a nyomtatott előfizetési lehetőséget választották. Ez azzal magyarázható – mondja Ariely –, hogy amikor nehezen összehasonítható alternatívákból kell választanunk, akkor öntudatlanul leszűkítjük a döntési mezőt azokra az opciókra, amelyeket egymással könnyebben összemérhetünk. A második kísérletben azon valóban gondolkodni kell, hogy a nyomtatott vagy az online változatot vegyük-e meg, vagyis az *(i)* és *(iii)* ajánlatok közti választásnak "tétje" van, a döntésbe "munkát kell befektetni". Amikor viszont az első kísérletben az a) mellett a *(ii)* és *(iii)* opciókat is össze lehet hasonlítani, akkor az utóbbi kettő közti döntés nagyon könnyű, és – akaratlanul is – ez a könnyűség határozza meg magát a döntés tartalmát. A második kísérlet tanúsága szerint abban az értelemben irracionális ez a döntés, hogy a többség – ha már egyszer hajlandó vagy épp rá van kényszerítve az alapos mérlegelésre, akkor – mégis inkább az első ajánlatot kéri, míg a "lustaságot megengedő" döntési helyzetben az ezzel ellentétes opciót választja. A példa szerint tehát irracionálisak vagyunk, de – teszi hozzá Ariely – ez az irracionalitás kiszámítható, hiszen meg tudjuk magyarázni, előre tudjuk jelezni. Hasonló eredményeket hoztak az olyan kísérletek, amelyekben Párizsba, illetve Rómába szóló, ingyen reggelivel összekötött hétvégi utazási ajánlat mellé betettek egy másik római utat ingyen reggeli nélkül. A döntő többség a csali ajánlattal támogatott római utat választotta (a "Róma vagy Párizs" nehéz döntés helyett a "Róma reggelivel vagy Róma reggeli nélkül" könnyű döntést választották, ami után már nem volt kérdés, hogy a "Róma reggelivel" opció jobb a reggeli nélküli változatnál). Egy másik kísérletből kiderült, hogy két ember arcképe közül is azt látjuk szimpatikusabbnak, amelyik mellé beteszik annak eltorzított, csúnyább változatát is. De ugyanezen viszonylagossági tényező miatt viselkedünk másként akkor, amikor hajlandók vagyunk egy utcasarokkal távolabb menni, hogy megvegyünk 18 dollárért egy tollat, ha tudjuk, hogy 7 dollár kedvezményt adnak a 25-ből, míg nem tesszük meg ugyanezt akkor, ha megtudjuk, hogy azt a 455 dolláros öltönyt, amelyet valahol már kinéztünk, megkaphatnánk kicsit távolabb 7 dollárral kevesebbért.

Jellemző ránk az ingyenességhez való viszonyunk is. Egy kísérletben két csokifajtát, egy nagyon magas és egy közepes minőségű csokit árusítottak a járókelőknek. A különleges csoki árát levitték 30 centről 15-re, az átlagos csokit viszont nagyjából a megszokott áron, 1 centért kínálták. Ilyen feltételek mellett a vásárlók háromnegyede a drága, egynegyede az olcsó csokit választotta (a 14 centes árkülönbség ellenére a jelentős árengedmény nagy csáberőt jelentett). Amikor azonban további 1-1 centtel csökkentették mindkét termék árát, a vásárlási hajlandóság teljesen eltolódott az olcsóbb, akkor már ingyenes csoki irányába. Az ingyenesség hatását sokféle módon vizsgálták, és mindig

[16] Dan Ariely magyarul megjelent könyvéről szóló ismertetőmben bővebben írtam erről [521].

ugyanazt tapasztalták: az ingyenesség mágnesként vonzza az embert. Egy másik egyszerű kísérletben egy bevásárlóközpont bejáratánál 7 dollárért osztogattak egy 20 dolláros ajándékutalványt, illetve ingyen egy 10 dollárosat a vásárlóknak. Annak ellenére, hogy anyagilag jobban jártak azok, akik fizettek a 20 dolláros utalványért (ők 13 dollárt nyertek az ügyleten), az elsöprő többség a 10 dolláros ajánlatot fogadta el, ahol nem kellett semmit sem fizetnie. Az ingyenesség varázserejét Ariely úgy magyarázza, hogy amikor a döntéseink során mérlegeljük a lehetséges előnyöket és hátrányokat, sokkal jobban félünk a rossz döntéseink nyomán esetleg bekövetkező veszteségektől, mint amennyire vonzanak a várható előnyök. Ha pedig valamit ingyen szerezhetünk meg, ez olyan érzelmi többletet ad az ingyenes alternatívának, ami legyőzi az előzetesen kalkulált veszteségérzetet.

Jól ismert az előzetes várakozások döntésbefolyásoló szerepe is: jobban ízlik ugyanaz a sör, ha úgy tudjuk, márkás gyártótól származik, jobban tetszik a mozifilm, ha nagyon jó előzetes kritikát kapott, egy szépen megterített asztalnál ülve finomabbnak ítéljük meg az elfogyasztott vacsorát stb. Az előzetes elvárások (vagy mondjuk így: sztereotípiák) gyakran előkészítik mind a cselekvőket, mind a véleményt formálókat, így azután az előzetesen várt eredmény be is következik (tehát az előítéletek sokszor önbeteljesítő próféciaként működnek). Hasonló a placebohatás esete: a placebo tényleg működik, a drágább gyógyszereket hatásosabbnak érezzük. Az előítélet, sztereotípia viszont hasznos is tud lenni, mert nem lehet mindig, mindent a nulláról indítva megítélni, dönteni.

Paul Slovic az "arány dominanciája" kifejezéssel illetve idézte fel C.K. Hsee kísérleti példáját amelyben ha a kísérleti alanyoknak fagyit kínáltak nagy pohárban, de nem teletöltve, vagy kis pohárban, de jól megpúpozva, akkor a legtöbb kísérleti alany az utóbbit választotta. Ez az arány megfordult, ha a két poharat egymás mellé tették, és összemérhetővé vált a két pohár nagysága [496].

Ezekről a heurisztikákról már volt szó korábban. Akkor, amikor az indulati-érzelmi cselekvések jellemzésekor bemutattam Zajonc, Epstein, Brunner, Slovic elméleteiben található megkülönböztetést a racionális-analiktikai és a tapasztalati-érzelmi gondolkodásra és megismerésre vonatkozóan (↬191.o.).

A viselkedés-gazdaságtan legfontosabb üzenete az, hogy nem minden helyzetben racionális az ember, és amikor nem az, akkor különféle heurisztikák szerint cselekszik. A kétféle viselkedés kétféle gondolkodást, kétféle beállítódást feltételez, és az a nagy kérdés, hogy meg tudjuk-e mondani, mikor melyik válik dominánssá. Hámori Balázs ezt a termelési logikák, amunkavégzés minőségének átalakulásával véli magyarázhatónak.

> „Addig talán rendben is volt a racionalitás axiómaként, modellen kívüli, a priori adottságként kezelése, amíg a gondolkodás, az értelem működése, az információk feldolgozása csak másodlagos kísérőjelenség volt az alapvetően materiális természetű gazdasági folyamatokban. Ma azonban, amikor egyre több ember számára a gondolkodás, a döntés nem a termelést megszakító epizód, hanem maga a munka, már nem kielégítő minden további vizsgálódás nélkül elfogadni az emberi gondolkodásra vonatkozó feltételezéseket." [210]

Bizonyos mértékig helytállónak gondolom ezt az érvet, de azt hiszem, hogy ennél többről és részben másról van szó. Szerintem különböző helyzetek az ember másfajta képességeit aktivizálja és így másfajta cselekvésre készteti. Ebben a döntő külünbséget az a kétféle megismerési, gondolkodás szisztéma közli választás vagy inkábbb választódás jelenti. Bizonyos helyzetekben az érzelmi-tapasztalati séma, máskor meg a racionális-analitikai séma mentén működünk. Lehet, hogy tendenciaszerűen korszakolni lehet az emberiség történelmét, hogy milyen helyzeteket állít elő, és ebből fakadóan milyen működési logikák

aktiválódnak bennünk. Egy erre tett javaslatként fogom fel Hámori fenti gondolatát. De a jelenség jobb megértéséhez még szükségünk van arra, hogy a helyzettípusok és az aktiválódó magatartási sémák közti kapcsolatot jobban le tudjuk írni.

Összefoglalás

A weberi cselekvéstípusok elemzése után szükségesnek tartok néhány észrevételt tenni. Azt, hogy miért a racionalitás fogalma van a tipológia középpontjában, hogy a racionalitás fogalma mentén "sorba lehet " fűzni a négy weberi cselekvéstípust, sokan megírták már, még maga Weber is [530]. Bertalan László figyelmeztett rá, hogy a weberi racionalitásfogalomnak van egy szűkebb és egy tágabb értelmezése [49]. A szűkebb értelmezés szerint racionálisnak csak a célracionálist tarthatjuk, és a további három típus , tehát az értékracionális, az érzelmi és a habituális valamilyen mértékig irracionális. A tágabb értelmezés szerint racionális a célracionális és az értékracionális, míg irracionális az érzelmi és a habituális. De ezt Weber is pont így látta.

> „A cselekvés »racionalizálásának« *egyik* lényeges eleme, hogy a meggyökeresedett szokáserkölcshöz való belső alkalmazkodást az érdekviszonyokhoz való tervszerű alkalmazkodás váltja fel. De ez persze nem meríti ki a cselekvés »racionalizálásának«fogalmát. Ezen kívül beszélhetünk még értékracionalizálásról, ha tudatos értékek kerülnek a szokáserkölcs, és nemcsak a szokáserkölcs, hanem az indulati-értelmi cselekvés helyére. Végül lejátszódhat a racionalizálás úgy is, hogy az értékekbe vetett hittől mentes, tisztán célracionális cselekvés kerül az értékracionális cselekvés helyére." [575, 4.§.3.]

Abban viszont már koránt sincs ilyen egyetértés, hogy az **érdek**, az **érdekvezérelt cselekvés** fogalmának vajon milyen a viszonya a weberi négyes cselekvésfelosztáshoz. Elég széles körben népszerű az a nézet, hogy egyfelől az érdekvezérelt cselekvés megegyezik a célracionális cselekvés fogalmával, másfelől az érdek kategóriája az önzés fogalmával. Az első állítást akár még el is tudnám fogadni, de hogy mégsem teszem azt, annak az az oka, hogy a második azonosságot teljesen elfogadhatatlannak tartom, és a célracionális vagy érdekvezérelt cselekvést semmi esetre sem tartanám kizárólag egoista motívumokon alapuló cselekvésnek. Mégha sokan is gondolják ezt így. Albert O. Hirschman például az érdekek és a szenvedélyek viszonyáról, az érdek fogalmának megjelenéséről és elterjedéséről szóló kitűnő könyvében azt a kitételt használta, hogy az "érdek mint ésszerű, szándékos önzés" [235, 51], és könyvében végső soron "azonosult" az alábbi idézetben foglalt véleményekkel.

> „Sokan úgy vélték, hogy az érdek egyesíti a két kategória jobbik természetét, hiszen az értelem felértékeli és kordában tartja az önzés szenvedélyét, s ez a szenvedély irányt és erőt ad az értelemnek. Az ilyen módon elegyített emberi cselekvés, úgy tartották, mentes a szenvedélyek romboló hatásától és az értelem tehetetlenségétől." [235, 51]

A következő fejezetben alaposabban elemzem majd a társas motivációk körét, akkor szó lesz az agoizmus, altruizmus viszonyáról, helyéről a társas motivációk egészén belül (↪234.o.), ezért itt csak azt erősítem meg újra, hogy a Weber által bevezetett célracionális cselekvés vagy a sokak által használt érdekvezérelt cselekvés fogalma nem kapcsolható kizárólagos módon össze az egoista motiváción alapuló cselekvés fogalmával.

A célracionális cselekvéssel kapcsolatos további fontos észrevételem még, hogy bár a célracionális cselekvés a weberi tipológia középpontjában helyezkedik el, mert ez a fogalom kézenfekvő viszonyítási alapot ad a többi elemzéséhez is, de a fogalmi elemzés szempontjából vett központi szerep nem jár együtt szükségszerűen a valós társadalmi gyako-

rlatban való központi szereppel, elterjedtséggel. A racionalitás fontos az életünkben, fontos a társadalmak fejlődésében is, de a racionalitás sosem volt, most sincs és sosem lesz egyeduralkodó. A racionalitás kizárólagosságának, de talán még a racionalitás dominanciájának az elképzelése csak egy hamis mítosz. Ennek több oka is van.

> "A racionalitás elve két teljesen eltérő módon sérülhet. Egyrészt az emberek minden igyekezetük dacára gyakran követnek el logikai következtetési hibákat. Másrészt a cselekvéseik sokszor 'meggondolatlanok': amikor a szokás vagy a szenvedély vezérli őket, akkor még csak nem is törekednek a racionális önkontrollra. (...a racionalitás effajta hibái, csakúgy, mint az önérdekvezéreltség posztulátumának megsértései, funkcionálisan adaptívnak bizonyulhattak az emberi faj genetikai és kulturális fejlődésében.)" [237]

Az előző részekben írtam sokat az aracionalitás gazdaságosságáról, racionalitásáról, a döntési paradoxonokról a döntéselmélet területéről, a korlátozott racionalitásról, a kilátáselmélet által megmagyarázható irracionálisnak ítélt jelenségekről, az érzelmeink, szokásaink ésszerűségeiről. Cselekvéseinket, magunkat csak úgy értjük meg, ha nem engedjük meg a leegyszerűsítő megközelítéseket.

Az összefogalás végén bemutatok egy kicsit talán meglepő ábrát, ami a weberi cselekvéstípusok és néhány további fogalom viszonyát ábrázolja. Mondhatnám azt is, hogy az áttekintő ábrán a különböző motivációjú cselekvéstípusok egymáshoz való viszonyát a logika hatszög segítségével mutatom be, de ez csak formai értelemben igaz, tartalmilag nem. Hat (pontosabban: nyolc) Weber által használt fogalmat ugyanis be lehet mutatni egy olyan rendezésben, amely formailag megegyezik a logikai hatszög struktúrájával. A logikai négyszög sarokpontjaiba lehet betenni a négy weberi cselekvéstípus kategóriáját. A köztük levő viszonyt azon az alapon lehet megmagyarázni, hogy mindegyik cselekvéstípust az alapján ítéljük meg, hogy mennyi benne a tudatos mérlegelés. Az (**A**)-sarokban levő célracionális cselekvés mindent mérlegel, ebben az értelemben univerzális kategória. A másik univerzális kategória az (**E**)-sarokba tett habituális cselekvés, amikor is nincs semmi tudatosan mérlegelve. A két partikuláris kategória az (**I**)-ben levő értékracionális cselekvés, illetve az (**O**)-ban levő indulati-érzelmi cselekvés, amikor vagy egy szempont nincs (mert nem lehet) mérlegelve, vagy egyetlen szempont jelenik csak meg a tudatban, tehát csak azt lehet "mérlegeni".

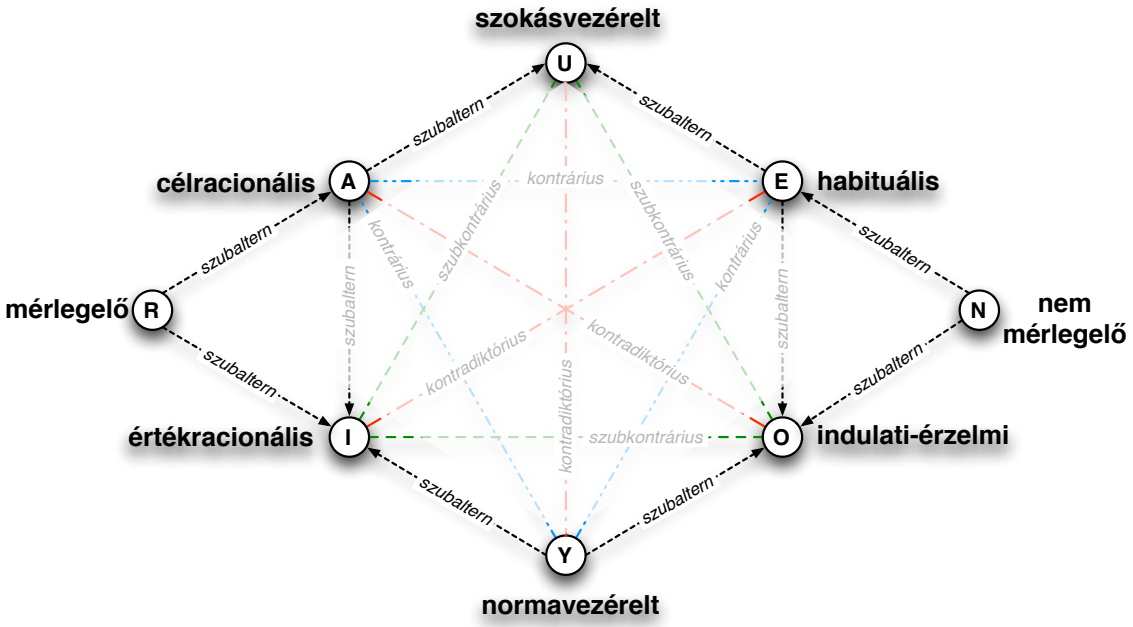
A négy cselekvéstípus mellett a **szokáshoz igazodó cselekvést**, illetve az e fogalom alá sorolt **szokáserkölcs** és **érdekhez igazodó szokás**, valamint a **rendhez igazodó cselekvés** fogalmát kell figyelembe vennünk. Ez nyolc kategória, de ezek közül a szokáserkölcs fogalma megfeleltethető a habituális cselekvés, az érdekhez igazodó szokás pedig a célracionális cselekvés fogalmának. A szokás fogalmának két altípusát maga Weber definiálta így, és ő volt az is, aki az érdek fogalmát a szokás kategóriája alá rendelte. Azt, hogy a két-két fogalom megfeleltethető egymásnak, azt én állítom [519], de úgy vélem, hogy ez a weberi definíciók alapján könnyen belátható. Ha viszont ezt elfogadjuk, akkor a logikai hatszög (**U**)-pontjába behelyettesíthetjük a szokásvezérelt cselekvés kategóriáját úgy, hogy közben még az az összefüggés is fennmarad, miszerint a szokás és a célracionális (érdekvezérelt) és habituális cselekvés diszjunkciójaként írható le (↪48.o.). Azt is könnyen elfogadható javaslatnak érzem, hogy a normához igazodó cselekvés kifejezését felcserélhetjük a normavezérelt cselekvés teminusával. Weber ugyan másként tipizálta a rendhez igazodó (normatív) cselekvéseket, itt mégis megfontolhatónak érzem, hogy ezt a kategóriát betegyük a logikai hatszög (**Y**)-sarokjába. A normához igazodásban ugyanis mindig van egy kicsi értékracionalitás és mindig van egy kicsi érzelem. Az értékracionális

cselekvés furcsa keveréke az elköteleződésnek és a mérlegelő beállítódásnak – ez talán könnyebben elfogadhatónak látszik. Annak támogatására, hogy a normakövetésben ott lappang mindig az érzelem, Jon Elstert hívom segítségül, aki szerint:

> „ . . . a szégyen(érzet) nem pusztán egyik támogatója a társadalmi normáknak, hanem *a* támogatója.” [146, 145]

Megkockáztatom tehát, hogy a normavezérelt cselekvéshez szükséges mind az értékracionális, mind az indulati-érzelmi cselekvés együttes minősége (ahogy ezt a logikai hatszög (**Y**)-pontjától “el lehet várni”).

A hat cselekvéselméleti típuskategóriát tehát összefűzhetjük egymással a logikai hatszög sémája szerint, de azt nem állítom, hogy mindemellett a logikai hatszög esetén fennálló logikai kapcsolatokat is tételezhetjük köztük, sőt, ezek létezését tagadom. A hatszög szerkezete tehát csak bizonyos mértékig és bizonyos szempontból érvényes itt, de ha másra nem, legalább szemléltetési célokra alkalmas az a párhuzam. Az ábrába felvettem még két új kategóriát, amelyek a mérlegelés minősége mentén fognak egybe két-két kategóriát. Az (**R**)-pontba tett, mérlegelésen alapuló cselekvések alá tartozik a két racionális cselekvéstípus, míg a (**U**)-sarokban levő nem mérlegelő kategória két altípusa a habituális és az indulati-érzelmi cselekvés.[17]



[17]Bertalan László részben másfajta felosztást javasolt, amikor a ‘mérlegelő/nem mérlegelő’ dimenzió mellett az ‘önértékű beállítódás’, illetve ‘nem önértékű beállítódás’ pólusai mentén osztályozta a weberi cselekvéstípusokat [49].

Chapter 4

Társadalmi kapcsolat

„*Sir Dannis Robertson* [szerint] a közgazdász feladata megmutatni, hogyan lehet minimalizálni a legszűkösebb erőforrás, a szeretet iránti igényt. Azt követelte közgazdásztársaitól, hogy azonnal kezdjenek figyelmeztető csaholásba, ha olyan javaslatokkal találkoznak, melyek hatékony megvalósításához szeretetre van szükség.” [85]

„Egy racionális ágenstől általában elfogadjuk, ha saját érdekei szerint cselekszik. Egy mesterséges ágenstől azonban azt várjuk el, hogy valaki más (a felhasználó vagy a tervező) érdekében járjon el, aki az elvárásait valahogy közvetítette az ágens felé.” [69]

A társadalmi cselekvés fogalmának elemzésekor kiderült, hogy a cselekvés fogalma annyiban szűkebb terjedelmű az egyszerű cselekvés fogalmához képest, hogy itt “megköveteljük” azt is, hogy a cselekvőn túl legyen még valaki, akihez a cselekvő – valamilyen módon – igazodik a cselekvésével (↪**??**.o.). Ekkor azonban ez a másik személy még csak közvetítő szerepet játszik: a társadalom potenciális hatását fejezi ki. Ez a másik személy a társadalom jelzése. Ahogy a cselekvő igazodik hozzá, úgy már a társadalmiságot érhetjük tetten. De ez még csak hatás, és nem kölcsönhatás. A társadalmiságban benne rejlő kölcsönösséget, kölcsönhatást, kölcsönös egymáshoz igazodást a **társadalmi kapcsolat** fogalmával lehet igazán jól megragadni.

> „Társadalmi »kapcsolatról« beszélünk, ha többek magatartása értelmi tartalmának megfelelően kölcsönösen egymáshoz igazodik, és magatartásukat ez a kölcsönös igazodás irányítja.” [575, 54]

A cselekvés és a társadalmi cselekvés fogalmába már bele kellett építeni az **ágencia**, valamint az **intencionalitás** minőségét, a most következő feladat a **kölcsönösség**, és egy kicsit később majd az **interdependencia** összetevéjének felvétele az elméleti keretrendszerbe. A következő fejezetben a kölcsönös igazodásnak, egymásra figyelésnek egy különösen fontos esetét vizsgálom meg. Azt a beállítódást elemzem, amely szabályozza bennünk, hogy miként viszonyulunk magunkhoz (Egohoz) és másokhoz, a másikhoz (Alterhez) a cselekvéseink során.

4.1 Ágensre irányultság: társas preferencia

Egyszer New Yorkban egyedül sétáltam a Central Parkban, amikor elém termett egy fiatal gyerek, akiről lerítt, hogy pénzt akart kérni tőlem. Már nem emlékszem pontosan, mi, hogyan történt. Olyasmit kérdezhetett, hogy honnan jöttem, mire azt feleltem neki, hogy Magyarországról ('from Hungary'). Lehet, hogy a kiejtésem nem volt tökéletes, lehet, hogy a fiú nem hallott még arról, hogy van ilyen nevű ország, de az biztos, hogy félreértett engem. Félreértett és megszánt. Nem mondott semmit, csak nézett rám egy ideig, miközben azt gondolta rólam, éhes lehetek ('hungry'). Ezt onnan tudom, hogy a zsebébe nyúlt, elővett egy csokit és odanyújtotta nekem.

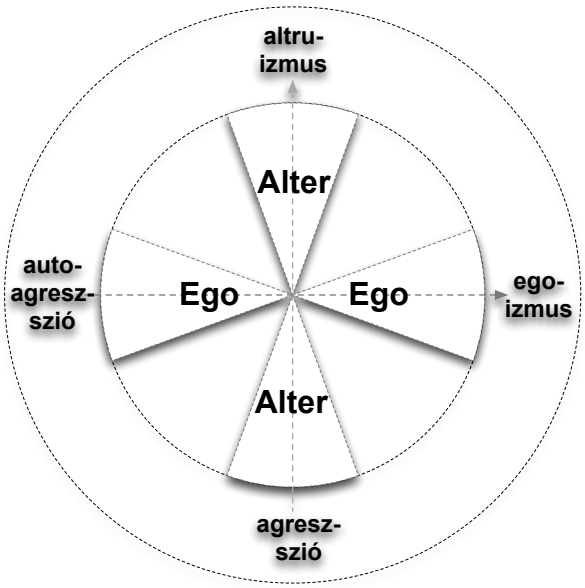
Ezzel a történettel a közgazdászok nem igazán tudnának mit kezdeni. A közgazdaságtan követőinek túlnyomó részére ugyanis nagyon jellemző az a beállítódás, amely szerint az ember – gazdasági – cselekvését az 'önző ember' hipotézisére támaszkodva lehet (kell) megmagyarázni. Pedig ez nem igaz. Tagadhatatlan, hogy sokszor, nagyon sokszor valóban működőképes a hipotézis. Lehet, hogy az esetek nagyobb részében. Talán. De ugyancsak sokszor, sőt, nagyon sokszor találkozhatunk olyan helyzetekkel, amelyekben önzetlen beállítódások mentén cselekszenek az emberek. A társadalmi cselekvés modelljébe ezt a kétfajta beállítódást mindenképpen fel kell vennünk. De ha már beszélünk **önző** és **önzetlen** beállítódásokról, akkor érdemes azt is megvizsgálni, nem lehet-e általánosabban kezelni ezt a kérdést, nem lehet-e egy olyan elméleti keretet felállítani, amelyben az egoizmus és altruizmus is értelmezhető, de a modellbe fel lehet venni további fogalmakat is. A beállítódás, a motiváció fogalmát kell tehát pontosítani, és ezen belül lehet majd értelmezni a cselekvők egymáshoz való viszonyát kifejező fogalmakat. Utóbbiak terjedelme szűkebb a beállítódás általánosabb fogalmának terjedelméhez képest, meg kell tehát mondanunk azt, hogy milyen jellemzők mentén tudjuk meghatározni, tipizálni az ágenciával kapcsolatos beállítódásainkat.

A beállítódásaink, motivációink sok esetben abban az értelemben (és irányban) szűkülnek le a cselekvéseink során, hogy gyakran "csak" arra figyelünk, hogy a cselekvéseink következményei hogyan érintik a társadalmi kapcsolat résztvevőit, az ágenseket (Egot és Altert). A kérdés itt az, hogy a cselekvő mennyire veszi figyelembe azt, hogy a döntése (cselekvése) hat(hat) részben saját magára, részben a társára (a másikra, a többiekre). Kézenfekvőnek tűnik erre felállítani azt a fogalmi kettősséget, amely szerint a cselekvő (Ego) lehet önző vagy önzetlen. Ezek a fogalmak azt ragadják meg, hogy a döntés/cselekvés következményei mennyire jók, előnyösek vagy károsak, előnytelenek az Ego, illetve az Alter számára. Ha az Ego a hasznok maximalizálása, illetve a károk minimalizálására törekszik (és csak arra), akkor egoizmusról ('Ego'-izmusról), egoista beállítódásról beszélhetünk, ha döntései során figyelembe veszi Alter szempontjait is, akkor altruizusról ('Alter'-izmusról), altruista beállítódásról. Ha így gondolkozunk, akkor az önző, egoista (selfish, egoist) és az önzetlen, altruista (unselfish, selfless, altruist) beállítódás egymás ellentétének, komplementerének tekinthetjük.

Ez az ellentét azonban megszűnik, pontosabban átalakul, ha a másikhoz (Alterhez) való értékviszonyulást úgy kezeljük, hogy eközben figyelünk az értékelés (mint olyan) szükségszerűen **bipoláris** jellegére. Ha feltételezzük, hogy a döntés/cselekvés nem mindig az Ego és/vagy Alter javára irányulhat, de előfordulhat az is, hogy bizonyos helyzetekben Egonak vagy Alternek (vagy mindkettőnek) kára keletkezik, és ezt a lehetőséget is

értékelni akarjuk, akkor a modellünket bővíteni, általánosítani kell ezen a ponton. Ha a bipolaritást figyelembe akarjuk venni, akkor a **jóindulatú** (benevolent) és **rosszindulatú** (malevolent) beállítódást kell szembe állítanunk egymással. A jóindulat az ágens javát, a rosszindulat az ágens kárát preferálja.

A bipolaritás modellbe emelése magával vonja azt, hogy az egoizmus-altruizmus ellentétpár helyett másfajta kettősséget kell alkalmaznunk. Az **ágensre irányultság** általános fogalmából kell kiindulnunk. Ennek segítségével a **magára irányuló** (self-regarding, self-serving), illetve a **másikra irányuló** (other-regarding, other-directed) fogalmakat használva megragadhatjuk a kétfajta ágens (Ego és Alter) irányába történő viszonyulás mozzanatát.[1] Ha ki akarjuk fejezni az ágensre irányultság pozitív és negatív típusait, akkor további fogalmakat kell felvennünk. Később részletesebben kifejtem, mivel, miért és hogyan érdemes bővíteni a fogalmi modellünket, de már előzetesen érdemesnek tartom itt jelezni, hogy az Alter felé irányuló rosszindulatú beállítódás jelzésére az **agresszió**, az Egora vonatkozó rosszindulatra pedig az **autoagresszió** (önagresszió) fogalmakat vehetjük igénybe. Ha így teszünk, akkor lesz két dimenziónk, az ágensre irányultság, illetve e viszony polaritása, valenciája. A két dimenzió mentén négyfajta beállítódást fejezhetünk ki: ha a jóindulat az Egora irányul, akkor egoizmusról, ha Alterre, altruizmusról, ha a rosszindulat az Egora, agresszivitásról, ha Alterre, autoagresszivitásról beszélhetünk. Kiindulásként felrajzolhatunk egy ábrát, amelyet persze a későbbiekben még sok minden mással ki kell egészíteni.



Az ágensre irányuló beállítódások, a szociális preferenciák ábrájára tekintve egyből hiányérzetünk támadhat amiatt, hogy ez a keret olyan fogalmakat határoz meg, amelyek a szociális preferenciák tiszta végpontjait jelentik egy-egy dimenzió mentén, de nem mutat

[1] Stilisztikai szempontok miatt, ha nem zavaró, használni fogom még a *másra irányuló beállítódás* kifejezést is. Az ágensre irányuló beállítódásokat szokták még **szociális preferenciáknak** (social preferences), **társas motivációknak** is nevezni, ezen belül pedig a magára irányulást a **proszelf** (pro-self), a másikra irányulást pedig **proszociális** (prosocial) beállítódásként értékelik. Bár utóbbi kettő fogalom értelmezésében nincs igazán konszenzus, előfordulhat, hogy a továbbiakban – megint csak stilisztikai szempontok miatt, és csak akkor, ha nem okoz félreértést – olykor használni fogom ezeket a terminusokat is.

semmit az átmenti állapotokból. Márpedig jogos az a feltételezés, hogy vannak, sőt, sokkal gyakrabban vannak olyan esetek, amikor valakinek a másikhoz való viszonya egy kicsit altruizmusból, egy kicsit egoizmusból áll össze, és van úgy, hogy az ember társas irányultsága olykor a pozitív, máskor inkább a negatív pólushoz áll közelebb valamelyik dimenzióban. Nagyon steril az az elképzelés, ami szerint vagy a más javát (kárát) vagy a saját javamat (káromat) akarhatom. Sokkal életszerűbb azt feltételezni, hogy a döntéseink során valamilyen módon és mértékben egyaránt figyelembe vesszük a magunk és a másik javát (kárát).[2] Ehhez viszont arra van szükség, hogy a fenti ábrát (és az eddigi modellünket) további fogalmakkal kiegészítsük.

Magasabbrendű beállítódás: büntetni akarás

Az előző fejezetben a szociális preferenciák osztályozásához használható két dimenziót különítettem el. Az értékirányultságok bipoláris jellegének megragadására a jóindulatú, illetve rosszindulatú beállítódások szembeállítását javasoltam, hogy ezáltal definiálhassuk az agresszió (másiknak kárt okozni) és az autoagresszió (magunknak kárt okozni) fogalmát. Később még tárgyalni fogom az agresszivitás jelenségét, és akkor kitérek arra a kérdésre, hogy hol érhetjük tetten és hogyan magyarázhatjuk az agressziót az állati és emberi társadalmakban. Szükségesnek érzem azonban, hogy egy fontos kiegészítést, vagy inkább valamiféle pontosítást tegyek ezen a ponton. Ha ugyanis az agresszivitáson azt értjük, hogy nem az ágens javát, hanem kárát akarja valaki, akkor e jelenség létezését sok esetben csak úgy tudjuk megmagyarázni, ha figyelembe vesszük azt a tényt, hogy az ember büntetni akarja magát és/vagy a másikat. Ha azonban a büntetés jelenségét is figyelembe vesszük itt, akkor az agresszivitás és a rosszindulat kapcsolatáról is kicsit másként kell gondolkoznunk.

A társadalmi cselekvésekről szólva korábban már említettem, hogy a cselekvés fogalmát magasabb rendben is értelmezhetjük (↰**??**.o.). Amikor az emberek szankcionálják, büntetik vagy jutalmazzák a másikat, akkor az ilyen cselekvéseket magasabbrendű fogalmakkal írhatjuk le, mert önmagukban nem értelmezhetők, csak akkor, ha valamilyen másik cselekvésre "vonatkoztatjuk". A büntetés viszont a legtöbb esetben valamiféle agresszió alkalmazását jelenti. Ebben nem különbözik a "sima" agressziótól. A büntető beállítódásában viszont lényeges különbség figyelhető meg: a büntető jóindulat mentén büntet, azaz saját büntető cselekvését jónak tartja. Kárt okoz a büntetett félnek, de ez a károkozás más. Erről a fontos szemléletváltásról később még hozok példákat, a fogalmi elemzésben is mélyebben kitérek a problémára, itt csak annyit akarok jelezni, hogy az agresszivitás jelenségét, az agresszív beállítódások létezését elég gyakran (bár nem mindig) a büntetési szándékkal tudjuk megmagyarázni. Azt is később fogom elemezni, hogy a büntetés, akár úgy értelmezzük, mint egyfajta beállítódást, akár úgy, mint egyfajta cselekvést, logikai értelemben magasabbrendű fogalomnak számít.

4.2 Társas értékorientáció

A korábbiakban felvettem a modellbe azt a két dimeziót, amelyek mentén leírhatjuk a szociális preferenciákat. Az ágensre irányultság, valamint az értékpolaritás tengelyei men-

[2] A későbbiekben tárgyalandó játékelmélet számára sokkal természetesebb, használhatóbb ez a megközelítés.

tén négy "szélsőértéket" definálhatunk: egoizmus, altruizmus, agresszió, autoagresszió. Következő lépésként bővíteni szeretném a társas motivációk fogalmi modelljét. Ezen a ponton azonban egy – talán – szokatlan módszertani probléma nehezíti a továbblépést. A szociális preferenciák modelljének kiegészítése során a **társas értékorientáció** (social value orientation, SVO) elméletének eredményeit szeretném használni úgy, hogy a fogalmi építkezés logikáját nem törjem meg ezzel. A társas értékorientáció elmélete alkalmas erre, mert zökkenőmentes továbblépést biztosít. Egy olyan módszertant ad a kezünkbe, amely segítségével kísérleti úton lehet mérni az emberek társas beállítódásának irányát és mértékét (minőségét, mennyiségét). Történelmi tény viszont, hogy az egész elmélet azzal a céllal jött létre, hogy valahogy meg lehessen szabadítani a kísérleti mérési szituációkat az ún. interdependencia jelenségétől azért, hogy csak az egyéni cselekvők beállítódásait tudjuk tisztán mérni. Hogy az interdependencia pontosan mit jelent, és miért kell kilépni az ilyen helyzetekből, azt a játékelmélet alapfogalmainak ismeretében lehet elmagyarázni. Ezt viszont – a fogalmi modellezés építkezési logikája miatt – csak jóval később tudom megtenni. Ezt a problémát úgy tudom csak feloldani, hogy nagyon röviden itt is el fogom mondani azt, amit tudni kell a játékelméleti helyzetek leírásáról, hogy érteni lehessen a társas értékorientáció elméletének kapcsolódását a játékelmélethez.

A társas értékorientáció elméletének alapjait David M. Messick és Charles G. McClintock fektették le 1968-ban megjelent cikkükben [341]. A tanulmányból két fontos eredményt érdemes megemlíteni. Egyfelől elkülönítettek háromféle motivációt, társas értékorientációt egymástól, amelyeket **individuális**, **kooperatív** és **versengő** irányultságnak neveztek. Az individuális beállítódású ember csak a saját hasznát (own gain) nézi a döntései során, a kooperatív ember a maga és a társa együttes hasznát akarja maximalizálni (joint gain), míg a versengő típus arra tör, hogy minél nagyobb legyen kettejük közti különbség a maga javára (relative gain). Egy példát mutatva erre az alábbi táblázatban azt ábrázolhatjuk, hogy a vizsgált alany (Ego) két döntési lehetőség közül választhat (X és Y), mindkét alternatívában a maga és a társa számára lehetséges **kifizetéseket** hasonlítja össze:

	X	Y
Ego kifizetése	5	8
Alter kifizetése	2	6

Ha a táblázatban szereplő számok megnyerhető pénzt jelentenek, akkor az X alternatíva választása azt jelenti, hogy Ego jobbnak tartja azt az állapotot, amikor 5 egységnyi pénzt kap, miközben Alter 2-t ahhoz az Y alternatívához képest, amely során Ego 8, Alter 6 egységhez jut. Egy ilyen helyzetben az individualista irányultság mentén Egonak az Y alternatíva a jobb, hiszen 8-at kap 5 helyett, és ugyancsak az Y-t fogja választani, ha kooperatív beállítódással rendelkezik, mert $(8+6) > (5+2)$. A versengő értékorientáció esetén azonban változik a helyzet, és ekkor már Ego az X alternatívát fogja preferálni, hiszen $(5+2) > (8+6)$. Egy ilyen kicsi táblázat arra is alkalmas, hogy különböző számokat írogatva bele, segítségével kísérleti helyzetekben egyszerűen mérni lehet az emberek a társas beállítódását.

A tanulmány másik eredménye az, hogy a három irányultság elkülönítése után a szerzők egyfelől megmutatják, hogy összesen hatféle módon lehet két-szereplős, kétalternatívás helyzetekben a háromféle beálíítódást reprezentálni, másfelől – és számunkra most ez a fontosabb – rámutatnak arra, hogy izomorf megfeleltetést lehet teremteni a

szociális motivációk mérésére alkalmas táblázatok, illetve a játékelméletben használatos preferenciamátrixok között. Erre pedig azért van (volt) szükség, mert a játékelmélet azokat a társadalmi interakciókat (játékokat) próbálja megragadni és megmagyarázni, amelyekben a résztvevők (játékosok) együttes döntéseitől függ a helyzet megoldása. A játékok mindig olyan helyzeteket írnak el, amelyekben hiába tudjuk, hogy külön-külön mit szeretnének elérni a játékosok, a végeredmény mindkettejük döntésétől függ. Erre mondjuk azt, hogy **interdependencia**, kölcsönös függőség jellemzi ezeket a kapcsolatrendszereket. Mindez azért érdekes a társas értékorientáció elmélete szempontjából, mert a játékelméleti helyzetekben a kölcsönös függőség létezése miatt sosem tudhatjuk biztosan, hogy miért döntött valaki így vagy úgy a játékban (mert a játék végeredménye mindig a játékosok döntéseinek eredőjeként alapján alakul ki). Ezt lehet feloldani a fent jelzett izomorfia segítségével. Egy adott játékelméleti játszma helyett ugyanis elegendő a neki megfeleltethető egy-személyes kísérleti vizsgálatot (az SVO-vizsgálatot) elvégezni, hogy megtudjuk a kísérleti alany beállítódást. A két-személyes játékelméleti játszmákat így vissza lehet vezetni egy-személyes SVO-kísérletekre. Erre az egyszerűsítésre vezették be a szerzők a dekomponált játék (decomposed game) fogalmát. Hatféle dekomponált játékot különítettek el egymástól, amelyek mind izomorfak valamelyik hagyományos játékkal. A megfeleltetést az alábbi kapcsolatrendszer mentén végezték el.

	X	Y
Ego kifizetése	c_1	c_2
Alter kifizetése	d_1	d_2

$\rightarrowtail$

	X_{Alter}	Y_{Alter}
X_{Ego}	$(c_1+d_1),(c_1+d_1)$	$(c_1+d_2),(c_2+d_1)$
Y_{Ego}	$(c_2+d_1),(c_1+d_2)$	$(c_2+d_2),(c_2+d_2)$

Az izomorfia bemutatására vegyük a fogolydilemma játékot. A bal oldali SVO-táblázatba beírt négy érték – fenti "kulcs" alapján – pontosan megfeleltethető a jobb oldali táblázatban látható preferenciamátrixnak, ami pedig a jól ismert fogolydilemmához tartozik.

	X	Y
Ego kifizetése	1	2
Alter kifizetése	2	0

$\rightarrowtail$

	X_{Alter}	Y_{Alter}
X_{Ego}	3,3	1,4
Y_{Ego}	4,1	2,2

Messick és McClintock konverziós megoldását használva a hetvenes évektől kezdve rengeteg kísérletet végeztek, amelyekben a társas motivációkat vizsgálták sokféle szempont alapján. Az induló cikk megjelenése után pár évvel McClintock bővítette a modelljüket és a 'motivációs sémák' közé felvette az 'altruista' irányultságot is [331], ami addig nem volt benne a modellben. Ekkortól tehát már négy motívumot (orientációt) különítettek el egymástól. A négyféle beállítódást úgy mutatom be alábbi táblázatban, hogy (a táblázat utolsó oszlopában) melléjük rendelem azt a döntési alternatívát, amelyeket az adott irányultságú kísérleti személy választana magának az ott látható négy lehetőség közül (a zárójelbe tett két érték közül az első az Ego, a második az Alter nyereményét mutatja).

Az (Ego+Alter) kifejezés a két ágens együttes hasznát jelenti, míg az (Ego-Alter) kifejezés az Ego relatív hasznát Alterhez képest. A 'max' függvényjel pedig azt fejezi ki, hogy az adott irányultság a jelzett haszon maximumát szeretné elérni.

individualizmus	saját nyeremény maximuma	max(Ego)	(100,70)
kooperáció	együttes nyeremény maximuma	max(Ego+Alter)	(90,90)
versengés	nyereménykülönbözet maximuma	max(Ego-Alter)	(80,40)
altruizmus	másik nyereményének maximuma	max(Alter)	(70,100)

A versengés fogalmát McClintock a bevett felfogáshoz – főleg a játékelméletben megszokott értelmezéshez – képest rendhagyó módon értelmezi, amikor a nyereségek különbségére teszi a hangsúlyt. Ez az orientáció kisebb nyereményt ér el, mint a tisztán önző, individualista beállítódás, viszont az lehet az "előnye", hogy a másikhoz képest jobban jön ki a helyzetből (hiszen így többet szerez a partnerénél, míg az individualista döntés után). Ez a hozzáállás értelmezhető úgy is, mint a kisebb veszteségre való törekvés. A játékelméletben versengésen az itteni felfogás helyett sokkal inkább az individualista beállítódást értik, de ezt többértelműség nem lesz annyira zavaró a későbbiekben, mert a játékelmélet modellje nem képes ezt a finom társas motivációs felosztást "kezelni". Az SVO jól elkülönített versengés és individualizmus fogalmai gyakorlatilag egybemosódnak a játékelmélet világában. Érdemesnek tartom viszont a társas értékorientáció individualizmus terminusát kicserélni az *egoizmus* terminusra a fogalom jelentésének meghagyása mellett. Az indokom csak annyi, hogy az individualizmus fogalmához sok más konnotációs réteg is tapad, ami az egoizmussal kapcsolatban nem jelentkezik, valamint az egész cselekvéselméleti modell legfontosabb ágensszerepéhez, az 'Ego'-hoz sokkal jobban illeszkedik hangulatában az egoizmus terminusa akkor, amikor a magára irányuló beállítódást akarom kifejezni.[3]

A társas motivációk elméletében két változásról kell még említést tennem. Az első változás az elmélet modelljét érintette, ugyanis 1973-ban Donald W. Griesinger és James W. Livingston azt javasolták [198], hogy az addigi motivációs sémákat érdemes lenne kibővíteni. Mivel a bővítés szemléltetésére geometriai módszert alkalmaztak (a nyolc elemű bővített motivációs listát egy körön elhelyezve mutatták be), ezért pár évvel később Wim B. G. Liebrand ezt a reprezentációs technikát beemelte az elmélet központjába [307]. Az új geometriai keretrendszer, a McClintock-féle motivációs gyűrű (Ring measure) már lehetővé tette, hogy ne kategoriális, hanem már mértéken alapuló nominális változóként lehesen a vizsgált beállítódásokat kezelni.

Amikor Griesinger és Livingston négy új motiváció beemelését tartotta szükségesnek, akkor arra hivatkoztak, hogy a szociális irányultságok lehetnek negatív valenciájúak, és az elmélet már elfogadott négy beállítódástípusa mellé (azokkal szemben) fel kell venni további négy típust, a **szadizmus** (sadism), a **mazochizmus** (masochism), az **önfeláldozás** (martyrdom) és a **szadomazochizmus** (sadomasochism) kategóriáit. Ezekben az esetekben a döntéshozó (Ego) negatív módon "áll a dolgokhoz", tehát a haszon helyett az ágens(ek) kárát akarja. Ezt úgy fejezhetjük ki, hogy a haszon maximalizálása helyett a haszon minimializálása (a kár maximalizálása) a "célja".[4] A modell bővítésének szükségességét elfogadva, de az alkalmazott terminusokat kicserélve a korábban már bemuta-

[3] További érvem a terminuscsere mellett, hogy az individualizmus fogalma az evolúcióbiológiában széles körben használt, ami szemben áll a társadalomtudományok által preferált egoizmus terminusával. A tudományok közti terminológiai versengésben – ahol csak lehet – saját szakterületem gyakorlatát követem.

[4] A kármaximalizálás vs. haszonminimalizálás fogalmai között ekvivalenciát teremthetünk ilyenfajta megfeleltetéseken keresztül:

max(Ego-Alter) = max(Ego) & min(Alter)

max(Alter-Ego) = min(Ego-Alter) = min(Ego) & min(-Alter) = min(Ego) & max(Alter)

tott, általam preferált terminusokra, az alábbi táblázatba rendezhetjük az új kategóriákat:

autoagresszív (mazochizmus)	saját nyeremény minimuma	min(Ego)	(10,50)
destrukció (szadomazochizmus)	együttes nyeremény minimuma	min(Ego+Alter)	(20,30)
önfeláldozás (mártíromság)	nyereménykülönbözet minimuma	min(Ego-Alter)	(30,80)
agresszió (szadizmus)	másik nyereményének minimuma	min(Alter)	(50,20)

A táblázatban zárójelben feltüntettem az eredeti névjavaslatokat is, és négy újabb alternatíva feltüntetésével jeleztem azt is, hogy az egyes irányultságok mentén melyik lehetőségeket választanák a kísérleti alanyok.

Ha egyesítjük az eredeti (négyes) és az új motivációs sémakészletet, akkor egyszerre láthatjuk őket a melléjük rendelt, rájuk jellemző kifizetésekkel együtt. A könnyebb összehasonlítás végett a kifizetéseket teszem a táblázat első oszlopába:

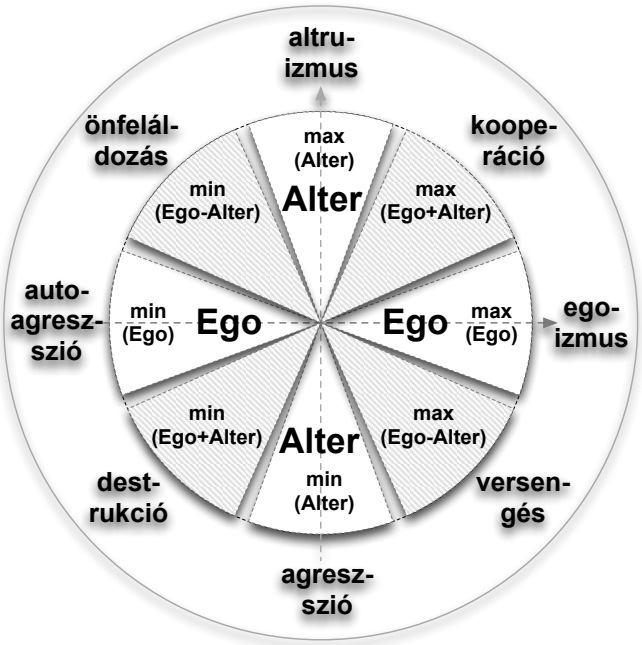
(100,70)	egoizmus	max(Ego)
(90,90)	kooperáció	max(Ego+Alter)
(80,40)	versengés	max(Ego-Alter)
(70,100)	altruizmus	max(Alter)
(10,50)	autoagresszió	min(Ego)
(20,30)	destrukció	min(Ego+Alter)
(30,80)	önfeláldozás	min(Ego-Alter)
(50,20)	agresszió	min(Alter)

A nyolc szociális preferencia közül a Griesinger-Livingston-féle negatív motivációk modellbe emelését közel sem minden kutató helyeselte. Az agressziót mára már a legtöbb kutató elfogadta, de a további három negatív beállítódás mérését sokan fölöslegesnek, megalapozatlannak tartották (és tartják). Ezzel szemben a társas értékorientáció elmélet egyik jeles képviselője, Paul A. M. van Lange egy új irányultságot is felvett a McClintock-féle motivációs gyűrű elemei közé. A kooperáció mellé javasolta betenni az egyenlőségre törekvő *egalitarianizmust* (egalitarianism) mint a proszociális beállítódás másik megnyilvánulási formáját [556], azonban egy, a munkatársaival írott, 2007-es cikkében a kooperatív és egalitárius irányultságot már összevonja [557]. Viszont ők is csak 5 orientációról beszélnek (a négy pozitív motiváció mellett vizsgálják még az agressziót).

A SVO-kísérletek egyik nagy hozadéka az az ismeret, amit a különböző szociális preferenciák elterjedtségéről tudnak mondani a kutatók (ilyen eloszlásra vonatkozóan adatokat közöl például[558]). A kísérletek tanusága szerint a proszociális (kooperatív, az egalitárius és altruista) beállítódás a teljes népesség majd 60%-ra jellemző, akik között több a nő, és az életkorral együtt nő az proszociális irányultság aránya is. A sorozatos felmérések ismétlődő tapasztalata, hogy elenyésző számban vannak azok, akik tisztán altruista szempontok szerint döntenek. A proszelf beállítódású emberek durván 35-40 %-nyian vannak, akik közül az egosita (individualista) motivációjú személyek közel 30 %-ot tesznek ki. Az egoisták között többségben vannak a férfiak. A versengő típusú emberek alkotják a népesség 10-15 %-át, míg az agresszív orientáció aránya gyakorlatilag nulla, bár az ilyen esetek olykor előfordulnak kísérleti szituációkban is. Az agresszió fakadhat bosszúvágyból, felháborodásból is. A kísérleti személyeket például feldühítheti az, ha valaki más sosem kooperál, ha kihasználja a többieket, és a felháborodott alanyok ilyenkor a saját érdekeiket is félretéve igyekeznek neki minél nagyobb kárt okozni a másiknak. Meg kell még itt jegyezni, hogy az agresszív motivációt eleve kevesebbet is vizsgálták,

többek között azért, mert vannak, akik nem sorolják a proszelf orientációk közé ezt a beállítódást mondván, hogy nem az ágens javára, hanem kárára tör.

Akárhogy is ítélik meg a különböző motivációkat a társas értékorientáció elméletén belül, a társadalmi cselekvések modelljébe a teljes motivációs gyűrűt érdemes felvenni – még akkor is, ha vannak olyan típusok, amelyek a gyakorlatban kevésbé elterjedtek, kevésbé jellemzik a cselekvőket. Az alábbi ábra azt mutatja, hogy a társadalmi cselekvőknek milyen társas motivációikat tulajdoníthatunk elméletileg.



Bár ezt a nyolcas tagolású felbontást nem tudom teljes mértékben kihasználni a későbbiekben, az elméleti teljesség miatt meghagyom így a modellt. A játékelméleti fejezetbe "csak annyit" viszek tovább belőle, hogy akkor majd megengedem, sőt, tudatosan alkalmazni fogom azt, hogy a játékosok nemcsak proszelf (egoista vagy versengő) beállítódással játszhassák a játékaikat, de proszociálisak (altruisták vagy kooperatívak) is lehessenek.

Ez a kettőség viszont felvet egy további komoly kérdést. Ha elfogadjuk, megengedjük, hogy a – játékelmélettel leírt – társadalmi helyzetekben különböző társas motivációkkal léphessenek be emberek, akkor ezt a különbözőséget, az irányultságok, a társas preferenciák különbözőségeit meg kell tudnunk ragadni a fogalmi modellünkben is. Ehhez – Franz Dietrich és Christian List javaslatát megfogadva – kettős ontológiát (esetleg többes ontológiát) kell kezelnünk, hogy egyszerre leírhassuk, értelmezhessük a játékosok (és a kívülálló modellezők, megfigyelők) különböző beállítódásait ugyanazon játékszituációra vonatkozóan:

> „ . . . a preferencia kialakulását leíró, tulajdonság-alapú beszámolónkat újraformálhatjuk mint kettős-ontológiai leírást, amelyben az ágens alternatíváinak ontológiáját elkülönítjük a megfigyelőétől." [131]

A nagy kérdés itt az, hogy miként akarjuk modellezni a társas preferenciákat. Ha azt állítjuk, hogy a cselekvő, az Ego proszociális értékekkel rendelkezik, akkor ezt hogyan írjuk le? Nem is az az igazi kérdés itt, hogy milyen súlyokat rendeljünk a magára irányuló,

illetve másikra irányuló beállítódások erősségének kifejezésére, hanem az, hogy mit fejezünk ki pontosan a proszociális motívumokkal. Az Ego vagy a megfigyelő vagy az Alter véleménye, preferenciarendezése számít-e ilyenkor? A különböző válaszok esetén mást és mást kell majd a proszociális irányultságokat leíró formulákba, képletekbe beírni.

A társas értékorientációval kapcsolatos fogalmakat két szinten vizsgálhatjuk: a tényleges cselekvések szintjén, illetve a cselekvésekkel kapcsolatos motivációk, irányultságok szintjén [503]. Az altruista minősítéhez hozzátartozik egyfelől Ego olyan cselekvése, amely Alter számára előnyös, Egonak költsége (vagy elmaradt haszna) volt vele, viszont azt is "elvárjuk", hogy másfelől Ego altruista motiváció miatt tegye meg azt, amit megtesz. Valaki csinálhat olyat, amivel másnak kedvez, magának árt, de ha ezt öngyűlöletből vagy önbecsülés hiányából teszi, akkor ezt nem nevezhetjük altruizmusnak [368].

Az egoizmus minden késztetést, minden vágyat magára irányulónak tekint, az altruizmushoz "elegendő" egy másikra irányuló késztetés, vágy léte, és már beszélhetünk róla. Ez módszertanilag fontos különbséget takar a két fogalom között. Az egoista beállítódás magára irányul, azzal nincs elméleti kiterjesztési probléma. A másikra irányuló orientáció esetében azonban felmerül a kérdés, hogy lehet-e, érdemes-e tágítani a 'másik' fogalmának terjedelmét. Az általam kifejtett megközelítésben a 'másik' egy 'ágens'. A cselekvéselmélet számára ez az értelmezés elegendő. Elméletileg lehetne tágítani ezen az értelemzésen, ha megengednénk, hogy a 'másikra irányulás' lehessen egy eszme, az emberiség egésze, a Föld stb. Angolul ezekre lehetne az irányultságokra a 'selfless' terminust alkalmazni, magyarul nem tudok ennek megfelelőjéről.

Korábban bemutattam, hogy miként definiálta Sober és Wilson a végső és instrumentális vágyak fogalmát (↰152.o.). Soberék a vágyfogalmak tisztázása után definiáltak három motivációfogalmat is [503], amelyek közül kettő már itt is szóba került. Az egoizmus és altruizmus mellett harmadik kategóriaként vezették be a **hedonizmus** fogalmát. Az egoizmust, illetve altruizmust magára irányuló (↰235.o.), az altruizmust másra irányuló végső vágyként definálták, a hedonizmust pedig – az egoizmus alá rendelve – olyan vágyként határozták meg, amely az élvezet megszerzésére és a fájdalom elkerülésére irányul.

Akkor beszélhetünk egyáltalán egoizmusról vagy altruizmusról, amikor az egyén saját érdekei és a releváns másik (mások) érdekei keresztezik egymást. Ilyen esetekben az altruizmus mások érdekét saját költségén támogató cselekvést jelent, az egoizmust viszont a saját érdeket mások költségén való követésével ragarhatjuk meg. Mivel ekkor két személy közti kapcsolatot elemzünk, mondhatjuk azt, hogy interperszonális önzésről, illetve interperszonális önzetlenségről van itt szó. Ennek analógiájára bevezethetnénk egy új fogalomkettőst, amit Pierre Le Morvan javasolt, és ami ugyanazon személy jelenbeli és jövőbeli érdekei között konfliktusok esetén használható, vagis intraperszonális kategóriákról, az intraperszonális önzésről, illetve intraperszonális önzetlenségről lehetne szó [368]. Bár megfontolandó maga ez a javaslat is, én itt inkább csak azt emelném át saját modellembe, amit Le Morvan a szükséges feltételek kapcsán dolgozott ki. Az egyszerűség és az összehasonlítás kedvéért mondhatjuk azt, hogy az önző saját, az altruista a másik érdekét elégíti ki, akkor az előbbi önérdekkövető, az utóbbi másérdekkövető, tehát van **önérdek** és **másérdek**. Ennek analógiájára nevezzük a jelenbeli érdeket **jelenérdeknek**, a jövőbelit pedig **jövőérdeknek** (ez utóbbi két kategóriát nyilván az intraperszonális attitűdökhöz rendelhetjük). Az új terminusok segítségével egyszerre fejezhetjük ki azt a három feltételt, amit Le Morvan rögzített az altruizmus és egoizmus elkülöníthetőségével kapcsolatban.

divergencia feltétel	legyen különbség, konfliktus az önérdek/jelenérdek és a másérdek/jövőérdek között
költségterhelési feltétel	az egyén előtt legyen választási lehetőség abban, hogy a önérdekét/jelenérdekét követi-e a másérdek/jövőérdek rovására vagy fordítva
értékelési feltétel	az egyén számára legyenek olyan esetek, amikor a másérdek/jövőérdek kárára hozott önérdek/jelenérdek melletti döntés morálisan rossznak ítélhető, illetve a másérdek/jövőérdek támogatása az önérdek/jelenérdekkel szemben morális jónak tartható

A társas értékorientációk nyolcas tagolásának általános bemutatása után a következő fejezetben kicsit részletesebben kitérek az egyes beállítódástípusokra. Nem fogok teljességre törekedni, már csak azért sem, mert a legfontosabb kategóriákkal kapcsolatos elméleteket bővebben elemzem később, a könyv játékelméleti fejezetében. Nehéz szétválasztani az egyes társas motivációkkal kapcsolatos mondanivalót, mert az egyes fogalmak értelmezésében, és különösen az egymáshoz közel álló fogalmak közötti határok kérdésében messze nincs egyetértés a kutatók között. A legnagyobb átfedés talán a kooperáció és altruizmus fogalmainak elemzései között van, de tulajdonképpen bármelyik fogalmat ragadjuk is ki a nyolcas felosztásból, minden esetben találhatunk közös részeket az egyes elméletek között.

Egoizmus

A társas értékorientációk áttekintését az egoizmussal (individualizmussal) kell kezdeni. Egyszerűen azért, mert mind az evolúcióbiológia, mind a társadalommagyarázat számára az egoizmus jelenti a természetes kiindulópontot – még akkor is, ha könnyű kiindulópontnak számít. Thomas Hobbes, Adam Smith és még sokan mások úgy gondolták, hogy az emberiség alapvetően és eredendően önző. önérdekkövető.[5] Hobbes szerint az ember alapállapotát a "mindenki harca mindenki ellen" kifejezéssel írhatjuk le, Adam Smith "láthatatlan kéz által irányított gazdasága" is azért működik, mert az egyéni cselekvők valamennyien saját önérdeküket követik. A gazdasági ember (homo oeconomicus) haszonmaximalizáló és kárminimalizáló lény, akit az önérdeken alapuló célok kitűzése és a célok eléréséhez szükséges eszközök racionális kiválasztása jellemez [237]. Főleg közgazdászok, de sokan mások is erősen hittek, hisznek abban, hogy ez az embermodell alkalmas lehet a társadalom működésének leírására. Úgy, ahogy Adam Smith gondolta.

> „Az ebédünket nem a mészáros, a sörfőző vagy a pék jóakaratától várjuk, hanem attól, hogy ezek a saját érdekeiket tartják szem előtt. Nem emberségükhöz, hanem önszeretetükhöz fordulunk, és sohasem a magunk szükségéről, hanem a rájuk váró előnyökről beszélünk nekik." [498]

Az Adam Smith által sugallt társadalomkép azonban egyoldalú, hiányos, szükség van a korrekciójára. Még akkor is szükséges módosítani ezen az elfogult szemléleten, ha közben elfogadjuk azt a tézist, miszerint a fogalmi építkezésben az egoizmus kategóriáját elsődlegesnek kell tartanunk az altruizmus fogalmához képest abban az értelemben, hogy az utóbbi csak az előbbi létezése esetén képzelhető el. Ezt Jon Elster a következőképpen indokolta.

[5] Ahogy azt korában már többször is jeleztem, az érdek fogalmát nem szabad az önzéssel azonosítani. Az önérdek terminusát azonan már felcserélhetjük az egoizmus kifejezésével.

„Bizonyos értelemben az önérdek ... alapvetőbb, mint az altruizmus. A természeti állapot – jóllehet gondolatkísérlet – logikailag koherens helyzet. Olyan világot viszont nem tudunk logikailag koherensen elképzelni, amelyben mindenkinek kizárólag altruista indítékai vannak. ... Ha senkinek sem lennének elsőfokú, önző örömei, senkinek sem lehetnének magasabb rendű, altruista indítékai sem. ... Ez tisztán logikai kérdés. Ahhoz, hogy egyesek altruisták lehessenek, másoknak legalább időnként önzőnek kell lenniük, de mindenki mindig lehetne önző. ... De ebből – sem általában, se bármely adott esetben – nem vonhatjuk le azt a következtetést, hogy az önzés a legelterjedtebb motiváció." [145, 61]

Nem tudom, mi alapján, hogyan lehetne mérni, vajon melyik lehet a legelterjedtebb emberi motiváció. Én mind a kérdésfeltevést, mind a válaszkísérleteket hanyagolnám. Nem vitatnám viszont az egoista beállítódás létét, sőt, elsődlegességét, inkább csak a kizárólagosságát. A tisztább kép kialakítása érdekében először az egoizmus fogalmát érdemes alaposabban megvizsgálni. Kiindulhatunk abból a meglepő és provokáló kérdésből, amit Andrew Oldenquist tett fel, hogy ti. hogyan lehetséges egyáltalán az önzés mint olyan [391]. Még mielőtt bármit is bemutatnék Oldenquist gondolatmenetéből, tisztázni szeretném, hogy Oldenquist nem kérdőjelezi meg az önzés létezését (és elterjedtségét), csak azt próbálja meg tisztázni, hogy milyen feltételek mentén beszélhetünk önzésről abban az értelemben, ahogy a hétköznapi beszédben használjuk ezt a fogalmat. Mindenesetre erős állítással kezdi tanulmányát.

„ ... a célkövető entitások "természetes állapota" a kantiánus, altruista vezérelv szerinti cselekvés, az egoizmusra jóval nehezebb magyarázatot adni, mint az altruizmusra." [391]

Oldenquist azzal indít, hogy képzeljük el két embert, Robert Kraut$_1$-ot (K_1) és Robert Kraut$_2$-ot (K_2), akik – klózozás eredményeként – minden tulajdonságukban megegyeznek, a környezetüktől is ugyanazokat az ingereket kapták és kapják. Mi történik – teszi fel a kérdés Oldenquist –, ha mindkét ágensnek fájni kezd a foga, de csak egyikük kezelésére van mód. Mi, akik "kívülről" figyeljük a dolgokat, nem tudunk egyik ágens érdekében sem dönteni, hiszen nem látunk semmi különbséget köztük. A kívülállónak Ha közöljük az érintettekkel is, hogy csak egyikük kaphat kezelést, akkor amennyiben K_1 úgy reagál erre, hogy neki kell kapnia a kezelést, akkor azt mondhatjuk, hogy egoista módon cselekedett. Ő különbséget tud tenni a saját és a társa fájdalma között, és saját fájdalmának a megszüntetését előbbre tartja, mint a másik kezelését. Eddig plauzibilisnek tűnik a dolog, de Oldenquist mégis ezen a ponton tartja szükségesnek, hogy tisztáza, mit is jelent pontosan az egoizmus és altruizmus fogalma, és a különbséget a hozzájuk kapcsolható **szelf** fogalom két különböző értelmezésében látja. A kétféle értelmezést így írja le.

„Az *egoizmus* azon alapul, amit a *partikuláris* szelf fogalmának fogok nevezni. Ez rendelkezik azzal a képességgel, hogy különbséget tudjon tenni partikuláris dolgok között, amelyek lehetnek akár teljesen hasonlóak is, azonosítani tudja magát mint egyetlent a többi közül, és képes erre a megkülönböztetésre alapozni saját akaratait, értékeléseit és viselkedési szabályait. ...

... az altruizmus ... vagy ... *imperszonális moralitás* azon alapul, amit valamely *fajta* szelf (egy példánya) fogalmának fogok nevezni. Ekkor, ha valamit értékelni kell, K_1 úgy tekint magára, mint akinek olyan tulajdonságai vannak, amelyekkel – elvileg – K_2 vagy bárki más is rendelkezhetne." [391, 26]

A kissé nehézkes definíció igazi értelme az, hogy elkülönítsük egymástól a szelf két felfogását. A két értelmezés abban tér el egymástól, hogy a szelfet egyediségében, partikuláris, indexikus jellegében fogjuk-e fel, vagy általános, univerzális értelemben kezeljük.

Az egoizmushoz kapcsolható szelf meghatározó minősége az, hogy nem hajlandó eltekinteni saját partikularitásától, míg az altruizmus szelfértelmezése nem tesz különbséget az általános fajta egyes – partikuláris – példányai között. Altruista alapon állva K_1 nem ítélheti úgy, hogy őt kell előnyben részesíteni K_2-vel szemben, mert nincs módja különbséget tenni kettejük között. Ilyenkor K_1 ugyanúgy viselkedik, mint a külső megfigyelők, akik szintén képtelenek dönteni a két teljesen hasonló ágens között. Emiatt az univerzális jelleg miatt nevezi Oldenquist az altruista vagy imperszonális értékelési irányelvet kantiánus politikának. Az egoismus tehát a szelfre mint egyedi példányra, míg az altruizmus a szelfre mint fajtára, tehát egyedi példányok halmazára gondol. Természetesen utóbbi esetben is a partikuláris szelf dönt, de úgy, hogy az értékítélet meghozatala során csak a fajta érdekét, szempontjait veszi figyelembe. Persze annak a fajtának az érdekét, aminek egyébként a partikuláris egyed is tagja.

A 'hogyan lehetséges az önzés' kérdésre Oldenquist válasz tehát az, hogy egoizmus akkor lehetséges, ha az ágens rendelkezik a **partikuláris szelf** tudatával. A partikuláris szelf megkülönböztetésében nem igazán a numerikus, partikuláris jelleg, hanem inkább az indexikus minőség a fontos, vagyis a saját tudat tudata vagyis az öntudat mozzata a meghatározó. Csak az öntudattal rendelekő lényeknek lehet fontos (és persze csak ők tudhatják), hogy melyik konkrét, partikuláris szelfről van szó a fajtába tartozó összes többi közül. Az altruista beállítódást vagy másként: a kantiánus politika követését is úgy határozhatjuk meg, hogy az egy olyan politika, amely kicserélhetővé teszi az indexikus jellegű 'én' terminusát általános jellegű terminusokra, hiszen ha így 'működik", akkor a döntéseiben nem a saját partikuláris szelf szempontjait, hanem a fajta szempontjait veszi figyelembe, és egyben kicserélhetőnek tartja magát bármelyik fajtatársával [391, 27].

Mind az altruizmus, mind az egoizmus fogalma megköveteli a szelf partikuláris fogalmát. Mindegyik esetben arra van szükség, hogy az individuum képes legyen megkülönböztetést tenni a sajátmagára, illetve másokra jellemző, esetleg teljesen megegyező állapotok között. Ilyen megkülönböztetésre nincs mód kizárólag az egyedek külső – más szemlélők által is felismerhető – tulajdonságaira támaszkodva. Ehhez a szelf tudata, vagyis öntudat szükséges. Az egoizmus és altruizmus közti különbséget az jelenti, hogy parikulárisan vagy univerzálisan értelmezik-e fel a partikulárisan létező szelfjüket. Az ember, amikor kanti értelemben univerzálisan cselekszik, akkor nem veszi figyelembe saját szelfjének létezését.

Az oldenquesti kérdésfeltevés azonban csak azután (és annak fényében) válik igazán érdekessé, miután kijelenti, hogy a tudatosság nélküli lények, például a robotok is perfekt kantiánusok. Oldenquist itt Kantra hivatkozik, aki szerint a tökéletesen racionális lények, mint például Isten, kizárlólag univerzális törvények szellemében cselekednének és nem lenne szükségük morálra. De Oldenquist szerint univerzális törvények szerint lehet (kell) cselekedni abban az esetben is, amikor a cselekvőnek hiányzik a szelf mint partikularitás fogalma. Az egoizmus ugyanis csak a partikuláris szelf képzetére tud rátelepedni, ezért annak hiányában nem létezhet (ugyanez mondható el a kizárólagos altruizmusról is, arról az altruizmusról, amelyik kizárólag a másik előnyeit követi, de a sajátját egyáltalán nem). Az egoizmushoz (vagy a kizárólagos altruizmushoz) szükség van arra a szelf-fogalomra (szelftudatra), amely alapján az ágens képes megkülönböztetni magát egy minden tulajdonságában vele teljesen megegyező másik ágenstől.

Ebben az értelemben az önző gén elmélet nem a partikuláris, invididuális gén önzőségén, hanem a géntípus önzőségén alapul. Értelmezhetünk tehát szuperindividuumokat,

csoportokat, olyan részrehajló szándékközösségeket, amelyek tagjai a típus bármelyik tagjának és ezáltal persze a közösség, a típus egészének az előnyére törekekdnek. Ezen a ponton bevezethetünk egy harmadik kategóriát, a **csoportönzés** vagy csoportegoizmus fogalmát, ami azt a furcsa kettőséget fogalja magába, hogy a csoport egésze másokkal szemben önző módon viselkedik, míg a csoport tagjai saját fajtatársaikkal szemben altruisták maradnak. A tényleges társadalmi gyakorlatokra utalva Oldenquist **törzsi moralitásnak** (tribal morality) nevezi el azt a jelenséget, amelyben az emberek valamilyen csoportképző elv szerint összetartozóak minősítve magukat önzőek a kívülállókkal szemben, és saját fajtájuk javára törekednek, de önzetlenek egyénileg a csoporttagok közötti viszonyokon belül. A csoport egészét az önzés jellemzi, a csoportot a **lojalitás** tartja össze, és a csoporttagoktól önfeladást, önfeláldozást kérnek (a családért, a vállalatért, a faluért, a csapatért, a pártért, az eszméért, a fajért, a hazáért). Ilyen esetekben mindig tetten érhető az indexikus jelleg, mert amikor ki kell állni a *mieinkért*, akkor sosem valamilyen tulajdonság miatt kell ezt megtenni (például azért, mert valamiben a legjobbak vagyunk), hanem azért, mert a csoport a **miénk**. Tartalmi értelemben a csoportönzés elméletéhez kapcsolhatjuk a diszkriminatív altruizmus fogalmát is, de ezt az altruizmus témakörén belül ismertetem (↬249.o.).

Altruizmus

Az emberiség különböző – tudományos, vallási, művészeti, erkölcsi – diskurzusaiban mindig is kiemelt helyet kapott az önzetlenség, az altruizmus kérdése. Itt én csak az altruizmus fogalmi elemzését, összetevőinek, tipizálási lehetőségeinek bemutatását szeretném elvégezni, és nem foglalkozom az altruizmus kialakulásának és fenntartásának biológiai, illetve társadalmi mechanizmusaival. Az altruizmussal kapcsolatos vizsgálatok, kísérletek első körének áttekintését nyújtja Dennis L. Krebs cikke [294], majd két évtized évtizeddel később teszi meg ugyanezt E. Midlarsky és E. Kahana [356], illetve Jane Allyn Piliavin és Hong-Wen Charng tanulmánya [413], arról pedig, hogy az altruizmus hogyan jelenik meg a különböző kultúrákban, Herbert Gintis ad alapos áttekintést egy cikkében [189].[6] A közgazdasági imperializmus egyik legszebb példájaként lehet hivatkozni Gary Becker vásott kölyök modelljére (Rotten kid theorem), amelyben Becker a családi kapcsolatokat jellemző altruizmus jelenségét írta le formális eszközökkel [40].[7] A méltányosságelméletekben (↬259.o.), az alkujátékok vizsgálataiban (↬298.o.) voltaképp az altruizmus létét, erősségét keresik, mérik, magyarázzák. Ezekkel az elméletekkel, elemzésekkel máshol foglalkozom, ezért itt nem térek ki rájuk. Nagyon változó, hogy mikor beszélnek kooperációról és mikor altruizmusról az ide tartozó biológiai jelenségek elemzéseiben, én a kooperáció kapcsán fogom röviden ismertetni ezeket az eredményeket, de az altruizmus fogalmához rendelten is ugyanúgy be lehetne mutatni ezeket a mechanizmusokat.[8]

Az altruizmus kapcsán mindenekelőtt arra hívnám fel a figyelmet, hogy ez a fogalom bizonyos értelemben elméleti konstrukció, nehéz elképzelni a tiszta altruizmus gyakor-

[6] A magyar nyelvű szakirodalomban az altruizmus problémakörét kiemelten kezelte Hámori Balázs az érzelemgazdaságtanról szóló könyvében [209], valamint Zsolnai László az etika hangsúlyos szerepének bemutatásával a gazdasági élet egészén belül [596].

[7] Becker modelljéről kiváló áttekintést ad magyar nyelven Janky Béla könyvében[258].

[8] Az önzetlenség kialakulásáról magyar nyelven ad kiváló áttekintést – alapvetően az evolúcióbiológiai eredményekre fókuszálva – Bereczkei Tamás egy könyvében[46], amiről én is írtam egy recenziót [518].

lati megvalósulását. Nem arra gondolok, hogy ideáltipikus fogalmi konstrukció lévén a tényleges gyakorlatban nem találhatjuk meg az elméletben elképzelt jelenséget, hanem arra, hogy én nem nagyon tudok elképzelni olyan esetet, amiben az Ego a cselekvésével úgy tud a másiknak, Alternek előnyt okozni, hogy eközben semmilyen értelemben nem változik a saját (Ego) helyzete, sem előnye, sem hátránya, költsége nem származik a dologból. Ha egyszer mindig adott értéktérben valósítjuk meg a cselekvéseinket, akkor nehezen tudom elképzelni az önmagunkra vetítve teljesen értékmentes cselekvést. Elméletileg persze elképzelhető, hogy nincs költsége valakinek, miközben a másik számára előnyt biztosít a cselekvésével. Ha olyan helyzetet képzelünk el, amelyben az Ego számára két döntési lehetőség van, és mindkettő alternatíva ugyanazt a eredményezi maga számára, viszont az Alter számára Ego egyik döntése előnyös, a másik hátrányos lenne, akkor mondhatjuk azt, hogy az Alter számára előnyös döntés Ego Alterral szembeni altruizmusát mutatná, hiszen saját szempontjait tekintve voltaképp nem lenne döntési helyzetben, a döntése igazából az Alterhoz való viszonyát érintené. Az ilyen elméleti lehetőségre azonban nem nagyon tudok tényleges gyakorlati példával szolgálni.

Az altruizmus másik fontos vonása az, hogy szükségszerűen másodrendű fogalom. Akárhogyan is definiáljuk az altruizmus fogalmát (az értékek, a hasznosságérzetek, a kifizetések fogalmair támaszkodva), a másikra irányultság lényegi mozzanatát csak úgy vehetjük figyelembe, hogy a az Ego értékelésébe (hasznosságérzetébe, kifizetésébe) belekalkuláljuk Alter értékelését (hasznosságérzetét, kifizetését). Ez a beágyazottság mindenképpen másodrendűvé teszi a fogalmat.[9]

Kristen Renwick Monroe azt javasolja, hogy az altruizmust olyan viselkedésként definiáljuk, ami valaki másnak okoz előnyöket úgy, hogy közben a cselekvő számára bizonyos költséggel, lemondással, esetleg valamilyen kockázattal jár [362]. Monroe azt is kiemeli, hogy melyek a definíciójának lényegi elemei, amelyek hiányában nem beszélhetünk igazi altruizmusról. Előszöris az altruizmushoz mindig cselekvés szükséges, tehát nem elégséges valamiféle jószándék, egy nemes gondolat kifejezése. Másodszor az altruista cselekedetnek elsősorban kell a másik javára irányulónak lennie, tehát ha egy egoista cselekedet másodlagos hatásaként előnyök keletkeznek a másik számára is, akkor Monroe szerint nem beszélhetünk altruizmusról. Harmadszor a szándékok fontosabbak, mint a tényleges következmények. Ha valaki jót akar tenni valakinek, de – mindegy, milyen okok miatt – balul sül el a dolog, vagy a kezdetben előnyösnek látszó következmények esetleg hosszútávon már hátrányosakká válnak, attól még nem lehet megkérdőjelezni az eredeti szándék és cselekedet altruista jellegét. Végül, negyedszer az altruista tettnek valamilyen módon csökentenie kell a cselekvő jólétét, hasznosságát, erőforrásait, tehát valamilyen költséggel kell járnia. Ha ilyesmi nincs, vagy még inkább, ha a cselekedet

[9] A preferencialogika úgy határozza meg a preferenciarendezés fogalmát, hogy csak a rendezés tárgyaira fókuszál (a lehetséges világállapotokra), így nem veszi figyelembe a rendezést végrehajtó személyt. Amikor a társas preferenciákat akarjuk megragadni, akkor szükségünk van arra, hogy a cselekvőt vagyis a döntést hozó, a preferenciarendezést végrehajtó személyt is beemeljük a leírásba. Mondhatjuk azt is, hogy perszonifikálni kell a fogalmat. Első lépésben fel kell venni a fogalom összetevői közé a preferenciarendezést végrehajtó személyt, Egot. Ez azonban nem elégséges a társadalmi preferenciák fogalmának leírásához, mert így még csak annyit tudunk kifejezni, hogy ki az értékelést végző személy. Azt is jelezni kell tudnunk, hogy Ego társas preferenciájába hogy épül be Alter társas preferenciája. Ha kettős ontológiát használunk, akkor nem lesz feltétlenül minden modellben másodrendű az altruizmus fogalma. Ha ugyanis csak az Altert érintő kifizetéseket tesszük be Ego preferenciájának paraméterei közé, akkor attól még nem lesz fogalmi beágyazódás. Amennyiben azonban Ego szociális preferenciájába beépül Alter preferenciája (vagy hasznosságérzete), akkor azzal már "felmegyünk" a másodrend szintjére.

nemcsak a másik, de a cselekvő számára is előnyökkel jár, akkor az a kölcsönös előnyök, a kollektív jólét, a kooperáció, a harmónia világába tartozik.

Elias L. Khalil egy nagyon alapos tanulmányában úgy definiálja az altruizmus fogalmát, hogy azt a **jótékonyság**, önkéntes (és névtelen) adományozás (charity) fogalmával kapcsolja össze, miközben kizár egy sor olyan értelmezési lehetőséget, amelyeket elég sokszor és elég sokan az altruizmus fogalma alá szoktak sorolni [278]. Előbb ez utóbbiakat veszem sorba. Khalil szerint létezik három – interakció alapú – értelmezés, ami nem tartható altruizmusnak. Az első az **egoista altruizmus**, ami az altruista cselekvést a várható jövőbeni haszonnal magyarázza. Erre lehet példa a direkt reciprocitás jelensége, amit a játékelmélet is sokat elemzett (↬304.o.). Ez a értelmezés párhuzamba állítható az evolúcióbiológiából származó, triversi reciprock altruizmus fogalmával (↬252.o.). Khalil szerint ez a jelenség azért nem tartató altruizmusnak, mert itt csak arról van szó, hogy a jövőbeni viszonzásért cserébe, tehát a jövőbeni előnyökért hajlandók vagyunk a jelenben áldozatokat hozni a másikért. A második – Khalil által az altruizmusból kizárt – felfogás az **egocentrikus altruizmus**. Akkor beszélhetünk erről, ha az Ego abból eredő hasznosságérzete, hogy élvezi Alter hasznát/örömét valami jószág elfogyasztásakor, nagyobb, mint az a hasznosság, mint amit Ego az adott jószág saját fogyasztása révén érhetne el. Ilyen az anya öröme a gyereke boldogsága miatt. Többet ér az anya számára az, ha a gyerekét látja örülni, ezért nem magának vesz meg egy új cipőt, hanem a gyerekének egy játékot. Ez a helyettesítő élvezet (vicarious pleasure), az Alter örömából Egonak jutó rész – Khalil szerint – ugyanúgy működik, mint a biológiában a rokoni kiválasztás hamiltoni mechanizmusa (↬252.o.). A harmadik téves altruizmusfelfogásként Khalil az **altercentrikus altruizmust** nevezi meg. E felfogás szerint vannak proszociális jegyeink, amelyek arra késztetnek minket, hogy támogassuk a velünk valamiért egy csoportba sorolt társainkat. E jelenség biológiai párhuzamaként a csoportszelekvciót jelölhetjük meg (↬253.o.). Nem kell itt genetikailag örökölt morális gén létezését feltételezni, elég csak a tanult, kulturális mechanizmusokon keresztül belénk nevelt viselkedési mintákra gondolni. Ezzel, Khalil szerint, az a baj, hogy egybemossa az altruizmus éa a becsület(esség) jelenségét (erre még később visszatérek).

A második körben Khalil bemutat három olyan altruizmusértelmezési lehetőséget, amelyeket közösen normatív felfogásnak nevez, és amelyek mindegyikét ki is zárja az elfogadható altruizmusértelmezések köréből. A különböző interpretációk mögötti elméletek közül az első a kantiánus megközelítésen alapuló etikai elmélet, a második a szimbolikus interakcionizmushoz köthető szocializációs vagy kulturalizációs elmélet, a harmadik pedig a pszichológiai megközelítésű melegségvágy elmélet. Az **etikai megközelítés** pédájaként Khalil Amitai Etzioni könyvére hivatkozik, ami a szerző által a közgazdaságtan számára ajánlott új szempont, a morális dimenzió figyelembevételének a szükségességéről szól [148]. Ez az elmélet sokban hasonlít ugyan az altercentrikus megközelítésre, de az megkülönbözteti őket, hogy az altercentrikus hozzáállás mindig valamilyen választott szempont szerint (például a másik jólétét szem előtt tartva) igazodik a másikhoz, az Alterhez, míg a kantiánus elv minden szempontból, tehát szempontoktól függetlenül egyenlően kezeli a másik embert, amikor nem engedi meg nekünk, hogy bárki mást eszközként lehessen kezelni saját cselekvésünk során. Etzioni ezt az elvet úgy modellezi, hogy a szokásos értelemben vett, az élvezettel összekapcsolható hasznosság fogalma mellé felvesz egy másik kategóriát is, amit morális hasznosságnak nevez el. Ez a kétféle hasznosságérzet egyszerre hajtja az embert. A **szocializációs elméletek** legfőbb

képviselőjeként George Herbert Meadre hivatkozik Khalil, aki az altruista szemléletmód kialakulásának okát a nevelésben, szocializációban látja [337]. A **melegségvágy elmélet** kidolgozója, James Andreoni szerint nem a másik jólétének növelése, hanem az adakozás vágya hajtja az embereket az altruista cselekedetek felé. Azért tesznek így, mert a segítségnyújtás után érzett belső melegségre (warm-glow) vágynak [15]. Khalil e háromféle altruizmusmagyarázat egyikét sem fogadja el mondván, hogy ezek nem minden mástól független, belső motiváción, hanem valamilyen másfajta ösztönzésen alapulnak.

Utolsó körben Khalil példákat mutat arra, hogy mit is tart igazából altruizmusnak, és – ahogy azt már említettem – kijelenti, hogy akkor és csak akkor beszélhetünk altruizmusról, ha az jótékonyságból fakad. Végezetül – a félreérthetőség elkerülése érdekében – elkülöníti, pontosabban leválasztja az altruizmus jelenségéről egyfelől a családon belüli szülői altruizmus, másfelől a filantrópia, emberszeretet kérdését.

Khalil is azt mondja, sokan másik is vallják ezt, hogy az altruizmus nem más, mint hosszútávú egoizmus. Az anya azért segít gyermekének, mert arra számít, hogy ezt majd visszakapja, amikor öreg lesz, és támogatásra szorul. Bár úgy vélem, hogy az emberi motivációk nem képesek ilyen hosszú léptékekben érdemben befolyásolni minket, ezért én értelmesnek tartom "igazi" altruizmusnak tartani ez ilyenféle hosszútávú egoizmust is, de van egy ennél még fontosabb érvem arra vonatkozóan, hogy érdemes, sőt, el kell különíteni két fogalmat egymástól. Mégha hosszútávú egoizmusként is értelmezzük az altruizmust, akkor is szükségünk van arra, hogy a rövidtávú és hosszútávú egoizmusokat el tudjuk különíteni egymástól. Az ugyanis tény, hogy ugyanazokba a helyzetekbe kétféle beállítódás mentén lépnek bele az emberek, jelen időben egoista és jelen időben altruista módon, és ha utóbbira azt mondjuk, hogy ez a jelenideű altruizmus valójában csak hosszútávú egoizmusként értelmezhető, akkor is meg kell tudnunk ragadni a jelenben eltérő kétféle beállítódást. Ez pedig azt követeli tőlünk,hogy továbbra is fenntartsunk kétféle fogalmat erre a kétféle jelenségre. Ha az adott kontextusban nem zavaró, akkor továbbra is megtarthatónak vélem az eredeti altruizmus-egoizmus fogalomkettőst, és csak akkor alkalmazom a hosszútávú minősítéssel ellátott egoizmus terminusát, ha ezt valamely kontextus megkívánja.

Amikor Oldenquist a csoportönzés fogalmáról írt (↪245.o.), akkor azt állította, hogy a csoport tagja egymással szemben altruista módon viselkednek, viszont a csoport határain kívüli kapcsolataikban egoistákká válnak. Egészen más irányból tekintett ugyanarra a jelenségre, mégis hasonló eredményre jutott Garret Hardin, aki bevezette a **diszkriminatív altruizmus** fogalmát [220]. Jardin alaptézise az volt, hogy nem lehet univerzális értelemben altruizmusról beszélni, az altruizmus megnyilvánulásához mindig szükség van arra, hogy diszkrimináljunk, vagyis valamilyen hatókörön, valamilyen csoporton belül érvényesítsük ezt a politikát, amihez természetesen hozzátartozik az is, hogy a hatókörön, a csoporton kívül felhagyunk az altruizmusunkkal. A diszkriminatív altruizmus fogalma abban az értelemben is termékeny kategória, hogy a fogalomra támaszkodva könnyen képezhetünk egy nagyon plauzubilis tipológiát. Bár Hardin olykor talán suta, kicsit nehézkes terminológiai megoldásokat választott, de tartalmilag nagyon kézenfekvő, a valós társadalmi gakorlatokat jól jellemző tipizálást javasolt.

- individuum
- család
- bajtársiság, kronizmus
- tribalizmus, törzsiség

- hazafiság, patriotizmus

Az egyes típusokra az jellemző, hogy a lista sorrendjének megfelelően a típusokhoz egyre növekvő számosság, illetve egyre csökkenő mértékű altruizmus kapcsolódik. A legerősebb altruizmust a legindividuálisabb kapcsoaltainkban a szerelemben, az igazi barátságban valósítjuk meg, de itt a "kedvezményezettek" száma a legkisebb. A családon belül megvalósuló altruizmus is erős ugyan, de már kevésbé, mint a lángoló szerelemben, viszont több (2-5-10) emberre is kiterjeszthető. A kronizmuson alapuló kapcsolatok jelentik a következő lépcsőfokot, amikor a már megint csökkent az önfeláldozási hajlandóságunk, de itt már több tízes, esetenként százas nagyságrendű emberrel szemben állunk altruista beállítódással. A kronizmus példái lehetnek az iskolai, egyetemi években, a katonai szolgálatban, a munkahelyeken közösen eltöltött idő által kovácsolt közösségek. A törzsiség kategóriája valamilyen közös vonás, általában valamilyen élmény- vagy értékközösség alapján összetartozó emberek közti altruizmust írja le. Ifjúsági zenakarok, zenészek, művészek, sportegyüttesek, sportolók stb. iránti rajongás, elköteleződés vagy épp olykor az etnicizmus kötheti így össze az embereket. Végül a leggyengébb erősségű, de azért még létező, viszont mindenképpen a legnagyobb számosságú emberre kiterjedő önzetlenséget figyelhetünk meg a patriotizmus katagóriája alá sorolható jelenségek esetében.

Az altuizmust persze megpróbálhatjuk a tartalma alapján is tipizálni. Egy ilyn felosztás lehet az, amikor elkülönítjük a megosztás (sharing), a segítségnyújtás (helping), az adás (giving), a kooperáció (cooperation) cselekvéstípusait egymástól. De nyilván ki lehete még taláni további osztályozásokat is.

Kooperáció és versengés

Az egoizmus és az altruizmus fogalmainak tisztázására irányuló diskurzusban is felbukkan a kooperációs és a versengés kategóriája, de az igazi "karriert saját jogon" érte el ez utóbbi két fogalom, amihez a fókuszt és a keretet a játékelmélet biztosította (269.o.). A dolgokat kissé leegyszerűsítve mondhatjuk, hogy az elindulása után elég hamar a játékelmélet egyik fő kérdésévé vált, hogy mikor, milyen társadalmi helyzetekben (játékokban) várhatunk kooperatív vagy versengő magatartást a résztvevőktől (játékosoktól), illetve miként lehet biztosítani a kooperációt azokban a játékokban, amelyekre alapértelmezésben a versengés jellemző. A játékelméletben azonban nem vagy csak kevesebbet foglalkoztak a kooperáció és versengés, valamit az altruizmus és egoizmus egymáshoz való viszonyával. Mivel a játékelméleti fejezetben sok szó esik majd a kooperáció és versengés problémaköréről, az ezeket biztosító mechanizmusokról, itt inkább a fogalmak jelentésének tisztázására törekszem.

A kooperáció értékelésekor kétféle téves nézet tapasztalható. Mivel az altruizmus és az egoizmus "között" helyezkedik el félúton, ezért mindkét irányból a kooperációra tekintve könnyen az ellentétes pólushoz tartozónak lehet minősíteni azt. Az altruista szemszögéből nézve a kooperáló fél valamilyen mértékben önérdeket is követ, ezért erre az mozzanatra hivatkozva egoistának lehet tartani őt. Ez persze fordítva is igaz, vagyis amilyen mértékben valaki elmozdul a tiszta egoizmus felől az altruizmus irányába, olyan mértékben már altuistának is lehet minősíteni őt (elfeledkezve a "még megmaradt" egoista hányadról).[10] Tulajdonképpen nem ad semmi újat az a felosztás, ami a nemzetközi

[10] Ez mást jelent, mint amikor az altruizmust egoizmusnak értelmezi valaki mondván, hogy az altruista is azt csinálja, ami neki jó, tehát egoista. Erről korábban kicsit bővebben írtam (249.o.).

kapcsolatok elemzésének világából származik [277], mégis segíthet minket abban, hogy a kooperáció igazi értelmét megtalálhassuk. A nemzetközi kapcsolatokban három kapcsolattípust különítenek el egymástól: a **viszályt** (dischord), a **harmóniát** (harmony) és a **kooperációt** (cooperation). A három fogalom értelmezése eléggé adja magát: harmónia esetén a felek érdekei konfliktusmentesen illeszkednek egymáshoz, míg a viszály esetén az érdekek ellentétesek, vagyis teljes az érdekütközés, végül a kooperáció azt jelenti, hogy a feleknek vannak egymással összeegyeztethető érdekeik, és ezek fontosabbak a többi – konfliktusos viszonyban levő – érdekhez képest. Ezt a hármas tagolás egy dimenzióban kiegészíthetjük Anatol Rapoport javaslatával [430], aki szerint a konfliktus (vagy az előző felosztás szerint a viszály) lehet **harc** (háború, verekedés), amikor erőszakos eszközöket alkalmazva próbáljuk meg a másikat legyőzni, lehet **játszma** (játék, önérdekkövetés), amikor erőszak nélkül, a másikkal versengve próbáljuk meg legyőzni, megelőzni őt, illetve lehet **vita**, amikor meggyőződéseket ütköztetve próbáljuk meg a másikat meggyőzni (ekkor már csak kommunikatív cselekvésről lehet szó). Ezt a kiegészítést egyelőre csak az elméleti teljeség igénye miatt tettem meg, gyakorlati segtséget nem jelent a kooperáció fogalmának tisztázása érdekében. A kooperáció jelenségét jobban megérthetjük, ha azt a harmónia vagy a kölcsönyös előnyök (mutualizmus), illetve a viszály (konfliktus) jelenségeihez képest próbáljuk meg elhelyezni a fogalmi térben. Ha a kooperációt úgy értelmezzük, hogy abban mind az Egonak, mind az Alternek előnyei származnak (hiszen a kooperáció az egoizmus-altruizmus tengelyek pozitív értéktartományában helyzkedik el), akkor a kooperációt a harmóniához közelebb álló fogalomnak kell tartanunk. Ha viszont arra gondolunk, hogy bár a kooperációban mindkét ágens számára előnyös, de Egonak egy ilyen helyzet kevesebb nyereséggel, s ebben az értelemben veszteséggel jár a "tiszta" egoizmust megvalósító magatartáshoz képest, akkor meg azt mondhatjuk, hogy a kooperáció nem is annyira a harmóniához, hanem inkább a viszályhoz áll közelebb. A kooperációban lényege tehát az, hogy egyszerre találhatjuk meg benne a harmóniát (kölcsönös érdekegyezést) és a – potenciális – viszályt (érdekkonfliktust).

A fogalmi tisztaságot jelentős mértékben nehezíti az, hogy sokan és sokszor az egoizmus és az altruizmus úgy kezelik, mint egymással teljesen szembenálló fogalmakat, holott a társas preferenciák nyolcas tagolású felosztása egyértelműen mutatja, hogy ez az értelmezés leszűkítő jellegű, és az altruizmus és egoizmus tengelyen ábrázolni kell a pozitív és negatív polaritást egyaránt, hogy teljesebb képet kaphassunk. Ekkor viszont nagyon összezavarhatja a gondolkodást, ha nem helyezzük el pontosan ezen a két tengelyen, hogy hol tudjuk értelmezni, helyén kezelni az olyan fogalmakat, mint előny, haszon, elmaradt kár, hátrány, kár, elmaradt haszon, indokolt és indikolatlan költség, ártalom, veszteség. A játékelméleti fejezeten belül az indirekt reciprocitásról szóló részben (↬315.o.) például a kooperáció fogalmánank még a megszokott értelmezést tulajdonítjuk (kooperáció = segítés a másiknak (Egonak költség/hátrány, Alternek előny), ám a versengés, dezertálás tartalmát némiképp átértelmezzük, amikor az "nem segítésként" fogjuk fel (Egonak nincs költség/hátrány, Alternek nincs előny).

A kooperáció biológiai mechanizmusai

A társadalmi kapcsolatokon belül a kooperációt kialakító és fenntartó mechanizmusokkal sokat foglalkozom majd a könyv játékelméleti fejezetében (↬269.o.). A játékelméleti apparátust a biológiai tudományok képviselői is hamar használni kezdték, részben az ál-

latvilágban megfigyelhető kooperációs jelenségek modellezésére. A biológiai modellek eredményeinek bemutatását beilleszthetném a játékelméletet tárgyaló fejezetbe is, de ezt inkább itt teszem meg. Röviden bemutatom a biológiai elemzésekben leírt legfontosabb kooperációs modelleket. Alaposabb elemzésbe nem tudok, nem is akarok belekezdeni, csak egy rövid kitekintésre vállalkozom (a rokoni altruizmus, a reciprok altruizmus, a zöld szakáll hatás, az rosszindulat/spite, a csoportszelekció és a költséges jelzés vagy hátrány-elv fogalmakkal jelölhető eredményeket próbálom meg vázlatosan felmutatni.)

A biológusok közül William Donald Hamilton tette meg az első komoly lépést azzal, hogy a rokonok közt létező altruizmus jelenségét megmagyarázta a **rokoni altruizmus** fogalmára alapozott elméletével Hamilton [207], [208]. Ennek lényege röviden annyi, hogy azok a gének lehetnek sikeresebbek, amelyek olyan viselkedési mintákat örökítenek tovább a szülő gazdaállatokban, amelyek az utódok támogatását biztosítják. Ennek sikerességének az az egyszerű evolúciós magyarázata, hogy a szülőállatban levő gén egyre csökkenő valószínűséggel ugyan, de valamilyen mértékben továbbélhet a leszármazottakban is, így az ő fennmaradásuk is a gén replikációját szolgálja.[11]

Az evolúcióbiológián belül a gének által meghatározott kooperatív, altruista jelenségek magyarázataban legtöbbször a Hamilton szabályt alkalmazzák [207], [208]. Ez a szabály két ágens várható költsége és nyereménye közti kapcsolatot írja a köztük levő rokonsági fok erősége szerint. A fokális ágens az, aki cselekvő szerepben van (akit neveznek donornak is), a recipiens pedig a potenciálisan kedvezményezett ágens, aki arra számíthat a kapcsolatban, hogy altruista, kooperatív beállítódással fognak hozzá viszonyulni. A kérdés mindig az lesz, hogy a cselekvő ágens cselekvése, ami mindig költséggel jár számára, milyen eredményeket képes kiváltani a recipiens számára. A donor költségét c-vel, a recipiens előnyét b-vel, a köztük levő rokonsági fokot pedig r-rel jelölve a Hamilton szabály a következő formulával írható le:

$$rb > c$$

A képletet értelmezve látható, hogy ha adottnak vesszük a b előny és c költség nagyságát, akkor egy bizonyos rokonsági fok felett az egyenlőtlenség teljesülhet. vagyis ugyanazt a dolgot csak adott közelségben levő rokonoknak tesz meg egy ágens. De a képlet úgy is értelmezhető, hogy ha adottnak veszünk egy rokonsági fokot és egy előnyt a recipiens oldalán, akkor a cselekvő ágens csak egy adott c költségig hajlandó elmenni a recipiens megsegítése érdekében.

Az altruizmus, a kooperáció biológiai modellezésében a következő lépés a **reciprok altruizmus** fogalmával leírható jelenségek elemzése következett az 1970-es évektől kezdve. A kiindulópontot Robert L. Trivers 1971-es tanulmánya jelentette [546]. A reciprok altruizmus a kooperatív viselkedés egyik fajtájaként értelmezhető és a kölcsönös előnyök cseréjén alapul. Nemcsak azonos fajba tartozó állatok között jöhet létre, és a tartalmát tekintve nagyon sokféle tevékenységre kiterjedhet. A reciprok altruizmus fogalmával megragadható jelenségekre idővel egyre inkább használni kezdték a direkt reciprocitás fogalmát is. Trivers tanulmányában felsorolt néhány fontosabb tevékenységterületet, ahol megjelenhet a reciprok altruizmus (direkt reciprocitás).

„1. Segítségnyújtás veszély esetén (pl. baleset, ragadozók, fajon belüli agresszió)
2. Élelem megosztása

[11] Hamilton elméletét kicsivel bővebben bemutattam máshol [517].

3. A beteg, a sebesült, a nagyon fiatal és az öreg megsegítése
4. A szerszámok megosztása
5. A tudás átadása" [546]

A triversi felsorolás koránt sem teljes. A későbbi megfigyelések során kiderült, hogy még elég sokféle tevékenységtípusra kiterjeszthető az elmélet hatálya,[12] mint ahogy az is, hogy lehet többszemélyes altruizmusról (multiparty altruism) is beszélni, amikor azt tapasztaljuk, hogy az altruista, kooperatív ágens a partnere rokonával szemben is "viszonzási politikát" folytat [546]. A reciprok altruizmus elméletével kapcsolatban fontos kiemelnem, hogy vannak, akik úgy vélik, hogy szigorú értelemben véve ebben a reciprocitástípusban nincsen altruizmus, ez csak a jövőbe eltolt viszonzást jelentigazából , tehát egoizmuson alapul. Ez valóban így van, de korábban már jeleztem (↪249.o.), hogy ha ezt a fajta altruizmust hosszútávú egoizmusnak tartjuk, akkor is szembe kell tudni állítani ezt a rövidtávú egoizmussal, és így kétféle fogalmat kell továbba is használatban tartanunk. Vannak, akik másfajta kritikával élnek a direkt reciprocitás elméletével szemben. Tim Clutton-Brock szerint a kooperáció, a reciprocitás sokszor nem az időben eltolt kölcsönösséget, viszonzást, hanem inkább a kölcsönös előnyök, érdekeltségek azonnali kielégítését jelenti, és tételes listát ad az állatok közti direkt reciprocitás jelenségre megmutatva, hogy sokszor azonnali kölcsönös előnyök kihasználása jellemzi a kapcsolatokat [100].

A 2000-es évek második felében elméleti alapozású kritikai támadások érték a rokoni vagy a recoprok altruizmus elméleteit. Voltak, akik az evolúció hordozói közé a csoportokat is beemelve a szocioóbiológia egész rendszerét újragondolandónak vélték. Edward O. Wilson és David Sloan Wilson a szociobiológia elméleti alapjainak újraformálására tettek kísérletet, miközben gondolatmenetük egyik kulcselemévé tették a következő mondatot: „Az önzés legyőzi az altruizmust egy csoporton belül, de a az altruista csoportok legyőzik az önző csoportokat." [584]. Az evolúcióbiológia egyik nagy alakja, Martin A. Nowak kutatótársaival 2010-ben felvetett, hogy a rokoni altruizmus (kin selection), az inkluzív fittnesz hamiltoni magyarázósémája helyett lehetséges eredeti evolúciós magyarázatot adni, vagyis fölösleges az inkluzív fitnesz elmélet. A javasolt új elmélet középpontjában a "eusociality" fogalma áll [383].

Szélesebb kontextusban nem lenne megengedhető, de a biológiai keretek között a **rosszindulat** vagy a károkozás (spite) fogalmát összevonhatjuk az **agresszió** fogalmával. Ez utóbbi kategóriát külön tárgyalom (↪255.o.), itt csak azt mutatom meg, hogy milyen elméletek születtek a rosszindulat és a kooperáció közti kapcsolatra vonatkozóan. Első pillanatra talán meglepő állítás, de a kooperáció kialakulását bizonyos feltételek mellett meg lehet magyarázni a rosszindulatú, agresszív hozzáállás feltételezésével is [262], [324]. Egy háromszereplős helyzetet feltételezve modellezni lehet a kooperációt a rosszindulatú magatartás jelenlétével. Ekkor a modellbe fel kell venni harmadik szereplőként a másodlagos recipienst, akinek előnyös lehet, ha a donor árt a recipiensnek. Azért lehet előnyös, mert a másodlagos recipiens helyzete javulhat azzal, hogy kisebb lesz a versenyképessége a recipiensnek. Ezt a jövőbeni versenyképesség-növekedést egyébként ki lehet terjeszteni akár a donorra is. [579]

A kooperáció olyan egyedek között is kialakulhat és fennmaradhat, akik nincsenek szoros rokoni kapcsolatban egymással. A **csoportszelekció**, illetve a hozzá kapcsolható **zöldszakáll hatás** (greenbeard) fogalmával írják le ezt a jelenséget [125, 115], [579]. Ha

[12] A reciprok altruizmussal kapcsolatos példákról lásd: [517].

nem rokoni kapcsolatban álló egyedek között megjelenik egy beazonosításra (és ezzel persze elkülönítésre) alkalmas biológiai jegy, akkor a jegyet hordozók között megjelenhet és tartósan megmaradhat egymás segítésének szándéka. A 'zöldszakáll' megnevezés Richard Dawkins egyik könyvéből ered, és arra a feltűnő jegyre utal, ami ha megjelenik egy populáción belüli kisebb csoport tagjai között, akkor a körön belüliek (tehát a zöld szakállal rendelkezők) el tudják különíteni magukat a többiektől, és egymást támogatva ki tudják használni a kooperáció adta előnyöket. E mechanizmus sikerességéhez tehát csak annyi szükséges, hogy olyan gén(ek) jelenjen(ek) és terjedjen(ek) el az egyedek valamely csoportjában, amelyek egyszerre ösztönöz(nek) a kooperációra egyfelől, illetve produkál(nak) látványos csoportazonosító jegyet az egyedekben. Az azonosító jelek (zöld szakáll) és a segítségre való hajlam kialakulása persze elszakadhat egymástól, vagyis megjelenhetnek azok a csaló egyedek, akik csak a csoportba tartozás külső jegyét hordozzák magukon, és ezzel élvezik a többiek altruista hozzáállását, de maguk nem kooperálnak. Kiderült azonban, hogy ilyen esetekben is képesek tartósan megmaradni az altruista zöldszakállak, sőt a kutatók találtak olyan csoportszelekciós mechanizmusokat is, amelyek a károkozással összefonódva fejlődtek ki [579].

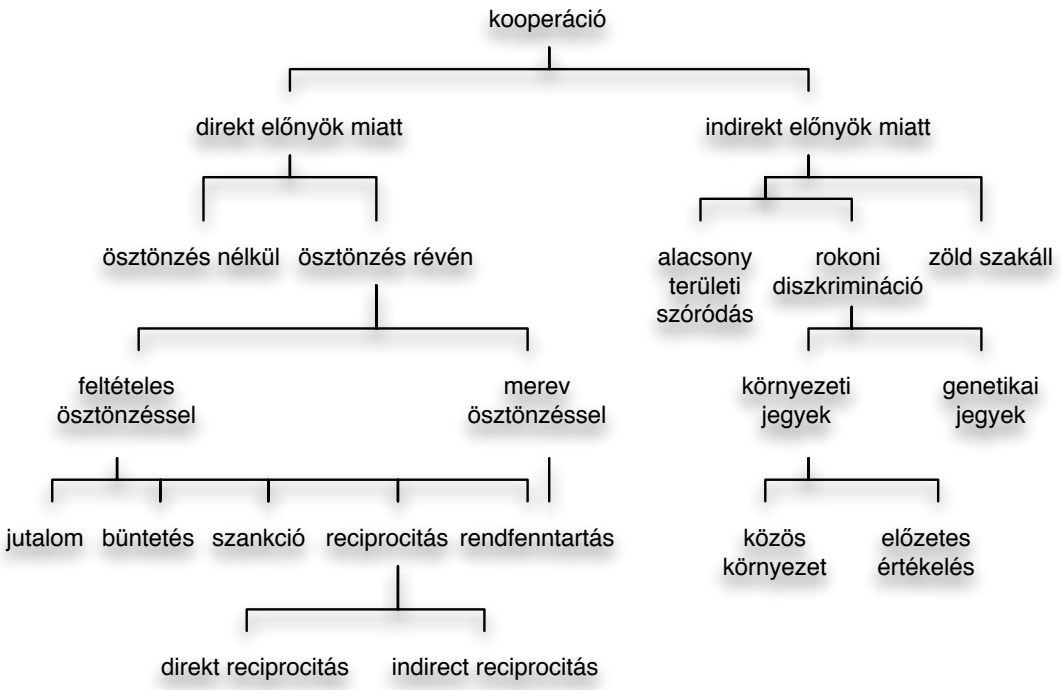
A **költséges jelzés** (costly signaling) vagy **hátrány-elv** (handicap principle) elmélet alapjait Amotz Zahavi dolgozta ki [591].[13] A handikep-elméletét részben a rokoni altruizmus és a reciprocitáselméletek hiányosságainak pótlására dolgozta ki, amit egyik tanulmánya címével is kifejezett: "Az altruizmus mint hátrány – a rokoni kiválasztás és a reciprocitás [elméletek] korlátai" [592].[14] A költséges jelzés vagy hátrány-elv elmélet lényegét akkor érthetjük meg igazán, ha arra a helyzetre gondolunk, amire vonatkozóan Zahavi felépítette elméletét (még akkor is, ha azóta a költséges jelzés elméletet más területekre vonatkozóan is alkalmazhatónak ítélték, ítélik). A párválasztás előtt álló egyedek számára érdemes lehet minél feltűnőbb külső jegyekkel felhívni magukra a potenciális partnereik figyelmét. Ezek a **jelzések**, amelyek minél feltűnőbbek (nagyobbak, színesebbek, hosszabbak stb.), annál inkább jelezhetik az egyed nagyságát, erősségét, rátermettségét, vagyis annál inkább felkelthetik a többiek figyelmét, és ezzel növelhetik a párosodás(ra való kiválsztódás) esélyét. De ha ez így van, akkor érdeke lehet minden egyednek csalnia e téren, és ekkor a jelzést adó minőség (a farok, az agyar, a sörény, a taréj, a tollak nagysága, színe stb.) nem feltétlen lesz összhangban az egyed valódi minőségével. Úgy akadályozhatja meg ezt az evolúció, hogy a jelzést költségessé tesz. Csak az a jelzés lesz hiteles, megbízható, amelynek "kifejlesztése", fenntartása sokba kerül az egyednek, ami sok esetben abban is megnyilvánul(hat), hogy hátrányt okoz hordozójának. A nagy szarvasagancs nehezebb, könnyebben beleakadhat a növényzet, nehezebb fák közt közlekedni vele. De az az állat, amelyik képes fenntartani a költséges jelzésre szolgáló vonásait, testrészeit, amelyik viselni tudja hátrányok következményeit, az méltó a kiválasztásra. És ebben van egy – a szaporodés érdekét szolgáló – jelenbeli önfeladás, altruizmus, amit abban az értelemben nevezhetünk kooperációnak, hogy nem csal saját biológiai jegyeinek "kommunikálásakor".

A kooperáció (altruizmus) biológiai mechanizmusainak rövid bemutatását egy áttekintő ábrával zárom. Stuart A. West, Claire El Mouden és Andy Gardner a kooperáció jelenségének gyakori félreértelmezési hibáit számba véve, és ezáltal a kooperáció bi-

[13] A kérdéskör magyar áttekintését lásd: [46].

[14] Nem véletlenül rímel Zahavi cikkcímére Jonathan Wright tanulmányának címe sem: "Az altruizmus mint jelzés: Zahavi alternatívája a rokoni kiválasztás és reciprocitás [elméletekkel] szemben[587].

ológai mechanizmusait is sorban bemutatva az alábbi ábrában összegezték a kooperáció kialakulásának lehetőségeit [580].



Az ábrán szerepelnek olyan jelenségek is, amelyekről itt nem beszéltem, így vnnak megmagyarázatlan fogalmak, de nem gondolom, hogy nekem kéne bemutatnom az összes jelenséget. Inkább csak egy jelzést akartam adni arra, hogy az olyan, az emberi társadalmakban fontos jelenségek, mint altruizmus, kooperáció, hogyan modellezhetők az evolúcióbiológia eszközeivel.

Agresszió

A tudományos vizsgálatok sokáig ugyanúgy, sőt, talán még inkább elhanyagolták a rosszindulatú, károkozásra törő cselekvések jelenségét, mint tették ezt az altruista cselekvések esetében.[15] Pedig más (például vallási, erkölcsi, művészeti) diskurzusokban mindig is jelen volt a rosszindulat, az agresszió, a gonoszság kérdésköre (akárcsak az altruizmusé). A tudományos magyarázatok hosszabb időn keresztül az egoizmus kizárólagosságára támaszkodtak, és idő kellett ahhoz, hogy a kutatók előbb beemeljék az altruista motívumokat a társadalmi cselekvést leíró modelljeikbe. Még többet kellett várni a rosszindulat jelenségének tudományos vizsgálatára.

W.D. Hamilton tágította ki először az elemzési fókuszt, amikor vizsgálandónak tartotta a **rosszindulat** vagy **károkozás** (spite) jelenségét az állatvilágban [206].[16] Hamilton

[15] Brian Skyrms 2007-ben így írt erről: „Both spite and altruism on their face appear to violate the rational-choice paradigm. Both have an evolutionary explanation in terms of correlated interactions. Yet spite rarely receives the attention given to altruism. A search on Google scholar for »evolution of altruism« gets 1,570 hits, while one for »evolution of spite« gets 32.” [495].

[16] Nehéz az angol ‘spite’ terminust magyarra fordítani, mert a ‘rosszindulat’ valamiféle beállítódásra

rövid cikke elején még azt a kérdést tette fel, vajon lehetséges-e, hogy egy állat kárt okozzon magának úgy, hogy eközben egy másiknak még nagyobb kára keletkezzék, és ha ez előfordul, akkor hogyan lehetne modellezni. Hamilton a kérdésbe foglalt állítást nevezte el rosszindulatnak. Később azonban, írása egy másik pontján már kicsit másként határozta meg ugyanezt a fogalmat, amikor azt állította, hogy "rosszindulatú az a viselkedés, amely kárt okoz másokat, de nem jár előnnyel az egyén számára". A felállított modell szempontjából végülis az utóbbi értelmezés volt a fontos, a modell eredménye pedig az lett, hogy az átlagos rokonsági szint alatti egyedekkel szemben evolúciós szempontból olykor előnyös lehet a rosszindulatú magatartás. Hogy miként lehet pontosan meghatározni az átlagos rokonsági szint alatti kapcsolatot, az most számunkra nem fontos (a lényeg, hogy a biológusok meg tudják határozni), nekünk csak azért fontos ez a minősítés, mert ezzel lehet "elérni" azt, hogy a Hamilton szabály (↬252.o.) képletében a rokonsági fokot jelző r tényező értéke lehessen negatív. Ebben az esetben már értelmezni lehet formálisan is a károkozás jelenségét is, mert így már elképzelhető olyan helyzet, amikor az egyenlőtlenség teljesül, ha a károkozás eredményeként előálló b előny negatívba vált [183].

Rosszindulattal kapcsolatos jelenségeket figyeltek már meg az állatvilágban, bár nem túl sokat, és mindig csak fajok közti kapcsolatokban. Hamilton maga is azt állítja saját modelljéről, hogy az alapján a rosszindulatú magatartás megjelenését csak ritkán lehet elvárni, mert egyrészt minden rosszindulatú tevékenység költségekkel jár, másrészt nehezen lehet felismerni az átlagos rokonsági szinthez viszonyított "távolságot", harmadrészt a modell olyan, hogy relatíve kis méretű populációk esetében "teszi lehetővé" a rosszindulat megjelenését és elterjedését, ami viszont könnyen a populáció gyors kihalásához is vezethet.[17] Az állatok közti konfliktusokat korlátozott háborúként értelmezve John Maynard Smith és G.R. Price megpróbálta tipizálni azokat, és elkülönített egymástól több stratégiát (az egér, a héja, a bika/szájhős, a megtorló és a szondázva megtorló stratégiákat), de ezeket itt nem mutatom be [327]. Edward O. Wilson szociálbiológiáról szóló könyvében – példákat is mutatva rájuk – hosszabban ír a károkozás és az agresszió jelenségéről [585], és vannak kutatók, akik a károkozás segítségével próbálják meg modellezni az indirekt reciprocitás kialakulását [265].

Az, hogy az állatvilág vizsgálata során kevesebb figyelmet szenteltek a károkozás, az agresszió jelenségének, sokkal inkább érthetőnek tűnik számomra, mint az a tény, hogy milyen kevés elemzés irányult az emberi agresszió jelenségére (ez alól csak a pszichológia szakterülete jelent kivételt). Természetesen vannak nagyon jó elemzések az agresszivitás jelenségéről [427], [332]. Az agresszióra fogalmára szerkezetét tekintve nagyjából az altruizmus definíciójához hasonló meghatározást adnak a kutatók.[18]

> „Agresszióként definiálhatunk minden cselekvést, amely egy másik emberrel szemben ártó szándékkal valósul meg, és az elkövető abban a tudatban van, hogy egyfelől kárt fog okozni a másiknak, másfelől a célszemély igyekszik majd kikerülni ezt a viselkedést." [88]

A definíció fontos eleme, hogy nem elég csak a károkozás tényét megállapítani, a szándékot, tudatosságot is hozzá kell rendelnünk. Ennek hiányában például az orvosi tevékeny-

utal, és nem tudja kifejezni a fogalomban benne levő cselekvési tartalmat, a károkozás viszont nagyon erős, túlzottan direkt jellegű.

[17] A. Gardner és S.A. West cikkében van néhány konkrét példa a Spite jelenségével kapcsolatban [183].

[18] Ranschburg Jenő félelemről, agresszióról szóló könyvében is hasonlóképpen definiálta a fogalmat (bár ő figyelembe vette a tárgyra irányuló agresszivitást is): „Agressziónak nevezünk minden olyan szándékos cselekvést, amelynek indítéka, hogy – nyílt vagy szimbolikus formában – valakinek vagy valaminek kárt, sérelmet vagy fájdalmat okozzon. [428, 90].

ség egy részét is agressziónak kéne tartanunk. A fogorvos ugyanis fájdalmat okoz a páciensének a gyökérkezelés során, de mivel – tegyük fel, hisz ez a tipikus – nem ez a szándéka, ezért nem minősíthetjük agressziónak. Ebben az értelemben a szadomazo kapcsolatban sincs agresszió, mert az "áldozat" szerepében játszó fél nem akarja elkerülni a fájdalmat. Az agresszióval sokszor együtt jár(hat) a harag érzelme, és ilyenkor indulati-érzelmi cselekvésként kell minősítenünk, de nincs szükségszerű kapcsolat a két jelenség között. Elképzelhető, hogy valaki puszta számításból, hideg fejjel okoz kárt, fájdalmat, sérelmet valakinek. Az agresszivitás jelentkezhet kín- vagy fájdalomokozás formájában, nagyobb károk esetén már szadizmusról beszélhetünk, ami megnyilvánulhat mások megalázásában, fenyegetésében, tényleges fizikai vagy verbális bántalmazásában.

Hogy miért cselekszik agresszívan valaki, az nem tartozik e fejezet tárgyához. A biológusok válasza "üss vagy fuss" ("fight or flight") szellemes ugyan, talán még szemléletes is, de az ember motivációs rendszerét azért tarthatjuk ennél bonyolultabbaknak. Alfred Adler volt az, aki az ember legfőbb mozgatórugójaként a hatalmi törekvéseket tette meg, ami a mások fölé kerekedés és a mások fölötti uralkodás vágyától egészen a mások leigázásának, totális maga alá rendelésének vágyáig húzódhat [3]. Az adleri emberképben nyilván helye van az agresszivitás, károkozás fogalmának, de sezt sem kell itt kibontanunk.

A játékelméleti, közgazdasági gondolkodás világába is bekerült a rosszindulatú viselkedés problémája, igaz nem önmagában, hanem inkább a jóindulatú cselekvésekkel való viszonyában. David K. Levine empirikus kísérletekben vizsgálta az altruizmus és rosszindulat jelenlétét [303], Jack Hirshleifer pedig fogalmi szinten elemezte a gonoszság, a másik kárára irányuló motiváció, a jóindulat, illetve rosszindulat (benevolence vs. malevolence) kérdését, és megállapította, hogy a közgazdaságtannak sokkal többet kellene foglalkoznia ez e fogalmak terjedelme alá tartozó jelenségekkel [237].

Autoagresszió

Az autóagresszió, az ágens sajátmaga ellen irányuló rosszindulata a tiszta egoizmus poláris ellentéte. Ekkor (akárcsak az egoizmus esetében) a tudatban csak az Ego van jelen, Alter nem jelenik meg, tehát ilyenkor minden rossz, ártó szándék az Egora irányul. Vannak, akik éppen ezért nevezik ezt mazochizmusnak. Kevés igazi példát lehet erre a jelenségre hozni. Ez nyilván nem véletlen. Az élet, az evolúció, az ember lényegével ellentétesnek tűnik, hogy egy ágens úgy okozzon kárt önmagának, hogy eközben senki, semmi más javát nem szolgálja. Amilyen természetesnek tűnik az "öncélú" – egyéni vagy csoportos – egoizmus, annyira temészetellenes az autoagresszió lehetősége. Az önmagára irányuló agresszió, károkozás legnyilvánvalóbb példái az öngyilkosság, öncsonkolás, önbüntetés különböző formái lehetnek, de ezek túlnyomórésze inkább az önfeláldozás fogalma alá sorolható be inkább. Az önbüntetés azon példái tartozhatnak ide, amelyekben nem találunk semmiféle olyan motívumot, amelyben a más – egyéni vagy csoportos – ágens is feltűnne. Nem nagyon hiszem, hogy az állatvilágból lehetne erre példát találni, de az ember esetében ezt el tudom képzelni.

Önfeláldozás

Az önfeláldozásról akkor beszélünk, ha a cselekvő magának kárt, költséget, a másiknak előnyöket okoz. A tiszta altruizmus esetén Ego, a cselekvő ágens nem foglalkozik saját-

magával, csak a másik ágensre, Alterre figyel. Ha az altruizmust a "másiknak kedvezés" jelenségeként értelmezzük, akkor az önfeláldozás jelenti a legerősebb altruizmust.

A játékelmélet világából több olyan helyzetet azonosíthatunk, amely az önfeláldozás példájaként minősíthető, bár sajnos abban a világban terminológiai szempontból elég nagy fogalmi pongyolaságot találhatunk. Az aszimmetrikus játékok közül az 'anya-fia bankrabló' nevű játék mindenképpen az önfeláldotás jelenségébe sorolható [517]. Nem érzem feladatomnak, hogy kitérjek itt az emberi öngyilkosság kérdésére, de annyit nyilván érdemes megemlítenem, hogy Durkheim az öngyilkosság típusai között megemlíti az *altuista öngyilkosság* fogalmát, amit úgy jellemez, hogy erős társadalmi integráció létezésekor a közösség elvárása olyan erős az egyénnel szemben, hogy utóbbi maga megteszi az önfeláldozó lépést[140]. Az viszont egyáltalán nem igaz, hogy az öngyilkosság kizárólag emberi sajátosság volna. Bár nem túl gyakori, de előfordul az állatvilágon belül [266, 113]. A kutatók adaptív öngyilkoságnak nevezik, és a létezését úgy magyarázzák, hogy az egyed a saját génjének sikeres továbbörökítése érdekében pusztítja el magát. Előfordul, hogy a borsótetű (Acyrthosiphon pisum) testébe darazsak beszúrják a tojásaikat, hogy az utódaik parazitaként nevelkedjenek fel a borsótetű testéből táplálkozva. Ilyen esetekben előpordul, hogy az így megfertőzött borsótetű ledobja magát a levélről, ahol addig a társaival együtt élt. A földön biztosan áldozatává válik bogaraknak, de önfeláldozásával saját génjeit hordozó rokonai számára előnyöket biztosít, mert a parazita darázslárvák nem vándorolhatnak tovább a társaira. Az ausztráliai vöröshátú pók (Latrodectus hasseltii) nősténye is megeszi a hím párját, de ezt a kutatók nem a nőstény agresszivitásával, hanem a hím önfeláldozásával magyarázzák [266, 114]. Az önfeláldozásnak ugyanis adaptív előnye van, mert a magát halálba kopuláló hím saját génjeit továbbörökíti, míg a halálos párzástól megmenekülő hímtársai erre nem képesek – mindaddig, amíg nem vállalkoznak saját halálos kalandjukra. Ha figyelembe vesszük Oldenquist csoportegoizmus fogalmát (↰245.o.), akkor ezt a biológiai jelenséget úgy értelmezhetjük, hogy ez az öntudat nélküli altruizmus fogalmával írható le. A vallási áldozat, az emberáldozat, a katonai önfeláldozás (Dugonocs Titusz példája) is ide tartozik, mert olyan önkárokozások, amelyekben mindig megtalálható valamilyen általánosított másik, valamilyen csoport, szuperindividuum.

Destrukció

A romantikus emberfelfogás a destrukciót[19] károsnak, elítélendőnek, és ezzel együtt azonban (és talán éppen ezért) ritka jelenségnek tekinti. Első ránézésre úgy tűnhet, hogy még elméletileg is alá lehet ezt támasztani, hiszen a destrukció mindkét fél számára, tehát Egonak is, Alternek is kárt, hátrányt okoz, és nehéznek tűnik érveket szállítani a kettős hátrányokozás értelmére. Pedig nem így van. Mielőtt az okokról és mechanizmusokról beszélnék, előbb hivatkoznék arra, hogy a romantizicmustól mentes közgondolkodás is "tisztában van" azzal, hogy a mindennapi életünkben is van destrukció. Ezt bizonyítja az a mondás, miszerint „A szellem kiszabadul a palackból, és azt mondja a parasztnak, hogy egyet kívánhat, de akármit is kíván, a szomszédja kétszer annyit fog kapni. – *Szúrd ki az egyik szemem!* – jön a válasz." A tanmese tartalmát még rövidebben kifejezi a

[19] Vannak, akik a destrukció helyett a szadomazochizmus terminusát használják, aminek én nem tartanék túl szerencsésnek a terminusra rakódott egyéb konnotációk miatt, így inkább kitartok a destrukció terminusa mellett.

dögöljön meg a szomszéd tehene is – az előbbinél is gyakoribb – mondásunk. Szóval igenis van destrukció az életünkben. A kérdés az, hogy milyen "érvek", milyen indokok, milyen mechanizmusok képesek kiváltani ezt a jelenséget. Legalább két érvet, két mechanizmust meg tudok nevezni erre. Egyfelől az érzelmi mozgásaink miatt kerülhetünk ilyen helyeztebe, amikor valamilyen frusztrációt az indulati-érzelmi mechanizmusaink a destrukción keresztül tudnak csak megoldani. A "dögöljön meg a szomszéd tehene is" felkiáltás mögé könnyen ellképzelhetjük azt a helyzetet, amelyben az elkeseredettség úgy vakítja el az embert, hogy csak egyfajta totális romboláson keresztül látja csak feloldhatónak belső feszültségeit.

Ha jobban belegondolunk, akkor már az indulati-érzelmi döntések mögött is ott dolgozik bennünk valamifajta belső büntetési szándék, és ezt, mármint a büntetés jelenségének hatókörét kiterjeszthetjük. Ha így teszünk, akkor felismerhetjük, hogy érzelmileg kevésbé meghatározott helyzetekben is tetten érhetjük a büntetésen keresztül megvalósított destrukciót. Az alkujátékok elemzése során kiderül (↪298.o.), hogy bizonyos játékhelyzetekben (például az ultimátumjátékokban) az emberek gyakran olyan döntéseket hoznak, amelyekkel mind magukat, mind a társukat elütik a várható nyereségtől, tehát destrukciót hajtanak végre. Ennek magyarázata pedig egyszerűen az, hogy a felajnálott nyeremények kölcsönösen káros visszautasítását a méltánytalannak tekintett ajánlatokkal szembeni megtorlásként értelmezhetjük [6]. Amennyiben ez az értelmezésem helytálló, akkor viszont az állatvilágban, ahol a magasabbrendű cselekvésként értelmezhető büntetés jelenségét nem feltételezhetjük, nem számolhatunk a destrukció jelenségére sem.

Méltányosság, szolidaritás, reciprocitás

A bemutatott társas preferenciák a fogalmi építkezés alapfogalmai közé tartoznak, amely persze felbukkanhatnak összetett fogalmak komponenseiként. Vannak gyakran elemzett fogalmak, amelyeket a társadalmi kapcsolat kategóriája alá kell besorolnunk, de már összetett fogalmaknak kell tartanunk őket. Ilyen a **reciprocitás** fogalma annak minden altípusával együtt, ilyen a **méltányosság** vagy igazságos (fair) beállítódás vagy cselekvés kategóriája, és ilyen a **szolidaritás**, a másokkal való együttérzés fogalma is. Ezek mindegyikéről tengernyi kutatás, elemzés, tanulmány, könyv született, melyek bemutatása meghaladja a lehetőségeimet és képességeimet. A reciprocitás fogalmát máshol már bemutattam [517], itt csak néhány fontos vonását akarom kiemelni. A reciprocitás mint időben eltolt viszonzások rendszere sokrétű fogalom, és ezért nem meglepő, hogy sokféleképpen próbálták meg tipizálni. Számomra kétféle szempont szerint felosztás fontos, amelyek azonban nem ortogonálisak egymásra, van átfedés közöttük. Szokás **pozitív reciprocitás** és **negatív reciprocitás**, illetve **direkt** és **indirekt reciprocitásról** beszélni. Fehr és Gächter a pozitív és a negatív reciprocitást aszerint különbözteti meg, hogy valaki kooperál, jutalmaz vagy dezertál, büntet [157]. Herbert Gintis ugyanerre a kettősségre a gyenge, illetve erős reciprocitás fogalmakat alkalmazza [188]. Az altruizmus és reciprocitás kapcsolatát elemzi még Gintis társaival az 'Explaining altruistic behavior in humans' című cikkében [190], amelyből kiderül, hogy a kutatók között abban a tekintetben nincs vita, hogy a pozitív és negatív reciprocitás fogalmát a jutalmazó, kooperatív, illetve a büntető viselkedéshez kötik. Ez a szembeállítás persze korántsem újkeletű. Még jóval a játékelmélet, a kísérleti közgazdaságtan "megjelenése" előtt a szociológus Alvin W. Gouldner [195] vagy az antropológus Marshall Sahlins [457] is hasonló felosztást java-

soltak. A könyv játékelméleti részében külön fejezetet szentelek a direkt (↬304.o.) és indirekt reciprocitás (↬315.o.) fogalmainak, ezért itt nem térek ki rájuk. A különbség annyi köztük, hogy a viszonzás egymást ismerő vagy ismeretlen emberek között valósule meg. Egy rövidebb részben kitérek a méltányosságelméletekhez köthető alkujátékok kérdésére is, de ennek a kérdéskörnek lényegesen kevesebb figyelmet szentelek. Egyáltalán nem foglalkozom viszont a szolidaritás problémakörével. Nem azért, mert nem tartom fontosnak, inkább a kellő felkészültség hiányzik. Magyar nyelven a méltányosságelméletekről jó áttekintést adnak Gulyás Attila cikkei [199], [200], a szolidaritás témakörében Janky Béla könyvét ajánlom az érdeklődőknek [258].

4.3 A társadalmi kapcsolat típusai

Az előző fejezetben a beállítódásokkal kapcsolatos fogalmakat elemeztem. Különösen sokat foglalkoztam az ágensre irányultság különböző formáival. A fejezet bevezetőjében két dolgot is jeleztem. Egyfelől azt, hogy ezek a fogalmak mind a társadalmi kapcsolat fogalma alá sorolhatók be, másfelől azt, hogy az ágensre irányultsággal kapcsolatos fogalmak szűkebb terjedelműek a társadalmi kapcsolat fogalmához képest. Természetesen minden generikusan alárendelt fogalom szűkebb terjedelmű a fölöttes kategóriához képest, de itt nem erre a trivialitásra utalok. Az ágensre irányultság mint beállítódás figyelembe vétele fontos a társadalmi kapcsolat tipizálása szempontjából, de a szükség van valami másra is. Amikor Max Weber meghatározza a társadalmi kapcsolat típusait, akkor két dimenzió mentén különíti el az új kategóriákat. Egyrészt használja azt a dimenziót, amit én az ágensre irányultság fogalmával ragadtam meg, másrészt hivatkozik a társadalmi cselekvések típusaira (racionális, illetve aracionális cselekvések két nagy osztályára), amelyeket korábban én is bemutattam. Mivel nagyon fontos kategóriákat vezet be ezen a ponton, érdemes megnéznünk, hogyan teszi mindezt. A társadalmi kapcsolat meghatározása után Max Weber a fogalom három altípusát különíti el egymástól. Először a **harc** (konfliktus) fogalmát definiálja, majd kicsivel később pedig megadja a *közösség* és a *társulás* meghatározását is.

> „*Harcnak* nevezünk egy társadalmi kapcsolatot, amennyiben a cselekvőt az a szándék vezeti, hogy saját akaratát a másik vagy a többi féllel szemben keresztülvigye.” [575, 64]
>
> „ *Közösségnek* nevezzük a társadalmi kapcsolatot, ha a társadalmi cselekvést . . . a résztvevők szubjektíve átérzett (érzelmi-indulati vagy tradicionális) összetartozásán alapuló beállítódás jellemzi.”
>
> „ ... *társulásnak* nevezzük a társadalmi kapcsolatot, ha a társadalmi cselekvést racionálisan (érték- vagy célracionálisan) motivált érdekkiegyenlítődésen vagy ugyanígy motivált érdekkapcsolódáson alapuló beállítottság jellemzi.” [575, 66].

Ez a weberi tipizálás bár nagyon fontos, de terminológiai szempontból közel sem teljes, érdemes tehát kiegészíteni.[20] Bár Weber erre tudatosan nem igazán reflektál, mégis mondhatjuk azt, hogy a weberi kategóriarendszeren belül a fenti három fogalom egy szembenállást fejez ki. A harc fogalma van az egyik oldalon, míg a közösség és társulás fogalmai a másikon. A szembenállás lényege az, hogy a harcot az ellenállással szembeni akaratérvényesítés, tehát az “kizárólagos” önérdekérvényesítés, a közösséget

[20] Ezzel a kiegészítéssel korábban másutt már megpróbálkoztam [519], de itt érdemesnek tartottam újra neki futni a problémának. Egyrészt azért, mert itt részletesebben és alaposabban meg tudom indokolni a mondanivalómat, másrészt most fogalmilag általánosabb módon oldom meg ugyanazt a feladatot.

és a társulást pedig az összetartozás, az érdekkiegyenlítődés, érdekkapcsolódás jellemzi. Ezt a dichotómiát az ágensre irányultság fogalmának két pólusával jól megragadhatjuk, bár a terminológiai konzisztenciához szükségünk van egy-két új fogalom bevezetésére. Mondhatjuk ugyanis azt, hogy a harc a magára irányuló, míg a közösség és kapcsolat a másikra irányuló beállítódással kapcsolható össze, de ha ezt tesszük, akkor az is nyilvánvalóvá válik, hogy nem ugyanazon a fogalmi szinten vannak ezek a kategóriák. A weberi fogalomkészletből hiányzik a harc fogalmi párja a másra irányuló beállítódás területén, illetve alá kell bontani a harc kategóriáját azon logika mentén, ahogy a közösség és kapcsolat kategóriái különülnek el egymástól. A három weberi fogalom az alábbi módon rendezhető táblázatba (kérdőjelek jelzik a hiányzó kategóriákat vagy rendezési dimenziókat).

	magára irányuló	másikra irányuló
?	harc	?
?	?	társulás
?	?	közösség

Szükség van tehát egy új kategóriára a közösség és társulás fölöttes fogalmaként, illetve szükség van két új kategóriára a harc fogalma alatt. Kicsit egyszerűbb az első feladat megoldása. Ha a magára irányuló beállítódás főfogalma a harc, akkor ennek párja (ellentéte) az **együttműködés** (kooperáció) lehet a másikra irányuló beállítódás oldalán.

Kicsivel nehezebb a harc fogalmának alábontása – egyszerűen azért, mert a magyarban nem áll rendelkezésre megfelelő terminológia a feladat megoldásához. Tartalmilag, szemantikailag az **ellenfél** és az **ellenség** kategóriái megfelelőek lennének, de szintaktikailag, konnotációjukat tekintve nem illeszkednek jól a kiegészítendő fogalomkészlet elemeihez. Igaz, a harc és a közösség és társulás fogalmi közt is érezhető már ez a fajta diszkrepancia, ugyanis míg az utóbbi kettő valamiféle csoportra utal (nyilván azokra a csoportokra, amelyek tagjai a másikra irányuló beállítódás "kedvezményezettjei"), addig a harc inkább a magára irányuló beállítódás következményeként kialakuló állapotot írja le. Ugyanez a baj az ellenfél és az ellenség kategóriáival is. Ha viszont túllendülünk ezen a problémán, akkor már nyugodtan párba állíthatjuk az ellenfél kategóriáját a társulás fogalmával, míg az ellenségét a közösségével. Az összekapcsolási igényt pedig úgy védhetjük meg, ha megnézzük, hgy Weber hogyan tett különbséget a közösség és a társulás fogalmai között. Sok segítséget jelent, hogy Weber maga is jelezte, hogy a közösséghez hozzá lehet kapcsolni az érzelmi és a tradicionális (habituális) beállítódást, a társuláshoz pedig a kétfajta racionalitást. Ha a közösség és társulás fogalmai között a mérlegelő, illetve nem-mérlegelő beállítódás szerint lehet tehát különbséget tenni, akkor ugyanezt a felosztást használhatjuk az ellenfél és ellenség fogalmai között is. A mérlegelő vagy nem-mérlegelő beállítódás helyett a racionális vagy aracionális beállítódás kategóriáit használva a fenti hiányos táblázatot az alábbi módon tehetjük teljessé.

	magára irányuló	másikra irányuló
	harc	együttműködés
racionális	ellenfél	társulás
aracionális	ellenség	közösség

A weberi felosztás explikációja (illetve kiegészítése) után tehetünk még egy lépést a társadalmi kapcsolatok tipizálásában. Az emberi konfliktusok természetét vizsgálva az el-

lenség és ellenfél, barát vagy ellenség fogalmáról is értekezett George Schwab [466].[21] Schwab a **közellenség**, a köz, közösség ellensége (enemy), illetve a **személyes ellenség** (foe) fogalmát állította szembe egymással, és e két fogalom jelentésváltozásainak történeti bemutatása közben kiemelte, hogy fontos mozzanat volt a semleges szereplő megjelenése a szembenálló felek között. Ezt a figyelmeztetést megfogadva nekünk is érdemes felvenni egy harmadik pólust, a semlegességet a társdalmi kapcsolatok közé.[22] Ha így teszünk, a következő táblázathoz juthatunk.

	magára irányuló	másikra irányuló	egyikre sem irányuló
	harc	együttműködés	semlegesség
racionális	ellenfél	társulás	kívülálló
aracionális	ellenség	közösség	kívülálló

4.4 Interdependens cselekvések

A társadalmi cselekvésről szóló fejezetben eddig azt a kérdést jártam körül, hogy mit jelent a cselekvés fogalma, milyen típusai vannak a cselekvésnek, és kik a társadalmi cselekvések ágensei, a megelőző fejezetben pedig foglalkoztam az intencionalitás, az akarat problémájával. A cselekvés fogalmi lényegének megragadásakor az intencionáltságra fókuszáltunk. Eddig azonban az egyéni cselekvőre koncentráltunk, és nem figyeltünk arra a lehetőségre, amikor az egyéni cselekvések megvalósulása, azok sikeressége a kapcsolatba került felek mindegyikétől függ, vagyis a felek kölcsönös függőségben vannak. Ezt a társadalmi állapotot kölcsönös függésnek vagy interdependenciának, interdependens helyzeteknek nevezhetjük. A továbbiakban az interdependens helyzetek, és az ilyen helyzetekbe került cselekvések alapvonásait szedem össze. Ha már tisztázva van, hogy:

- a cselekvés elméletén belül mit jelent a tevés és a tartózkodás fogalma,
- az intencionalitás elméletén belül mit jelent az akarat fogalma,
- a cselekvés elméletén mit jelent az ágens fogalma,

akkor a cselekvők saját és mások cselekvésére irányuló akaratait, és ezek egymáshoz való viszonyait táblázatba foglalhatjuk. Valamelyik ágens cselekvésére Ego és Alter akarata vagy nem-akarata irányulhat (mindkettő intencionális!).

	Alter akarja	Alter nem akarja
Ego akarja	ágens cselekvését	ágens cselekvését
Ego nem akarja	ágens cselekvését	ágens cselekvését

A kölcsönös függőség itt abban érhető tetten, hogy a táblázatot elemeit úgy kell értelmezni, hogy egy cselekvésre vonatkozóan két akaratot akarnak "érvényesíteni" ('Ego akarja, Alter akarja, hogy ágens cselekedjen' stb.), és e két akarat metszetéből, egymáshoz való viszonyából kerekedhet ki valahogy a cselekvés tényleges kimenete, eredménye. A fenti táblázat azonban még többértelmű, hiszen mind az ágens, mind a cselekvés fogalmába

[21] Magyar nyelven Orthmayr Imre adott egy alapos áttekintést a konfliktus problémaköréről [399].

[22] Ez a triádok elemzésében is fontossá válhat, amikor a semleges fél békítő, közvetítő, esetleg ügydöntő szerepet tölthet be, aki nélkül nem lehet megérteni a társadalmi kapcsolatok dinamikáját, de akkor is hasznát vehetjük ezt a fogalmat, amikor felhatalmazottsági konfliktusokkal állunk szemben.

két-két értéket helyettesíthetünk (az Ego/Alter, illetve a tenni/tartózkodni párokat), ezért konkrét értékekkel kell kitöltenünk a táblázatot. Ha így teszünk (beírjuk, hogy 'Ego akarja, Alter akarja, hogy Ego megtegye' vagy 'Ego akarja, Alter nem akarja, hogy Ego megtegye' stb.), akkor az alábbi tizenhat állapothoz jutunk.

	Alter akarja	Alter nem akarja
Ego akarja	Ego megtegye (1)	Ego megtegye (2)
Ego akarja	Ego tartózkodjon (3)	Ego tartózkodjon (4)
Ego akarja	Alter megtegye (5)	Alter megtegye (6)
Ego akarja	Alter tartózkodjon (7)	Alter tartózkodjon (8)
Ego nem akarja	Ego megtegye (9)	Ego megtegye (10)
Ego nem akarja	Ego tartózkodjon (11)	Ego tartózkodjon (12)
Ego nem akarja	Alter megtegye (13)	Alter megtegye (14)
Ego nem akarja	Alter tartózkodjon (15)	Alter tartózkodjon (16)

A tizenhat állapot közül a felében a két fél akarata megegyezik (vagy mindketten akarják vagy egyikük sem akarja valamelyikük cselekvését), tehát konszenzus (akaratkonszenzusról) van közöttük (az 1., 3., 5., 7., illetve a 10., 12., 14., 16. esetekről van itt szó). Az ilyen helyzetekben nyilván az várható, hogy a közös akarat szerint valósul meg (marad el) a cselekvés, hiszen mindkét fél egyaránt akarja (van nem akarja) azt. A másik nyolc esetben viszont, ahol a felek akaratai nem egyeznek meg, ott konfliktusról (akaratkonfliktusról) beszélhetünk. Az ilyen helyzetekben a cselekvés feltétele az, hogy a felek valamilyen módon rendezik a konfliktust (ez a helyzet a 2., 4., 6., 8., illetve a 9., 11., 13., 15. esetben). Hogy hogyan, azt egyelőre nem vizsgálom. Érdemes kicsit átrendezni a táblázatot, hogy jobban látszódjék benne az, hogy kinek a cselekvéséről van szó.

	Alter akarja	Alter nem akarja	Alter akarja	Alter nem akarja
Ego akarja	Ego megtegye (1)	Ego megtegye (2)	Alter megtegye (5)	Alter megtegye (6)
Ego akarja	Ego tartózkodjon (3)	Ego tartózkodjon (4)	Alter tartózkodjon (7)	Alter tartózkodjon (8)
Ego nem akarja	Ego megtegye (9)	Ego megtegye (10)	Alter megtegye (13)	Alter megtegye (14)
Ego nem akarja	Ego tartózkodjon (11)	Ego tartózkodjon (12)	Alter tartózkodjon (15)	Alter tartózkodjon (16)

Ebben az elrendezésben a táblázat első két oszlopában az Ego, a következő kettőben pedig az Alter cselekvésére irányuló akaratok egymáshoz való viszonyát láthatjuk. Mindkétszer létezik 4-4 olyan helyzet, amikor konszenzus, illetve konfliktus van a felek között. Ego cselekvése esetében konszenzus van az 1.,3.,10.,12., konfliktus van a 2.,4.,9.,11. esetben. Alter cselekvése esetében konszenzus van a 5.,7.,14.,16., illetve konfliktus van a 6.,8.,13.,15. esetben.

A játékelméleti fejezetben tárgyalt társdalmi kapcsolatokat másfajta interdependencia jellemzi, mert ott a cselekvők akaratai sosem a másik cselekvőre, hanem csak saját maguk cselekvéseire irányulnak. Akkor a kölcsönös függőség abban a vonásban fejeződik ki, hogy a játék végeredményét sosem az egyik vagy a másik játékos cselekvése, hanem mindig a kettejük (vagy többük) cselekvéseinek összessége, tehát azok egymásra hatása határozza meg. Adott tárgyalási kontextuson belül könnyen el tudjuk különíteni az interdependencia itt jelzett kétféle értelmét – mégha erre tudatosan nem is igazán szoktunk reflektálni.

4.5 Hatalom

Amennyire meghatározza életünket a hatalom jelensége, annyira nem tudunk mit kezdeni vele. Nem tudjuk igazán megragadni, nem tudjuk leírni tudományos eszközökkel. Mindenhol ott van, de rejtőzködik, bújkál, letagadja magát. Azok tagadják le leginkább, hogy van, akik élnek vele, és azok érzik leginkább, hogy van, akikkel szemben élnek vele. Nem véletlen az a rengeteg értelmezési, definiálási kísérlet, ami a fogalom megragadására irányult.[23] Nem érthetjük meg igazán a cselekvéseinket, ha nem tudjuk, mit jelent ez a fogalom, ha nem értjük, maga a jelenség hogyan befolyásolja életünket. A sokféle értelmezési kísérlet közül még mindig Max Weber meghatározását tartom a legpontosabbnak, azt fogadom el, és abból kiindulva próbálom meg explikálni a fogalmat. Nem véletlen, hogy Weber a kategóriatanában rögtön egymás után definiálta a **hatalom** és az **uralom** fogalmát, hiszen a két fogalom sok szempontból nagyon közel van egymástól, de az sem véletlen, hogy két kategóriát tartott szükségesnek elkülöníteni, nem egy fogalmat definiált. Persze nem könnyű a pontos határvonalat meghúzni a két weberi fogalom között, ez sokaknak nem is sikerült, és nagyon sok félreértés, félreértelmezés született a két fogalom jelentésével, egymáshoz való viszonyával kapcsolatban. Itt egyelőre csak a hatalom fogalmát próbálom meg körbejárni, az uralom kategóriáját máshol elemzem (↪397.o.). Az induláshoz idézzük fel Max Weber meghatározását.

> „Minden olyan esetben hatalomról beszélünk, ha egy társadalmi kapcsolaton belül van rá esély, hogy valaki saját akaratát az ellenszegülés ellenére is keresztülvigye, függetlenül attól, hogy min alapul ez az esély.” [575, 77]

A hatalom fogalmának igazi értelme az uralommal szembeállítva fedhető fel igazán, de erre még várnunk kell. Az viszont rögzíthetjük, hogy a hatalom egymással ütköző akaratok esetén létezik, vagyis a akaratkonfliktust, mégpedig a nyílt akaratkonfliktust szükséges előfeltételként kell értelmeznünk. És akkor beszélhetünk hatalomról, ha az a konfliktus valamilyen kényszer alkalmazásával oldódik fel. Ezt fejezi ki az “ellenszegülés ellenére is” passzus a weberi definícióban. A hatalomban mindig van egy **domináns** és egy **alárendelt** “szerep”, mindig van egy, a hatalmat gyakorló és egy, a hatalomgyakorlást elszenvedő fél. Az is nyilvánvaló még a meghatározásból, noha tételesen nincs kimondva, hogy a hatalomgyakorlást elszenvedő fél akarata nem teljesül, hiszen akaratkonfliktus esetén a domináns fél akarata csak akkor érvényesíthető, ha a szembenálló másik akarat nem teljesül. Márpedig az akarattal rendelkező ember (lény) azt nagyon nehezen viseli, ha az akarata ellenére kell valamit megtennie. A szabad akarat kérdése itt a másoktól független, mások által nem blokkolt vagy nem direkt módon befolyásolt akaratképzés lehetőségét jelenti. Szándékosan fogalmaztam ennyire kiterkert módon. Nem minden akaratbefolyásolás tartozik a hatalom jelenségébe, csak azok, amelyek a nyílt akaratkonfliktusokat kényszer alkalmazásával törik fel. Az akaratszabadság tehát itt a másoktól független akaratképzés lehetőségét jelenti, és az akarat ilyen értelemben vett szabadsága – úgy tűni – az egyik legfontosabb emberi minőségünk.[24] Az akarat szabadságának pozitív

[23] A kandidátusi értekezésemben én is foglalkoztam a hatalom kérdésével, bemutattam a jelenség tipizálási lehetőségeit, és megpróbálkoztam a fogalom explikálásával is [524]. Kifejezettem a hatalom jelenségét mutattam be egy egyetemi tankönyv egyik fejezetében [528], és külön fejezetet szenteltem a hatalom-uralom párosnak a korábbi cselekvéselméleti könyvemben is [517].

[24] Nem tudom, hogy az állatok számára mit jelent a másik állat általi erőszakalkalmazás, ezért ezt a tételt nem terjesztem ki az akarattal rendelkező lények halmazára, de ezzel nem még tagadom meg a kiterjeszthetőség lehetőségét.

értékéből fakad, hogy az akarat megtörése negatív érzelmeket vált ki. A hatalom lényegi összetevőjét jelentő kényszer alkalmazása után mindig megtapasztalhatjuk a "szenvedő" fél frusztráltságát, méltánytalanságérzetét. A hatalom pont azért rejtőzködő természetű, mert a tényleges hatalomgyakorlást, azaz mások akaratának megtörését társadalmi szinten elítéljük, és az emberi társadalomban való létünk óta ki akarjuk szorítani a békés hétköznapjaink világából. Ez a tartós társadalmi konszenzus azonban eddig sem tudta, és elég nyilvánvalóan a jövőben sem tudja megakadályozni a hatalmon keresztüli befolyásolás jelenségét. Ha a hatalmat olyan jelenségként értelmezzük, amelyben a két ágens közti akaratkonfliktust valamilyen kényszer alkalmazásával oldja fel az egyik fél, akkor ezt a jelenséget megtalálhatjuk az állatvilágban is. Amikor szűkösek az erőforrások, amikor meg kell harcolni az élelemért, a területért, a szállásért, a szexuális partnerért, akkor ez a harc egymással szembenálló állatok között zajlik, ahol legtöbbször a nyers erő dönt. Az alfahímet nem megválasztják a többiek, hanem az éppen legerősebb egyed harcolja ki magának ezt a pozíciót. A csípkedési sorrend nem megállapodás kérdése, nem is az érkezési sorrendet fejezi ki, hanem az aktuális erőviszonyok puszta leképezését jelenti. A szűkös erőforrásokhoz való hozzáférés sorrendjét jelzi mindenki számára, ami persze az időben folyamatosan változik, hiszen az erősek elgyengülnek, a korábbi gyengék erősekké válnak. A sorban elfoglalt pozíciót mindig meg kell védeni, és persze mindig lehet támadást indítani egy jobb helyért. És mindig az erő dönt. Az erőszakalkalmazás a hatalom legelemibb és legősibb forrása. Amikor tehát a hatalom eszközét alkalmazza az ember, akkor az egyik legősibb biológiai késztetést (és technikát) hagyja érvényesülni.

Korábban tételesen nem mondtam ki, de úgy használtam az erőszak fogalmát, mintha a hatalom egyfajta forrásaként értelmeznénk. A szembenálló fél akaratának megtöréséhez mindig szükség van valamilyen kényszerítő eszközre, és ezt az eszközt nevezhetjük a hatalom forrásának. Eddig annyi derült ki, hogy ez lehet az erőszak(alkalmazás). A kérdés az, hogy elképzelhető-e más ilyen eszköz, vagyis létezik-e másfajta **hatalomforrás** is. A válasz nyilván igenlő. A tudás (információ), a pénz (gazdasági erőforrás), a pozíció (hatáskör) olyan erőforrásként is értelmezhetőek, amelyek egyenlőtlen eloszlása aszimetrikus viszonyokat teremt a felek között, és a hatalom gyakorlását teszi lehetővé. Az a társadalmi konszenzus, hogy a hatalomgyakorlás nem kívánatos jelenség, amit ki kell szorítani a társadalmi gyakorlatokból, megnyilvánul abban a tényben is, hogy a hatalomgyakorlás minden formáját, vagyis a hatalom mindenféle eszközének, forrásának alkalmazását bünteti a jogrend. Az erőszak, a zsarolás, a pénzzel, gazdasági erőfölénnyel való befolyás, a bennfentes kereskedés, a monopolhelyzettel való visszaélés, a szervezeti pozícióval való visszaélés, a hatáskörtúllépés, a politikai befolyás tilalma mind valamilyen hatalomgyakorlási helyzetet akar megakadályozni a jog eszközével – miközben a hatalom potenciális forrásait "nevezi meg". A társadalmi választások elmélete rengeteg olyan problémát vetett fel és dolgozott ki (↪411.o.), amelyek a hatalomgyakorlás különböző lehetőségeiként is értelmezhetők egyben. A háború művészete, a hatalommérnökség, a harci konfliktusok feloldásának, megnyerésének technikái is – legalább részben – a hatalom eszközeiről, forrásairól és ezáltal kicsit a hatalom típusairól is szólnak, de ezek tudományos feldolgozása még várat magára. Weber azt írta a hatalom kategóriájáról, hogy amorf fogalom. Részben ez magyarázhatja azt, hogy nagyon kevés értelmes tipizálás jelent meg a hatalom fogalmával kapcsolatban. Talán az egyetlen régóta létező és számomra is elfogadható tipizálás azt a kettőséget ragadja meg, hogy az akaratérvényesítés során a hatalmat gyakorló fél mire tudja rávenni a másikat: arra, hogy megtegyen, vagy

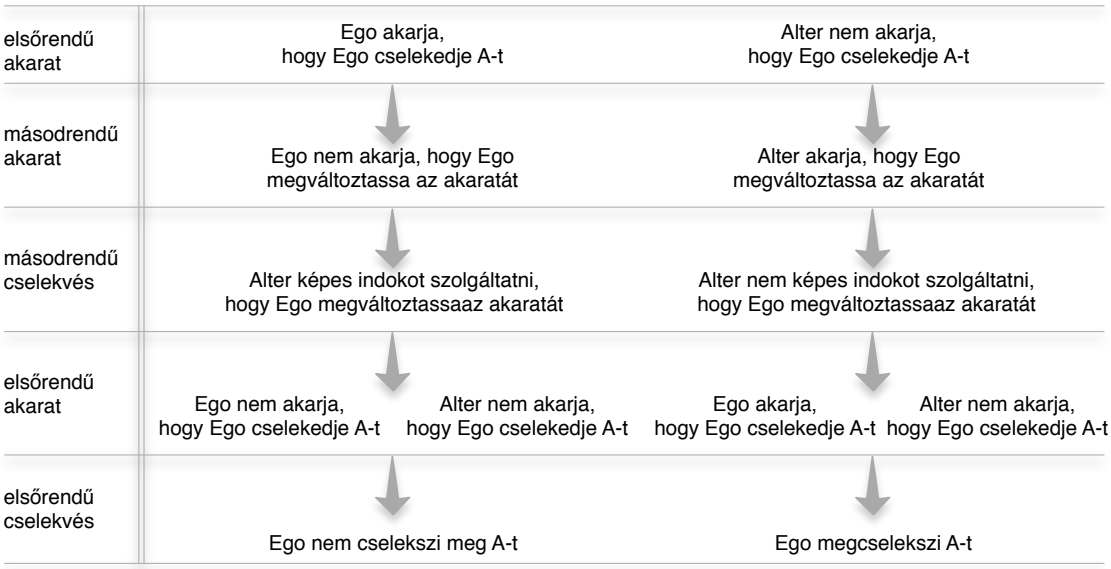
arra, hogy ne tegyen meg valamit. Ennek megfelelően beszélhetünk **pozitív hatalomról**, amikor az alárendelt fél megtesz valamit, amit egyébként önszántából nem akarna, illetve **negatív hatalomról**, amikor az alárendelt egyén nem teszi meg azt, amit egyébként megtenne magától.

Bármennyire is száműzni szeretnénk ezt a jelenséget a társadalmi térből, a hatalom mindenhol jelen van, sokszor ez dönti el a konfliktusaink kimenetét. Ez a felismerés könnyen kialakíthatja azt a pesszimista hozzáállást, hogy ez ellen nem lehet semmit sem tenni, mindig minden a hatalmasoké volt a múltban, van a jelenben, és lesz a jövőben, a hatalom mozgatja a világot. Bár én is úgy vélem, hogy a hatalom az egyik legmeghatározóbb jelensége a társadalom működésének, a hatalom mindenhatóságának tézisében nem hiszek. Említettem már, hogy a hatalmat csak az uralomhoz való viszonyában érthetjük meg teljesen. De azt még nem mondtam, hogy van egy másik fontos szempont is, amit figyelembe kell vennünk, ha teljes mélységben meg akarjuk érteni a hatalom fogalmát. Jack Hirshleifer hívta fel a figyelmet egy cikkében arra [237], hogy a hatalom nem mindenható. Hirshleifer a katonai konfliktusról értekezik, de érvelését nyugodtan kiterjeszthetjük a hatalomgyakorlás mindenféle formájára. A földrajzilag kicsinek mondható csatatéren a nyertes mindent visz, ami azt jelenti, hogy ha – akárcsak egy picinyke katonai erőfölény birtokában, de – megnyeri a csatát, akkor utána – a katonai erőfölényéhez képest aránytalan mértékben – megszerez mindent magának. A közgazdaságtanban ezt a jelenséget a **növekvő mérethozadék** vagy skálahozadék (increasing-return) elvével írják le. Ez az elv azonban nem terjeszthető ki a legyőzött ország teljes területére, nem érvényes már a háború egészére nézve sem, nem beszélve a békés időszakokról. Minél nagyobb földrajzi területre szóródnak szét az emberek (a csata után), annál inkább egy másik logika, a **csökkenő mérethozadék** vagy skálahozadék (decreasing-return) elve kezd érvényessé válni. Ekkor az egyre több katonai/hatalmi erőforrás biztosítása is csak relatíve egyre kevesebb eredményt képes produkálni. Minél inkább szészóródnak az emberek a földrajzi (vagy a társadalmi) térben, annál nehezebben lehet hatalomtechnikákat alkalmazni velük szemben. Hiershleifer megjegyezi még, hogy a növekvő mérethozadék elve miatt lehet kialakítani (és fenntartani) az állami erőszak-monopóliumokat, és a csökkenő mérethozadék logikájából érthető meg, miért nem alakul ki a világállam, vagy miért esnek szét a nagy birodalmak, és miért képtelenek minden téren érvényesíteni a nagy hatalomközpontok saját akarataikat.

A hatalom fogalmának explikációja

Korábban bemuttam, hogy milyen interdependens társadalmi kapcsolatba kerülhet két cselekvő (↪263.o.). Ezek közül voltak konszenzusos, voltak konfliktusos helyzetek. Definiáltam a harc (konfliktus) fogalmát is (↪260.o.). Az akaratkonfliktus lényege, hogy az egymásnak feszülő (egymásnak ellentmondó) akaratok csak egymás kárára valósulhatnak meg. A kérdés tehát az, hogy miként "győzhet" az egyik akarat a másik "fölött", vagy fordítva: miért nem érvényesül valaki saját akarata a másik ágens akarata ellenében. Ezen a ponton nagyon körültekintőnek kell lennünk. A pontos ítéletkezéshez arra van szükség, hogy figyelembe vegyük azt, hogy az éppen vizsgált helyzetben melyik ágens cselekvésével kapcsolatos akaratkonfliktusról van szó. E szempont figyelembe vételével kétféle akarattípust különböztethetünk meg. Beszélhetünk saját cselekvésre irányuló 'cselekedni akarás'-ról, illetve a másik cselekvésére irányuló 'befolyásolni akarás'-ról.

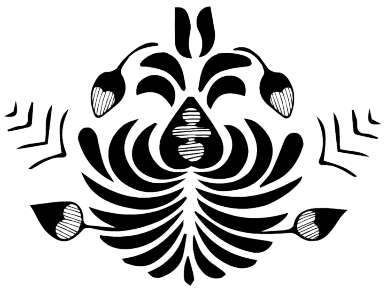
Az akaratkonfliktus ténye még csak az ellentétes akaratok létezését fejezi ki, semmi többet. A nagy kérdés az, hogy milyen módon oldódik fel ez a konfliktus. A konfliktus kétféle lehetséges kimenete közül az egyik az, amikor a cselekedni akarás érvényesül (tehát a befolyásolási akarat kudarcot vall), másik az, amikor a befolyásolási akarat érvényesül. Az első esetben a befolyásolási akarat sikertelen marad, a másodikban sikeres. Ha valaki az utcán elém áll, és pénzt követel tőlem, és én nem akarok adni neki, kitérek az útjából, tovább megyek, akkor a másik nem volt képes érvényesíteni az akaratát velem szemben. Ha viszont a másik fél egy fegyvert fog rám, és azzal fenyeget, hogy lelő, akkor odaadom a pénzemet neki, saját – eredeti – akaratom ellenére. Konfliktus esetén a mindkét fél elsődleges akarata valamelyikük cselekvésére irányul, de a két akarat eltér egymástól. Az akarat jelzése azonban még nem elég a megvalósuláshoz. A saját akarat érvényesítéséhez szükség van egy második akaratra is, mégpedig arra az akaratra, hogy a másik fél változtassa meg a saját elsődleges akaratát, illetve a partnerben az a másodrendű akarat keletkezik ilyenkor, ami tartalma nem akarni a saját elsődleges akaratának megváltoztatását. A konfliktushoz ez a kétfajta és kétszintű akaratkülönbség elengedhetetlen. Lehetséges, hogy valaki az elsődleges akaratkonfliktus konstatálása után inkább visszalép, nem akarja érvényesíteni saját akarát a másikkal szemben. Viszont ha mégis erre törekszik, akkor az akaratok kétszintű konfliktusa feloldásához szükség van valamire. A befolyásoló akarattal rendelkező ágens oldaláról szükség van arra a pótlólagos cselekvésre, amelynek eredményeként a cselekedni akaró fél megváltoztatja a saját eredeti akaratát, és a befolyásoló akarat szerint fog cselekedni. A befolyásoló akarattal a kapcsolatba belépő ágensnek elfogadható (**(gyakorlati) indokot**) kell szolgáltatnia arra vonatkozóan, hogy a cselekedni akaró (vagy nem akaró) ágensnek miért kellene megváltoztatnia a saját cselekvésére irányuló eredeti akaratát. A befolyásoló akarat kifejtése, kifejezése elsőrendű cselekvés, az megfelelő indok szolgáltatása viszont már másodrendű cselekvés (ami egy másodrendű akarattal kapcsolható össze), hiszen a befolyásolni akaró ágens indokot szolgáltató cselekvése (illetve az erre irányuló akarata) a másik fél elsőrendű cselekvésére irányuló akaratának megváltoztatására irányul. A folyamatot így ábrázolhatjuk.



Háromféle módszer, háromféle indok szolgáltatása lehetséges a cselekvő eredeti akaratának megváltoztatására. Az egyik módszer a — weberi értelemben vett – **hatalom** (POWER), a másik a – weberi értelemben vett - **uralom** (AUTHORITY) gyakorlása, a harmadik a "vitán keresztüli" **meggyőzés** (BELIEF-REVISION). A hatalom esetében az akaratot megtörik ("ellenszegülés ellenére viszik keresztül az akaratot"), az uralom esetében az ellenkezés úgy iktatják ki (↰397.o.), hogy egy norma által előírt kötelességre vagy tilalomra hivatkoznak. A vita során érveket hoznak fel egymással szemben a másik akaratának megváltoztatása érdekében. Akkor beszélhetünk hatalom jelenségéről, ha a hatalmat gyakorolni kívánó fél valamilyen kényszert alkalmaz gyakorlati indokként a vele szemben levő akaratának megváltoztatására. Ahogy azt már korábban jeleztem, ez lehet erőszakkal, tudással, pénzzel, gazdasági erőforrással, szervezeti pozícióval, de akár valamilyen a másik személyes elköteleződéssel való (vissza)élés. A lényeg az, hogy az adott társdalmi kapcsolatban alkalmas legyen az alávetett fél akaratának megtörésére. Ha hatalmi helyzeteket akarunk megérteni, akkor mindig azt kell megtalálnunk, hogy milyen hatalomforrásra tudott támaszkodni a hatalmat gyakorló fél, illetve ez miért volt alkalmas arra, hogy az alávetett személy megváltoztassa az akaratát.

Chapter 5

Kényszermentes társadalmi kapcsolat



„Két ember megy a szavannában, amikor meglátnak egy oroszlánt. Az egyik elkezdi ledobálni magáról mindenét, hogy gyorsabban tudjon futni az oroszlán elől. Ezt látva a társa megszólal:
– *Hova erőlködsz? Az oroszlán gyorsabb az embernél, úgysem tudsz elfutni előle.*
Mire a válasz:
– *Nekem nem is az oroszlánnál kell gyorsabban futnom, hanem a nálad.*" (Pesti vicc)

Ebben a fejezetben azokat a társadalmi kapcsolatokat elemzem, amelyekben nincs kényszeralkalmazás, tehát a kapcsolat résztvevői szabadon, mások befolyásától mentesen cselekedhetnek. Ezt a világot írja le a klasszikus értelemben vett **játékelmélet**. A játékelmélet azokkal a társadalmi kapcsolatokkal (helyzetekkel, interakciókkal, szituációkkal) foglalkozik, amelyekre a kölcsönös függőség, egymásrautaltság, az **interdependencia** jellemző. A kapcsolat kimenete sosem csak az egyik vagy másik szereplő (játékos) cselekvésétől (döntésétől) függ, hanem mindig a két (vagy több) játékos döntésének eredője határozza meg a végkifejletet. Nem szokás hangsúlyozni, inkább evidenciaként tekintenek arra, hogy kényszermentes társadalmi kapcsolatokról van szó, de ez a reflektálatlanság sajnos nehezíti annak tisztázását, vajon milyen kényszereket kell/szabad kizárnunk a játékelméleti elemzésekből. Ebben a fejezetben nyugodt szívvel kizárom a külső kényszer alkalmazásának lehetőségét, tehát hatalommentes helyzeteket feltételezek, és kicsit bizonytalankodva kizárom a belső kényszer lehetőségét is, tehát norma- és uralommentes világot feltételezek. Ez utóbbiról azonban azt gondolom, hogy csak módjával fogadható el, nem nagyon tudok elképzelni olyan társadalm kapcsolatot, amelybe ne lehetne normatív és uralmi összetevőket találni. Ezzel szemben a hatalom kizárását elméletileg is, gyakorlatilag is teljesen el tudom fogadni. Az eddigiekből következően azt mondhatjuk, hogy a játékelmélet az egymással interdependes társadalmi kapcsolatokban levő ágensek érdekvezérelt cselekvéseinek elmélete.

A játékelméleti építkezés során a társadalom lehető legegyszerűbb modelljéből indulok ki. A legegyszerűbb itt egyszerre sok feltétel teljesülését jelenti. Egyrészről a lehető legkisebb társadalmat tételezem fel, ami két főből áll: Ego és Alter alkotja. Ego az a személy, akinek a cselekvését meg akarjuk érteni, magyarázni, Alter pedig az a személy, aki szükséges a társadalmi kapcsolat létezéséhez (és ezzel a társadalom legkisebb egységének létrejöttéhez).[1]

A játékok bemutatása során három korábbi játéktaxonómiára hivatkozom. A 2x2-es játékok első teljességre törekvő leírását Rapoport és Guyer adták meg, akik a preferenciamátrix sajátosságai alapján végezték az elemzésüket, de nem foglalkoztak komolyabban játékosztályok kialakításával, a játékok tipizálásával, illetve nem adtak neveket sem a játékoknak (sőt, nem is mutatták be őket – például történetek segítségével). A játékok számosságára vonatkozóan Rapoport és Guyer 1966-os cikke alapján közel negyven éven keresztül elfogadott volt az a nézet, hogy 78-féle játékot lehet elkülöníteni egymástól. Ez a szám úgy jött ki, hogy a 2x2-es preferenciamátrix 2x4 értékét 576 (4!x4!) féle módon lehet elrendezni, amelyek közül viszont sokatat ekvivalensnek lehet tekinteni egymással. Rapoporték úgy gondolták, hogy a mátrix sorait és oszlopat felcserélhetjük anélkül, hogy a játék logikáját, belső viszonyrendszerét megváltoztatnánk. Ezért 4-4 mátrix összekapcsolhatóvá vált azon az alapon, hogy ezek csak formai reprezentációjukat tekintve különböznek egymástól. Ezentúl azonban úgy gondolták Rapoporték, hogy attól is eltekinthetünk, hogy melyik játékos szemszögéből "nézzük" a játékot, vagyis ugyanazt a játékot kapjuk, ha a sorjátékos (Ego) vagy az oszlopjátékos (Alter) felől tekintünk a preferenciamátrixra. Ez szimmetrikus játékok alapján mindenképpen igaz, de másra nem. Rapoporték végülis 12 szimmetrikus és 66 aszimmetrikus játékot különítettek el.

Ugyan kevesebb játékra figyelt, viszont a legtöbb játékleírást Steven J. Brams adta meg egy 1994-ben megjelent könyvében. Brams kizárta az elemzésből azokat a játékokat, amelyek nem konfliktusosak (amikor a kölcsönösen legjobbnak rangsorolt kimenet a játék végeredménye), így csak 57 játékot vizsgált [74], viszont nagyon értékes jellemzéseket adott a vizsgált játékokra.

A közel negyven éves konszenzust rúgta föl 2005-ben David Robinson és David Goforth könyve, amit a 2x2-es játékok topológiájáról írtak [443]. A két szerző 78 helyett 144 játékra kiterjesztve mutatta be azt, hogy milyen topológiai leírás adható a 2x2-es játékokra, illetve milyen általánosabb játékosztályokat lehet elkülöníteni közöttük. Az alapjátékok számosságára vonatkozó eltérés azzal magyarázható, hogy Robinsonék megjelenéséig 66 aszimmetrikus játékkal számoltak azon az alapon, hogy felcserélhetőnek vélték a sor és oszlop játékosokat a preferenciamátrixban. Rapoporték szerint azonban az eltérő nézőpont miatt igenis van különbség a játékok értékelésében, ezért meg kell különböztetni ezeket egymástól. Én is elfogadom ezt az értelmezést, mert a Rapoporték által feltételezett ekvivalencia csak matematikai értelemben tartható, társadalmilag nem. Közel sem mindegy ugyanis, hogy egy aszimmetrikus helyzetben melyik oldalán állunk a "dolognak". Márpedig – a cselekvéselmélet részeként – a játékelméletet is úgy kell felfognunk, mint ami az Ego, a fokális ágens cselekvését akarja megérteni, ami itt azt jelenti, hogy egyáltalán nem mindegy, hogy Egóról vagy Alterrről beszélünk, tehát a játékosok nem cserélhetők fel.

A játékokat több szempont szerint jellemezhetjük. Figyelhetünk arra, hogy:

[1] A játékosok megragadására sokféle jelölési konvenció létezik, amelyek közül olykor használom a 'Sor' és 'Oszlop' játékos megnevezést.

- hányan játszák a játékokat (lehet két- vagy többszereplős a játék)
- hány döntési alternatíva van a játékosok előtt (lehet két- vagy többalternatívás játék)
- hányszor játszanak egymással a játékosok (lehet egyszer vagy többször játszott, tehát egyszeri vagy ismételt a játék)
- azonosan értékelik-e a játékosok a játékot vagy sem (lehet szimmetrikus vagy aszimmetrikus a játék)
- a preferenciarendezések szigorú rendezésen alapulnak-e vagy sem, vagyis megengedhető-e az, hogy két döntési alternatívát azonosnak minősítsen egy játékos, esetleg egy alternatívát ne értékeljen (lehet szigorúan rendezett vagy nem szigorúan rendezett a játék)
- ismerik-e egymást a játékosok vagy sem, aminek az ismételt játékok esetében van kiemelt jelentősége (lehet direkt vagy indirekt reciprocitás játék)
- kommunikálhatnak-e a játékosok a döntések meghozatala során és ennek során tehetnek-e egymásnak ígéreteket, elköteleződéseket vagy sem (lehet nem-kooperatív, "ígéretmentes" vagy kooparatív, "ígéretes" a játék[2])
- egyszerre döntenek-e a játékosok vagy egymás után, egyik a másik döntésének ismeretében (lehet szinkron, szimultán, statikus vagy aszinkron, szekvenciális, alternáló, dinamikus a játék, de ide tartoznak a méltányosságelméletekhez használt alkujátékok is)

A társas értékorientáció elmélete olyan eszköztárat fejlesztett ki, amely alkalmas a játékelméleti szituációkba került játékosok valós szociális preferenciáinak feltárására (↬236.o.). Erre történetileg azért volt szükség, mert a játékok interdependens mivolta miatt nem lehetett kísérleti eszközökkel vizsgálni a résztvevők tényleges motivációit, mivel az interdependencia ténye "elfedte" a motivációkat. Ezek a kísérletek számszerűen is kimutatták, milyen arányban vannak a társadalomban különböző társas beállítottságú emberek. Ez a tudás azonban a játékelmélet keretei között kevéssé érvényesült, részben azért, mert a játékelmélettel foglalkozó kutatók nem figyeltek a máshol keletkezett eredményekre, részben azért, mert a játékelméleti keretek leszűkítették, megnehezítették a szociális preferenciákra vonatkozó tudás alkalmazásának lehetőségeit. A játékelmélet magjában levő kétszemélyes, kétalternatívás játékokban például csak kétféle döntést lehet hozni a játékosoknak, így az elméletileg lehetséges nyolcféle szociális beállítódás közül igazából csak a kooperácó és a versengés típusait használták, használják a játékelmélet művelő.

A játékelmélet középpontjában a játékosok helyzetértékeléseit rögzítő preferenciamátrix áll. A játékok dinamikáját, belső logikáját, a végeredmény meghatározottságát az adja, hogy a játékok lehetséges kimeneteire vonatkozó egyéni rangsorolások milyen viszonyban állnak egymással. A kulcs tehát a preferencia fogalma. Ebből fakadóan, bár nem szokás hangsúlyozni, a játékelmélet formális leírásához szükség van a preferencialogikára. Emellett bizonyos játéktípusok jellemzéséhez használni kell deontikus logikai operátorokat is, tehát a deontikus logikát is fel kell venni a játékelmélet megragadására alkalmas nyelvcsaládba.

(IL.0) $\mathscr{L}_{IL}(\mathscr{L}_{BL}, \mathscr{L}_{DL})$, ahol

$\mathscr{L}_{BL}$	preferencialogika
$\mathscr{L}_{DL}$	deontikus logika

[2]Ezen a ponton sajnos nagyon szerencsétlen a 'kooperatív' minősítés, mert a terminusnak egész más jelentése van itt ahhoz képest, ahogy a játékelmélet egészében szóba kerül a kooperativitás kérdése.

5.1 Magjátékok

A társadalmi kapcsolatok játékelméleti modellezését úgy építhetjük fel, hogy kiindulunk a lehető legegyszerűbb, legkisebb méretű társadalom modelljéből, majd lépésenként egyre több kezdeti feltételt "elengedünk", hogy a társadalom egyre összetettebb, egyre bonyolultabb modelljéhez juthassunk. Ebben az építkezésben a preferenciamátrix van a központban. Kiindulásként rögzítjük, hogy egyszerre érvényesül a kétszereplős, kétalternatívás, egyszeri, szimmetrikus, szigorúan rendezett, szinkron, nem kooperatív játékok feltételegyüttese. Ezeket a játékokat korábban **magjátékoknak** neveztem el [517], és – ahol lehetett – alapos leírást is adtam rájuk (sok példával), ezért itt csak a legfontosabb vonásait mutatom be a játékoknak.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

A játékokat a preferenciamátrix alapján jellemezhetjük. A preferenciamátrix a játékosok alternatívákra vonatkozó rendezéseit fejezi ki, amit számokkal jelölünk. A preferenciamátrix segítségével kifejezhető strukturális kapcsolatokat (kényszereket) ábrázolhatjuk grafikusan is. Kétféle geometriai (topológiai) reprezentációt is figyelembe veszek.

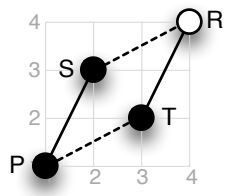
Egyfelől átveszem Robinson-Goforth könyvéből a játékokat azonosító (és egyben jellemző) topológiai ábrákat, amelyek úgy reprezentálják az egyes játékokat, hogy a két játékos preferenciaértékeit egy közös 4x4-es koordinátarendszerben tüntetik fel. A sorjátékos (Ego) értékeit a vízszintes, az oszlopjátékos (Alter) értékeit a függőleges tengely mentén fejezik ki, és az értékek metszetébe egy-egy pontot rajzolnak. Ezután összekötik azokat a pontokat egymással (folytonos, illetve szaggatott vonallal), amelyek egy-egy játékos döntései által összekapcsolódnak, ami másfelől egyfajta döntési kényszert is jelent a másik játékos számára [443].

4,4	2,3
3,2	1,1

$\rightsquigarrow$

R	S
T	P

$\rightsquigarrow$



Az ábrában az egyensúlyi kimenet(ek) létezését a fehér közepű köröcske jelzi. A játékok képi azonosítására dolgozott ki Pierre Dragicevic egy olyan vizuális jelölési rendszert, amely azzal a megoldással különíti el egymástól az egyes játékokat, hogy a preferenciamátrix négy cellájában szereplő preferenciaértékeket összeköti egymással növekvő sor-

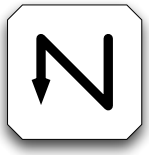
rendben, és így mindig a játékra jellemző, egyedi, vonalszerű alakzatot kap [134]. Dragicevicnek ezt az egyszerű, de látványos megoldását (még tovább egyszerűsítve, a "végletekig stilizálva") szintén alkalmazom az egyes játékok jellemzésekor.

3,3	1,4
4,1	2,2

$\rightarrowtail$

3	1
4	2

$\rightarrowtail$



A játékok leírásakor feltüntetem még a preferenciamátrixukat, valamint a széles körben elterjedt cellamegnevezési konvenció alapján kialakítható SRPT-rendezéseiket. A preferenciamátrix celláit megnevezhetjük és azonosíthatjuk az S, R, P, T betűk segítségével – sorrendben a Sucker, Reward, Punishment, Temptation (szívás, jutalom, büntetés, kísértés) kifejezések kezdőbetűit használva. Mivel a játékokat egyértelműen jellemzi az, hogy a négy cellához milyen preferenciaértékek vannak rendelve, tehát a négy betű sorrendje egyedi módon megragadja a játékot, így megfelel azonosítónak is.

A 12 szimmetrikus játék között mindegyiknek van egy párja, amellyel szorosabb kapcsolatban van a többiekhez képest. Amennyiben a kisátlóban szereplő értékeket felcseréljük, miközben a főátló értékeit változatlanul hagyjuk, akkor mindig az adott játék komplementer párjához jutunk. Robinson és Goforth antijátéknak nevezi az így kapott játékokat, amit én nem tartok szerencsés megnevezésnek.[3] A komplementer játékokat úgy is lehet értelmezni, hogy a kisátló preferenciaértékeinek felcserélése voltaképpen azt fejezi ki, hogy a játékot két olyan játékos játssza, akik az alapjáték játékosaihoz képest ellentétes ágens-beállítódással rendelkeznek (azaz a játékban két egoista beállítódású helyett két altruista játékos van).

Mivel az alapjátékokat máshol korábban alaposan elemeztem [517], itt nem fogok sok példát adni a játékokra, csak akkor, ha valamilyen fontos vonás érzékeltetése végett van szükség egy új példa bemutatására, elemzésére, vagy olyan játékot elemzek, amelyre a korábbiakban nem tudtam példával szolgálni.

Ahhoz képest, hogy milyen magas mértékben formalizált tudomány, a játékelméleti terminológia több szempontból is meglepően pongyolának mondható. Az egyes játékok elnevezési logikája messze nem mondható konzisztensnek. Zavaró, hogy ugyanazt a játékot teljesen eltérő néven hivatkozzák meg a különböző kutatók (egy-egy játéknak sokszor négy-öt neve is van). A játékokhoz kapcsolható szerepekhez olykor negatív, máskor pozitív hangulatú neveket kapcsolnak, és bár többféle elnevezést is intuitívnak érezhetünk, gyakran félrevezetőek, zavaróak lehetnek ezek az eltérő konnotációk. A fogolydilemma játékban például a két "ellentétes" szerepre lehet a 'kizsámányoló' vagy a 'mártír', illetve a 'potyautas' vagy a 'balek' neveket akasztani, a gyáva nyúl játékban lehet valaki 'gyáva nyúl' vagy 'fenegyerek', illetve 'önkéntes' vagy 'haszonleső'. Bármennyire is jó lenne – a magasabb konzisztencia reménye – miatt egységesebb terminológiát bevezetni, nem lehet, nem érdemes eltérni a már kialakult szóhasználati gyakorlattól.

A játékelméleti szakirodalom nagy részében központi szerepet tulajdonítanak annak, hogy hogyan alakul(hat) ki a játékok végeredménye, egyensúlyi helyzete. Én a játékelmé-

[3] Például azért nem, mert a koordinációs játék antijátékát anti-koordinációs játéknak kéne nevezni, ami azért nem lenne jó, mert ezen kifejezéssel már régóta a koordinációs helyzetek egyik típusát írják le.

let tárgyalása során sokáig mellőzhetőnek tartom ezeket a fogalmakat, oly' annyira, hogy a könyvemben gyakorlatilag nem is használom őket. Úgy érzem azonban, hogy a "tisztesség" kedvéért legalább említést kell tennem róluk a bevezetőben. Két fogalomról kell feltételenül szólnom. Az egyik a Pareto-optimalitás fogalma. Egy játék valamely kimenetét Pareto-optimálisnak nevezzük, ha nem létezik egy olyan másik elosztás, amely szintén megvalósítható, amelyben senki sem jár rosszabbul és legalább egy szereplő jobban jár, mint a szóban forgó elosztásban. A Nash-egyensúly fogalmával pedig azt az állapotot írhatjuk le, amikor egy játékosnak nem érdemes változtatnia saját stratégiáján, ha a másik (többi) játékos sem változtat az aktuális döntésén, mert nem járna jobban a változtatással. Szokták ezt úgy is mondani, hogy minden játékos a legjobb választ adja a többiek lépésére. Egy játékos domináns stratégiájának nevezik azt a legjobb lépést, ami független a többiek stratégiáitól. Itt kell még megjegyeznem, hogy a stratégia terminusát sajnos túlhasználja a játékelmélet, és a különböző játéktípusok leírása során egyre több jelentésréteget húznak rá erre a terminusra. A kezdetekben a stratégia mint egyszeri döntés, lépés, cselekvés, aztán mint cselekvésekre vonatkozó döntéssorozati szabály, végül az egyéni cselekvésekre és a közösségi értékelésre vonatkozó szabálykettősség. Ezen a többértelműségen segíteni már csak azzal lehet, ha reflektálunk rá, egy új terminológia alkalmazásának itt sem lenne értelme.

A játékelmélet egy komoly hiányossága, hogy nem tudja kezelni azt a tényt, hogy a játékosok többfélék lehetnek. A személyiség, neveltetés, kapcsolati tőke, társadalmi beágyazódás másfajta személyiségtípust alakít(hat) ki, és ezekből a különbségek megnyilvánul(hat)nak a döntési helyzetekben is. A kockázatkerülő, kockázatkövető, kockázatdomináns. kifizetésdomináns személyiséggel rendelkező játékosok másként játszhatják le ugyanazt a játékot a többiekhez képest. A teljesen racionális döntéshozó létezését a játékelmélet keretei között sem feltételezhetjük, tehát mindazt, amit leírtam a racionális és aracionális cselekvés viszonyáról (↫222.o.), azt itt is figyelembe kéne venni. Erre azonban nem vagy csak korlátozott mértékben képes ez a diszciplina. Pozitív példaként utalok itt arra a kísérletre, amelyben a csomagolás hatását lehet bemutatni – játékelméleti környezeten belül [492, 33]. A kísérlet résztvevőinek az alábbi táblázat alapján kellett döntenie arról, hogy kooperálnak-e vagy dezertálnak. Ha a játékos a C gombot nyomja meg (kooperál), akkor magának egyet ad, a másiknak kettőt, míg a versengést jelentő D gombbal magának kettőt ad, a másiknak nullát.

	magának ad	másiknak ad
C	1	2
D	2	0

Ha ebből az egyszemélyes döntési helyzetből kétszemélyes játékot csinálunk, akkor azonnal kiderül, hogy ez a helyzet a fogolydilemmával egyezik meg, amelyben a kétszeres dezertálás a várható végeredmény. Ennek ellenére ezt a játékot a kísérletekben sokkal kooperatívabban játsszák, mint az eredeti fogolydilemma játékot. Ennek pedig az az oka, hogy ebben a "csomagolásban" nyilvánvaló, hogy csak a másik játékostól jöhet az igazi nyereség.

Korábban esetszétválasztás módszereként hivatkoztam rá [517], amire alkalmazni lehet még a **függetlenségi axióma**, a lényegtelen alternatívák függetlenségének elve, a feltételes valószínűségek függetlenségének elve, biztos dolgok elve (sure thing principle) kifejezéseket is [243]. A játékelmélet szempontjából vett jelentősége mindezeknek az, hogy a

játékokkal szembeni elvárásként mindig érvényesítenünk kell azt, hogy a döntési teret fel kell tudni bontani egymástól független világállapotokra. Amennyiben ez nincs így, akkor a játékokra könnyen téves magyarázatokat adhatunk. Ennek szemléltetésére nézzük meg a következő példát [243]. Tegyük fel, hogy a vizsgája előtt egy diáknak két lehetősége van (tanulni vagy nem tanulni), és a vizsga kimenete is kétesélyes (vagy átmegy a diák vagy nem). Ezt a döntési teret felrajzolhatjuk a következő mátrixba.

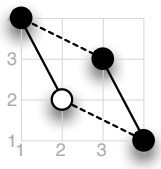
	tanul	nem tanul
átmegy a vizsgán	2	3
megbukik a vizsgán	0	1

A játék elemzéséből az jöhetne ki, hogy nem érdemes tanulnia, hiszen mindkét lehetséges kimenet (megbuki vagy átmegy) esetén jobban járna. Ez az intuíciónkkal ellentétes eredmény annak köszönhető, hogy ebben a játék(leírásban) az egyes kimenetek nem függetlenek egymástól.

*

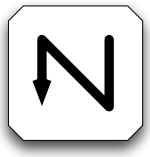
A bevezető megjegyzések után következzen a tizenkét magjáték bemutatása.

Fogolydilemma játék



3,3	1,4
4,1	2,2

TRPS



a játék száma	
Rapoport-Guyer:	12
Brams:	32
Robinson-Goforth:	111

A játékelmélet története is azt mutatja, és sok kutató elméletileg megalapozott véleménye szerint is a fogolydilemma áll a szimmetrikus játékok (sőt, többek szerint, az egész játékelmélet) középpontjában. Sok szempontból osztom én is ezt a véleményt, de azt azért meg kell jegyeznem, hogy van egy olyan elfogultság ebben a minősítésben, amely a konfliktusos jelleg fakad. Amikor John Platt a társadalmi csapdákról szóló, híres, 1973-as cikkében "metaforát keres" a fogolydilemma számára, akkor Adam Smith 'láthatatlan kéz' (invisible hand) fogalmát állítja szembe az által kreált 'láthatatlan ököl' (invisible fist) kifejezéssel [414]. Az "áldást hozó láthatatlan kéz" fogalmával Adam Smith azt az állapotot ragadja meg, amikor az egyéni érdekek megegyeznek és a közösségi érdek nem sérül. Ezzel áll szemben John Platt szerint a "láthatatlan ököl, amely lesújt a közjóra", amikor az egyéni érdekek teljesen ellentétesek, és a közösségi érdek is sérül. Mindez azért érdekes itt, mert a láthatatlan kéz harmóniát kifejező fogalmához is rendelhető egy játék (a harmóniajáték), amit persze a játékelmélet szintúgy mostohán kezel, mivel a konfliktus hiányában, a teljes harmónia miatt "nincs benne semmi érdekes".

A fogolydilemma legfontosabb tanulsága, hogy vannak olyan helyzetek, amelyekben a kétszeres dezertálás lesz a játék várható kimenete, mert a bizalmatlanság és/vagy erős

egoista beállítódás miatt senki meri/akarja a kooperációt felvállani. Ekkor viszont mindenki kisebb nyereséggel jön ki a játékból a kettős kooperáció esetén elérhető kifizetéshez képest. A játék várható szuboptimális kimenete adja a játék társadalmi dillema jellegét.

A fogolydilemma játék eredeti meséje, bár életszerű, de nem mondható túl gyakorinak, viszont számtalan társadalmi helyzetről lehet állítani, hogy ebbe a játékosztályba tartozik. A közös társadalmi javak, a közlegelők egyéni túlfogyasztása (szemetelés a köztereken, az ingyen parkolóhelyek "feltöltése"), a potyautasság jelenségének számtalan megnyilvánulása (bliccelés a tömegközlekedésben, lakótelepi lakásokban a fűtés "lopása" a szomszédoktól), a piaci csökkenő árspirál jelenségének kialakulása, vagy az árucsere helyzetek mélyén mindig is ott rejtőzködő késztetés a partner "átverésére" – olyan példák, amelyek jelzik, hogy a társadalmi cselekvések minden szegmensében felbukkannak fogolydilemma-szerű konfliktusok [517].

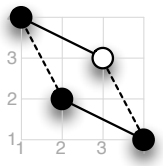
A fogolydilemma szereplőire több nevet is aggattak: potyautas, balek, mártír, kizsákmányoló. Azt már említettem korábban, hogy ezeknek vagy pozitív vagy negatív konnotációja van, ami kicsit nehezíti az egységes értelmezést. Érdemes viszont arra is felhívnom a figyelmet, hogy ezek a névadások abban az értelemben sem pontosak, hogy olyan fogolydilemma-helyzetek olyan feloldásaira utalnak, amelyek – a szokásos értelmezés szerint – atipikusak. Mind a potyautas-balek, mind a mártír-kizsákmányoló szerepmegnevezés ugyanis olyan helyzetre utal, amelyben az egyik játékos dezertált, a másik kooperált. Ez a többszereplős fogolydilemma esetében sokkal inkább elfogadható, mint a kétszereplős játékban.

A fogolydilemmában benne rejlő **társadalmi dilemma** vagy John Platt – az 1970-80-as években népszerű – kifejezésével élve a vele reprezentálható **társadalmi csapdák** [414] szuboptimális megoldása megfogta a kutatókat, és nyugodtan mondható, hogy ez a játék lett a játékelmélet kiindulási, viszonyítási pontja. Ez megnyilvánult abban is, ahogy a játékok jellemzéséhez kapcsolható elnevezések általánossá váltak a szakirodalomban. Ezek a fogolydilemma esetére találó nevek sok esetben pontatlanok, néha egyenesen félrevezetőek vagy értelmetlenek voltak, mégis megmaradtak közös terminológiai alapnak. A kifizetési és preferenciamátrix négy celláját a fogolydilemma logikájának megfelelően nevezték el – egyébként valóban nagyon szemléletesen – kísértésnek, jutalmazásnak, büntetésnek és szívásnak (temptation, reward, punishment, sucker), és ezek az állapotmegnevezések más játékok esetében tartalmi értelemben használhatatlanok voltak.

Ugyancsak a fogolydilemma logikájának felelt meg az a megoldás is, ami a játékosok előtt álló alternatívákat próbálta meg a lehető legnagyobb mértékben általánosítani. Ha a játékosok cselekvési lehetőségeit a végletekig absztraháljuk, akkor az együttműködés, kooperálás (C), illetve a nem-együttműködés vagy másként a dezertálás (D) dichotómiájához jutunk. Mindkét játékos kooperálhat vagy dezertálhat, így a játéknak négyféle kimenete lehetséges: kölcsönös kooperálás (CC), kölcsönös dezertálás (DD), az egyiknek balekság, a másiknak potyázás (CD), illetve fordítva (DC). A fogolydilemma-helyzetek tanulsága az, hogy hiába lenne a kölcsönös kooperáció mindenkinek előnyösebb a kölcsönös dezertáláshoz képest, a bizalmatlanság miatt a dezertálás (a potyautasság, a másik kizsákmányolásának szándéka) válik uralkodó motívummá. Ezt az az összefüggést fejezi ki a következő egyenlőtlenség (preferenciarendezés): DC > CC > DD > CD (amit a cellaelnevezési konvenciónk szerint felírhatunk más alakba is: T > R > P > S vagy ezt rövidítve: TRPS). Ez az absztrakció és az erre épülő jellemzések hasznosak, sok játék esetében teljesen adektvát módon használhatók, de néhány játék esetében nagyon ne-

héz a játékosok előtt álló alternatívákat a kooperativitás kontra versengés dimenziójában értelmezni – ilyen például a koordinációs játék (↬283.o.).

Mártírjáték



3,3	4,1
1,4	2,2

SRPT

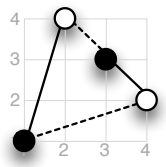
a játék száma	
Rapoport-Guyer:	9
Brams:	9
Robinson-Goforth:	166

A mártírjáték a fogolydilemma komplementár párja, aminek a kétszeres kooperálás a várható kimenete. Szokták holtpontjátéknak (Deadlock) is nevezni [421], Robinsonék az 'Anti-PD' nevet használják, Ken Binmore pedig a 'Prisoner's Delight' nevet húzta rá [56, 131]. A mártírjáték legfontosabb vonása az, hogy egy fogolydilemma-szerű szituációból képes kihozni a "közösségi optimumot" a játékosok számára. Ezzel egyben arra is például szolgál, hogy nemcsak az önzés nem mindig racionális, de előfordulhat az is, hogy az altruizmus nagyobb társadalmi nyereséget biztosít az önzéshez képest [517]. A játék alapmeséjét úgy értelmezhetjük, hogy a fogolydilemma-helyzetbe egy szerelmespár kerül, akiknek első számú preferenciája a szerelmüknek kedvezőbb kimenet biztosítása lesz (vagyis altruista beállítódás mentén döntenek), és ebből fakadóan mindketten a kooperálás mellett döntenek. Nem a kooperáció mint olyan önértéke vagy érdekkonformitása miatt, hanem azért, mert a kooperáció logikája mentén lehet a legjobb kimenetet biztosítani a partnerük számára.

A mártírjáték megnevezés egyszerre találó és kicsit félrevezető. Azért lehet ezt mártírjátéknak hívni, mert amikor a két altruista beállítódású személy meghozza saját egyéni döntését, akkor valóban vállalják a mártíromságot, a teljes önfeláldozást a másikért. Mindkét szereplő beállítódására alkalmazható tehát a mártír minősítés, ám ez a kettős mártíromság a játékból végülis a társas optimumot hozza ki, ezért hangzik kicsit furcsán a mártírjáték elnevezés.

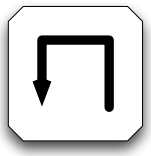
A fogolydilemmánál is el lehetett volna mondani, de itt is ugyanolyan releváns lehet felhívni a figyelmet egy fontos kísérleti eredményre. Robert Frank, Thomas Gilovich és Dennis Regan 1993-as tanulmányukban leírták, hogy amikor fogolydilemma játékot játszattak egyetemistákkal, akkor a végzős közgazdászhallgatók között sokkal nagyobb arányban fordult elő a kölcsönös dezertálás, mint más hallgatók körében, amiből azt a következtetést vonták le, hogy a dezertálás – legalább részben – tanult lehet [174]. A mártírjáték mint a fogolydilemma komplementer párja arra ad lehetőséget, hogy a játékelmélet segítségével is meg tudjuk ragadni azt a – társas preferenciák elmélete által vizsgált – jelenséget, hogy az emberek nem kizárólag egoista beállítódásúak. Ha két altruista játékos belekerül egy fogolydilemma szituációba, akkor az ő altruista beállítódásuk a játékot "átértelemezi", aminek eredményeként ugyanazt a helyzetet egy másik játékként látják, és így ők már mártírjátékot fognak játszani.

Gyáva nyúl játék



3,3	2,3
4,2	1,1

TRSP



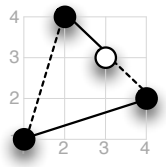
a játék száma	
Rapoport-Guyer:	66
Brams:	57
Robinson-Goforth:	122

Ha a fogolydillemmában a két legkisebb preferenciaértéket felcseréljük egymással, akkor jutunk a gyáva nyúl játékhoz. Ez azért fontos, mert ezáltal a kétszeres dezertálás lesz mindkét játékos számára a legrosszabb alternatíva, és emiatt a játékosok választása eltolódik a kétszeres dezertálás felől a kisátlóban levő két cella felé. A játék ezáltal döntetlen jellegű lesz, vagyis két ekvivalens egyensúlyi kimenettel fog rendelkezni, amelyek közül nem lehet előzetesen kiemelni egyiket sem, vagyis nem lehet megjósolni, melyik kimenet lesz a játék tényleges végeredménye. Ez a minőség **koordinációs problémaként** tekinthető, amikor az a játék alapkérdése, hogy milyen módon lehet úgy koordinálni a játékosok döntését, hogy ugyanazt a kimenetet válasszák mindketten. Lesz még több is ilyen, a tizenkét magjáték fele koordinációs helyzetként értelmezhető.

A gyáva nyúl játékra rengeteg valós példát lehet találni. A "névadó" filmbeli történet az egymással szembe (vagy a szakadék felé) haladó galerivezérek versengéséről szól, amikor az a kérdés, ki fog előbb kitérni a másik útjából (vagy ki ugrik ki előbb a szakadékba tartó kocsiból). Aki tovább kitart, az lesz a fenyegyerek, a másik pedig a gyáva nyúl. De ebbe a játékosztályba tartozik a hidegháború korszakát jellemző kölcsönös elrettentés politikája, vagy mindazok a hétköznapi történetek, amikor az a kérdés, hogy ki vállalja fel az önkéntes szerepét (ki az, aki leviszi a szemetet, aki elmosogat, aki kitakarít, aki lemegy a boltba. aki felkapcsolja a villanyt stb. [517]). Az ilyen helyzeteknek az a legfontosabb vonása, hogy egyvalakinek kell elvállalnia valamely cselekvés végrehajtását, és azzal a másiknak nagy jót tesz. E közös vonás felismerését – magyar nyelven – kicsit nehezíti a játék 'gyáva nyúl' megnevezése, illetve az ebből származó 'fenegyerek' és 'gyáva nyúl' szerepnevek használata. A gyáva nyúl játékban az "önkéntes" játékos negatív minősítést kap, míg sok más játékleírásban az valódi értelemben vett önkéntes szerep pozitív konnotációval rendelkezik (már maga az 'önkéntes' jelző is ilyen). Ezt a játékot egyébként – a hozzá kapcsolt meséktől függően – nevezik még héja-galamb (hawk-dove) vagy hótorlasz (snowdrift) játéknak, valamint olykor elhasználódási háborúnak (war of attrition).

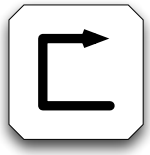
A játék lényegét az adja, hogy mindegyik szereplő a másikra vár, pontosabban arra, hogy ki adja fel előbb, ki nem tud tovább kitartani a másikkal szemben. Ez a logikája a piaci alkufolyamatoknak is, amikor a potenciális vevő és eladó vételi és eladási ajánlatai fokozatosan közelednek egymás felé, és az anagy kérdés, végül is ki lesz az, aki nem bírja tovább és elfogadja a másik ajánlatát, vagyis nem próbál tovább alkudni. Ennek a problémának a leírására dolgozta ki Thomas C. Schelling az alkuelméletet, ami nagyban hozzájárult ahhoz, hogy később megkapja a Nobel-díjat [462].

Önkéntes játék



3,3	4,2
2,4	1,1

SRTP

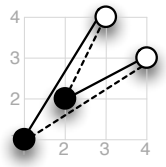


a játék száma	
Rapoport-Guyer:	7
Brams:	7
Robinson-Goforth:	155

Az önkéntes játék a gyáva nyúl játék komplementer párja, és legfőbb vonása az, hogy a gyáva nyúl játék döntetlenjét képes feltörni, és ezzel az optimális közösségi megoldást (a második legjobnak preferált kimenetet) képes nyújtani a játékosok számára. A játék kevés figyelmet kap a szakirodalomban. Az ok ugyanaz, mint a mártírjáték esetében: sokak számára a konfliktusmentessége miatt érdektelennek tűnik az egész.

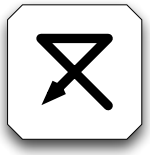
Robinson és Goforth anti-gyávanyúl játéknak (Anti-Chicken) nevezi a játékot. Az általam preferált önkéntesjáték megnevezés arra utal, hogy ezt a játékot a gyáva nyúl játék keresett/remélt önkéntesei játsszák. Amíg azonban a gyáva nyúl játékban az önkéntes magára vállalja a második legrosszabb kimenetet biztosítva, hogy ezzel a partnere a legnagyobb nyereménnyel zárhassa a játékot, addig itt két önkéntes játékos kerül egymással szembe, akik kölcsönösen a második legjobb eredménnyel szállhatnak ki a játékból. Az önkéntes jelző alkalmazása itt ugyanazzal a problémával ját, mint amit már említettem a mártír jelző jogosságával kapcsolatban a fogolydilemma és a mártírjáték viszonylatában. A gyáva nyúl játék csak annyit képes megmutatni, hogy az általa leírható helyzetekben kellene egy önkéntes, de maga a játék nem “tud” ilyet produkálni, sőt a végtelen másikra várás kimenetét vetíti előre. A gyáva nyúl játék komplementere, az önkéntesjáték viszont valóban arról szól, hogy az altruista beállítódás valóban kiváltja az önkéntes viselkedést mindkét fél részéről, és ezek egymásra hatásaként a társas optimumot tudják kihozni a játékból. Az ‘önkéntes’ minősítés legjobban arra az aszimmetrikus játékra illene rá, amelyben valóban van egy fenegyerek (vagy egy haszonleső) és egy gyáva nyúl (vagy egy önkéntes), de ez nem lehetne sem a gyáva nyúl, sem az önkéntes játék.

Vezérjáték



2,2	3,4
4,3	1,1

TSRP



a játék száma	
Rapoport-Guyer:	68
Brams:	54
Robinson-Goforth:	133

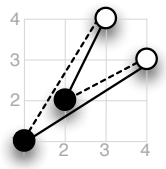
Talán ennél a játéknál (illetve a komplementer párjánál) tapasztalható a legnagyobb kavarodás a nevek haszálatával kapcsolatban a játékelméleti szakirodalomban. A nemek harca megnevezést alkalmazzák erre a játékra és a párjára egyaránt. Bár ez a keveredés önmagában mutatja, hogy valóban értelmes és érdemes a komplementer játékokat úgy felfogni, hogy azokban a játékosok másikra irányuló beállítódásai vannak felcserélve, hiszen

amíg a nemek harca játékban az önző beállítódású játékosok a vezér szerepre törnek, addig az önzetlenek az áldozatvállalásra hajlandó hős szerepét vállalják magukra. Az egyértelműség kedvéért – követve Rapoport javaslatát [432] – erre a játékosztályra a vezérjáték megnevezést fogom használni, míg e játék komplementer párjára a hősjáték nevet alkalmazom. A vezérjátékra magyarul használják még a vezérürü megnevezést is, de ezt én hanyagolnám.

A vezérjáték is koordinációs problémák közé tartozik, akárcsak a gyáva nyúl játék, de annyiban azért más a játék mintázata, hogy a két ekvivalens egyensúlyi kimenet itt a kisátlóban levő két cellában van. Ez annyit tesz, hogy az egyensúlyi állapotok valamelyikét úgy lehet csak elérni (és ezáltal a sikeres koordinációt megvalósítani), ha a két résztvevő eltérő cselekvést hajt végre (az egyik kooperál, a másik dezertál). Ezért van az, hogy ezeket a játékokat valamilyen szerepmegosztás (nemi, családi, foglalkozási munkamegosztás) modellezésére tartják alkalmasnak. A játéknak azért javasolta Rapoport a vezér megnevezést, mert a játékba "kódolt" döntetlent egy dominanciára törekvő játékos tudja feltörni azzal, hogy képes elérni a partnerénél azt, hogy fogadja el őt vezérnek, aki így a játék legjobb eredményét szerezheti meg magának, de ezzel párhuzamosan "képes" biztosítani a második legjobb kifizetést a partnerének. A komplementer párjához képest érdekes és fontos vonása még a játéknak, hogy a kétszeres versengés adja a legrosszabb kifizetést mindkét játékos számára (míg a hősjátéknál ez pont fordítva van, ott a kétszeres kooperáció a legrosszabb).

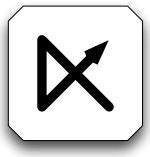
A játékra sok hétköznapi példa hozható. Ide tartoznak azok a helyzetek, amelyekbe az útkereszteződésbe érő, a kanyarodásban egymást zavaró gépkocsivezetők vagy az egy nyomvonalat, két irányból használni akaró közelekedők (sofőrök, gyalogosok) kerülnek, de a vezérjáték írja le azt a helyzetet is, amikor két ember között megszakad a telefonbeszélgetés, és el kell dönteniük, ki hívja vissza a másikat.

Hősjáték



2,2	4,3
3,4	1,1

STRP



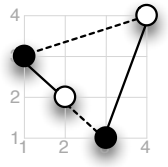
a játék száma	
Rapoport-Guyer:	69
Brams:	55
Robinson-Goforth:	144

A hősjáték a nemek harca, a kultúrák harca, a szerepmegosztást modellező játék altruista beállítódású játékosok által játszott változata (és egyben a vezérjáték komplementere). A mintázata ugyanolyan, mint a vezérjátéké: két egyensúlyi kimenete van. A döntetlent itt akkor lehet feltörni, ha a két altruista hős közül az egyik "felvállalja" az egoista szerepét. Ez a játék idézi elő azt a furcsa helyzetet, hogy a kétszeres kooperálás a legrosszabb mindenki számára.

A játék illusztrálására a klasszikus nemek harca történetét érdemes legelőször meghivatkozni. A játék leírása szerint a férjnek és feleségnek külön-külön kell eldöntenie, hogy hova menjenek estére, boxmeccsre vagy színházi előadásra. A két legrosszabb kimenet az lehet, ha kooperálnak, vagyis odamennek, amit a partnerük szeret, vagy dezertálnak, vagyis odamennek, amit maguk szeretnek, hiszen mindkétszer elkerülik egymást. A két

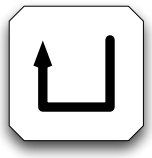
legkívánatosabb kimenetek egyikét csak akkor érhetik el, vagyis csak akkor találkozhatnak, ha egyikük dezetrál, míg a másikuk kooperál. Azt persze ez a játék nem tudja megmondani, hogy ki lesz az egyik vagy a másik szereplő, sőt, a játék pont azt állítja, hogy ezt sosem lehet a játék (pontosabban csak a játék) alapján tudni. A gazdaság területéről származó példaként lehet említeni azt a helyzetet, amikor két szabványt használó vállalatnak kellene megegyezniük egy közös szabvány követésében de ide sorolható még az a helyzet is, amikor két tudósnak kell megegyeznie arról, melyikük anyanyelvén írják meg a közös cikkük egy-egy fejezetét, vagy amikor két cég egy hosszú szerződéselőkészítési folyamat végén meg kell egyezzen a kötbér nagyságáról.

Szarvasvadászat játék



4,4	1,3
3,1	2,2

RTPS



a játék száma	
Rapoport-Guyer:	61
Brams:	-
Robinson-Goforth:	322

A szarvasvadászattal kezdődő hatos sorozat tagjai abban mutatnak közös vonásokat, hogy mindegyik játékosztály preferenciamátrixában a kétszeres kooperációról elnevezett R-cellában érheti el mindkét játékos a legnagyobb nyereményt. Emiatt ez a cella mindegyik játékban az egyik lehetséges egyensúlyi kimenetet adja (amely mellé három esetben még egy másik egyensúlyi kimenet is "bekerül").

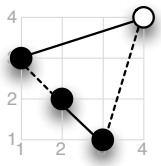
A szarvasvadászat játékot több szerző szerint a fogolydilemmából lehet "származtatni" úgy, hogy a két legmagasabb preferenciaértéket kicseréljük egymással [494]. Ebben az esetben mindkét játékos számára a kölcsönös kooperálással lehet megszerezni a játék legnagyobb nyereményét. Bár ebben a játékban a dezertáló fél még mindig alaposan "ki tudja használni" a kooperáló játékost, hiszen a második legnagyobb kifizetést szerezheti meg, miközben a kooperáló partnere a legrosszabb nyereményhez juthat csak hozzá, a játékot mégsem ez az asszimmetria jellemzi igazán. Első látásra talán meglepő, de azt a lehetőséget, hogy a kölcsönös kooperálással mindkét játékos megszerezhetné a legmagasabb kifizetést, nem feltétlen ragadják meg a játékosok. Bármennyire is csábítónak tűnhet ezt feltételezni, mégsem ez a kimenet lesz játék végeredménye. A játéknak két egyensúlyi kimenete van, mert a kétszeres dezertálás is Nash-egyensúlynak felel meg. A kölcsönösen legjobb kifizetéshez szükséges a kölcsönös kooperálás, de ha a kooperálás nehéz, költséges a felek számára, akkor ezzel szemben "kényelmes" alternatívát jelenthet mindkét játékos számára az, ha a nagyobb erőfeszítés nélküli, biztos alternatívát részesíti előnyben (a "jobb ma egy veréb, mint holnap egy túzok" mondást beigazolva). Ezt a logikát persze lehet úgy is értelmezni, hogy a felek arra játszanak, hogy a másik kooperálása mellett ők dezertáljanak (hiszen ekkor a második legjobb kifizetéshez juthatnak hozzá). Ez a másik kárára szerezhető "könnyű nyeremény" elfogadható teszi a legnagyobb nyereményről való lemondást, ezért a játékban a kétszeres dezertálás is egyensúlyi helyzetté válik. Ebben az értelemben és ezen a ponton még mindig a fogolydilemma "hatása" érezhető. Lehet azonban egy harmadik értelmezést is adni a szarvasvadászat játék származtatására vonatkozóan. Kiindulhatnánk ugyanis a harmóniajátékból (↪284.o.),

amelyben ha megcserélnénk a két legkisebb értéket (az 1-t és 2-t), akkor a szarvasvadászat játékhoz jutnánk. Ezt úgy lehetne interpretálni, hogy ha a kölcsönösen legjobb kifizetést biztosító harmóniajátékban a kétszeres dezertálás kifizetését növeljük "eggyel" (már nem a legrosszabb, hanem a második legrosszabb lesz a jutalom), akkor ez elég lehet ahhoz, hogy a kétszeres dezertálás"összemérhető" és vállalható alternatíva legyen a kétszeres kooperáláshoz képest mindkét játékos számára. Ez az elmozdulás pedig kibillentheti a játékot az egyszeres egyensúlyi helyzetből és két Nash-egyensúlyi pontot eredményezhet.

A példák között a "névadó" szarvasvadászat játék mellett említeni érdemes a matrózlázadás történetét, ami egyébként minden autokrata munkahelyi vezetővel szembeni fellépés modelljeként is értelmezhető egyben [517].

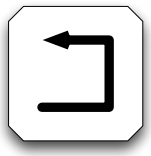
Brian Skyrms egy egész könyvet szentelt annak a tézisének, miszerint igazából a fogolydilemma áll a játékok központjában, belőle vezethető le a fogolydilemma, és nem fordítva, ahogy sok más szerző ezt gondolja [494]. Robinson és Goforth említi, hogy bizalmi dilemmaként (trust dilemma), illetve koordinációs játékként (coordination game) is felfogható a játék [443], Amartya Sen 1967-es cikke óta nevezik olykor biztosítási játéknak (Assurance Game [475]), Sam Perlo-Freeman pedig Deterrer játéknak (Deterrer) nevezi [408].

Háztulajdonos játék



4,4	3,1
1,3	2,2

RSPT

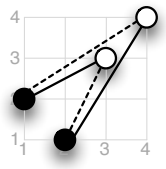


a játék száma	
Rapoport-Guyer:	5
Brams:	-
Robinson-Goforth:	355

A játék a szarvasvadászat komplementere, és ezt is úgyanúgy többféle módon származtathatjuk, mint ahogy azt a szarvasvadászat játék esetében már bemutattam. A Robinson-Goforth könyv azt a társadalmi helyzetet kevert harmóniajátéknak (Mixed Harmony) nevezi. A játéknak azért adtam a 'háztulajdonos' nevet, mert azt a logikát írja le, ami a vigalmi negyedek kialakulásának hátterében áll. Kétszereplős játékként kicsit életszerűtlen ugyan a történet, de még így is jól visszaadja azokat a belső kényszereket, amelyek a játék preferenciamátrixában meghúzódnak. Van két egymás mellett fekvő ház, a tulajdonosaik dönthetnek úgy, hogy bordélyházat csinálnak a házából vagy kiadják bérlőknek a házaik lakásait. A legvonzóbb kimenet mindkettejük számára az, ha bordélyházat nyitnak, mert sokat kereshetnek vele, nem zavarják egymást, sőt a két háztömbből álló "mini vigalmi negyed" kialakulásával még nagyobb forgalomra is számíthatnak, hiszen nagyobb híre megy a szolgáltatásuknak. Ha csak az egyikük vág bele ebbe az üzletbe, akkor neki anyagilag még mindig nagyon megéri (ez lesz a második legjobb opció számára), de a másik háztulajdonosnak ez válik a legkevésbé kívánatos lehetőséggé, mivel a szomszéd bordélyház zavarná a háza lakásait bérbe vevő lakosokat, tehát alacsonyabb lakbért tudna tőlük beszedni. Ehhez képest a második legrosszabb alternatíva az lenne, ha mindketten bérlőknek adnák ki a lakásokat, mert akkor a jól szituált környéken magasabb lakbért kérhetnének a bérlőiktől. Ez a logika egyfajta szegregációs logikaként is értelmezhető,

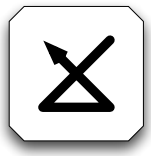
és ezzel magyarázhatjuk hasonló profillal rendelkező üzlet-, buli- vagy művésznegyedek kialakulását a nagyobb városokon belül.

Koordinációs játék: ügyfél



4,4	1,2
2,1	3,3

RPTS



a játék száma	
Rapoport-Guyer:	63
Brams:	-
Robinson-Goforth:	333

Ahogy azt már korábban említettem, ennél a játéknál (illetve a komplementer párjánál) valójában nem értelmes kooperálásról vagy dezertálásról beszélni, itt ugyanis pusztán arról van szó, hogy két cselekvési alternatíva közül érdemes ugyanazt választania mindkét játékosnak. A cselekvési lehetőségeknek abban az értelemben nincs értékvonzata, hogy az egyiket kooperálásnak, tehát kívánatosnak, a másikat dezertálásnak, tehát nem kívánatosnak tartanánk, mi, a játék megfigyelői. Mégis használom a kétféle minősítést, de csak azért, mert így a már megszokott módon tudok az egyes cellákra, illetve kimenetekre hivatkozni.

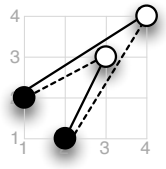
A játék két legmagasabb kifizetése a kétszeres kooperálással vagy kétszeres dezertálással érhető el. Ezért nevezik ezt koordinációs játéknak. Ennek a játéknak is két egyensúlyi helyzete van. A kétszeres kooperáláshoz szükséges kooperálás választása azzal is együtt jár, hogy ha mégsem jön össze a koordináció, akkor a legkisebb kifizetéshez lehet csak hozzájutni, míg ha a másik egyensúlyi állapot eléréséhez szükséges dezertálást választva nem jön össze a koordináció, akkor a második legrosszabb nyereményt lehet elérni. Az első egyensúlyi állapot elérése hatékonyabb, a másiké kockázatmentesebb. Az ilyen játékok nagy kérdése, hogy kockázatdominán/kockázatérzékeny vagy kifizetésdomináns/kifizetésérzékeny (risk dominant vagy payoff dominant) egyensúly alakul-e ki bennük [536], ami a helyzeten túl erősen függ(het) attól is, hogy milyen személyiségű játékosok kerülnek össze egymással.

Bár a játék elterjedt megnevezése a koordinációs játék, én mégis az ügyféljáték nevet választottam. Az olyan fogyasztói vásárlások esetén, amikor a hálózati hatás érvényesül (minél többen választják az adott terméket, szolgáltatást, annál jobban nő a szolgáltatás önértéke) vagyis az ügyfélkör nagysága hozzájárul a jószág értékéhez. Ilyen esetekben koordinációs kérdéssé válik, hogy a versengő szolgáltatások közül melyiket választják a vásárlók, kiknek lesznek az ügyfeleik, és ebből következően melyik vállalatnak lesze nagyobb ügyfélköre (forgalma, bevétele, nyeresége stb.).

Példa lehet erre a szituációra, amikor két mobilszolgáltató (ATel és BTel) van a piacon [536]. Ha az egyik szolgáltató percdíja 10, a másiké 20 egységnyi pénz, az egyiktől a másikhoz irányiló – tehát az ATel felől a BTel felé tartó– hívások ára 70, fordítva pedig 40 egység, akkor az ezt a piaci feltételrendszert a tiszta koordinációs játák logikája szerint rendezik el a vásárlók (vagyis tiszta koordinációs játékot kapunk, ha e feltételek alapján előállítjuk a két szolgáltató közti koordinációs verseny preferenciamátrixát). Az igazán tiszta koordinációs játékot úgy lehet modellezni, hogy a két kívánatos és a két elkerülendő kimenetet ugyanolyan preferenciaértékkel látjuk el. Ezt azért nem teszem itt meg, mert ez

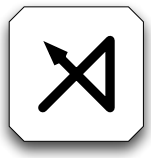
ellentmondana annak a kezdetekben rögzített elvárásnak, miszerint a preferenciabesorolásoknak szigorú rendezéseknek kell lenniük. Az igazán tiszta koordinációs játék klasszikus példája az a kérdés, hogy a jobb- vagy a baloldali közúti közlekedést kell-e választanunk magunknak.

Koordinációs játék: balszélső



4,4	2,1
1,2	3,3

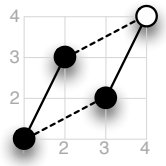
RPST



a játék száma	
Rapoport-Guyer:	60
Brams:	-
Robinson-Goforth:	344

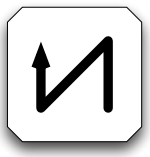
Ez a játék az ügyféljáték (tiszta koordinációs játék) komplementer párja. Ha egyáltalán említik a szakirodalomban, akkor csak annyit mondanak róla, hogy lényegében megegyezik a tiszta koordinációs játékkal, ami igaz is a játék várható kimeneteit tekintve. Olyan példát fabrikáltam rá, ami bár életszerű, szokatlannak mondható szokásos játékleírásokhoz képest. Ha egy focimeccsen a bal oldali középpályás szöktetni akarja a balszélsőt, akkor vagy a bal vagy a jobb szélre rúghatja a labdát, és a balszélső is ugyanúgy vagy bal vagy a jobb szél irányába kezdhet el rohanni. Ha mindketten ugyanazt az irányt "választják", akkor mindkét esetben gólhelyzetbe kerülhet a csatár. Mivel balszélső ballábas, azért egy kicsivel jobb számára a bal oldali támadás a jobb oldalihoz képest. Ha nem sikerül az akció, akkor a balszélső számára kicsivel jobb az a helyzet, amikor a bal oldalra fut, mert bár nem kapja meg a labdát, de legalább ott van, ahol otthonosan érzi magát, míg a jobb oldalon idegenebbül mozogna. A bal oldali középpályásnak az a legrosszabb opció, amikor jobbra küldi a labdát, de a szélső balra fut, mert mert ekkor neki kell a jobb oldalra a labda után futni és a neki "nem fekvő" területen kijavítania a hibáját.

Harmóniajáték: csónakos



4,4	2,3
3,2	1,1

RTSP



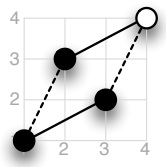
a játék száma	
Rapoport-Guyer:	6
Brams:	-
Robinson-Goforth:	311

A benne rejlő harmonikusság, konfliktusmentesség miatt kevés figyelmet szenteltek eddig ennek a játéknak, pedig Adam Smith láthatatlan kéz elméletéhez ez illeszkedik a legjobban. A játékban mindkét játékos számára a kölcsönös kooperálás biztosítja a legnagyobb nyereséget. Ezért nevezték Robinsonék harmóniajátéknak, amit korábban – Olson nyomán – én privilégiumjátékként azonosítottam [517], míg Skryms Prisoner's Delightként hivatkozik rá [495]. Kicsit más nézőpontból tekintve a játékra Sam Perlo-Freeman az Appeaser (békéltető, lekenyerező) nevet használja [408].

A csónakos játék meséje Brian Skyrmstől származik [495]. Két halász egy szigeten horgászik egész nap, és nap végén egy csónakkal indulnak haza úgy, hogy mindkettejüknek van egy-egy pár evezője, és mivel éhesek, minél gyorsabban szeretnének hazajutni. Ekkor mindketten a kölcsönös kooperációt tartják a legjobbnak, mert ekkor érnek haza leggyorsabban. A második alternatíva mindkettejük számára az, ha maguk nem eveznek, de a parnerük igen, mert akkor is hazaérnek, bár lassabban, viszont nem kell erőt kifejteniük, a harmadik legrosszabb ennek a fordítottja lenne. A legrosszab kimenetet számukra a kétszeres dezertálás jelentené, hiszen ekkor sosem érnének haza.

A Perlo-Freeman által javasolt békéltető (appeaser) megnevezés azért találó, mert azt a beállítódást fejezi ki, amely arra törekszik, hogy a partner kooperációját (békés hozzáállását) minden áron fenn tudja tartani. Ez abban mutatkozik meg, hogy a legjobban preferált kölcsönös kooperálás (béke) után a második legfontosabb alternatívának a partner kooperálását tartják a felek. Így jellemezték a második világháború előtti angol diplomácia stratégiáját a németekkel szemben. Winston Churchill (aki ezt a politikát tagadta meg) azt mondta erről, hogy "An appeaser is one who feeds a crocodile – hoping it will eat him last" [408]. Más megközelítésből tekintve a játékra, észrevehetjük, hogy egy jól ismert másik mondás ("Ha békét akarsz, készülj a háborúra!") szintén ehhez a játékhoz köthető, csak ekkor a hangsúly eltolódik úgy, hogy a legjobban preferált kölcsönös kooperálás (béke) után a másodikra helyre az az alternatíva kerül, amikor a játékos már dezerál a partner kooperálásával szemben.

Harmóniajáték: pacifista



4,4	3,2
2,3	1,1

RSTP

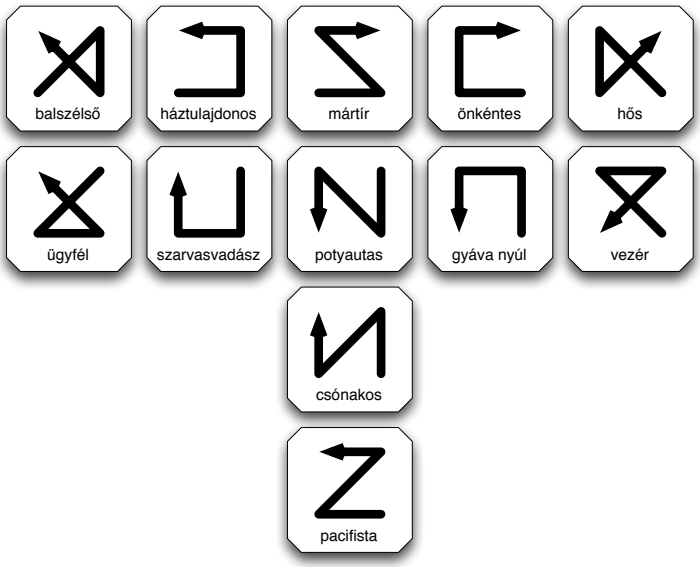
a játék száma	
Rapoport-Guyer:	
Brams:	
Robinson-Goforth:	366

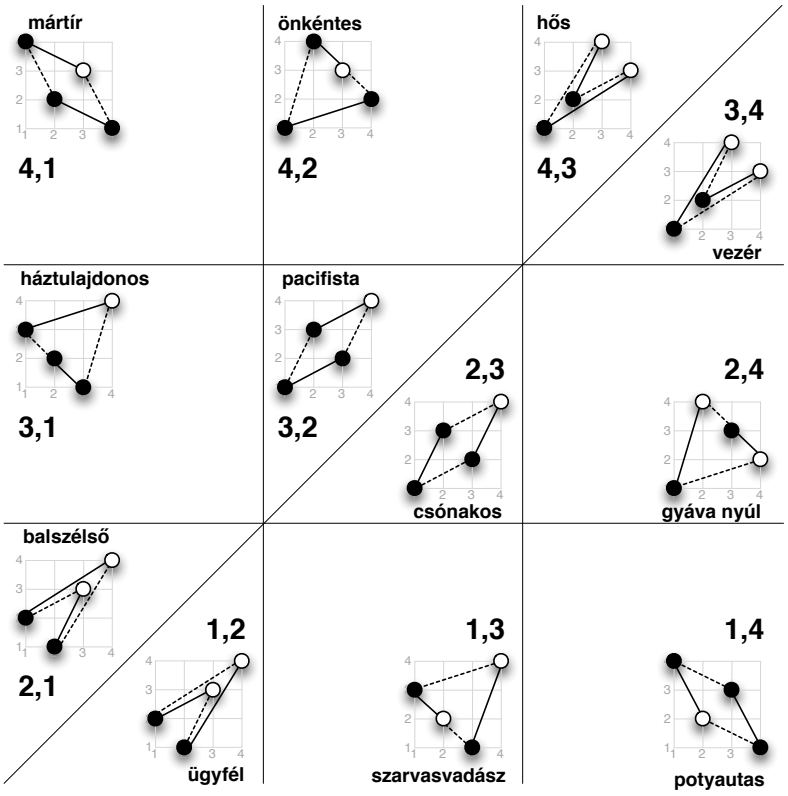
Ez a játék a csónakos játék (tiszta harmoniajáték) komplementere, és a sajátosságát a csónakos játékhoz képest a második legjobb alternatívával kapcsolatos preferenciaváltozás adja. Azon túl, hogy mindkét fél a kölcsönös kooperálást tartja a legmagasabbra, a második helyre mindketten a dezertálással szembeni kooperálást teszik, vagyis a saját kooperálást preferálják a leginkább. Ez az igazi pacifista hozzáállás: a pacifista békét akar, még akkor is, amikor a partnere már dezertálna. A pacifista úgy akarja a békét, hogy maga mindent meg is tesz érte. Ez fejeződik ki abban, hogy a két – pacifista – játékos magasabbra értékeli azt az kimenetet, amikor maga kooperál, míg a társa dezertál a fordított felálláshoz képest.

Erre a játékra is elmondható, hogy kevesebbet elemzik, nyilván megintcsak a konkfliktusmentessége miatt. Sam Perlo-Freeman adta a játéknak a pacifista elnevezést [408], míg – más szerzőkkel ellentétben – Amir N. Licht ezt nevezi harmóniajátéknak [306].

A magjátékok áttekintése

A tizenkét magjáték között érdekes összefüggések vannak, amelyek még inkább észrevehetőek, ha valamilyen összesítő ábrában mutatjuk meg őket. Tanulságos áttekinteni egyszerre is az egyes játékoknál egyenként bemutatott vizuális reprezentációkat.

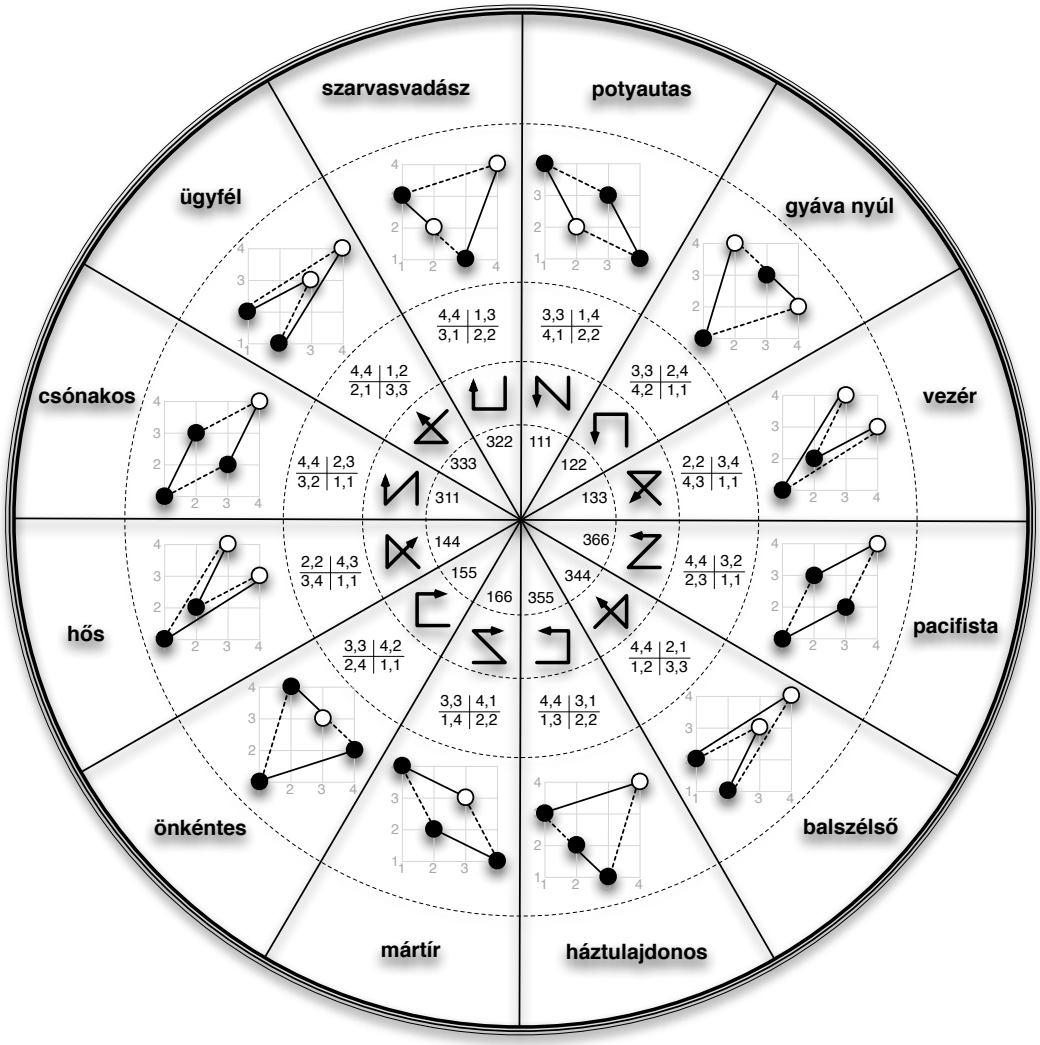




Az “útjelzős” ábrán jól látszik, hogy háromféle mintázat van, a z-, az n- és az x-betűs. A függőleges tengelyen van négy játék, melyek középpontjában a fogolydilemma (potyautas) és a harmónia (csónakos) játék van (mellettük ott vannak még a tengelyen a komplementer párjaik). A fogolydilemmaból (és komplementer társából, a mártírjátékból) kaphatjuk meg a gyáva nyúl és a szarvasvadászat játékokat (illetve ezek komplementereit), majd még egy “lépés téve” jobbra , illetve balra eljuthatunk a kétféle nemek harca, illetve kétféle koordinációs játékhoz.

A játékok topológiai sajátosságait mutató kis ábrákat egy közös térben feltüntetve és ezáltal az összes magjátékot egyszerre nézve még jobban szembeötlővé válik az a tény, hogy a fogolydilemma (potyautas) játék adja, ígéri a legkisebb nyereményt a játékosok számára. Még a szarvasvadászat játékban fordulhat elő az, hogy a szuboptimális egyensúlyi pontot (2,2) választják a résztvevők, egyébként minden más esetben csak magasabb nyereményekkel lehet a játékokból kijönni.

Ha a legfontosabb adatokat mind egy helyen tüntetjük fel, akkor adódik egy harmadik lehetőség is a játékok egyidejű ábrázolására. Valahogy így:



Az ábra elkészítéséhez segítségül hívtam azt a vizuális technikát, amelyet már évezredek óta használ az emberiség személyiségtípusok megjelenítésére. A zodiákus tizenkét ál-

latövi jegyét, tizenkét személyiségtípusát sokszor jelenítik meg kördiagram segítségével.[4] Ezt a megoldást vettem át én is a tizenkét magjáték ábrázolására. Természetesen egy percig sem állítom, hogy bármilyen tartalmi kapcsolat lenne a kétfajta reprezentáció között, kizárólag a vizuális megjelenítés áttekinthetőséget és megérthetőséget könnyítő szerepe miatt veszem igénybe ezt a technikát. Az összes fontos vizuális információt, a játékok preferenciamátrixait és neveit, valamint a Robinson-Gofort könyvben használt azonosítószámokat feltüntettem az ábrában.

*

A magjátékok bemutatása után úgy léphetünk tovább a társadalmi cselekvések játékelméleti modellezésében, hogy egymás után "elengedünk" egyet-egyet a kezdetben megfogalmazott feltételeink közül. Ezáltal a társadalmi kapcsolatok egyre összetettebb modelljéhez jutunk. A modell fokozatos kiterjesztését az alábbi lépésekben fogom elvégezni. Először megmaradok annál a feltételnél, hogy a játékosok csak egyszer találkoznak egymással, vagyis egyszeri játékot játszanak. Ezeket nevezem **statikus játékoknak**. Ezután a játékok logikai rendjében engedem meg a változtatást azzal, hogy figyelembe veszem a büntetés mint cselekvési alternatíva lehetőségét is. A büntetni csak akkor tudunk, ha van egy azt megelőző – büntetendőnek ítélt – cselekvés. Ehhez el kell engedni a szinkronitás feltételét, vagyis **aszinkron játékokat** kell elemeznünk. Ha így teszünk, még a statikus játékok világán belül maradunk, hiszen lehetnek egyszeri aszinkron játékok (például az alkujátékok). A következő lépésben lehet dinamikát adni a játékoknak, amit az ismételt játékok segítségével tudunk modellezni. A játékok ismétlésére, a dinamikus szemlélet alkalmazására akkor van módunk, ha a modellünkbe beemeljük a társadalmi emlékezet fogalmát. A játékok ismétlődésének ugyanis az a feltétele (és persze hozadéka is egyben), hogy a játékosok tudják azt, mi történt velük az előző forduló(k)ban, vagyis ki hogyan cselekedett korábban. Első körben a **direkt reciprocitás játékokat** tárgyalom, amelyekben a játékosok ismerik egymást. Ezt követi az **indirekt reciprocitás játékok** elemzése, amikor egyfelől a játékosokról nem kell feltételeznünk, hogy ismernék egymást, másfelől viszont a döntési szabályok alátámasztásához szükség lesz az értékelési rendszerek, reputációképző szabályok modellbe emelésére is. A legvégén pedig röviden kitérek majd arra, hogy a koordinációs problémáknak, a koalíciós helyzeteknek milyen különös vonásaik vannak.

5.2 A játékok statikus kiterjesztése

A magjátékok tárgyalása utáni modellbővítést a statikus játékokon belül maradva végzem el. Sorban bemutatom, hogy milyen módokon lehet leírni, jellemezni az aszimmetrikus, a többalternatívás, a többszereplős, végül a nem szigorúan rendezett játékokat.

5.2.1 Aszimmetrikus játékok

Az aszimmetrikus játékoknak aszimmetrikus preferenciamátrixuk van, minden másban megegyeznek a szimmetrikus játékokkal. A kétszemélyes, kétalternatívás játékok közül

[4] Érdekességként jegyzem csak meg, hogy vannak még további 12-osztatú ábráink: a nap 12 órából, az év 12 hónapból áll.

csak 12 szimmetrikus, az összes többi aszimmetrikus. Korábban már elmondtam, de nem baj ha itt megismétlem, hogy hosszú időn keresztül széles körben elfogadott volt Rapoport és Guyver nézete, amely szerint 66 aszimmetrikus alapjátékot különíthetünk el egymástól. Ez a szám úgy jött ki, hogy Rapoporték ekvivalensnek minősítették azokat a játékokat, amelyek a két játékos felől nézve ugyanúgy "néznek ki". Ha arra gondolunk, hogy mindegy, hol is "áll" a két játékos, a Sor vagy az Oszlop játékos szerepében játszik-e, akkor ez az ekvivalenciatételezés elfogadható. Ha azonban Egoról és Alterről beszélünk, vagyis elkülönítjük egymástól azt a két szerepet, amelyek közül az egyiket (Egot) kitüntetett helyzetbe tesszük, hiszen mindig az ő cselekvését, döntéseit akarjuk megérteni az adott játék értelmezésekor, akkor már igenis "láthatunk" különbséget két olyan játék között, amelyek közül az egyikben Ego a Sor, a másikban Ego az Oszlop játékos szerepébe kerül. Egonak nem mindegy, hogy az aszimmetrikus helyzet melyik pólusán van. Ezt a szemléletváltást hozta el Robinson és Gofort 2005-ös könyve [443], és ezt fogadom el én is. Ebből következően az alapjátékok száma 144, az aszimmetrikus játékoké pedig 132.

magjátékok	
aszimmetrikus játékok	aszimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

A szimmetrikus játékok ismeretében már könnyebb dolgunk van az aszimmetrikus játékok modellezésekor. Ha ugyanis elválasztjuk egymástól, és külön tekintünk a két játékos, Ego és Alter preferenciarendezéseire, akkor ezek alapján mindkettőjükhöz hozzárendelhetünk egy-egy magjátékot. Mivel az aszimmetrikus játékoknak aszimmetrikus a preferenciamátrixuk, ezért a szétválasztás eltérő magjátékokat eredményez. Azt mondhatjuk tehát, hogy az aszimmetrikus játékok lényege az, hogy a játékosok – ugyanazt a játékhelyzetet másként értékelve – egyszerűen különböző magjátékokat játszanak egymással.

Korábban bemutattam több példát az aszimmetrikus játékokra [517], úgyhogy itt csak utalnék rájuk, nem mutatom be őket alaposabban. Aszimmetrikus játékokat, mégpedig nagyonis életszerű játékokat kapunk akkor, ha az ismert magjátékokat, illetve azok komplementer párjait tesszük össze egymással. Ilyenkor és így tudjuk az olyan élethelyzeteket modellezni, amelyekben valóban értelmesen használhatjuk az olyan szerepminősítéseket, mint a kizsákmányoló, a haszonleső, a vezér, vagy a másik oldalon a mártír, az önkéntes, a hős. Ha ugyanis az első játékos fogolydilemmaként, a második mártírjátékként látja ugyanazt a társadalmi helyzetet, akkor az első tényleg mindent visz a másikkal szemben, vagyis valóban kizsákmányolja a mindenét feláldozó mártírt (vagy a haszonleső valóban learatja az önkéntes által egyedül elvégzett munka gyümölcsét).

Természetesen aszimmetrikus játékat egymástól távolabb álló játékok összerendelésével is képezhetünk. Salamon király jólismert bibliai története is aszimmetrikus játékként írható le. Arról a meséről van itt szó, amelyben az igazi anya és egy hazug asszony egyszerre igényli magának egy csecsemő anyaságát, mire a király úgy akar igazságot tenni köztük, hogy majd kettévágja és egyenlő arányban elosztja köztük a gyereket. Ekkor az

igazi anya lemond az anyaságról, de ez a gesztus elég ahhoz, hogy a király neki ítélje a gyereket. A játékelmélettel az első – tervezett – salamoni döntés kihirdetése utáni pillanatig lehet modellezni ezt a helyzetet, de addig viszont nagyon szépen rá lehet illeszteni egy aszimmetrikus játékot.[5] Glenn H. Snyder és Paul Diesing a nemzetközi kapcsolatok és a háborúk történetéből rengeteg epizódot kiemelve játékelmélet segítségével próbálták meg modellezni azokat – nagyon megyőző módon [502]. E diplomáciatörténeti játékok nagyobb része aszimmetrikus volt, melyek közül a felvett kesztyű (called bluff) játékkal lehet leírni az 1962-es kubai rakétaválság vagy az 1982-es Falkland-szigeteki háborút Argentína és Nagy-Britannia között, de érdekes történelmi és kortársnak mondható példákat hozott a szerzőpár a pártfogó vagy az agresszorjáték kapcsán is [502], [517]. Agresszorjátékkal írható le az 1898-as Fashoda incidens a franciák és brittek között vagy épp az Usa és Észak-Korea közti viszony, pártfogó játékként értelmezhetjük a franciák és csehek szövetségét a németekkel szemben 1938-ban, az USA és Tajvan közti viszonyt 1958-ban vagy a Szovjetunió Kubának nyújtott támogatását 1962-ben. A felvett kesztyű játékról akkor beszélhetünk, amikor egy eredetileg gyáva nyúl játékként értelmezett helyzet egyik szereplője átértelmezi a kereteket, és fogolydilemmaként játszik tovább, a pártfogó játékot egy önkéntes- vagy egy mártírjáték és egy vezérjáték egymásra hatásaként foghatjuk fel, míg az agresszorjáték egy mártírjátékból és egy gyáva nyúl játékból rakható össze. Az aszimmetrikus játékok tárgyalását talán azzal a megjegyzéssel érdemes zárnom, hogy akár 66, akár 132 aszimmetrikus játékot tartunk elkülönítendőnek, messze vagyunk attól, hogy mindegyikükre legyen életszerű leírás, konkrét gyakorlati példa. Az általam ismert szakirodalomban jó, ha harminc aszimmetrikus játékhoz tudtak eddig példát rendelni.

5.2.2 Többalternatívás játékok

A kétalternatívás játékok abban az értelemben nagyon egyszerű társadalomképet feltételeznek, hogy ebben a világban csak "igen-nem" döntések hozhatnak a játékosok. Sok esetben elég ez a megoldás. De természetesen nem kevésszer kerülhetünk olyan helyzetbe, amelyekben a döntéseink nem ilyen egyszerű szerkezetűek, amikor több alternatíva közül kell választanunk. Gondoljunk csak a kő-papír-olló néven ismert, egyébként nagyon egyszerű játékra, amelyben három döntési lehetőség közül kell választaniuk a játékosoknak. Az ilyen helyzeteket a többalternatívás játékokkal modellezhetjük.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	többalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

A többalternatívás játékok modellezését még könnyebben és még rövidebben el lehet in-

[5] Pontosabban kettőt, attól függően, hogy milyen megfontolásokkal él az igazi anya a gyereke vélt jövőjére vonatkozóan, de ez a különbség egyrészt életszerűen modellezi a különböző lehetséges személyiség- vagy gondolkozástípusokat, másrészt nem lényegbevágó a játék kimentét illetően [517].

tézni, mint ahogy tettem azt az aszimmetrikus játékokkal. Más tudásterületeken gyakran alkalmazott technikát hasznosíthatunk itt újra. A kettőnél több alternatívát ugyanis úgy kezelhetjük, hogy azokat visszavezetjük kétalternatívás döntések egymás után fűzött sorozatára. Csak annyit kell tennünk, hogy az összes döntésre váró alternatívát besoroljuk két csoport valamelyikébe, majd a két csoportot "mérkőztetjük meg" egymással. A győztes csoportba tartozó alternatívákat újra két csoportra osztjuk szét, a csoportokat újra összemérjük egymással és így tovább. Az alternatívák csoportba sorolást és a csoportok összehasonlítását addig folytatjuk, ameddig két alternatíva marad a végére, amelyek közül egy utolsó összehasonítással már végső győztes lehet kihozni. A többalternatívás döntéseket – első körben – ennyire egyszerűen elintézhetjük, itt nem is foglalkozom velük többet.

5.2.3 Többszereplős játékok

A többalternatívás játékok modellezése nagyon könyen megoldható a kétalternatívás döntéssorozat technikájának alkalmazásával. Nem ez a helyzet a többszereplős játékok esetében. Sok minden megváltozik ugyanis akkor, amikor a társadalmat már úgy modellezzük, hogy két játékos helyett sok szereplő alkotja azt. Ezt a lépést meg kell tennünk, mert bár sokszor találhatjuk magunkat kétszemélyes játékszituációkban, de még többször kerülünk olyan helyzetekbe, amikor a helyzet minősítő sajátossága az, hogy sokan, nagyon sokan vagyunk benne.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	többszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

A többszereplős játékok több szempontból is jelentős mértékben átalakítják azt a döntési-cselekvési teret, amelyben a játékosok egymást és a többiek lehetőségeit értékelik. A problématér átalakulását nagyon alaposan bemutatja Robyn M. Dawes egy tanulmánya, az érdeklődőknek azt ajánlom elmélyülésre [124], de itt nem térek ki rá bővebben. A lényeg durván háom tézisben összefoglalható. Egyrészt minél több szereplő kerül interakcióba egymással, annál biztosabban tud anonímmá válni, "eltűnni" a többiek között az ember, ami annyit jelent itt, hogy nem lehet azonosítani, ki mivel járul hozzá a közöshöz, vagy épp ki mennyit vesz ki a közösből. Ez nyilván növeli a dezertálási hajlandóságot (és esélyt). Másrészt az egyre több szereplő jelenléte egyre csökkenti az egyén hozzájárulásának (vagy részesedésének) nagyságát az egészhez képest, és ez az összemérhetetlenség érzet megintcsak növeli a dezertálásra való késztetést. Harmadrészt a nagy tömegben elvész annak a lehetősége is, hogy az egyéni dezertálásokat észlelve az elkövetőivel szemben ilyen-olyan retorziókat, büntetéseket lehesen alkalmazni – fenntartva ezzel a kooperálási hajlandóságot.

A játékok többszereplős kiterjesztésének két magjáték esetén van különös hozadéka: az egyik a fogolydilemma, a másik a gyáva nyúl játék. Az előbbi a **közlegelők tragédiája**

néven vált ismertté és azóta is használttá, az utóbbi az önkéntes játék, a kollektív felelőtlenség vagy a **hiányzó hős problémájaként** terjedt el a köztudatban. A közlegelők tragédiája név és a Föld túlnépesedésének veszélyére vonatkozó problémafelvetés Garret Hardin 1968-as cikkében jelent meg [221], de ennek a cikknek semmi köze nem volt a játékelmélethez. Pár évvel később Russel Hardin volt az, aki névrokona hangzatos címe alá "bedolgozta" a sokszereplős fogolydilemma játékelméleti modelljét [222]. A közlegelők tragédiája arról szól, hogy a természetes úton előálló kollektív javainkat nagyon hamar és könnyen felemészthetjük, ha nem állítunk korlátokat a "mohó" egyéni fogyasztások elé. A közlegelők tragédiájával leírható helyzetet a társadalmi javak bevett tipizálására [517] hivatkozva el is nevezték **részesedés dilemmának** arra utalva ezzel, hogy ilyen esetekben az a nagy – és az érdekviszonyokon túli eszközökkel szabályozandó – kérdés, hogy kinek milyen részesedése lehet azokból a kollektív javakból, amelyek fogyasztásából gyakorlati értelemben nem zárható ki senki, viszont amelyek versengő fogyasztásúak, vagyis amit valaki elfogyasztott valamennyi részt belőlük, azt már más nem fogyaszthatja el. A közlegelők tragédiája modell húzható rá minden olyan társadalmi helyzetre, amelyben arról van szó, hogy a közös javainkat (többek között a természeti kincseinket) szabályozatlanul, a jövőnkre, a többiekre, a szomszédainkra nem figyelve túlfogyasztjuk, kizsákmányoljuk.

A másik fontos többszereplős kiterjesztés a gyáva nyúl játékra vonatkozik [129]. Az ennek érzékeltetésére kitalált játékot – sajnos nagyon szerencsétlen módon – többszereplős önkéntes dilemmának nevezték el. E játék lényege annyi, hogy három cellába bezárnak három embert, mindegyikük előtt ott van egy gázcsap, amit ha megnyitnak, mérges áramlik a szobába, és az ott levő egyén meghal. Viszont – és ez adja a csavart a játékban– a csap kinyitásával együtt a másik két szoba csapja automatikusan elzáródik, tehát saját halálával bármelyikük megmentheti a többiek életét. Az utolsó – az igazi késztetést jelentő – feltétel pedig az, hogy ha nem történik semmi, akkor egy óra múlva mindhárom szobába egyszerre bevezetik a gázt, tehát mindenki meghal. A kérdés az, ki lesz az önkéntes hős, aki magára vállalja ezt [517]. Ha valaki esetleg nem érzi túl életszerűnek ezt a példát, annak lehet akár valódi, megtörtént esetekre hivatkozni, akár ennél életszerűbb példákat is felmutatni. A megtörtént esetek talán leghíresebbje Kitty Genovese története 1964-ből. A new yorki lányt hazafelé tartva megkéselte egy támadója, aki a lány sikoltozására elmenekült a helyszínről. Viszont amikor észlelte, hogy a lánynak senki sem sietett a segítségére, miközben vonszolta magát a lakása felé, a támadó újra visszatért, és halálra szúrkálta áldozatát. Az eset hatalmas felháborodást keltett, miután kiderült, hogy 38 ember látta vagy hallotta az esetet, de senki sem sietett segíteni neki, sőt, nagyon sokáig a hatóságokat sem értesítette senki. Ez ügy kapcsán kerültek be a közgondolkodásba a kollektív felelőtlenség vagy hiányzó hős problémája kifejezések is. A hétköznapi életünkből vett példák között pedig nagyon sok – kisebb vagy nagyobb jelentőségű – példát említhetünk kezdve onnan, hogy ki szól rá a szemetelő vagy hangoskodó emberre, ki telefonál a szolgáltatónak, ha egy társasházban elmegy az áram, ki tesz bármit is a közterületek szépítéséért, addig bezárólag, hogy voltaképp az összes közpénzből fenntartott szolgáltatórendszerünket a közösbe egyénileg befizetett hozzájárulásaink tartják el. Ez utóbbi mozzanat alapján érthetjük meg, miért nevezik ezt a játéktípust **hozzájárulás dilemmának**. Amikor olyan kollektív jószágról van szó, amelynek fogyasztásából megintcsak nem lehet gyakorlatilag kizárni senkit, viszont a jószág nem versengő fogyasztású, azaz valaki más fogyasztása egyáltalán nem befolyásolja a

saját fogyasztásomat,[6] akkor az válik kulcskérdéssé, hogy ki(k) lesz(nek) az(ok), aki(k) majd tesz(nek) azért, hogy a mindenki által ingyen fogyaszható jószág előálljon.

A többszereplős játékok értelmezésekor és értékelésekor figyelembe kell venni azt a szempontot, hogy a többszereplős helyzetek gyakorlati szempontból időben szétcsúsznak, tehát aszinkronná válnak, így az interdependencia egyidejűségi feltétele nem lesz annyira szorító. Ez könnyíti a modellezési munkát, de bizonyos értelemben éles határvonalat húz a két- és többszereplős játékok közé. Azt is érdemes megemlíteni még, hogy bár a koalícióelméletek lényegüknél fogva több, mint kétszereplős helyzeteket írnak le, mégsem itt kell tárgyalni őket, mert a koalícióformáláshoz mindig valamilyen kommunikációra és megegyezésre van szükség a felek között, míg a modellkiterjesztés ezen pontján erre még nincs lehetőség. Innen indulva lehetne viszont megkezdeni azoknak a háromszereplős játékoknak a tárgyalását, amelyek a többszemélyes altruizmus (multiparty altruism) és a károkozás címkéivel írhatók le. Ezekre a könyv egy másik pontjám már kitértem röviden (☞253.o.), így itt csak utalok rájuk. Az altruizmus egyfajta megnyilvánulása lehet az, amikor egy ágens a rokonával szembeni dezertálást dezertálással viszonozza, valamilyen formában kárt okoz (ami maga számára költségekkel jár) azért, hogy a rokonát segítse ezzel. Azért említettem, hogy innen lehetne indítani az ilyen esetek tárgyalását, nem pedig itt lehete elvégezni az elemzést, mert ez a mechanizmus aszinkronitást vagy ismétlődő jelleget kíván meg. Viszont a többszereplős vonás is elengedhetetlen hozzá. Azt feltételezem, hogy ez a biológia által leírt kooperációs mechanizmus az emberek körében is tetten érhető lenne, noha konkrét eredményekről nem tudok.

5.2.4 Nem szigorúan rendezett játékok

Elméleti szempontból érthetőnek és könnyen elfogadhatónak tűnik az az elvárás, amely szerint feltételezzük, hogy a játékosok egyértelműen sorba tudják rendezni a játék lehetséges kimeneteit, alternatíváit. A preferenciarendezésre vonatkozó szigorú rendezési feltétel azonban a tényleges gyakorlatban sok esetben túlzottan szigorú elvárásnak minősül. Nem tudunk mindig minden alternatívát egyértelműen sorba rendezni. Van, hogy nem tudunk valamely alternatívát megítélni és egyszerűen kihagyunk a rendezésből, de az is előfordul, hogy nem vagyunk képesek dönteni két vagy több alternatíva esetében, és ilyenkor "döntetlenre hozzuk ki" a sorrendet köztük. Az ilyen esetek modellezésére az olyan játékok alkalmasak, amelyekben nem követeljük meg a szigorú rendezés feltételének fennállását.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	nem szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

[6] Egy koncerten, egy színházban nem csökkenti az előadás élvezi értékét az a tény, hogy velünk párhuzamosan még sokan mások is fogyasztják ugyanazt a társadalmi jószágot.

Az ilyen játékok modellezésekor azonban rögtön egy számíthatósági problémával kerülünk szembe. Ugrásszerűen megnő ugyanis a lehetséges rendezési mintázatok számossága, ami miatt – az eddigiekhez képet sokkal több játék áll a rendelkezésünkre [517]. A Robinson-Goforth szerzőpáros Matt Cargill-lel kiegészülve egy megpróbálták továbbgondolni, hogy lehet a játékok periódusos rendszereként is felfogható saját topologiai megközelítésüket ebben a nem szigorúan rendezett játékok irányban kiterjeszteni [444]. Ennek a problémának a tárgyalására itt nem vállalkozom.

5.3 A játékok magasabbrendű kiterjesztése büntetéssel

A játékelmélet első szintjén a játékosok stratégiáit úgy értékeljük, hogy a cselekvések (döntések) tartalmát absztraháljuk, vagyis lecsupaszítjuk a kooperálás-dezertálás dichotómiájára. A kooperálás a jó, a dezertálás a rossz, és az a nagy kérdés, hogy mikor, hogyan, miért lehet elérni, remélni, hogy kooperáljanak a játékosok. Ez a séma persze nem pontos. Hiszen a koordinációs játékoknál láthattuk, hogy a kooperálás vagy dezertálás fogalma nem használhatók tartalmas módon. Az összes alapjátékra igaz viszont, hogy ha van valamilyen feloldandó társadalmi dilemma vagy koordinációs probléma, akkor mindig van olyan kimenet, amely jobb, kívánatosabb a többinél, és a kérdés úgy szól, hogy milyen eszközzökkel lehet biztosítani a társadalmilag kívánatosnak tartott kimenetek követését. Az egyik válasz a döntések szankcionálása. Ismételt vagy aszinkron játékok esetében, amikor az egyik játékos ismerheti a másik döntését, szankcionálhatja a már ismert döntést, vagyis saját cselekedetét a másik játékos előző döntéséhez igazíthatja.

Olson vezette be a **szelektív ösztönző** (selective incentive) fogalmát [394], de tovább már nem pontosította. A fogalom alapos elemzését Pamela Oliver végezte el [393], [392], aki meggyőzően mutatta be azt, hogy milyen helyzetekben lehet hatásosan alkalmazni a pozitív ösztönzőket, azaz a **jutalmakat** (reward), és mikor a negatív ösztönzőket, azaz a **büntetéseket** (punishment). Több, más szempont mellett az egyik legfontosabb szempont a kollektív javak előállításához szükséges emberek számosságának kérdése. Amennyiben kevés ember hozzájárulása elegendő lehet (például egy kisvárosi újság finanszírozásához), akkor a pozitív ösztönzés elégségesnek bizonyulhat (kitüntetést adni a kevés támogatónak), de minél nagyobb létszámú hozzájárulóra van szükség, annál inkább csak a negatív ösztönzés lesz működőképes (például a sztrájktörés megakadályozására). Bár a hétköznapi szóhasználatban ritkán fordul elő, de legalábbis idegenül hangzik a negatív ösztönzés terminusa, de ugyanúgy szokatlannak lehet tartani a pozitív szankció kifejezését is, pedig a szankcióra is ráhúzhatjuk a pozitív vagy negatív minősítéseket, hogy azokkal a jutalmazás vagy büntetés jelenségét jelezzük. A továbbiakban én sz pozitív és negatív ösztönzés, illetve szankció, valamint a jutalom és büntetés terminusait egymással felcserélhető módon fogom használni. Azt is jeleznem érdemes azonban, hogy a későbbiekben gyakorlatilag csak a büntetés jelenségét fogom igénybe venni a modellek felépítése során, míg a jutalmazás fogalmát hanyagolni fogom.

Az ösztönzés, szankcionálás fogalmát, különösen annak negatív változatát, azaz a büntetést el kell tudnunk választani a “puszta” károkozás, “sima” agresszió jelenségétől. Ezt akkor tudjuk megtenni, ha felismerjük azt, hogy a büntetés (és jutalmazás) jelenségében van egy logikai értelemben magasabbrendű minőség.

Másodrendű játékok – költséges büntetés

A legáltalánosabb szempont szerint a szankcionálásnak két típusa lehetséges, a pozitív, illetve a negatív szankció. A negatív szankció vagyis a büntetés a rossz, nem kívánatos cselekvésre adott negatív tartalmú válasz. A pozitív szankció ennek ellentéte, vagyis a kívánatos cselekvésre adott pozitív tartalmú válasz. Pozitív és negatív válaszokon azt kell itt érteni, hogy a szankcionált fél számára előnyökkel vagy hátrányokkal jár-e az alkalmazott szankció. A szabkció alkalmazásának nyilván az lehet a célja, hogy a másik felet valamely cselekvés folytatásra vagy épp az attól való tartózkodásra lehessen bírni. Ha csak ennyiben állna a dolog, akkor egyszerű lenne a helyzet, hiszen a szankcionálás révén remélni lehetne a kívánatos állapotok elérését. Azonban a szankció mindkét fajtája költségekkel jár a szankciót alkalmazó ágens számára, és ez újfajta problémát eredményez a játékelméleti modellezés számára. Azzal ugyanis, hogy figyelembe kell venni a szankciónak, büntetésnek a költségeit, fel kell tenni a kérdést, hogy kik lesznek azok, akik magukra vállalják ezeket a szankcionálási költségeket, amelyeknek a közösség számára lehetnek ugyan előnyei, de a szankció költségeit viselő egyén számára biztosan olyan egyedi terheket jelentenek, amiért közvetlen hasznot nem remélhetnek. Ez a probléma vezette el a kutatókat, hogy megalkossák a **másodrendű fogolydilemma játék** fogalmát, amelyben az elsőrendű fogolydilemmát a szankcionálással feloldani próbáló másodrendű szankcionálási játékban ugyanolyan – immár másodrendű – dezetálási hajlammal szembesülhetünk, mint ami a "hagyományos", elsőrendű fogolydilemmát jellemzi [393], [403]. Ki fogja a költséges büntetést elvégezni a köz érdekében – ez a másodrendű fogolydilemma fő kérdése. Azért kell itt másodrendű fogalomról beszélnünk, mert a büntetésről csak úgy van értelme beszélni, ha azt egy előző, a büntetni kívánt cselekvésre vonatkoztatjuk. A büntetés leírásakor mindig meg kell tudnunk adni, hogy milyen korábbi cselekvést szankcionál, ami azt jelenti, hogy egy cselekvést (a büntetést) úgy tudunk leírni, hogy a paraméterei közt megadunk egy másik (a büntetendő) cselekvést, vagyis másodrendű fogalmat képzünk.

Az általános kép felfestése után érdemes picit a részletekben is elmerülni, és megvizsgálni azt, hogy milyen módon próbálják meg tipizálni a büntetések világát a szakirodalomban. Arra fókuszálva, hogy a büntetendő cselekvés valakinek valamilyen sérelmet okozhatott, jellemezhetjük aszerint a büntetéseket, hogy ki hajtja végre a büntetést. Amennyiben a sértett, akkor **sértetti büntetésről** (second-party punishment – 2PP) beszélhetünk, ha viszont egy harmadk személy, aki nem volt közvetlenül érintett az eseményben, akkor **kívülálló büntetésről** (third-party punishment) lehet szó [155]. Az, hogy milyen motívumok miatt büntetnek (altruista, agresszív stb.), itt másodlagos kérdés. Meg lehet különböztetni egymástól a **csoporton belüli** (within-group) és a **csoportok kívüli**, csoportok közötti (inter-group) büntetést. Az evolúcionalista magyarázat szerint a csoportok közti versengésben alakulhat ki és szilárdulhat meg az kooperáció (altruizmus) jelensége (↪251.o.). Egyes kutatók szerint a csoportok közti versengésben, konfliktusokban gyakran megjelenhet a **parokiális altruizmus** (parochial altruism) jelensége, ami azt jelenti, hogy valamilyen szempont szerint összetartozó csoport tagjai erősebben védelmezik magukat a kívülről érkező, nem kívánatos, ártó, büntetendő cselekvésekkel szemben a büntetés eszközével [48]

A büntetés motivációjára vonatkozóan versengő elméletek léteznek. N.J. Raihani és K. McAuliffe kísérleteik alapján azt állítják [426], hogy a csalás megbüntetésének szándékát nem pusztán a csalás ténye váltja ki, így nem annak viszonzásáról, tehát nem

negatív reciprocitásról van szó, hanem az ítélkezők a csalás megtörténte után azt is figyelembe veszik, sőt, elsősorban azt veszik figyelembe, hogy miként változott saját kifizetésük a csalóéhoz képest a csalás után (és következtében), és csak akkor büntetnek, ha az így észlelt egyenlőtlenséget nem tartják elfogadhatónak. Ernst Fehr, Urs Fischbacher és Simon Gächter megkülönbözeti a stratégiai és a nem stratégiai büntetéseket egymástól [156]. A **stratégiai büntetés** motivációját nem érzelmi tényezők adják, hanem inkább az a racionális megfontolás, hogy a büntetés által növelni lehet a saját haszon nagyságát a többi játékos nyereségének rovására. Fehrék megkérdőjelezik e büntetési forma széleskörű elterjedtségét. Ezzel szemben a **nem-stratégiai büntetés** motivációs hátterét a társadalmi normák be nem tartása miatt keletkező érzelmek biztosítják. Ilyen esetekben a büntetés célja a megsértett társadalmi normák betarttatása, az elvárt viselkedés kikényszerítése. Más szerzők szerint előfordulhatnak olyan esetek, amikor a büntetés motivációjaként rosszindulatot, agressziót (spite) vagy az irigységet (envy) lehet kimutatni. Gyakran olyan játékosok, kísérleti alanyok büntetnek így, akik a partnerüket túl "veszélyes" vetélytársuknak tekintik, akik hozzájuk képest sikeresebbek, és fenyegetik az érdekeiket [153], [158]. Egy másik fejezetben bemutatom a méltányosságelméletekhez kapcsolható kísérleti játékokat (↪298.o.), amelyek nagy részében megtalálhatjuk (sőt mérni lehet) a büntetés jelenségét, és ott bizony az egyik lehetséges magyarázat a destruktív beállítódás (↪258.o.). Ilyen esetekben könnyen előfurdulhat az, hogy valaki maga számára kis veszteséggel másoknak nagyon nagy károkat képes okozni [153], ugyanakkor ezen kísérletek másik nagy tanulsága az, hogy sok esetben igen nagy személyes veszteséget is hajlandóak az emberek magukra vállalni csak azért, hogy büntethessék partnerük méltánytalannak érzett cselekedeteit.

Szankcióparadoxon

A büntetés funkcióját azzzal magyarázhatjuk, hogy a kívánatos cselekvési mintákat azáltal (is) reméljük általánossá tenni és fenntartani a társadalomban, hogy büntetjük a nem kívánatos cselekvési formákat, amihez persze szükséges, hogy legyenek, akik magukra vállalják a dezertálók büntetését (annak költségeit). Ezen a ponton azonban egy furcsa paradoxonnal kell szembenéznünk. Ha ugyanis: *(i)* a büntetés "célja" általában a dezertáló, a nem-segítőkész magatartás szankcionálása, viszont *(ii)* a büntetés maga mindig valamilyen nem kívánt jelenség kiváltását, a segítésnyújtás megtagadását vagy testi, lelki erőszak alkalmazását jelenti, vagyis *(iii)* a büntetés tartalmát önmagában megítélve rossznak, dezertálásnak kell minősítenünk, akkor konstatálhatjuk, hogy e három tétel között ellentmondás feszül. Ezt többféleképpen lehet értékelni. Ha mindenekfölött álló értéknek tartjuk a segítést, a kooperálást, akkor nem szabad büntetni senkit, mert az nem kooperatív hozzáállás. Ezzel az elvvel azt lehet szembeállítani, hogy ha nem büntetünk, akkor ezzel "jóváhagyjuk" mások dezertálását, helytelen magatartását, ami a kooperáció fenntartása, a közösség egésze szempontjából nem előnyös. Ha erre az érvre hivatkozva mégiscsak büntetünk, akkor árnyalni kell a büntetést végrehajtó személy megítélését, hiszen ebből a szempontból mondhatjuk, hogy a büntető a közösség érdekét követve büntet. Ha így tesz valaki, akkor a büntetés "jogos, igazolt", amit szembe lehet állítani azzal a helyzettel, amikor valaki "jogtalan, igazolatlan" módon büntet. Ez pedig megkívánja, hogy beemeljük a modellünkbe a diszkrimináció jelenségét (és fogalmát), hiszen különböző irányban, különböző megítélésű szereplőkkel szemben másként kell alkalmazni a

büntetés eszközét.[7] Ez a paradox jelleg a büntetés másodrendű voltából fakad. Szemléletessé tehetjük ezt egy táblázat segítségével, ha abban feltüntetjük, hogy a cselekvések kétféle rendjében, kétféle szintjén milyen döntések elé kerülhetnek a cselekvők, és milyen cselekvési lehetőségeket lehet elkülöníteni egymástól.

	másodrendben kooperál, büntet	másodrendben dezertál, nem büntet
első rendben kooperál	kooperál és büntet	kooperál, de nem büntet
első rendben dezertál	nem kooperál, de büntet	nem kooperál és nem is büntet

Az elsőrendű cselekvések szintjén megvalósuló altruista/kooperatív magatartást segítségnyújtásnak, a nem segítést pedig egoista/dezertáló magatartásnak nevezhetjük. A büntetést viszont a segítségnyújtás megtagadásaként lehet értelmezni. Akkor érthetjük meg, hogy miért fontosak ezek a cserék, ha az új terminusokat beleírjuk a fenti táblázatba:

	másodrendben kooperál (büntet = nem segít)	másodrendben dezertál (nem büntet = segít)
első rendben kooperál (segít)	elsőrendben segít, másodrendben nem segít	elsőrendben segít, másodrendben segít
első rendben dezertál (nem segít = büntetés	elsőrendben nem segít, másodrendben nem segít	elsőrendben nem segít, másodrendben segít

Mivel két esetben egymásnak ellentmondó állítások állnak egymással szemben, ami nyilván nem követhető döntési/cselekvési alternatíva, csak két választás lehetséges: első rendben dezertálni és másodrendben kooperálni (azaz nem segíteni), vagy fordítva (azaz segíteni). Ha tehát az értelmezési keretrendszerbe beemeljük a büntetést mint másodrendű cselekvést, akkor ami első rendben önzetlenség, az másodrendben önzés, illetve fordítva. Mindez azért különösen fontos, mert ezt a két opciót ellentétes módon lehet értékelni még akkor is, ha egyértelműen az altruizmust, a kooperációt preferáljuk az egoizmussal, a dezertálással szemben.

Látható tehát, hogy a szankcionálás jelensége paradoxonhoz vezet. Ugyanazt a cselekvést elsőrendben kooperálásnak (jónak), másodrendben dezertálásnak (rossznak), vagy fordítva elsőrendben dezertálásnak, másodrendben kooperálásnak lehet minősíteni. Vagy-vagy. Értékválasztás kérdése. Az indirekt reciprocitás jelenségének tárgyalásakor pedig bemutatom majd, hogy milyen – egymásnak ellentmondó – válaszokat termeltek ki erre a különböző kultúrák az emberiség történelmében (↬315.o.).

5.3.1 Aszinkron játékok

A magjátékok kevésbé reflektált, kevesebbet elemzett feltétele a játékok szinkronitásának elvárása. Pedig vannak olyan esetek, amelyekben nem tartható fent ez a feltétel. A szimultán játszott, szinkron jellegű játékokkal szembe állíthatóak azok a helyzetek, amelyekben az aszinkron (szekvenciális, alternáló) döntéshozatal a jellemző. Ekkor a játékosok nem egyszerre, szimultán módon hoznak döntést, hanem egymás után, így a másik döntésének ismeretében választanak az alternatíváik közül.

[7] A diszkrimináció fogalmának itt nincs semmi negatív felhangja, egyszerűen csak annyit lehet vele kifejezni, hogy a károkozó büntetés, a segítségnyújtás, illetve annak megtagadása vagyis a büntetés melletti döntés szempontrendszere nem homogén.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	aszinkron
ismételt játékok	ismételt
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

Azért is érdekesek ezek a játékok, mert aszinkron döntési helyzetekben a büntetés jelensége is értelmezhetővé, lehetővé válik. A játékosok időben egymástól elváló döntési helyzetekbe kerülnek, és a döntéseikben figyelembe vehetik a partnerük már ismert döntéseit.

Az állatvilágból származó reciprok altruizmus egyik leghíresebb példájaként említik a vámpírdenevéreket, akik között a sikeres táplálékszerző körútról visszaérkező egyedek megosztják beteg vagy hoppon maradt társaikkal az élelmet, míg egy másik alkalommal ezek a szerepek cserélődnek [582]. Ebben a példában a reciprocitás időben eltolt módon jelentkezik, amit az aszinkron játékokkal lehet modellezni. Kétfajta modellt is lehet állítani az aszinkron játékokra [379]. Az alternálás, a szerepváltás lehet szigorú vagy laza. A szigorúan alternáló játékok során pontos rendben váltják egymást a játékosok, míg a laza változatban véletlenszerűen dől el, ki milyen szerepet kap egy adott fordulóban. Utóbbi sok eset életszerűbb, hiszen a reciprocitáson alapuló kapcsolatokban sokszor nem lehet biztosítani, hogy a segítségnyújtó és támogatott (donor és recipiens) szerepek pontos rendben kövessék egymást. Ahogy a fenti bekezdésben írottakból látszik, az aszinkronitás vagyis az időben eltolt döntési helyzetek könnyen összekeverhetők, gyakran össze is csúsznak az ismétlődő játékhelyzetekkel, a reciprocitás jelenségével. Mivel a reciprocitás jelenségével később bővebben foglalkozom, ezért inkább ott mutatom be azt a stratégiát, ami aszinkron játékokban bizonyult eredményesnek (↬310.o.).

5.3.2 Alkujátékok

Az aszinkronitást azonban nem csak az ismételt játékok esetén lehet feltételeznünk. A kísérleti közgazdaságtan, viselkedés-gazdaságtan vizsgál olyan helyzeteket is, amelyek csak egyszer játszott játékoknak felelnek meg, ám a játékosok nem egyszerre, hanem egymás után döntenek, így a későbbi döntéshozók már ismerik a partnereik választásait. Ezeket is aszinkron helyzeteknek minősíthetjük.

Az ilyen játékokat a kísérleti közgazdaságtan régóta és elég nagy intenzitással vizsgálja, igaz, nem elsősorban a helyzetek aszinkronitása miatt, hanem inkább azért, mert ezekben a játékokban jól lehet vizsgálni a kísérleti alanyok beállítódásait, értékelköteleződéseit, normatív struktúráit. Ezért tárgyalhatnám máshol is ezeket a játékokat,[8] de mivel vannak érvek amellett, hogy itt szerepeljenek, ezért itt mutatom be őket. Problémásnak érzem a játékok általános megnevezését is. A szakirodalomban gyakran **alkujátékoknak** (bargaining games-nek) nevezik az ilyen játékokat, ami több szempontból is vállalható, megmagyarázható, de az **alkuelmélet** kifejezéssel párba állítva már kicsit fél-

[8] Például a társas preferenciák elmélete környékén (↬236.o.).

revezetőnek, könnyen félreérthetőnek tartom az alkujátékok terminusát. Az 'alku' fogalmának kétszeres használatával "sugallt" szemantikai közelség itt félrevezető, mert amíg az alkuelmélettel egy folyamatot, egy procedurát írunk le, addig az alkujátékok az egyének beállítódásainak egyszeri kinyilvánításait "mérik". Mivel azonban az alkufolyamat mélyén is ott lapulnak az alkudozó egyének beállítódásai, értékelkötelződései, ezért – fenntartásokkal ugyan, de csak – elfogadhatónak tartom az alkujátékok megnevezés használatát (ezért szerepel már a fejezet címében is). Az is amellett szól, hogy az alkujátékokat ebben a fejezetben tárgyaljam, hogy ezek aszinkron játékok. A játékosok itt egyszemélyes játékokat játszanak szekvenciális módon:

> „ ... az ultimátum (alku) játékokban a stratégiai interakció pusztán a jövőbeni döntések előrejelzésében testesül meg. Itt nincs az egyidejű lépések és végtelen játéksorozat tényéből fakadó kölcsönös függőség (interdependencia)." [201]

Ezekben a játékokban a méltányosság (fairness, equity) jelenlétét, mértékét, százalékokban is kifejezhető erősségét keresik. Azt vizsgálják, hogy milyen mértékű a másik személy iránti reciprok beállítódás mértéke, azaz mennyire hajlandó a kísérleti személy kooperálni a másikkal (lemondani saját előnyei egy részéről a másik javára), illetve hajlandó-e büntetni a másikat, ha az nem kooperál, még akkor is, ha a büntetés saját magának költségbe kerül. Mivel a személyiségben rejtőzködő reciprocitási hajlamot vizsgálják, gyakran nevezik 'erős reciprocitás' (strong reciprocity) elméletnek ezt a vizgálati irányt.

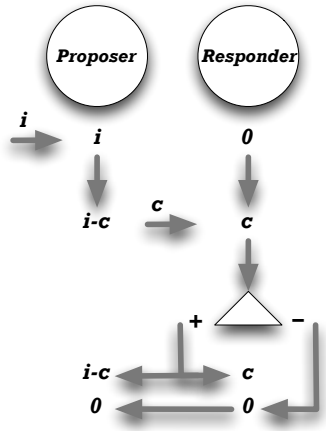
Egyszeri és ismételt játékok egyaránt előfordulnak, az a közös bennük, hogy empirikus kísérleteken, vizsgálatokon alapul az egész. Éppen ezért ezek az empirikus vizsgálatok nagyon szépen kiegészítik a formális játékelméleti modellek, gondolatkísérletek eredményeit. Nagyon sok kultúrában, szubkultúrában elvégezték, megismételték már a leghíresebb kísérleteket, és nagyjából hasonló eredményeket kaptak – akárhol is, akármilyen kultúrában is gyűjtötték az adatokat. Természetesen a különböző kultúrákban, szubkultúrákban felvett adatok között volt különbség, ami fontos vizsgáltai kérdéseket vet, vetett fel, de az erős reciprocitás elmélet körébe tartozó kísérletek talán legfontosabb tanulsága az az általánosíthatónak tűnő állítás, miszerint minden kultúra valamilyen mértékben beleteszi, beleneveli a tagjaiba a reciprocitásérzet, a méltányosságérzet valamilyen fokát. Ez a tétel pedid megintcsak azt a közgazdaságtan széles körben elterjedt vélekedést kezdi ki, amely szerint az ember – alapértelmezés szerint – önző beállítódással lép másokkal kapcsolatba.

Első körben bemutatom az ultimátumjátékot, a diktátorjátékot, a nyilvános javak játékát és a bizalmi játékot, majd utána röviden kitérek még néhány más játékra is.

Ultimátumjáték

Az alkujátékok talán legismertebb és egyben paradigmatikus példája az ultimátumjáték (Ultimatum Game – UG). A játék lényege a következő.[9] Adott összegből (i) az egyik játékos, az ajánlattevő (Proposer, Allocator) javaslatot tesz a másik játékos, a kedvezményezett (Responder) számára, hogy milyen arányban osztaná fel kettejük között a pénzt (i-c, illetve c). A második lépésben a másik játékos csak annyit tehet, hogy az ajánlatot elfogadja vagy visszautasítja. Amennyiben elfogadja, akkor mindketten megkapják az ajálattevő által javasolt összeget, visszautasítás esetén egyikük sem kap semmit.

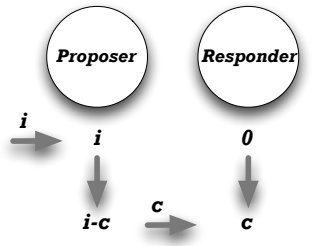
[9] Az ultimátumjátékról (és a többi alkujátékról) kiváló áttekintést ad [233], magyar nyelven a problémakört bemutatja [6], [199].



A játék azért keltett (és kelt a mai napig) meglepetést sokakban, mert tisztán önérdekkövető játékosokat feltételezve már a legkisebb ajánlatot is el kellene fogadniuk a kezdeményezetteknek, ám a kísérletek tanulságai szerint ennél jóval magasabb ajánlatot fogadnak csak el átlagosan a játékosok. Az előzetes "elméleti" megfontolások alapján ugyanis azt lehet feltételezni, hogy a racionális (és önző) beállítódású kedvezményezettek úgy gondolkozhatnak, hogy a legkisebb egységnyi ajánlat is több, mint a semmi (amit a visszautasítás esetén kapnának), ezért érdemes elfogadni bármilyen ajánlatot, ami a nullánál több. Ezt tudva az ajánlattevők is gondolkozhatnak hasonlóképpen és nyugodtan adhatják a legkisebb ajánlatot. A kísérletek azonban egészen más eredményeket hoztak, hoznak. Az eddigi felmérésekben (kultúráktól függően) 20-50 százalék között mozgott mindenhol az átlagos ajánlattétel nagysága. A kísérletek tanulsága szerint ezeket a helyzeteket nem lehet leírni a racionális és önző ember hipotézisére támaszkodva. Valami más magyarázó tényezőt kellett tehát találni. Ezen a ponton lehetett felvenni az ember, a társadalom modelljébe a méltányosságérzet fogalmát. Az emberben levő méltányosságérzettel tudjuk csak megmagyarázni az a büntetési szándékot, ami a visszautasítások tényében érhető tetten. A kedvezményezettek ugyanis saját anyagi érdekeikkel szemben vállalják a visszautasítást, ami így büntetésként fogható fel, hiszen így az ajánlattevő sem kap semmit a méltánytalan ajánlata miatt. A kedvezményezettben létezik valamiféle méltányosságérzet, ami meghatároz egy elfogadható ajánlati százalékot, és ha a konkrét ajánlat ennél az összegnél kisebb, akkor a kedvezményezett ezt a partner önzésének tudja be, és a méltánytalan ajánlatot visszautasítással bünteti. Az ajánlattevők pedig ezt a helyzetet racionálisan mérlegelve részben a büntetéstől (és ezáltal a teljes veszteségtől) való félelmükben, részben a bennük is létező méltányosságérzettől vezéreltetve elmozdulnak a fele-fele (fifty-fifty) elv irányába.

Diktátorjáték

A diktátorjáték (Dictator Game – DG) neve nagyon félrevezető, mert a játéknak nincs semmi köze a diktátorsághoz, a diktátori szerephez. Nyilván csak azt akarták kifejezni ezzel a névvel, hogy ebben a játékban egyszemélyi döntésről van szó, a másiknak nincs beleszólási lehetősége a döntésbe. A diktátorjáték hasonlít az ultimátumjátékhoz.

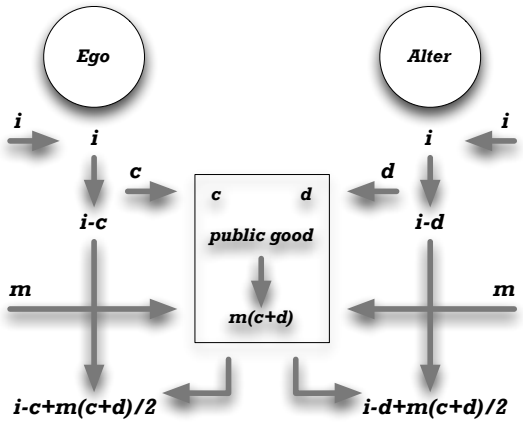


A különbség csak annyi közöttük, hogy a kedvezményezettnek nincs mérlegelési joga a második lépésben. Pontosabban nincs második lépés, vagyis az ajánlattevő ajánlata a játék végeredménye, a javak végleges megosztása is egyben. A kedvezményezetnek semmilyen formában nincs módja befolyásolni a végeredményt. Mivel ezt tudja az ajánlattevő is, ezért arra lehetne számítani az ilyen kísérletek előtt, hogy itt már igenis érvényesülhet az önző racionális ember hipotézise, és az ajánlattevők a minimális ajánlat környékén teszik majd meg ajánlataikat.

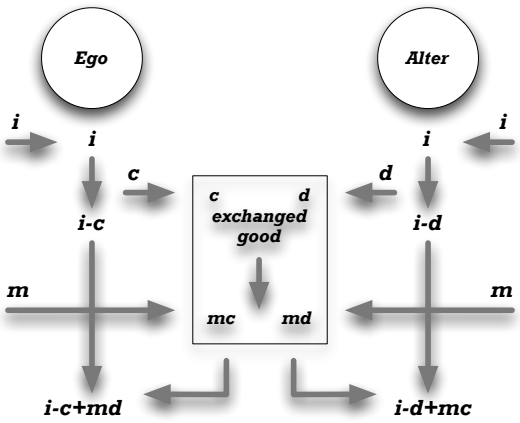
Az eredmények azonban mégsem ezt mutatják. Bár már kisebb mértékben, mint az ultimátumjátékban, de még mindig elég magas arányban ajánlanak fel összeget a kedvezményezett számára. Még akkor is (vagy inkább: annak ellenére is) így van ez, hogy ezekben a helyzetekben már azt sem lehet feltételezni, hogy a "büntetéstől" való félelem miatt lennének nagylelkűek az ajánlattevők. Ezek a kísérletek egyértelműen az ajánlattevő altruista beállítódását mérik (és mutatják egyben).

Nyilvános javak játék

A nyilvános javak játék (Public Goods Game – PGG) felfogható az előzőek sokszereplős kiterjesztéseként is (nevezik Simultaneous-Decision Prisoners' Dilemma játéknak is).

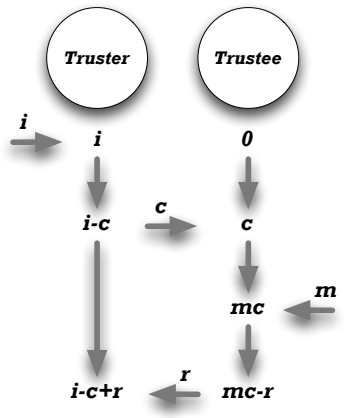


Van x (az egyszerűség és összehasonlíthatóság kedvéért legyen megint csak 2) játékos (Ego és Alter), mindegyiknek van valamennyi pénze (i), és mindenki beadhat a közösbe valamennyi pénzt (legyen ez c és d). Az egyéni döntések után a közös kasszába tett teljes összeget megtöbbszörözik (m), és a közös javakból egyenlő arányban visszakapják a játékosok a saját részüket. A kísérleti eredmények szerint kezdetben nagy az altruista hajlam a játékosokban, de idővel (vagyis a játékok számának növekedésével párhuzamosan) ez a kooperatív hajlandóság csökken. A kezdetben mért 40-50%-os felajánlások leesnek nulla közeli értékekre. Az is kiderült azonban ezekből a felmérésekből, hogy ha bevezetik a büntetés lehetőségét (amikor a játékosok saját pénzükből büntethetik a potyautasokat vagyis azokat a társaikat, akik keveset ajánlottak fel a közösbe, akkor a csoport egészén belül jelentősen megugrik a kooperálási hajlandóság. Erre a játékra is igaz, hogy a világ minden pontján, sokféle kultúrán belül elvégezték a kísérleteket, és mindenhol nagyjából hasonló eredményeket kaptak. Ez a játék sokban hasonlít a fogolydillema játékosztályba tartozó 100 dolláros játékhoz (bár nem azonos vele), amelynek a döntési folyamatábrája a következő.



Bizalomjáték

A bizalomjáték (Trust Game – TG, Investment Game – IG) név alatt más játékot értenek az alkujátékok, az erős reciprocitás elméletében, mint a kétszemélyes magjátékok körében, ahol az egyik magjátékot is ezzel a névvel illetik.



Ebben a játékban is két játékos tesz ajánlatokat egymásnak. Az ajánlattevő lépése előtt a kedvezményezett (Truster) kap egy összeget (i), amiből lehetősége van egy részösszeget (c) "visszaajánlani" a partnerének (Trustee). Ezt az előajánlott összeget megtöbbszörözik (m), és a megnövelt összegű pénzt az ajánlattevőnek meg kell osztani kettejük között, amiből a kedvezményezett visszakap valamit (r). A játék konstrukciója olyan, hogy leginkább a viszonzási hajlandóságot, a negatív reciprocitást lehet mérni benne, hiszen az ajánlattevő (akinek a viselkedését vizsgálják) az őt ért impulzus (az előzetes ajánlat) hatására dönt (legalább részben). Ebben a kísérleti játékban is az előző játékokhoz hasonlóakat tapasztaltak, vagyis a mérési eredmények szerint az ajánlattevő személyekben – kisebb, nagyobb mértékben – jelen van az altruizmus.

A bevezetőben jeleztem, hogy magam is vitatható döntésnek tartom, hogy a méltányosságelméleteket az aszimmetrikus játékok fogalma alá sorolom. Vitathatóságánál tán csak a szokatlansága nagyobb e döntésnek. Nem lehet vita azonban azzal az elkülönítéssel (és besorolással), amivel a az eddig tárgyalt egyszer játszott, tehát statikus játékokat és a most tárgyalandó ismételt, tehát dinamikus játékokat elválasztjuk egymástól.

5.4 A játékok dinamikus kiterjesztése emlékezettel

Ha elengedjük az egyszer játszott játékok feltételét, akkor ismételt játékokhoz jutunk. Ekkor ugyanazt a játékot többször lejátszák egymással a játékosok, vagyis játék helyett játéksorozatot kell vizsgálni. Az egyes játékokat fordulóknak is nevezhetjük, és szükségünk lesz a stratégia fogalmának átértelmezésére is.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	ismételt
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

Az ismételt játékok bevezetésével a társadalmiság modellezése szempontjából nagyon fontos lépést tehetünk meg. Talán a legjelentősebb változást az jelenti, hogy a modellben be kell emelni a társadalmi emlékezet fogalmát, és ezáltal meg lehet ragadni a múlt és a jövő döntésbefolyásoló szerepét. Amíg csak egyszer játszott játékokról beszélünk, addig

nincs idő a modellben, a szereplők csak a "pillanatban élnek", döntéseikben nem befolyásolja őket sem a múlt ismerete, sem a "a jövő árnyéka". Ezt a nyilvánvaló modellezési hiányosságot korrigálja az a kiegészítés, hogy az ismételt játékokban a döntéshozatal során már figyelembe lehet venni mind a múlt megtörtént, mind a jövő várható eseményeit, és ehhez lehet igazítani a döntési stratégiákat.

Azáltal, hogy a játékból játéksorozat lesz, a stratégia fogalma módosul. Az egyszer játszott játékokban a statégia fogalma a döntés (vagy a lépés, a cselekvés) fogalmával volt azonos (de legalábbis összemosható). Az ismételt játékokban viszont stratégia fogalma alatt azt a döntési szabályt értjük, amely szerint a játékos dönt az egyes játékokban. A stratégia tehát egy döntéssorozatot meghatározó szabály. A döntés viszont cselekvésként is értelmezhető, így a stratégiát felfoghatjuk cselekvési szabályként is egyben.

Egy stratégia lehet **feltételes szabály** és lehet **feltétel nélküli szabály**. Feltételességen itt azt kell értenünk, hogy valamely játékban (tehát a döntéssorozat valamely pontján) a döntés során figyelembe veszik-e más – nyilván előző – játékok eredményeit, pontosabban: a játékoknak a játékosok cselekvéseiben megnyilvánuló eredményeit. Ha igen, akkor beszélhetünk feltételes stratégiáról, ha nem, az jelenti a feltétel nélküli stratégiát. Feltétel nélküli stratégia például az, amikor valaki mindig kooperál vagy mindig dezertál mindentől függetlenül, de ebben az értelemben feltétel nélküli az a stratégia is, amely minden fordulóban véletlenszerűen dönt, mert nem a játékosok korábbi döntésétől, azaz más játékok (más döntések) eredményétől függ az aktuális döntés maga, hanem csak a véletlentől. A feltételességet itt az ágensek cselekvéseihez, illetve az azokat meghatározó döntésekhez, szándékokhoz kötjük.

Ha megvizsgáljuk a stratégiák, szabályok logikai szerkezetét, észrevehetjük, hogy a feltételes stratégia másodrendű fogalom, hiszen az aktuális döntés változójában figyelembe kell venni egy korábbi döntés eredményét is. A stratégia itt cselekvési szabályt jelent. A stratégiák jellemzésekor *Ego* cselekvését írjuk le, de ennek során olykor figyelembe kell venni az *Alter* és az *Ego* korábbi cselekvéseit is. A cselekvést leíró $\textsc{do}(agent, a, t_i)$ predikátum paraméterei az ágens, a konkrét cselekvés "azonosítója" és a cselekvés ideje, időpontja. A cselekvők előtt két döntési, választási lehetőség van: vagy kooperálnak vagy dezertálnak. Ebben a dichotóm helyzetben a két alternatíva egymás tagadásának, komplementumának is tekinthető. A kooperáció és dezertálás értékminőségét (valenciáját, polaritását) egy értékfüggvénnyel fejezhetjük ki ($\textsc{value}(x, t_i)$), amelyhez kapcsolódó a pozitív és negatív valenciákat (*Good* és *Bad*) fogalmi primitívekként kell tételeznünk.

(IL.1) $\textsc{g} \equiv \textsc{g}(x, t_i) \equiv \textsc{good}(x, t_i) \stackrel{\text{def}}{=} \textsc{value}(x, t_i) = Good$

jó (HUN) – good (ENG)

(IL.2) $\textsc{b} \equiv \textsc{b}(x, t_i) \equiv \textsc{bad}(x, t_i) \stackrel{\text{def}}{=} \textsc{value}(x, t_i) = Bad$

rossz (HUN) – bad (ENG)

(IL.3) $\textsc{b}(x, t_i) \leftrightarrow \neg\textsc{g}(x, t_i)$

jó és rossz dichotómiája (HUN) – dichotomy of good and bad (ENG)

(IL.4) $\textsc{do}(agent, a, t_i) \rightarrow (\textsc{g}(a, t_i) \leftrightarrow \neg\textsc{b}(a, t_i))$

jó és rosz cselekvés dichotómiája (HUN) – dichotomy of good and bad action (ENG)

(IL.5) $\textsc{c} \equiv \textsc{c}(agent, t_i) \equiv \textsc{cooperation}(agent, t_i) \stackrel{\text{def}}{=} \textsc{do}(agent, a, t_i) \wedge \textsc{value}(a, t_i) = Good$

kooperáció (HUN) – cooperation (ENG)

(IL.6) $\text{D} \equiv \text{D}(agent, t_i) \equiv \text{DEFECTION}(agent, t_i) \stackrel{\text{def}}{=} \text{DO}(agent, a, t_i) \wedge \text{VALUE}(a, t_i) = Bad$

dezertálás (HUN) – versengés (HUN) – defection (ENG)

(IL.7) $\text{D}(agent, t_i) \leftrightarrow \neg\text{C}(agent, t_i)$

kooperáció és dezertálás dichotómiája (HUN) – dichotomy of cooperation and defection (ENG)

Az ismételt játék n lépésig (fordulóig) tart, és a stratégiát leíró szabálynak meg kell adnia minden i. lépésre (vagy másként: mindent t_i időpontra), hogy Ego mit lép. A feltételes cselekvési stratégiák olyan formulákkal írhatók le, amelyekbe valamelyik ágens korábbi cselekvése is be van ágyazva paraméterként, míg a stratégia szabály jellegét a deontikus operátor (**O**) fogja jelezni.

5.4.1 Reciprocitás

A feltételes stratégia (cselekvési szabály) összekapcsol két döntést, a jelen aktuális döntését a múlt vagy a jövő döntésével (esetleg döntéseivel). A feltételes stratégia így egymásra "feleselő", egymásra hivatkozó döntéseket, cselekvéseket köt össze, és ezzel lehetővé teszi azt, hogy modellezni tudjuk a reciprocitás jelenségét. A reciprocitás fogalmát cselekvések időben eltolt viszonzási rendszereként értelmezve a feltételes szabály fogalmával tudjuk megragadni a reciprocitás fogalmának legfontosabb összetevőjét. A reciprocitás jelenségében ugyanis összekapcsolódik az aktuális – jelenbeni – döntés/cselekvés a múltbeli, illetve a jövőbeni döntésekkel/cselekvésekkel, a jelen, a jövő és a múlt összerendelése adja az időben eltolt viszonzási rendszer lényegét (lásd Gouldner cikkét a reciprocitás értelmezéséről [195]).

Fontosnak tartom azonban megjegyezni, hogy egy szempontból nagyon félrevezető itt reciprocitásról beszélni. A reciprocitás fogalmának gouldneri értelmezése ugyanis normatív szinten zajlik (Gouldner összekapcsolódó jogosultságokról és kötelezettségekről beszél), míg a játékelméletnek ezen a szintjén még sehol nincsenek normák. A feltételes szabályokban megtestesülő szankciók, a jutalomként felfogható kooperációk, illetve a büntetésként felfogható dezertálások bár nyilván hasonlítanak a normativitás lényegi mozzanatát alkotó szankcionálási mechanizmushoz, de van egy nagy különbség. A játékelméletben itt még nincs norma, nincs előre rögzített előírás, nincs senkiben igény mások cselekvésére irányuló kötelezettségek vonatkozóan. És mindezen "hiányok" miatt természetesen a legitimitás kérdése sem vethető fel egyelőre.

5.4.2 Direkt reciprocitás játékok

A direkt reciprocitásról akkor beszélhetünk, ha a játékosok ismerik egymást. Kérdés persze, hogy miért is fontos mindez, mit is jelent ez az ismeretség. Amikor ismételt játékokat engedünk meg, akkor azon túl, hogy a modellezni tudjuk a társadalmi emlékezet jelenségét, valami más is lehetővé válik. A játékok ismételhetősége, az a tény, hogy a játékok sorban követik egymást, az teszi lehetővé, hogy a játékosok reagálni tudjanak a másik cselekvésére. Figyelembe vehetik a partnerük (és esetleg a saját) múltbeli cselekvéseit, illetve mérlegelhetik azt is, vajon a jövőben mit lépne a partnerük, ha ők a jelenben így vagy úgy cselekednének. Innentől kezdve a játékosok már alkalmazhatnak feltételes stratégiákat is.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	ismételt
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

Az a feltételezés, hogy az emberek ismerik egymást, nyilván csak kis méretű társadalmat feltételezve valószerű. A direkt reciprocitás érvényességének fontos feltétele ez, amit Anatol Rapoport már a hatvanas években rögzített [431].

A direkt reciprocitás jelenségét **ismételt fogolydilemma** játékként lehet modellezni. Az egyszer játszott fogolydilemma játékban a kölcsönös dezertálás a domináns stratégia. A sok-sok megerősítő példa mellett azonban gyakran tapasztalhatjuk azt, hogy tipikus fogolydilemma szituációkban valamiért mégiscsak megjelenik a kooperáció. Ezt azzal magyarázhatjuk, hogy ezek a szituációk ismétlődő jellegűek, vagyis a benne résztvevők ismerik egymást és gyakran kerülnek hasonló helyzetekbe. A játékelméletben ezt ragadhatjuk meg az ismételt játékok fogalmával. Az ismételt játékok azzal írják le pontosabban a társadalmi cselekvéseinket, hogy ebben a modellben már megjelenik a társadalmi emlékezet jelensége. Ezzel persze rögtön bonyolultabbá is válik a modell, hiszen innentől kezdve már számon kell tartani azt is, hogy a játékosok miket léptek (hogyan döntöttek) az előző fordulókban. Szerencsére az ismételt játékok (ismételt fogolydilemmák) helyzeteit a hetvenes évektől kezdve már lehetett számítógépes szimulációk segítségével is vizsgálni. Az ismételt fogolydilemma helyzetek elemzésének klasszikusa műve Robert Axelrod híres szimulációs tornájáról szóló beszámoló [30].[10]

A direkt reciprocitás játékelméleti modellezéséről a legtöbben a híressé vált tit-for-tat stratégia miatt és révén hallottak, írtak. Nem tagadható, hogy ez a stratégia valóban központi szerepet játszik a játékelméletben, de szeretném bemutatni azt is, hogy emellett vannak még más stratégiák is, amelyek egyrészt bizonyos feltételek között a tit-for-tat stratégiával "összemérhető" módon képesek teljesíteni, másrészt egy teljesnek mondható fogalmi tér hiányzó elemeit, részeit alkotják. Ezért a továbbiakban a direkt reciprocitás körülményei között elméletileg létező, értelmezhető összes másodrendű stratégiát elemezni fogom. Előtte azonban röviden ki kell térnem a stratégiák, pontosabban az evolúciósan sikeres stratégiák kérdésére.

Evolúciósan stabil stratégia

A stratégia fogalmát már bemutattam, és a későbbiekben is lesz még szó róla. Az **evolúciósan stabil stratégia** (Evolutionary Stabile Strategy – ESS) fogalmát azonban még nem vettem tárgyalásba, most viszont megteszem, mert szükségem lesz rá.[11] A fogalom az evolúciós elméletek területéről szivárgott át a játékelmélet vidékére.[12] Az evolúciósan

[10] A direkt reciprocitás játékelméleti modellezéséről, benne a tit-for-tat stratégiáról, bővebben: [517].

[11] A fogalmat John Maynard Smith emelte be a tudományos diskurzusba [500].

[12] A játékelmélet és az evolúcióbiológia kölcsönhatásáról, a tanulás és a kiválasztódás különbségéről máshol már elmondtam a véleményemet [517]. Megengedhetőnek tartom az evolócióbiológia fogalmait,

stabil stratégia fogalmával speciális minőséget lehet kifejezni a stratégiák egymáshoz való viszonyában. Az evolúcióbiológiától átvehető egyik fontos módszertani eszköz a populációdinamikai modellezés technikája. Ha vizsgálni akarjuk, hogy valamely viselkedésmód, magatartási stratégia hogyan képes elterjedni egy nagyobb populációban, akkor azt úgy tehetjük meg, hogy a "versengő" stratégiákat "szétosztjuk" a különböző csoportok között, és szimulációs számításokat végzünk arra vonatkozóan, hogy a csoportok, egyének interakciói során a különböző stratégiák követése mennyi nyereséget, veszteséget hoz a stratégiákat követő egyének, csoportok számára. A stratégiák sikerességét azzal kifejezhetjük ki, hogy a népesség egészében egyre több egyed követi a sikeres viselkedési szabályt. Erre a "kiválasztódási folyamatra" nyugodtan mondhatjuk azt, hogy társadalmi tanulással, nem pedig "biológiai" kiválasztódással valósul meg. A népesség egészében elterjedni látszó – ebben az értelemben – domináns stratégiák azonban még nem biztos, hogy hosszú ideig képesek lesznek stabilan fennmaradni. Fontos kérdés az, hogy mi történik, ha a populáción belül egy domináns stratégia mellett megjelenik egy másik stratégiát követő csoport. Meg tudja-e az új – mutáns – stratégia ingatni a korábban domináns stratégia helyzetét? Saját magával szemben hogy viselkedik a stratégia, megerősíti-e saját magát, vagy esetleg utat nyit más stratégiák számára, amelyek később 'lenyomhatják' őt? Ilyen kérdések merülhetnek fel a domináns stratégiákkal kapcsolatban. Az evolúciósan stabil stratégiák képesek ellenállni az ilyenfajta "támadásoknak", vagyis az adott viselkedési mintázatot tartósan fenn tudják tartani a közösség egészén belül.

A négy másodrendű stratégia

Az ismételt játékok modelljében az az egyik döntő mozzanat, hogy megjelenik benne az emlékezet. A feltétel nélküli stratégiák esetében nincs szükség még erre, de a tit-for-tat és a többi feltételes stratégia már az emlékezet segítségével tudja a stratégia feltételét szabályba foglalni. És ha van már emlékezet, akkor – érdekes módon – megjelenik a jövő is, hiszen ha megismerem a partnerem stratégiáját, akkor ez alapján saját döntésemet formálhatja a jövőre vonatkozó elvárás is. Axelrod szép kifejezésével élve: a jövő árnyéka visszahat a jelenre.

Nem mindegy azonban, hogy mit veszek figyelembe a múltból. A tit-for-tat arra figyel, hogy a partner mit lépett az előző fordulóban. Ez gazdaságos megoldás, mivel nem kell hozzá sok információt tárolni. Minél több patnerünk lehet, annál fontosabbá válik ez a szempont. Két irányban is tágíthatjuk a megjegyzendő múltbeli információk körét. Egyfelől hátranézhetünk messzebb az időben, vagyis több lépésre visszamenőleg figyelhetjük partnereink cselekedeteit. A tit-for-two-tat stratégia például két forduló adatait jegyzi meg: a partner első dezertálása után még kooperál és "csak" a sorban második dezertálást bünteti megtorlással. A távoli múltba nézés azonban nem igazán elterjedt a szimulációs elemzések világában – vélhetőleg azért, mert túlzottan nagy memóriaigénnyel járna, és – talán – nem is lenne annyira életszerű, hiszen a való életben sem tartjuk minden partnerünkről számon azt, hogy pontosan mit tett velünk szemben korábban. Sokkal inkább valamiféle általános benyomást alakítunk ki róluk, de ez már a reputációs elméletekhez vezetne minket el, amivel csak a következő fejezetben foglalkozom

módszereit igénybe venni a játékelmélet társadalomtudományi alkalmazása során, mert tudományos eszközként hasznosnak tartok őket, de mindez egy pillanatig sem jelenti, jelentheti azt, hogy elfogadnám és alkalmaznám a társadalomtudomány világában az evolócióbiológia determinisztikus szemléletmódját.

majd (↬315.o.).

A memória bővítésének másik útja az lehet, hogy továbbra is megmaradunk az előző forduló adatainál, de nemcsak a partnerünk lépését, hanem a sajátunkat is figyelembe vesszük. A tit-for-tat szabály ezt nem teszi meg, de vannak olyan stratégiák, amelyek figyelnek erre. Ha tudni szeretnénk, hogy a direkt reciprocitás feltételei mellett hány stratégia lehetséges elméletileg és mik ezek, akkor érdemes egy egyszerű formalizmust alkalmazni a stratégiák leírására [520].

Egy cselekvési szabály abból áll, hogy bizonyos feltételekhez kooperálást (C), másokhoz dezertálást (D) ír elő. Ezt kifejezhetjük úgy is, hogy a kooperáláshoz az '1', dezertáláshoz '0' értéket rendeljük. Mivel itt a stratégiákat mindig a játékos és partnere előző fordulóban tett lépései alapján határozhatjuk meg, egy cselekvési szabály úgy adhatunk meg, hogy az előző forduló négy lehetséges állapotához, a kölcsönös kooperáláshoz (CC) és dezertáláshoz (DD, illetve a két aszimmetrikus cselekvéshez (CD és DC) egy-egy cselekvésértéket rendelünk, vagyis a (CC,CD,DC,DD) állapotsorhoz hozzátesszük a szabály által előírtakat. Ezt 16-féleképpen tehetjük meg, tehát ennyi stratégia létezik. és az egyes stratégiákat a (0,0,0,0), (0,0,0,1), ..., (1,1,1,1) számsorozatok valamelyikével reprezentálhatjuk.

Nem minden elméletileg lehetséges stratégia "értelmes". Vannak olyanok, amelyek gyakran felbukkannak a szimulációs elemzésekben, és ami fontosabb: a gyakorlati életben. Ilyen a feltétel nélkül kooperáló, az **allC** (1,1,1,1), a feltétel nélküli dezertáló, az **allD** (0,0,0,0), a kockadobó, a **random** (0.5, 0.5, 0.5, 0.5) vagy az alternáló, az Alt (0,0,1,1). A stratégiák között előfordulnak olyanok is, amelyek sosem szerepelnek jól a szimulációk során és a gyakorlati előfordulásuk esélye is kicsi. Ezeknek nevet sem szokás adni, ezért csak a számnégyesükkel hivatkozhatunk rájuk. Ilyen például ez a stratégia: (0,0,0,1). A következő fejezetekben röviden megvizsgálom a négy legfontosabb stratégiát, amelyek valamilyen értelemben sikeresnek mondhatóak a szimulációs számítások és a gyakorlati élet tapasztalatai szerint.

Tit-for-tat, a tiszta viszonzó stratégia

A direkt reciprocitás elmélethez kapcsolható stratégiák legismertebb példája, a tit-for-tat a Robert Axelrod által meghirdetett számítógépes torna kapcsán került be a köztudatba. Axelrod nyilt versenyt hirdetett ismételt fogolydilemma játéksorozatban való részvételre. A versenybe benevezett stratégiák (pontosabban a stratégiákat megvalósító számítógépes programok) sok meneten keresztül egymás ellen játszottak, mindenki mindenki ellen megmérkőzött, a játékosok (stratégiák) minden játék után pontokat kaptak, ezeket folyamatosan összesítették, aztán a torna végén győztest hirdettek. A végeredmény mindenkit meglepett. Az általánosnak mondható elvárásokkal szemben, amelyek valamilyen dezertáló jellegű stratégia győzelmét jósolták, a **tit-for-tat** stratégia lett a torna nyertese, amely viszont tartalmazott kooperatív elemeket is. Mellesleg a tit-for-tat stratégia volt a legegyszerűbb az összes között, hiszen pusztán két szabályból áll: *(i) kezdj kooperálással*, illetve *(ii) mindig ismételd meg a partnered előző fordulóban megtett lépését*. A tit-for-tat terminusra nem született szélesebb körben elfogadott magyar fordítás. Nem azért, mert nincs a tartalmára megfelelő szöveg. Inkább az a gond, hogy túl sok van. Vannak olyan közmondásaink, amelyek egyértelműen megfelelnek a tit-for-tat stratégia lényegének és széles körben elterjedtek: *Jótett helyébe jót várj*, *Ki mint vet, úgy arat*, *Aki*

bottal köszön, annak doronggal felelnek, *Amilyen az adjonisten, olyan lesz a fogadjisten*, *Amilyen a kérdés, olyan a felelet*, *Amit nem kívánsz magadnak, te se tedd embertársadnak*, *Kölcsön kenyér visszajár*, *Valamit valamiért*, *Aki nem dolgozik, ne is egyék*, *Aki kardot ránt, kard által vész el*, *Ki gyomot vet, gazt arat*. A Bibliában is több helyen találhatunk olyan passzusokat, amelyek "ugyanerről beszélnek":

„A ki ember-vért ont, annak vére ember által ontassék ki; ..." [Mózes I. 9.6]

„Szemet szemért, fogat fogért, kezet kézért, lábat lábért,
Égetést égetésért, sebet sebért, kéket kékért." [Mózes II. 21.24-25]

„Törést törésért, szemet szemért, fogat fogért; a milyen sérelmet ő ejtett máson, olyan ejtessék rajta is. A ki barmot üt agyon, fizesse meg azt, de a ki embert üt agyon, halállal lakoljon." [Mózes III. 24.20-21]

„És a bírák vizsgálják meg jól a dolgot, és ha hazug tanú lesz a tanú, a ki hazugságot szólott az ő atyjafia ellen: Úgy cselekedjetek azzal, a mint ő szándékozott cselekedni az ő atyjafiával. Így tisztítsd ki közüled a gonoszt; Hogy a kik megmaradnak, hallják meg, és féljenek, és többször ne cselekedjenek te közötted ilyen gonosz dolgot. Ne nézz reá szánalommal; lelket lélekért, szemet szemért, fogat fogért, kezet kézért, lábat lábért. [Mózes V. 19.18-21]

Végül felidézek még egy utolsó, más kultúrkörből származó, az Edda énekekben található rövid versrészletet [387].

42.
Barátnak barátja légy,
adományt adománnyal
hálálj híven.
Gúnyt fizess meg gúnnyal,
s ne maradj fösvény,
fizetvén gonoszért.

43.
Barátnak barátja légy,
s barátod barátjának
szintén szívese;
ám aki ellened,
egyetlen hívét
se szíveld soha.

A magyar nyelvű játékelméleti szakirodalomban a tit-for-tat leginkább elterjedt fordítása a 'szemet szemért' kifejezés lett, amit azonban a leszűkítő – csak a negatív reciprocitást kifejező – értelme miatt nem tartok szerencsésnek. A tit-for-tat szabály lényege a viszonzás, ami miatt lehetne 'utánzó' vagy 'tükröző stratégiának' is nevezni, de még pontosabbnak érzem az **viszonzó stratégia** terminusát, mert ez fejezi ki leginkább a partner korábbi cselekvésének ismétlésében rejlő mechanikus mozzanatot. Pontossága ellenére sem akarom azonban az általam legjobbnak érzett terminust használni, és maradok inkább az eredeti kifejezés (tit-for-tat) mellett. A fentebb idézett közmondások, maximák értelmezésében egyébként használhatjuk a már hivatkozott megkülönböztetést a pozitív és negatív reciprocitás között. Ha a kooperációra kooperáló válasz a pozitív, míg a dezertálásra dezertáló válasz a negatív reciprocitás, akkor a *jó tett helyébe jót várj* vagy a *kölcsönkenyér visszajár* közmondások a pozitív, a *szemet szemért*, a *Ki kardot ránt, kard által vész el* vagy a *Ki gyomot vet, gazt arat* elvek a negatív reciprocitást írják le, míg a magyar fordítások közül az 'amilyen az adjonisten, olyan lesz a fogadjisten' (illetve kicsit gyengébb formában a *Ki mint vet, úgy arat*, illetve az *Amilyen a kérdés, olyan a felelet*) közmondások írják le a reciprocitás mindkét oldalát.

A szemet szemért elvről annyit még fontosnak tartok megjegyezni, hogy az – legalább Hammurapi törvénykönyve óta – a lex talionis elvének köznyelvi kifejeződéseként is fefogható. És ami még fontosabb: a lex talionis elv az arányos büntetés kívánalmát és gyakorlatát hozta el, amikor megjelent. Éppen ezért ez az elv, megjelenésekor humánus volt a korábbi állapotokhoz, a korábbi büntetési elvekhez és gyakorlatokhoz képest. A

"szemet szemért" elvet akkor úgy kellett érteni, hogy "csak szemet szemért", nem pedig úgy, ahogy korábban tették: "életet szemért". Ezt a történelmi tényt csak azért nehéz észrevenni és értékelni, mert a keresztény kultúrában a szemet szemért elv volt az, amely időben később szembe került egy másik híresssé vált maximával, a jézusi aktív szeretet parancsával. A tit-for-tat stratégia jelentősége abban rejlik, hogy ez egy kooperatív elemeket (is) tartalmazó stratégia, szemben a dezertáló stratégiákkal. A kooperálás mozzanata abban érhető tetten, hogy a stratégia szerint az ellenfél előző menetbeli kooperálását meg kell ismételni vagyis kooperálni kell. A tit-for-tat játékos egy barátságos, kooperatív partnerrel szemben maga is kooperatív módon fog mindvégig viselkedni. Igaz, a tit-for-tat tud barátságtalan is lenni, hiszen az ellenfele dezertálását azonnal megtorolja a saját dezertálásával. Mindezekért jellemezte úgy Axelrod a tit-for-tat stratégiát, hogy az:

- barátságos (mert az első menetben kooperál),
- megtorló (mert a dezertálást azonnal megbosszulja),
- megbocsátó (mert a kooperálásra azonnal kooperálással válaszol, bármi is volt azt megelőzően) és
- nem irigy (mert egy játékban sosem győzi le a másikat, maximum annyi pontot szerez, mint a partnere).

A tit-for-tat stratégia ezeken felül még feltételes stratégia is egyben, hiszen mindig a partner lépéséhez igazítja a saját lépését. Ebben a minőségében szemben áll a feltétel nélküli stratégiákkal (mint a mindig kooperáló allD, illetve mindig dezertáló allD szabály). Ez a feltételesség a stratégia reciprok jellegére utal, s mivel a feltételes kooperálásban legtöbbször tetten érhetünk valamiféle altruizmust is, a tit-for-tat stratégiát a reciprok altruizmus és a direkt reciprocitás) fogalmával is összekapcsolhatjuk. A korábban javasolt jelölési konvenció alapján a tit-for-tat stratégiát az alábbi táblázatos formában reprezentálhatjuk:

Ego és *Alter* cselekvései az előző fordulóban	(CC,	CD,	DC,	DD)
Ego cselekvése az aktuális fordulóban	(C,	D,	C,	D)
	(1,	0,	1,	0)

Ezt a (CC,CD,DC,DD) séma alapján lehet értelmezni, ami az előző forduló lépéseit jelzi. A CC azt jelenti, hogy az Ego (tehát a tit-for-tat stratégiát követő játékos) és Alter (tehát a partner) egyaránt kooperált, a CD azt, hogy Ego kooperált, Alter dezertált, DC esetén Ego dezertált, Alter kooperált és a DD a kétszeres dezertálásra utal. A tit-for-tat stratégia mindig a partner (Alter) lépéseit veszi figyelembe, és azt ismétli meg, így adódik a fenti séma szerint hogy a (1,0,1,0) reprezentáció. A cselekvési szabályra felírhatunk egy formulát is.[13]

(IL.8) $$\textsc{tft}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{c}(Ego,t_{i-1}) \rightarrow (\textsc{d}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i)) \wedge (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{c}(Ego,t_i))) \wedge (\textsc{d}(Ego,t_{i-1}) \rightarrow (\textsc{d}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i)) \wedge (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{c}(Ego,t_i))))$$

tit-for-tat (HUN) – szemet szemért (HUN) – viszonzó stratégia (HUN) – tit-for-tat (ENG) – TFT (ENG)

A formula azt fejezi ki, hogy mit tesz az *Ego* az *i*. lépésben. A viszonyzó stratégiát két állítással definináltuk, ezt fejezi ki a jobb oldali kifejezésben található részformula. Az első lépésben (t_1 időpontban) *Ego* kooperál (*C*), a többi fordulóban ($t_i > t_1$) a játékostársa

[13] Az első fordulóra minden stratégiának egy kivételszabályt kell megfogalmaznia, s mivel az összes vizsgált stratégia kooperálással kezd, ezért ennek jelzését elhagyom a formulákból.

($Alter$) előző (t_{i-1} időpontbeli) lépését ismétli. A tit-for-tat stratégia tehát sikeresnek bizonyult mind a szimulációban, mind a gyakorlatban. A kezdeti ragyogását azonban idővel kissé beárnyékolta az a tény, hogy kiderült, bár a stratégia kollektíve stabil, de evolúciós értelemben már közel sem az [71]. Ha a hibázás lehetőségét is beépítjük a modellbe, figyelembe vesszük a szimulációk során, akkor a stratégia jóval gyengébben teljesít. Ezt persze tudtuk már jóval a szimulációs számítások előtt is, hiszen a szemet szemért (tit-for-tat) elv kritikája kétezer év óta jól ismert a keresztény kultúrkörön belül. Hiszen a jézusi 'ki kardot ránt, kard által vész el' mondat ugyan egyfelől a negatív reciprocitás tényszerű megállapítás, de másfelől egy figyelmeztetés, normatív felszólítás gesztusa is egyben.

> "Szemet szemért-szemért-szemért ... ott ér véget, ahol mindenki vak." (Gandhi)
> "A szemet szemért ősi törvénye mindenkit vakon hagy." (Martin Luther King Jr.)

Bűnbánó stratégia

Ha aszinkron játékokból képezünk játéksorozatot, akkor meglehetősen sikeres stratégiának bizonyul a **szigorú, de igazságos** ('firm-but-fair' – **FBF**) szabály) [176]. Bár a 'szigorú, de igazságos' kifejezést a magyar közgondolkodás is ismeri, alkalmazza, mégis úgy gondolom, hogy jobban járunk, ha a **bűnbánó stratégia** (contrite tit-for-tat strategy) kifejezést használjuk helyette. Az FBF stratégia erősen kooperatív, mert csak egy esetben dezertál: ha az előző menetben a játékos kooperált, miközben a partnere dezertált. Ezt úgy értelmezhetjük, hogy nem is annyira a nagylelkűség, mint inkább a "bűnbánás" a stratégia igazi sajátossága, hiszen ha a játékos korábban dezertált, utána mindkét esetben kooperál (mintegy megbánva korábbi "bűneit"). A tit-for-tat stratégiától a kétszeres dezertálásra való reakcióban tér el azzal, hogy a bűnbánó ekkor kooperál, míg a tit-for-tat dezertál. A tit-for-tat szabály a partner minden dezertálását bünteti, ám a bűnbánó stratégia nem. A kölcsönös dezertálás helyzetét már "szégyenli" (hiszen ő is dezertált), ezért próbálja azt "kijavítani" a kooperálással. A bűnbánó stratégia táblázata a következő:

Ego és $Alter$ cselekvései az előző fordulóban	(CC,	CD,	DC,	DD)
Ego cselekvése az aktuális fordulóban	(C,	D,	C,	C)
	(1,	0,	1,	1)

A stratégiát leíró formula következőképpen néz ki:

(IL.9) $$\textsc{ctft}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{c}(Ego,t_{i-1}) \rightarrow (\textsc{d}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i)) \wedge (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{c}(Ego,t_i))) \wedge (\textsc{d}(Ego,t_{i-1}) \rightarrow \mathbf{O}\textsc{c}(Ego,t_i)))$$

bűnbánó stratégia (HUN) – szigorú, de igazságos (HUN) – contrite TFT (ENG) – Firm-But-Fair (ENG) – FBF (ENG)

A formula három mondatra tagolható: az első állítás az első fordulóban, a következő kettő részformula az összes többi fordulóban érvényes előírást rögzíti. Az első állítás azt fejezi ki, hogy az első fordulóban ($t_i = t_1$) Ego kooperál (C). A második részformula írja le azt a részszabályt, hogy amennyiben az előző fordulóban ($t_i = t_{i-1}$) az Ego kooperált (C), akkor az i. menetben a partnere ($Alter$) előző lépését ismétli. A harmadik összetevőben pedig az szerepel, hogy amennyiben az Ego az előző fordulóban dezertált (D), akkor most kooperál (C). Meg kell még itt jegyezzem, hogy amikor Frean bevezette az FBF-stratégiát, akkor a zajos játékokhoz igazodó sztochasztikus stratégiát tervezett,

így a kétszeres dezertálás esetében nem a biztos, csak a nagyon valószínű kooperálást írta elő, amit ezzel a mintázattal fejezhetünk ki: (1,0,1,p). A fent elmondottak logikája és érvényessége azonban azonban nem változik meg amiatt, hogy a bűnbánó stratégiát a FBF-szabály determinisztikus változataként értelmezem. Érdemes még ehhez azt is hozzátenni, hogy ugyanez a cselekvési szabály felbukkan a 'bűnbánó tit-for-tat' ('contrite TFT') [70], valamint a hírnévápoló ('standing') stratégia név alatt is [512]. Ezekkel azért nem foglalkozom itt, mert mindkettő stratégia az indirekt reciprocitás körébe tartozik.

Számító stratégia

Ahogy Axelrod tornáján Anatol Rapoport pszichológus volt az, aki beküldte a nyertes tit-for-tat programot, úgy az ő nevéhez köthető annak a stratégiának az első megfogalmazása is, amit **együgyű** (simpleton) stratégiának nevezett el [440]. Ez a stratégia később más neveket is kapott, és inkább ezek terjedtek el szélesebb körben. A 'nyer-marad-veszít-vált' (angolul win-stay-lose-shift – WSLS) stratégia [378], valamint a Pavlov megnevezések [290][291] egyaránt a Rapoport-féle együgyű stratégiának felelnek meg, aminek lényegét a 'Győztes csapaton ne változtass!' mondás fejezi ki a legjobban. Ez a szabály a kölcsönösség után vagyis a szimetrikus cselekvések esetén (CC és DD) kooperál, aszimmetrikus helyzetekben (CD és DC) dezertál, amit így írhatunk fel:

Ego és *Alter* cselekvései az előző fordulóban	(CC,	CD,	DC,	DD)
Ego cselekvése az aktuális fordulóban	(C,	D,	D,	C)
	(1,	0,	0,	1)

A stratégia logikája, a 'győztes csapaton ne változtass' elv helyénvalósága azáltal érthető meg, ha belátjuk a stratégia egy fontos vonását. Ez a szabály ugyanis úgy működik, hogy mindaddig megismétli az előző fordulóban tett lépését, amíg nyer (a két legtöbbet fizető helyzetben), de amint veszít (a két legkevesebbet fizető helyzetben), ellenkezőjére váltja a cselekvését. Ha a játékok lehetséges kimeneteinek kívánatosságát egyre növekvő számokkal jelöljük (1-2-3-4), akkor ez a stratégia akkor ismétli meg a korábbi cselekvését, ha 3-at vagy 4-et szerzett az előző fordulóban, és akkor vált cselekvést, ha előzőleg 1-t vagy 2-t kapott. A fogolydilemma esetében a négyféle (CC,CD,DC,DD) helyzetben a játékos (C,C,D,D) lépéseihez a (3,1,4,2) nyereményeket rendelhetjük hozzá. A cselekvési szabály az első és harmadik esetben megismétli az előző lépését, vagyis (C,?D,?), a második és negyedik esetben magatartást vált, vagyis C helyett D-t, illetve D helyett C-t lép. Így alakul ki a teljes szabály mintázata: (C,D,D,C). Ezt a számokkal kifejezve a fentebb már jelzett formában írhatjuk fel: (1,0,0,1).
A stratégia formulája:

(IL.10) $$\textsc{pavlov}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{c}(Ego,t_{i-1}) \rightarrow (\textsc{d}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i)) \wedge (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{c}(Ego,t_i))) \wedge (\textsc{d}(Ego,t_{i-1}) \rightarrow (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i)) \wedge (\textsc{c}(Alter,t_{i-1}) \rightarrow \mathbf{O}\textsc{d}(Ego,t_i))))$$

számító (HUN) – győztes csapaton ne változtass (HUN) – Win-Stay-Lose-Shift (ENG) – Simpleton (ENG) – WSLS (ENG) – Pavlov (ENG)

A háromrészes formulából az első szabály írja le az első forduló teendőjét (*Ego* kooperál). A második részformula szerint, amennyiben az előző fordulóban ($t_i = t_{i-1}$) az *Ego*

kooperált (C), akkor az i. menetben a partnere ($Alter$) előző lépését ismétli. A harmadik részformulában pedig az szerepel, hogy amennyiben az Ego az előző fordulóban dezertált (D), akkor a jelenlegiben pont az ellenkezőjét csinálja annak, mint amit az $Alter$ csinált az előzó menetben ($\neg\text{DO}(Alter, t_{t-1})$).[14] Ez a szabály tanuló stratégia, mivel az előző forduló tapasztalatait értékelve alakítja ki az aktuális lépését. Mivel a szabály arra figyel, hogy a mindig a legtöbb pontszámot érje el, ezért **számító stratégia** megnevezés tűnik a legmegfelelőbbnek. Ez a stratégia is akkor igazán sikeres, ha a tit-for-tat játékosok előzőleg megtisztítják a terepet az önző allD stratégiáktól [440].

Haragtartó stratégia

Az elemzésekben időről időre felbukkan a Grim trigger (GT) stratégia [177], amit **haragtartó**, engesztelhetetlen, zord vagy zordra váltó stratégiának fordíthatnánk (Axelrod 1984-es könyvében ezt Friedman stratégiának nevezi [30]). Ez a cselekvési mintázat barátságosan kezd, és mindaddig kooperál, amíg a másik is ezt teszi, de a partner első dezertálása után ő is azonnal dezertálásba vált, és attól kezdve már sosem tér vissza a kooperáláshoz. Ez a stratégia engesztelhetetlenül haragtartó, sosem bocsát meg. Azok, akik ezt a stratégiát javasolják, az tartják az előnyének, hogy nagy a visszatartó ereje, hiszen a kooperálás feladását az örökös dezertálással fenyegeti, vagyis a legerősebb büntetést helyezi kilátásba. A haragtartó stratégiára a következő mintázat jellemző:

Ego és *Alter* cselekvései az előző fordulóban	(CC,	CD,	DC,	DD)
Ego cselekvése az aktuális fordulóban	(C,	D,	D,	D)
	(1,	0,	0,	0)

A stratégiát leíró formula következőképpen néz ki:

(IL.11) $$\text{GT}(Ego, t_i) \stackrel{\text{def}}{=} ((\text{C}(Ego, t_{i-1}) \rightarrow (\text{D}(Alter, t_{i-1}) \rightarrow \mathbf{O}\text{D}(Ego, t_i)) \wedge (\text{C}(Alter, t_{i-1}) \rightarrow \mathbf{O}\text{C}(Ego, t_i))) \wedge (\text{D}(Ego, t_{i-1}) \rightarrow \mathbf{O}\text{D}(Ego, t_i)))$$

haragtartó (HUN) – Grim Trigger (ENG) – Friedman (ENG)

A formula első állítása rögzíti, hogy Ego kooperál az első menetben. A második részformula szerint Ego előző fordulóbeli kooperálása esetén az aktuális fordulóban megismétli a partnere ($Alter$) előző lépését. A harmadik összetevő pedig azt írja le, hogy amennyiben az Ego az előző fordulóban dezertált (D), akkor most is dezertálni fog (D) függetlenül attól, mit tett korábban a partnere. A szimulációs eredmények azt mutatják, hogy ez a stratégia sosem tud igazán jól szerepelni, mert nagyon könnyen keveredik megtorlási spirálba [29]. Különösen így van ez zajos játékokban (erről később kicsit bővebben írok).

A négy alapstratégia

Ez előző fejezetekben bemutatott stratégiák mind olyanok voltak, hogy a szimulációs számítások és a gyakorlati tapasztalat szerint is a sikeres, bizonyos feltételek mentén gyakran alkalmazott cselekvési szabályoknak lehet minősíteni őket. Talán csak a haragtartó (Grim trigger) stratégia volt az, amire valamivel kevesebbszer hivatkoztak a többi

[14] Itt kihasználhatjuk azt, hogy a játékosok cselekvései csak két értéket vehetnek fel, vagy kooperálnak vagy dezertálnak, s emiatt a két cselekvésérték egymás tagadásának tekinthető.

háromhoz képest. A bemutatott stratégiáknak van egy érdekes közös vonása, ami talán nem véletlen. Mind a négy szabály megegyezik abban, hogy mit tesz azután, hogy a játékos kooperált az előző fordulóban (modhatjuk azt is, hogy a játékos jó volt):

- kooperál, ha a partnere kooperált,
- dezertál, ha a partnere dezertált.

Ezt fejezi ki az, hogy a számnégyesük első két helyén mindannyiszor az '1,0' érték szerepel. Ha a négy vizsgált stratégia megegyezik abban, hogy mit ír elő arra az esetre, amikor az előző fordulóban a játékos kooperált (jó volt), akkor a különbség csak abban lehet köztük, hogy mit ír elő a stratégia arra az esetre, amikor az előző fordulban a játékos dezertált (vagyis rossz volt). Ilyen esetből viszont pont négy van (nem lehet se több, se kevesebb), és az az érdekes, hogy pont ez alapján jellemezhetjük (különíthetjük el) a vizsgált stratégiáinkat is:

		(CC,	CD,	DC,	DD)
viszonzó	TFT	(1,	0,	1,	0)
bűnbánó	FBF	(1,	0,	1,	1)
számító	WSLS	(1,	0,	0,	1)
haragtartó	GT	(1,	0,	0,	0)

A direkt reciprocitáson alapuló kooperációs lehetőségeket keresve, a kooperatív stratégiákat értelmezve azt mondhatjuk, nem lehet véletlen, hogy a fenti négy stratégia szerepel a legtöbbet az elemzésekben (és tegyük hozzá: a mindennapi életünkben). A kooperációs minták elterjedésének feltétele lehet az a "félszabály", hogy ha előzőleg kooperáltunk (jók voltunk), akkor az viszonzás elve szerint, tehát a tit-for-tat stratégia szerint kell eljárnunk. Az előző menetbeli dezertálásunk (tehát a rosszaságunk, a rossz múltunk) esetén adott négyféle cselekvési lehetőség pedig azt a négy stratégiát írja le, amelyek kooperációs potenciálját a játékelmélet is alátámasztotta, és amiket a mindennapi életben mi is megtapasztalhattunk, alkalmaztunk évezredek óta, alkalmazunk a jelenben is.

A bemutatott formulák alapján szembetűnő, hogy olyan két stratégia van, amely teljesen a partner előző lépéséhez köti az aktuális cselekvését (a teljesen ugyanazt viszonzó tit-for-tat és a részben hasonló, részben ellenétes lépést viszonzó pavlovi szabály), és a két másik stratégia az *Ego* korábbi dezertálása esetén a partner előző lépésétől függetlenül cselekszik (a bűnbánó kooperál, a haragtartó dezertál). Természetesen észrevehető az a korábban már említett közös mintázat is, miszerint az *Alter* megelőző fordulóbeli kooperálása ("jósága") esetén mindig a partner lépését "kell" megismételni. A haragtartó és a bűnbánó stratégiák abban az értelemben merevebbek a másik kettő szabályhoz képest, hogy csak "egyszeresen" igazodnak a másik játékos korábbi lépéseihez (amikor az *Ego* kooperált az előző fordulóban). Ha úgy kerülnek a partnerükkel szembe, hogy az előző menetben dezertáltak, akkor mindketten "feltétel nélküli" stratégiát követnek (a bűnbánó stratégia kooperál, a haragtartó dezertál). Ehhez képest erősebben igazodik a tit-for-tat szabály a másikhoz, hiszen függetlenül attól, hogy *Ego* mit tett korábban, mindig az *Alter* lépését ismétli. Ez a szabály egy fokkal azonban még mindig merevebb a WSLW-stratégiához képest, hiszen az utóbbbi szabály hol a partner korábbi lépését ismétli, hol annak tagadását. Talán ez a "másság" vagy "képesség" magyarázhatja azt, hogy a szimulációk során – legalábbis bizonyos feltételek között, kiváltképp zajos környezetben – miért ez a stratégia bizonyult a legjobbnak. Amikor erre a pavlovi stratégiára azt mondják [290], [291], hogy tanul az előző fordulók eseményeiből, akkor ez a tanulási képesség talán épp ennek a "rugalmasabb" igazodásnak köszönhető.

Társadalmi hiba – zajos játékok

Az előző fejezetekben megírtam, hogy az ismételt fogolydilemma játék (Iterated Prisoner's Dilemma Game – IPDG) bevezetésével modellezni tudjuk a társadalmi emlékezetet és ezzel a cselekvéseink viszonzási rendszeren keresztüli összekapcsolását. Az egymásra hivatkozó, egymásba kapcsolódó cselekvési láncok megfelelő stratégiák követésével alakíthatók ki és tarthatók fenn. Ezen stratégiák között a legismertebb a tit-for-tat szabály, aminek megtalálhatjuk a nyomát minden kultúrában. És – legalább a keresztény-zsidó kultúrából – ismert a tit-for-tat stratégia talán legfontosabb hátránya is: a szemet szemért elv könnyen "rálöki" az embert a megtorlási spirál pályájára. A történelem, a mindennapi élet tapasztalatai azt mutatják, hogy ha mindig a tit-for-tat szabályhoz ragaszkodnánk mereven, akkor egyetlen tévedés, kicsiny kis hiba is elegendő lenne ahhoz, hogy az egymásra feleselő cselekvések a dezertálási-büntetési ágra váltsanak át (vagyis a megtorlási spirálba keveredjenek bele). Axelrod tornájával szemben elég hamar megfogalmazták azt a kritikát, hogy ebből a szempontból nézve súlyos deficittel küzd, mivel nem képes kezelni a társadalmi hibázás lehetőségét. A számítógépes progamok sosem hibáznak. Az éppen adott szabályt minden lépésben pontosan végrehajtanak. Ebben az értelemben nagyon sterilek, és nem képesek modellezni azt a tényt, hogy a valós emberek – bár sokszor követnek valamely stratégiát – néha eltérnek a cselekvési szabály előírásaitól. Többféle oka is lehet ennek. Lehetséges, hogy a cselekvő maga téveszt, és eltér a szabály által előírtaktól. Ilyenkor – Reinhard Selten szép kifejezésével élve – a remegő kéz (trembling hand) jelenségéről beszélhetünk. De előfordulhat az is, hogy mások hibáznak abban az értelemben, hogy félreértik (akaratlanul vagy akarattal) a cselekvésünket. A téves percepció (fuzzy mind) ugyanúgy befolyással lehet játéksorozat kimenetére, hiszen végülis az a fontos, hogy a játékosok mit észlelnek egymás cselekvéseiből.

A számítógépes szimuláció során a tit-for-tat stratégia sosem kerülhet a megtorlási spirálba, míg a valós életben egy ember, aki általában követi a tit-for-tat szabályt, olykor mást léphet, mint amit a szabály szerint éppen lépnie kellene. Ezt a hibázási lehetőséget a játékelméleti modellekbe úgy "vehetjük fel", hogy bevezetjük a zajos játékok fogalmát [179]. Ez annyit jelent, hogy amikor az egyes stratégiákat követő játékosok közöti játéksorozatot szimuláljuk, akkor a bizonyos valószínűség mellett – mintegy "kívülről beavatkozva" –megváltoztatjuk egy-egy lépés tartalmát (tehát ha valakiken kooperálnia kéne, a lépését dezertálásra váltjuk, és fordítva). Ezzel beviszünk egy hibázási faktort a rendszerbe, és a nagy kérdés az lesz, vajon milyen stratégiák lesznek sikeresek ilyen feltételek között. A válasz tulajdonképpen nem meglepő. Olyan stratégiák képesek kezelni a megváltozott – tegyük hozzá: a valóságot jobban modellező – helyzeteket, amelyek egyrészt jóindulatúbbak a tit-for-tat stratégiához képest, másrészt tanulni képesek a múlt – jó és rossz – döntéseiből. Az előzőre a nagylelkű tit-for-tat (Generous Tit-For-Tat – GTFT) szabály a példa [382], az utóbbira a pavlovi ("győztes csapaton ne változtass" elvet megvalósító) stratégia [378]. A GTFT-t azért mondják nagylelkű tit-for-tat szabálynak, mert – a stratégia követésébe egy belső valószínűségi változót csempészve – bizonyos pontokon, amikor dezertálnia kellene az előírás szerint, mégis kooperál. Néha nagylelkűen "elnézi" a partner dezertálását. Ez a szabálymódosítás elég lehet ahhoz, hogy a társadalmi hiba miatt beinduló - és egyébként végtelenné váló – megtorlási spirált meg lehessen szakítani, és vissza lehessen téríteni a kooperálási ágra a játék menetét. A pavlovi (vagy más néven Win-Stay-Lose-Shift) stratégiát már bemutattam korábban, a szabály logikáját egy korábbi fejezetben leírtak alapján lehet megérteni (↫311.o.).

5.5 A játékok közösségi kiterjesztése reputációval

Említettem, hogy a direkt reciprocitás jelensége csak kis létszámú társadalmakban működik, hiszen az a feltétel, hogy a résztvevőknek ismerniük kell egymást, nem enged meg nagyobb létszámot. Viszont a helyi közösségek növekedés sok helyen nagyobb létszámú társdalmakat eredményezett, ám ezekben a társadalmakban is megtalálhatóak a kooperáció mintázata, itt is van reciprocitás. A kérdés tehát az, hogy miként, milyen formában, miért alakulnak ki mindezek? A társadalmi modellünket ezen a ponton újra csak bővítenünk kell, mégpedig egy nagyon fontos társadalmi jelenséget kell felvennünk: a **reputáció** mint **társadalmi értékelés** jelenségét. Ahhoz pedig ezt modellezni tudjuk, szükségünk lesz még arra is, hogy a 'közösséget' mint az értékelést végző új "szereplőt" is felvegyük a modellbe a játékosok mellé.

5.5.1 Indirekt reciprocitás játékok

A direkt reciprocitás jelensége kialakítható és fenntartható azzal, hogy kooperálás és dezertálás (büntetés) valamilyen egyvelegéből kialakított stratégiát érvényesítünk. Az ilyen feltételes stratégiák azonban csak akkor és azon az alapon működnek, ha feltesszük, hogy a játékosok ismerik egymást. A kooperációt azonban megtalálhatjuk olyan ismétlődő társadalmi helyzetekben is, amelyekben idegenek kerülnek kapcsolatba egymással. Meg kell tudnunk magyarázni az olyan eseteket is, amikor a játékok ismétlődnek, de a játékosok ismeretlenek egymás számára. Az ilyen helyzeteket nevezzük **indirekt reciprocitásnak**.

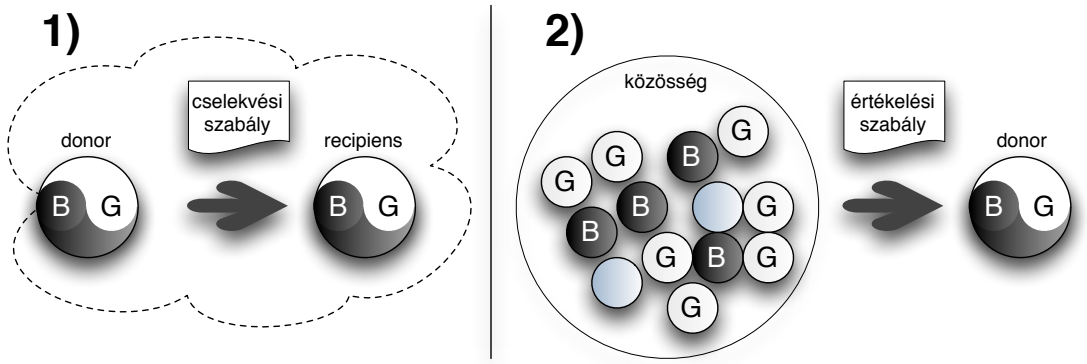
magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	ismételt
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	nem kooperatív (ígéretmentes)

Az indirekt reciprocitás megragadásához nem elég az eddig fogalomkészlet, szükségünk van egy rendkívül fontos társadalmi jelenség, az értékelés, a reputációképzés modellbe emelésére. Ez az értékelési jelenség a reputáció, ezért az indirekt reciprocitás elmélet megnevezés mellett előfordul, hogy a kutatók a reputációs elméletek kifejezést használják.

Innentől kezdve némiképp megváltozik a terminológia is. A cselekvők jelzésére eddig használt *Ego* és *Alter* nevek helyett a **Donor** és a **Recipiens** terminusokat alkalmazzuk. Ez a terminológiai váltás arra utal, hogy a reputációs elméleteken belül a kooperáció voltaképp az ismeretlenek megsegítését jelenti. Eddig az volt a kérdés, hogy 'miért kooperál *Ego* az Alterrel szembe kerülve', mosttantól viszont azt vizsgáljuk, hogy 'miért nyújt segítséget a *Donor* a *Recipiens* számára'. A *Donor* terminus aktívabban fejezi ki a kooperálás, segítségnyújtás tényét, elvárását, lehetőségét, míg a *Recipiens* terminus jóval inkább utal e szerep passzivitására az *Alter* terminushoz képest. *Donornak* és *Recipiensnek* speciálisabb a jelentése, viszont "messzebb vannak" egymástól, a két szerep sokkal kevésbé kicserélhető *Ego-Alter* szereppárhoz képest.

Az indirekt reciprocitás, a reputáció modelljében fel kell vennünk egy teljesen új szereptípust is, az értékelést, a reputációképzést végző közösség fogalmát. Erre a reputációs elméletek kevéssé, de legalábbis nem eléggé reflektáltak, pedig nagyon fontos ez a szerep. Hogy miért is? A direkt reciprocitás feltételei között a kooperáció kialakulásához még "csak" annyi kellett, hogy a játékosok emlékezzenek mind a saját, mind a társaik korábbi cselekedeteire, majd ezek alapján végrehajtsák azt, amit a stratégiájuk "előír". A gond az, hogy az indirekt reciprocitás helyzetében semmilyen információ nem áll rendelkezésre a partner korábbi cselekedetéről, hiszen ismeretlenekről van szó. Itt is kell valami fogódzó a döntés meghozatalához, a kérdés csak az, hogy mi lehet az. Megoldás lehet, ha valamilyen értékjelzést ragasszunk minden játékoshoz, ami orientálhatja a *Donort* abban, segítsen-e vagy sem partnerének. Ez az értékjelzés a **reputáció**, amit minden játékoshoz egyedi módon hozzá kell rendelni, hiszen ismeretlenek között sosem tudhatjuk, kivel hoz össze a sors, *Donorként* kikkel kerülhetünk olyan helyzetbe, amikor döntenünk kell arról, segítsünk-e neki. Ha mindenkin van értékjelzés, ha mindenkinek van reputációs címkéje, akkor ennek segítségével már képesek lehetünk dönteni. A *Donor* cselekvési stratégiájában nem a korábbi cselekvéseket, hanem a reputációs értékeket kell figyelembe venni.

A reputáció kezelésére egy "fekete-fehér világképen" alapuló mechanizmust működtethetünk. Elég azt mondani, hogy az emberek vagy jók (Good – G) vagy rosszak (Bad – B), és minden értékelés során e két értékből lehet választani és valamelyiket a *Donorhoz* rendelni. Talán meglepő, de majd látni fogjuk, hogy ez a primitív megoldás meglepően jól modellezi a társadalmi értékelés valós folyamatait. Egy fogalom jelentését azonban át kell értelmeznünk: a stratégia fogalmát ki kell terjeszteni. Stratégia fogalma alatt eddig azt cselekvési szabályt értettük, amely alapján a játékos meghatározhatta, hogyan döntsön, miket lépjen a játéksorozat bármely pontján. Erre a viselkedési szabályra továbbra is szükség lesz, viszont emellett a stratégiának ki kell terjednie még arra is, hogy miként lehet meghatározni a játékosok reputációját. Ehhez szükség van valamilyen értékelési szabály (mechanizmus, módszer) alkalmazására. Ezt az értékelési szabályt azonban már nem a *Donor*, hanem a közösség veszi birtokba. Az a közösség, amely 1) "figyeli" azt, hogy a *Donor* saját cselekvési szabálya alapján miként dönt arról, segíti-e a *Recipienst* vagy nem, és ez alapján a közösség 2) az értékelési szabály segítségével értékeli *Donor* cselekedetét, vagyis felülvizsgálja a *Donor* reputációs értékét.



Az indirekt reciprocitás stratégiafogalmát kiterjesztve mindig tudnunk kell, hogy milyen viselkedési szabály szerint dönt a *Donor*, illetve milyen értékelési rendszer alapján

értékeli (újra) a *Donor* reputációját a közösség a *Donor* cselekvése alapján. A cselekvési szabályokról bővebben írtam másutt, itt elegendőnek tartom hivatkozni az ott leírtakra [517]. Az értékelési mechanizmusok áttekintését viszont annyira fontosnak tartom, hogy itt röviden bemutatom a legfontosabb reputációs rendszereket.

Első körben a négy másodrendű értékelési rendszert veszem sorba, majd utána bemutatom azt, hogy a harmadrendű értékelési rendszereket hogyan lehet modellezni a már ismert másodrendű rendszerek segítségével. A másodrendű értékelési szabály fogalma azt jelenti, hogy az értékelés során két dolgot veszünk figyelembe: a *Donor* cselekvését, illetve a *Recipiens* reputációját. A harmadrendű értékelés ezeken felül figyel a *Donor* reputációjára is [517]. Az indirekt reciprocitás, a reputáció típusainak megragadásához további formulákat kell felvennünk a formális modellünkbe. Valamilyen logikai elemmel meg kell tudnunk ragadni a reputáció, vagyis a közösségi értékelés jelenségét. Erre a célra alkalmas lehet egy **A** deontikus operátor felvétele a leíró nyelvünkbe, amivel – a már ismert $\textsc{value}(agent, t_i)$ értékelési függvényünkkel együtt – kifejezhetjük azt, hogy egy értékelési szabály szerint jó és rossz minősítést (reputációt) kell kiosztani az ágensekre.[15]

(IL.12) $\mathbf{A}\{\textsc{value}(agent, t_i) = Bad\}$

az ágens rossz reputációja (HUN) – bad reputation of agent (ENG)

(IL.13) $\mathbf{A}\{\textsc{value}(agent, t_i) = Good\}$

az ágens jó reputációja (HUN) – good reputation of agent (ENG)

A *Donor* saját cselekvési szabálya szerint cselekszik, ami a most vizsgált esetekben mindvégig egy nagyon egyszerű, másodrendű szabály lesz. A **diszriminatív kooperálás** (Disc:Co) stratégiáját fogja az öszes vizsgált módszer alkalmazni (a jó *Recipienssel* szemben kooperál, a rossz partnerrel szemben dezertál).[16] Az értékelő közösség saját reputációs szabálya szerint ítéli meg a *Donor* aktuális cselekvését, és nyilván különbözni fognak egymástól a most bemutatandó értékelési módszerek. Egy valamit kell még tisztázni, mielőtt a reputációs szabályok bemutatásába kezdenék: meg kell adni, hogyan lehet összekapcsolni a cselekvéseket és az értékeléseket egymással. Ezt a következő szabállyal (formulával) fejezhetjük ki.

(IL.14) $\textsc{do}(Ego, a, t_i) \rightarrow \textsc{value}(Ego, t_i) = \textsc{value}(a, t_i)$

a cselekvés és értékelés összekapcsolása (HUN) – connection of action and assessment (ENG)

Ez a formula azon az alapon teremti meg a kapcsolatot a cselekvés és az ágensek minősítése között, hogy a cselekvésekre, illetve ágensekre vonatkozó értékelés valenciája megegyezik (például a kooperációra mint cselekvésre és a jó cselekvőre mint ágensre irányuló értékelés valaencája mindkét esetben pozitív).

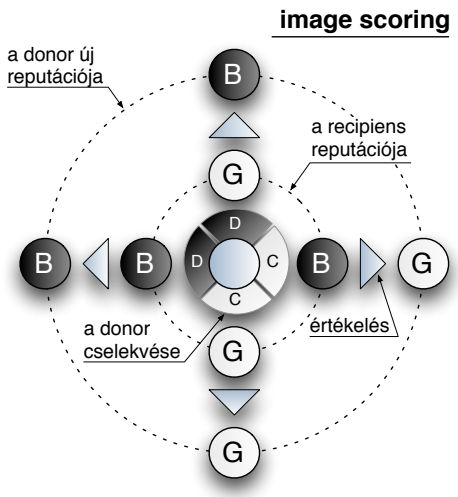
Kép-mutató (Image Scoring) értékelési rendszer

Az indirekt reciprocitás első elmélete az image scoring módszertant mutata be, mint a reputációs rendszerek prototípusát [380]. Szigorú módszertani szempontok szerint ez az

[15] Mivel mindegyik itt vizsgált értékelési szabályra igaz az, hogy az első fordulóban egységesen jó minősítéssel kezd, ezért ezt a kivételszabályt nem veszem itt figyelembe.

[16] Erről bővebben lásd: [517]

értékelési rendszer nem is másod-, hanem elsőrendű, de ez nem változtat azon a tényen, hogy ebben a minőségében is az elméletileg lehetséges négy másodrendű értékelési rendszer egyik típusát jelenti.



A többihez hasonlóan ez a módszertan is azzal indít, hogy mindenkinek 'jó' minősítése van a játékban. A cselekvési szabály a már említett diszkrimanítív kooperálás szabálya: Jóval szemben kooperálni, Rosszal szemben dezertáni kell. Az Image scoring módszer lényege az, hogy minden egyes döntés után a *Donor* képe (imidzse) aszerint változik, hogy segítette-e a *Recipienst* vagy nem. Ha a *Donor* kooperált, vagyis segített, akkor onnatól Jó lesz (akármi is volt addig), dezertálása esetén pedig, vagyis amikor nem segített, a *Donor* 'rossz' minősítést kap (akármi is volt addig a reputációja). Az első reputációs elméletről szóló cikkben a szerzők (Nowak és Sigmund) azt hozták ki, hogy ez a módszertan képes stabil kooperációt kialakítani egy közösségen belül, de a későbbi szimulációs számítások árnyaltak ezen a képen. Egyesek olyan – katalizáló – szerepet tulajdonítanak ennek a stratégiának, mint amilyet szerepet a TFT-stratégia tölt be a direkt reciprocitást szimuláló modellekben [381]. Mindenesetre születtek később olyan publikációk is, amelyek az Image scoring módszer tényleges érvényesülését mutatták ki az állatvilág bizonyos szegmenseiben.[17] A módszer formális reprezentációja:

(IL.15) $$\textsc{is}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{g}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) = \textsc{value}(a,t_i)\}) \wedge (\textsc{b}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) = \textsc{value}(a,t_i)\}))$$

kép-mutató (HUN) – Image Scoring (ENG)

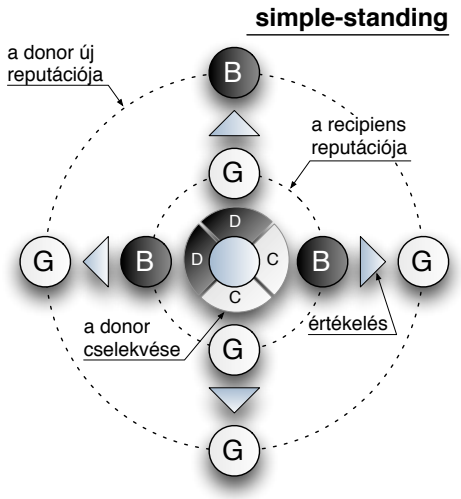
Az Image scoring módszert nemcsak azzal támadták később, hogy sok esetben nem képes a kooperáció tartos kialakítására, de "érték" alapon is kritizálták sokan. Ez a módszer ugyanis "érzéketlen" a *Recipiens* reputációjára.[18] Márpedig – szóltak az ellenérvek - – ha nem segít valaki a *Recipiensnek*, az nem feltétlen jelent rosszat, amiért a *Donort* reputációját "le kéne húzni". Mert mi van akkor, ha a *Donor* büntetésből nem segít a rossz *Recipiensnek*? E kérdésből kiindulva kezdték el keresni a további értékelési rendszereket.

Egyszerű hírnévápoló (Simple Standing) értékelési rendszer

A Simple Standing értékelési rendszer abban tér el az Image Scoring módszertől, hogy a rossz *Recipienssel* szemben megengedi a büntetést. Ez abban nyilvánul meg, hogy ha a *Donor* egy rossz *Recipiensnek* nem segít (büntetésből), akkor ezt a Simple Standing módszer helyesli és a *Donort* ezek után Jó minősítéssel látja el.

[17] Íme egy beszédes című cikk erről: "Image scoring and cooperation in a cleaner fish mutualism": [83].

[18] Ezen az alapon mondható egyébként elsőrendűnek ez a módszertan, hiszen csak a *Donor* cselekvését veszi figyelembe.



Ez a módszertan az egyik legtöbbet sulykolt értékelési rendszerünket modellzi a valós társadalmi életünkből. A gyerekeinket úgy neveljük fel (és felnőttkorunkban is ezerszer elmondjuk magunknak, másoknak, egymásnak) azt a mondatot, amely a Simple Standing módszertant megalapozó elvnek tekinthető:

a rossz elnyeri méltó büntetését

A büntetés igazolását adja meg ez az értékelési rendszer, ezért olyan fontos. A mindennapi narratíváink egyik legfontosabb összetevője ez az elv. Ezáltal tudjuk nagy létszámú társadalmakban is fenntartani a kooperációs mintákat, mert lehetővé teszi a büntetés intézményét az ismeretlenekkel, idegenekkel szemben úgy, hogy a büntetésben megnyilvánuló – elsőrendű – dezertálást társadalmilag igazoltnak, megengedhetőnek minősítsük.[19] Ez a stratégia a Standing metódus másodrendű változata [388], [389], [390]. A különbség köztük abban van, hogy ez az egyszerűsített változat jónak minősíti a rossz személy visszautasítását akkor is, ha a *Donor* maga előtte rossz volt (az eredeti hírnév-gyűjtő stratégia ezt a diszkriminációt csak a jó *Donorokkal* szemben alkalmazta). Ez annyit jelent, hogy ez a szabály jónak tartja azt is, ha a hipokrita játékos bünteti a *Rosszat*. A másodrendű változat alkalmazásának "célja" az értékelési feladat egyszerűsítése lehet, hiszen ebben az esetben kevesebb információt kell beszerezni és feldolgozni. A szabály formulája:

(IL.16) $$\textsc{ss}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{g}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) = \textsc{value}(a,t_i)\}) \wedge (\textsc{b}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{g}(Ego,t_i)\}))$$

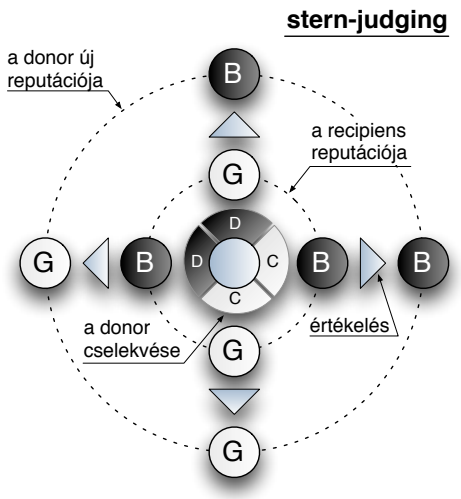
egyszerű hírnévápoló (HUN) – Simple Standing (ENG)

A Simple Standing értékelési módszer lényege tehát az, hogy "bevezeti" a büntetés intézményét. Ebben azonban nem még csak az első lépést teszi meg, hiszen ez a stratégia azzal ismeri el a büntetést, hogy jutalmazza azokat, akik megbüntették a *Rosszakat*. Kitart viszont azon értékelés mellett, hogy a kooperáció feltétlen jó dolog, ha tehát valaki kooperál, akkor őt mindenképpen jónak kell minősíteni.

Kérlelhetetlen (Stern Judging) értékelési rendszer

A kérlelhetetlen (Stern Judging) értékelési rendszer még egy lépést tesz a büntetés irányába a Simple Standing módszertanhoz képest. Az utóbbi ugyanis egy ponton még "elnéző" a Rosszakkal szemben, és ezen a ponton lehet még szigorítani a rossz *Recipiensekkel* szembeni fellépésen.

[19] Idegeneken itt azokat kell érteni, akiknek a cselekvéseire nem látunk rá, ezért csak az általános reputációs címke alapján döntünk arról, hogy segítsünk-e nekik vagy sem.



A Stern Judging [401] megegyezik egy jóval korábban megfogalmazott elképzeléssel, amit kandori-módszernek neveztek el [272]. A stratégia két alszabályból áll:

(i) nő a *Donor* reputációja, ha segíti a Jót és nem segíti a *Rosszat*, illetve

(ii) csökken a *Donor* reputációja, ha segíti a *Rosszat* és visszautasítja a *Jó* támogatását.

Ez a módszertan is egy tényleges társadalmi gyakorlatban megtalálható elv reprezentációjaként fogható fel. Vannak társadalmi állapotok, amikor sokan a *rossz elnyeri a mető büntetését* elvnél is erősebb elvárásaink vannak a *Donorokkal* szemben, és megköveteljük tőlük, hogy ne csak büntessék a *Rosszakat*, de ne is segítsék őket. Ez utóbbi új mozzanat azt jelenti, hogy ez a módszertan elítéli a rossz *Recipiens* támogatását, vagyis a cselekvő reputációját rosszra változtatja, ha az kooperált egy rossz partnerrel. Ezt fejezi ki az alábbi mondás:

vétkesek közt cinkos, aki néma

Ez az elv azért szigorúbb az előzőnél, mert a kooperációt (ami önmagában eddig jónak minősült) nem tartja elfogadhatónak, helyeselhetőnek akkor, ha az rossz *Recipiensre* irányul. A mondás azt fejezi ki, hogy vannak olyan helyzetek, amelyekben a hallgatás a *Rosszak* között beleegyezést, elfogadást, tehát kooperációt jelent, és ez – e módszertan szerint – elfogadhatatlan, tehát büntetendő. A stratégia másodrendű értékelésen alapul, és csak annyiban tér el a hasonló – harmadrendű – 'ítélkező' (judging) stratégiától, hogy azt is jutalmazza, ha *Rossz* büntet *Rosszat*, ami annyit jelent, hogy a hipokrita magatartást is jónak ítéli. Mindegy, ki büntet (*Jó* vagy *Rossz*), csak büntesse meg a *Rosszat*. Formulája a következő:

(IL.17) $$\textsc{sj}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{g}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) = \textsc{value}(a,t_i)\}) \wedge (\textsc{b}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) \neq \textsc{value}(a,t_i)\}))$$

kérlelhetetlen (HUN) – Stern Judging (ENG)

Jól látszik az értékelési szabály formuláján, hogy mindkét feltétel (jó vagy rossz *Recipiens*) esetén reciprok módon alakul ki a *Donor* új reputációja, csak amíg a jó *Recipienssel* szemben az új reputáció megegyezik, addig a *Rossz* esetében ellentétes lesz a *Donor* cselekvésének értékelésével.

Elkerülő (shuning) értékelési rendszer

A Shunning értékelési rendszer próbálja még tovább növelni a "büntetési faktort", ami első pillanatra talán nehezebben vehető észre, ha a módszertan ábrájára tekintünk, mert a Stern Judging módszertanhoz képest annyi a változás benne, hogy a *Rosszal* szemben alkalmazott büntetést (nem segítést) most rossznak minősíti (ezzel "törli" a Simple Standing módszer által bevezetett szabályt a rossz megbüntetésének helyességére

vonatkozóan). A módszertan kevesebb figyelmet kapott a többi három másodrendű értékelési rendszerhez képest, pedig fontos társadalmi jelenség modellezésére alkalmas [404], [537], [390]. A Shunning stratégia logikája (két alszabálya) a következő:

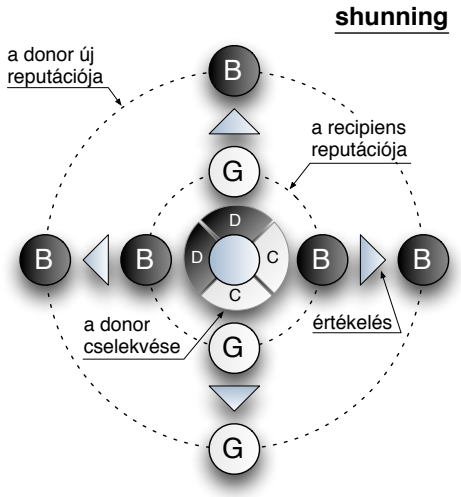
(i) mindig bünteti a dezertálást és

(ii) minden esetben bünteti a *Donort*, ha az *Rossz* személlyel kerül kapcsolatba – még akkor is, ha egyébként a *Donor* segíti a *Recipienst*

A stratégia másodrendű értékelést használ és a szimulációs kísérletek eredményei szerint hatékonyan hozzájárul a kooperációs minta elterjedéséhez. Ez a stratégia arra a közkeletű mondásra emlékeztet, amely szerint

aki korpa közé keveredik, azt megeszik a disznók

vagyis tehet az ember bármit, akár kooperál, akár dezertál, ha Rosszak közé kerül (értsd: rosszz *Recipiensekkel* kerül kapcsolatba), menthettlenül Rosszá válik ő maga is.



A módszer névválasztásával azt akarták kifejezni, hogy a stratégia a Rosszak kizárására, elkerülésére ösztönözhet, mivel a velük kapcsolatos bármilyen érintkezés "káros" lehet a *Donor* számára. Első hallásra talán meglepő, nehezebben értelmezhető, elfogadható ez az értékelési módszer, pedig a Rosszakkal való érintkezés büntetése meglehetősen gyakran előfordult a történelemben [517], sőt, várhatóan a jövőben is találkozni fogunk ezzel a jelenséggel. Ez az értékelési rendszer a kizárásra törekvő, a szegregációs szándékok mélyén meghúzódó logika. Vannak társadalmi állapotok, amikor társadalmi csoportok között túl naggyá válnak a társadalmi, kulturális távolságok, és elérkezhet egy pillanat, amikortól az egyik közösség hirtelen átvált a Shunning módszertanra, mert a közösségközi konfliktusokat úgy vélik feloldhatónak, ha a két közösséget teljesen elszeparálják egymástól. Ilyenkor veszik elő ezt a módszertant. A szimulációs számítások egyébként ezt az értékelési rendszert kooperációs minták kihordására alkalmatlannak szokták minősíteni (és a történelmi tapasztalatok sem igazán arról szólnak, hogy a szegregáció képes lett volna megoldani a közösségközi konfliktusokat). Formulája:

(IL.18) $$\textsc{sh}(Ego,t_i) \stackrel{\text{def}}{=} ((\textsc{g}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{do}(Ego,a,t_i) \rightarrow \textsc{value}(Ego,t_i) = \textsc{value}(a,t_i)\}) \wedge (\textsc{b}(Alter,t_i) \rightarrow \mathbf{A}\{\textsc{b}(Ego,t_i)\}))$$

elkerülő (HUN) – Shunning (ENG)

A kirekesztés, a szegregáció szabályának formuláján érezni a *Rossz* dominanciáját, hiszen a jó *Recipiens* esetén a reputációs értékelés megyezik a cselekvés értékvalenciájával, ám a *Rossz* partnerrel szemben csak rossz minősítést lehet kapni (bármit is tegyen az ember).

A négy értékelési szabály áttekintése

Az értékelési szabályok bemutatása után érdemes egyetlen táblázatba rendezve összegezni a fenti négyféle szabályt, hogy jobban szembetűnővé válhassanak a köztük levő hason-

lóságok és különbségek.

értékelési szabály ↓	VALUE'$(Alter, t_i) \rightarrow$	G		B	
	DO'$(Ego, t_i) \rightarrow$	C	D	C	D
Image Scoring	VALUE'$(Ego, t_i) \rightrightarrows$	G	B	G	B
Simple Standing		G	B	G	G
Stern Judging		G	B	B	G
Shunning		G	B	B	B

A táblázatból könnyen kiolvasható a négy szabály közös vonása. A *G* jó partnerrel szemben mindegyik értékelési módszer ugyanazt írja elő: ha a jó *Alterrel* szemben *Ego* kooperált, akkor *G* jó minősítést kap, ha dezertál, akkor *B* rosszat. A különbség a *B* rossz partnerrel szembeni cselekvés utáni értékelésben rejlik.

Harmadrendű stratégiák dekomponálása

A stabil kooperációképes másodrendű stratégiákat még inkább fontosnak tarthatjuk, ha felismerjük, hogy a sikeres harmadrendű stratégiák mind visszavezethetőek a most tárgyalt másodrendű stratégiákra [517]. Elméletileg 4096 harmadrendű stratégia létezik, és ebből összesen nyolc olyan van, amely kooperáció kialakítására és fenntartására képes. Ám ezek – egy egyszerű dekomponálási technika alkalmazásával – könnyen értelmezhetőek a korábban elemzett négy másodrendű értékelési szabály és egy-egy egyszerű cselekvési szabály kompozíciójaként.

5.5.2 A direkt és indirekt reciprocitás közös mintázatai

Korábban bemutattam, hogy mind a direkt, mind az indirekt reciprocitás feltételei közt négy-négy másodrendű stratégia (cselekvési, illetve értékelési szabály) képes valamilyen formában és mértékben hozájárulni a kooperáció kialakulásához. A stratégiákhoz rendelt formulák hasonlósága alapján a cselekvési és értékelési szabályok párba kapcsolhatók egymással.[20] Az egyszerű *TFT* és az *Image Scoring*, a *bűnbánó* és a *Simple Standing*, a *Pavlov* és a *Stern Judging*, valamint a *haragtartó* és a *Shunning* szabályok ugyanazt a mintázatot mutatják, logikájukat tekintve hasonlóképpen működnek [445]. Ha a direkt reciprocitáshoz tartozó stratégiák esetén a feltételes cselekevési szabályokban kicseréljük a kooperálásra, illetve dezertálásra utaló *C*, illetve *D* jelzéseket *1*, illetve *0* értékekre, valamint az indirekt reciprocitás értékelési szabályait összegző táblában szereplő *G*, illetve *B* jelek helyett szintén *1*, illetve *0* értékeket írunk, akkor a kétféle szabálynégyest egy táblában reprezentálhatjuk.
A párba kapcsolható szabályok közötti azonos mintázat alapja a büntetéshez való viszony hasonlósága. Mind az *Image Scoring*, mind a *TFT* elsőrendű szabály. Saját "terepükön" ugyanaz a szerepük, mindkettő ugyanazt a funkciót tölti be:

> „Úgy véljük, hogy a Scoring szerepe az indirekt reciprocitásban hasonló, mint a Tit-for-tat szabályé a direkt reciprocitáson belül. Egyik stratégia sem evolucionárisan stabil, de az a képességük, hogy katalizálják a kooperáció kialakulását kedvezőtlen körülmények között, valamint az egyszerűségük jelentőséget kölcsönöz nekik." [381].

[20] Lásd erről magyarul [522], illetve angolul [523] írt tanulmányaimat.

ACRs ↓	C		D		$\leftarrow$ DO'(Ego, t_{i-1})
	C	D	C	D	$\leftarrow$ DO'$(Alter, t_{i-1})$
Simple TFT	1	0	1	0	Image Scoring
Contrite TFT	1	0	1	1	Simple Standing
Pavlov	1	0	0	1	Stern Judging
Grim Tigger	1	0	0	0	Shunning
DO'$(Ego, t_i) \rightarrow$	C	D	C	D	↑ ASRs
VALUE'$(Alter, t_i) \rightarrow$	G		B		

Mindkét szabály "ellenzi" a büntetést, bár ezt hipokrita módon teszik. Közös elvük az, hogy nem fogadják el büntetést, tehát ha valaki büntet, akkor azt büntetik: 'büntesd, aki büntet'.

Amikor Sugden 1986-ben javasolta saját T_1 módszerét (a *Standing* szabályt) [512], akkor ő nem beszélt reputációról, bár ahhoz közel álló fogalmat használt: 'kiállás' ('standing'). Sugden nem igazán reflektált a kérdésre, de tény, hogy cselekvési és nem értékelési szabályokról beszélt. Jóval később Sugden szabálya *Standing* (*Simple Standing*) néven mint értékelési szabály az indirect reciprocitás elméletek részévé vált. Amikor 1989-ben Boyd újracímkézte Sugden szabályát *bűnbánó tit-for-tat* stratégiának [70], akkor ő is cselekvési szabályként hivatkozott rá. Ez a kettős kezelésmód persze valamiféle terminológiai zavar, fogalmi pongyolaság jele, de mindebben az a fontos most, hogy ez az "összemosás" nem véletlen. Az összekeverhetőség ugyanis a *bűnbánó* és a *Simple Standing* szabályok közötti szoros kapcsolatot jelzi. Amikor Frean bevezette a szigorú, de igazságos *Firm-But-Fair* szabályt, akkor a következőpen értelmezte a szabályt:

> „. . . it is 'firm' in that it retaliates by defecting if it was a sucker in the previous round. It is also 'fair' in that it does not retaliate against a defector if it defected itself, and it cooperates with suckers rather than continuing to exploit them." [176].

Ez az indoklás ahhoz hasonlít, amit a *Contrite TFT* stratégia bemutatásakor már elmondtunk egyszer – bár az érvelésben tetten érhető egyfajta elmozdulás a büntetés irányába. *Ego* úgy értékeli *Alter* előző fordulóbeli dezertálását, hogy az igazolt büntetés volt *Ego*'s saját korábbi dezertálásával szemben, és mivel *Ego* egyfelől egyetért az igazolt büntetés intézményével ('firm'), másfelől igazságos ('but fair'), így *Ego* kooperál *Alter* lépésére válaszul. Közös szlogenjük ez lehetne: 'ne büntesd, aki büntet'.

A *Pavlov* vagy *WSLS* stratégia jellemzésekor többen kiemelték annak tanulási képességét, mi több, ezzel magyarázták a stratégia sikerességét is. Most, amikor ezt a cselekvési szabályt állítjuk párba a *Stern Hudging* értékelési szabállyal, akkor – a büntetéshez való viszonyra fókuszálva – kicsit átértelmezhetjük a *Pavlov* szabály logikáját, hatásmechanizmusát. Rosas úgy értékeli, hogy a *Pavlov/Stern Judging* szabálykettős hízelgő (flattery), alázatos (obsequious) stratégia [445]. Szerintem inkább arról van szó, hogy a *Pavlov* esetében is a *Stern Judging* módszer logikája érvényesül. Mindkét szabály hasonlóképpen értelmezhető. *Ego* úgy értékelheti *Alter*'s kooperálását *Egoval* szemben, hogy ezzel *Alter* tartózkodott megbüntetni *Egot*, aki pedig dezertált az előző fordulóban. A dezertálás büntetése viszont *Alter* kötelessége lett volna, s mivel ezt elmulaszotta, ezért *Ego* ezt a kooperálás helytelennek tartva böntetéssel (dezertálással) sújtotta. Ez volt a *Stern Judging* értékelési szabály logikája, és ez az indoklás érvényes, alkalmazható a *Pavlov* cselekvési szabály esetében is. Éppen ezért lehet mindkettejük receptje a következő szabály: 'büntesd, aki nem büntet'.

Mind a *haragtartó* (*Grim Trigger*), mind a *Shunning* szabály erősen, vagy inkább túlzottan büntetéspártinak mondható. Nem véletlen, hogy a szimulációk során kiderült, hogy ezek a stratégiák nehezen, ritkán vezettek a kooperációs minták elterjedéséhez. A haragtartó stratégia az első dezertálásra végnélküli büntetéssel reagál, azaz bármit is tesz a partnere, a stratégia képviselője válogatás nélkül dezertál. A Shunning módszer is ugyanezt a logikát követi, csak ebben az esetben a partner "rosszasága" váltja ki ezt az egyértelmű elkülönülést. A közös vonásuk tehát az, hogy büntetik mind a kooperációt, mind a dezertálást, ami úgy is értelmezhető, hogy nem az – ilyen vagy olyan – cselekvést, hanem magát a cselekvőt utasítják el. Elvük lehetne: 'büntess mindent'.

A négy stratégiapáros bemutatása alapján megállapíthatjuk, hogy négy stratégiapárhoz négy alapelvet társíthatunk, amelyek mindegyike a büntetéshez való eltérő viszonyulást fejez ki. Ezt összegzem a következő táblában.

szabálypár	közös elv	
egyszerű TFT & Image Scoring	büntesd, aki büntet	$P(P)$
bűnbánó TFT & Simple Standing	ne büntesd, aki büntet	$\neg P(P)$
Pavlov & Stern Judging	büntesd, aki nem büntet	$P(\neg P)$
haragtartó & Shunning	büntess mindent	$P(P \vee \neg P)$

A négyféle fenti logika állandóan megnyilvánul a mindennapjainkban, amikor cselekszünk, döntünk, amikor értékeljük mások cselekvését. Ez a négy logika beépül a személyiségünkbe, a társadalmi helyzeteinkben, a kulturális, szubkulturális emlékezetbe, a közösségek közös tudatába, és helyzetről helyzetre változóan, tehát sohasem determnisztikus módon, de általában valamelyik elv dominanciája mellett befolyásolja a cselekvők magatartását. A tudósi szereptől egy pillanatra elszakadva, a bulvár irányába lépve egyet mondhatjuk, hogy kicsit azok vagyunk, amelyik elvet választjuk magunknak a fenti négyből.

5.6 Kooperatív játékok: koalícióelmélet

A kooperatív és nem-kooperatív játékok fogalmát John Nash vezette be. Kétféle módon szokták magyarázni, mit is jelent ez a megkülönböztetés. Egyfelől nem kommunikálhatnak egymással, ami azért fontos, mert nem így nem köthetenek egyezséget a partnereikkel a játékosok annak érdekében, hogy kikerüljenek a társadalmi csapda helyzetekből.

magjátékok	
aszimmetrikus játékok	szimmetrikus
többalternatívás játékok	kétalternatívás
többszereplős játékok	kétszereplős
nem szigorúan rendezett játékok	szigorúan rendezett
aszinkron játékok	szinkron
ismételt játékok	egyszeri
direkt reciprocitás játékok	ismerős játékosok
indirekt reciprocitás játékok	ismeretlen játékosok
kooperatív (ígéret) játékok	kooperatív (ígéretmentes)

Harsányi János hívta fel a figyelmet arra [225], hogy az igazán fontos mozzanat ebben nem az, hogy a kommunikáció révén biztosítható az információcsere, és így egy közösen

jobb megoldásra számíthatnak a felek, mert a másik szándékának ismerete, tehát a többletinformáció még nem feltétlen elegendő a jobb megoldás eléréhez. Sokkal fontosabb az, hogy a kommunikáció révén kikényszeríthető elköteleződéseket tehetnek egymás felé a játékosok, amire a nem-kooperatív játékok során nem volt lehetőségük. Fontos, hogy ezek az elköteleződések kikényszeríthetőek, azaz az elköteleződések betarttatására vannak eszközeik a feleknek. Enélkül az egész semmit sem érne. A kikényszeríthető elköteleződések azonban már átvezetnek minket a normativitás területére, amit viszont – fontossága miatt – külön fejezetekben kell tárgyalnunk.

Chapter 6

Pozíció, norma, uralom

„A jogtudományt tanuló angol diákok döntő többsége a jog alapvető kategóriáival való ismerkedést Hohfeld 'Alapvető jogi fogalmak' című tanulmányával kezdi." [226, 171]

„Egy rossz normától megszabadulni mindannyiunk érdeke, mégis sokszor nagyon nehéz elérni ezt." [53]

„A karizma az a képesség, hogy hatni tudjunk másokra logika nélkül." (Quentin Crisp)

Ebben a fejezetben a normativitás jelenségét szeretném megalapozni úgy, hogy még nem lesz szó a norma fogalmáról. Nem tudom persze elkerülni, hogy ne használjam olykor a norma, a jog, normativitás, legalitás fogalmait, de ezeket majd csak a következő fejezetben fogom pontosan definiálni. A norma jelenségének leírásához, a fogalom megalapozásához arra van szükségünk, hogy a deontikus logikát bizonyos értelemben kiterjesszük, és az eddig megismert imperszonális deontika helyébe felépítsük a perszonális deontikus logika elméletét. Ehhez fel kell használni a Wesley Newcomb Hohfeld jogértelmezését, és fel kell építeni a hohfeldi elmélet formális logikai reprezentációját. Ehhez a skandináv logikai iskola által kijelölt úton kell végigmennünk, miközben olykor a fogalmi modell és a formalizálási javaslatok korrekciójára, olykor pedig kiterjesztésére van szükség.

6.1 Perszonális deontika

A logikai alapozó fejezetben jeleztem (↪116.o.), hogy a deontikus logika évtizedeken keresztül személytelen, imperszonális módon építkezett. Megkockáztatom, hogy az imperszonális megközelítés dominanciája is hozzájárult ahhoz, hogy a deontikus logika hosszú időn át nem tudta beváltani a hozzá fűzött kezdeti reményeket. Sokan várták

ugyanis, hogy a deontikus logika fejlődése során hamar eléri az a szintet, amikor már megfelelő alapokat képes nyújtani jogi, normatív szövegek formális leírásához, a jogi szövegek konzisztenciájának vizsgálatához és ezáltal jogi szövegek számítógépes feltárásához. Ez azonban nem következett be. Nem is nagyon következhetett be, mert amíg a deontikában nem lehetett személyközi viszonyokat kifejezni, kezelni, addig nem lehetett megragadni a jogban, normativitásban alapvető jogi, normatív viszonyokat. A társadalmi kapcsolatok feltárásához szükség volt arra, hogy az ágensek közti kapcsolatokat is kezelni lehessen a deontikán belül.

A deontikus logika a deontikus modalitások, a kötelezések, tiltások, megengedések, eltekintések logikájával foglalkozik. Amíg imperszonális deontikát "űzünk", azaz amíg egyetlen ágenst feltételezünk, amíg egyetlen deontikus operátoron van a fókuszunk, amivel ki tudunk fejezni egy normatív, deontikus állítást, addig nem tudjuk megragadni a jognak (normativitásnak) azt a minőségét, alapvonását, hogy a létező jogi (normatív) rendszerekben jogi (normatív) viszonyokat kell kezelnünk. Ez a deontika nyelvén azt jelenti, hogy a deontikus operátorokkal kifejezhető normatív aktusokat mindig "párosával", egymással összekapcsolódva értelmezhetjük.

A jogtudománynak már nagyon régóta az egyik alapvetése ez a tétel [275]. Ahol valakinek jogosultsága van valamire, ott van valaki más, akinek valamilyen kötelezettséggel rendelkezik az adott jogosultsághoz kapcsolódóan. Vagy fordítva: a kötelezettségek mindig mások jogosultságával járnak együtt. W.N. Hohfeld egyetértőleg idézi az alábbiakat:

> „Kötelesség vagy jogi kötelezettség az, amit valakinek meg kell tennie vagy nem szabad megtennie. A 'kötelesség' és a 'jog' korrelatív kifejezések. Ha megsértenek egy jogot, megszegnek egy kötelességet is." [241, 73]

A jogbölcseletben vita van arról, vajon csak általában vagy mindig érvényes ez a fajta komplementaritás. Vannak nagy hatású jogelméleti gondolkodók, akik vitatják a korrelativitás teljeskörűségének tézisét. Joseph Raz például így ír:

> „Egy népszerű nézet szerint minden joghoz tartozik egy korrelatív kötelezettség. Minthogy a jog a kötelezettség egyik alapja, jó adag igazság van az efféle korrelativitás-tézisben. A tézis legtöbb szokásos megfogalmazása mégis félrevezető." [435, 166]

Ebben a vitában nem tudok, nem is akarok állást foglalni, és mivel a fomális modell építésének jelen szakaszában nem is igazán lényeges ez a kérdés, el lehet tekinteni e szempont figyelembe vételétől. Lehet, hogy igaza van Joseph Raznak, és nem áll mindig a korrelativitás tézise, de azt Raz sem tagadja, hogy sok esetben igaz. Ha viszont azt állíthatjuk, hogy a jognak, normativitásnak vannak olyan területei, ahol diadikus viszonyokat, azaz jogi relációkat kell kezelnünk, akkor a jogi fogalmak fent hivatkozott kettősségét, korrelativitását valahogy be kell építeni a fogalmi modellünkbe. Ezt a relációs jelleget úgy tudjuk megragadni, ha a *imperszonális* szemléletmódot felváltjuk a *perszonális* megközelítéssel.

A továbblépés előtt azonban szükséges figyelmeztetnem arra, hogy az elemzendő fogalmak, terminusok mindennapi használatában sajnos sokfajta és elég széles körben elterjedt pongyolaság figyelhető meg. A terminusokat sokszorosan túlhasználják, vagyis egy-egy terminust – sokszor akaratlanul – többféle jelentéssel vesznek használatba, és ez sok esetben nagyon könnyen fogalmi zavarokhoz, félreértésekhez vezethet. Érdemes ezért itt egy-két előzetes megjegyzést tennem.

A jog a deontikus logika operátorait kifejező terminusok közül a kötelezést és a tiltást kiterjedt módon alkalmazza, ám a megengedés terminusát ritkábban veszi igénybe

(ha egyáltalán használja). Sokkal elterjedtebb a jog, jogosultság terminusa. Ennek ellenére nyugodtan rögzíthetjük, hogy a jog, jogosultság, megengedés terminusai mind ugyanahhoz a jelentéstartományhoz sorolhatók, mint ahová a megengedés is tartozik. Ebből persze az is következik, hogy akárcsak a megengedés kategóriája esetében, úgy a jogosultság mindennapi használatában is keveredik az a két lehetőség, hogy a nemkötelezés vagy a nem-tiltás állapotát fejezzük-e ki a terminussal. Mivel a hétköznapi fogalomhasználati gyakorlatba mindkét értelmezés mélyen beleivódott, értelmetlen lenne bármelyik interpretációt elsődlegessé tenni a másik ellenében, így csak arra törekedhetünk, hogy tudatosan figyelve erre a problémára, a terminusok használata során mindig egyértelművé tesszük, az adott pillanatban éppen melyik értelmezés tartjuk érvényesnek. A szinonimitásról beszélve megállapíthatjuk még, hogy a kötelezettség és kötelesség terminusait is egymással felcserélve lehet használni.

Sajnos létezik egy másik terminológiai pongyolaság jogi alapfogalmak területén, ami a laikusok számára még több problémát, félreértést okozhat – legalábbis a magyar nyelvben. A 'jog' terminusa ugyanis többféle értelemben használható. A jog egyfelől jelenthet jogosultságot, illetve másfelől normatív megnyilatkozást, és az utóbbi jelenthet egyetlen normatív kijelentést (mint jogszabály, jogtétel), de jelentheti akár az egész jogi rendszert is. Az igazi megkülönböztetést a jog mint jogosultság és a jog mint normatív rendszer (tételes jog) között kell megtennünk, a jogtétel és a tételes jog közti különbség inkább mennyiségi, nem annyira minőségi. Természetesen a jogtudomány réges rég elvégezte ezt a megkülönböztetést, hiszen az 'alanyi jog' (facultas agendi) és a 'tárgyi jog' (norma agendi) fogalmai lényegileg erre a kettősségre reflektálnak. Ami nagyon fontos (és ez sokak számára szokatlan, meghökkentő lehet), fogalmilag szükségünk van a 'jogi jog' és 'erkölcsi jog' elhatárolására egymástól, amit csak a "jog" fogalmának pontos explikációja, sokszálú értelmezése révén tehetünk meg. Amikor a későbbiekben a norma fogalmát fogom tipizálni a különböző fogalmi összetevők alapján (↪389.o.), akkor a különböző tipizálási lehetőségek egymás utáni kibontásával jobban lehet majd érzékeltetni, hogy valóban értelmes kategória akár a 'jogi jog', akár az 'erkölcsi jog' fogalma.

Másfajta fogalmi zavar jelentkezik a jogi viszonyok minősítésében. Itt nem terminológiai poliszémikus jellegű túlhasználásról, hanem fel nem ismert szinonimitásról van szó. Amikor a jogtudósok az összetartozó jogokról és kötelességekről beszélnek, akkor a köztük levő viszonyra a 'korrelativitás' terminusát alkalmazzák. Wesley Newcomb Hohfeld is a korrelativitás kategóriáját használta. A mai napig bevett ez a szóhasználat a jog világában. Később elemezni fogom a szociológusok által bevezetett társadalmi szerep fogalmát, szólok arról, hogy a társadalmi szerepek normatív összetevői közt ugyanúgy jogokat és kötelességeket találhatunk, mint a jog esetében. A szociológusok azonban az általuk vizsgált normatív komponensek közt megfigyelhető viszonyt más kategóriával ragadják meg, ők erre a komplementaritás terminusát használják (lásd Sir Ross [447], Gouldner, Parsons [405]). A korrelativitás terminusa helyett Stig Kanger már nem csak egy, de két másik kategóriát használ [275]. Később bemutatom, hogy Kanger milyen alapon minősíti hol inverz, hol konverz relációnak két jogviszony közt létező korrelativitást. Jogi szövegekben előfordul olykor a jogi viszonyok 'reciprocitása' kifejezés is, ami szintén a korrelativitás, komplementaritás jelenségéhez kapcsolódik.

Nagyon hasonló tartalom mentén van tehát négy-öt terminus, a korrelativitás, a komplementaritás, az inverzitás és konverzitás, illetve a reciprocitás, amelyek közül nekem is választanom kell majd. Mivel az empirikus társadalomtudományokban a korreláció

fogalmának a korrelativitás itteni tartalmához képest nagyon más jelentése van, ezért a korrelativitás helyett inkább a komplementaritás fogom használni, illetve ahol szükség lesz a nagyobb pontosságra, ott a konverzitás és inverzitás terminusait is alkalmazni fogom. És mindezek mellett – ha történeti vagy olykor stilisztikai okokból szükségesnek látszik – megengedhetőnek tartom a korrelativitás terminus használatát is. Mivel a reciprocitás fogalma nagyon erősen használt a társadalomtudományokban, nem tartanám szerencsésnek ezt a fogalmat az itt tárgyalt jelentésben is használatba venni.

6.2 Normatív pozíciók, jogosultságok és kötelezettségek

A társadalmi cselekvések indoklásai során rengetegszer hivatkozunk a jogainkra, a jogosultság fogalmára. Ilyeneket mondunk: *Jogom van a bolt előtti járdán sétálni. Jogában áll hallgatni. Jogában áll beszélni. Jogom van visszakapni a kölcsönadott pénzemet. Jogom van eladni a saját gitáromat. Jogom van nem eladni a saját gitáromat. Jogom van kidobni a nem használt ruháimat. Jogom van megváltoztatni a saját boltom nyitvatartási idejét. Nincs joga másnak arra, hogy eladja az én autómat valakinek. Jogom van szavazni. Jogom van nem szavazni. Nincs joga senkinek engem megverni. Ha megtámadnak, akkor jogom van megvédni magam.*

Ezek a példák azt jelzik, hogy egyrészt a jogaink irányulhatnak tevőleges cselekvésre (csinálásra), illetve nem-cselekvésre (tartózkodásra), másrészt a jogaink vonatkozhatnak a saját, illetve mások cselekvésére, harmadrészt a jogaink tetszőleges cselekvésre irányulhatnak, de olykor "csak" azt szabályozzák, hogy a jogi-normatív viszonyainkat megváltoztathatjuk-e vagy sem. De fenti példák azt is mutatják egyben, hogy a jogosultságainkhoz valóban hozzárendelhetők mások kötelezettségei. Amikor például a kölcsönadott pénzt várjuk vissza, akkor mindig valaki más kötelezettségére gondolunk (akinek vissza kell adnia a pénzt), vagy ha nekem jogom van erőszakmentes életre, akkor azt is jelenti egyben, hogy mindenki másnak tilos megvernie engem (kötelező tartózkodnia az erőszak alkalmazásától). Végül a példákból kitetszik az is, milyen sokféle jelentést sűrítünk a jog-jogosultság fogalmába. Itt az idő, hogy megnézzük, hogyan lehet szétszálazni ezt az összefonódó sokféleséget.

6.2.1 Hohfeld elsőrendű jogviszonyfogalmai

Wesley Newcomb Hohfeld volt az, aki miután észlelte, hogy a jog(osultság) és kötelesség fogalmát sokszorosan túlhasználják a jogászok, megpróbálta elválasztani a különböző jelentésrétegeket egymástól. Hohfeld két tanulmánya a jog és kötelesség fogalmainak többértelműségéről csak a halála után, 1913-ban és 1917-ben jelent meg.[1] A jogosultságok és kötelezettségek közti korrelatív kapcsolat tézise már Hohfeld előtt is nyilvánvaló volt a jogtudományon belül [275]. Hohfeld azzal hozott újdonságot, hogy rámutatott, a jogosultság, illetve kötelezettség fogalmaiban meghúzódó jelentésrétegek hogyan különíthetőek el egymástól.

[1] Hohfeld tanulmánya magyarul is olvasható [241], a Hohfeld-recepcióból kiemelném Arthur Corbin [107], [108], Albert Kocourek [283], Thomas D. Perry [410], Frederic B. Fitch [170], Walter J. Kamba [271] cikkeit, illetve Matthew H. Kramer, N.E. Simmonds és Hillel Steiner tanulmánykötetét [293].

Hohfeld jogfogalmainak struktúrája

Hohfeld szerint négy-négy jogosultság-, illetve kötelezettségtípust definiálhatunk, és ezek mindegyikét kétféle módon is kapcsolatba hozhatjuk egymással. Minden jogosultságnak van egy ellentéte, illetve van egy korrelatív párja a kötelezettségek között (és természetesen mindez igaz fordítva is, a kötelezettségekhez is rendelhető jogosultság). Hohfeld a következő kategóriákat vezette be:

right	privilege	power	immunity
duty	no-right	liability	disability

Fontos megjegyezni, hogy a Hohfeld-recepció az idők során pár ponton változtatott az eredeti terminusokon. Hohfeld követői az eredeti 'right' terminus helyett a 'claim' (illetve 'claim-right'), míg a 'privilege' helyett a 'liberty' és a 'freedom' kategóriáit is használták, illetve volt szerző, aki az 'immunity' helyett az 'exemption'-t használta, más a 'no-right' terminust javasolta lecserélni az 'exposure'-re [273], [274].[2] A hohfeldi kategóriák magyar fordítására több esetben olyan javaslat született, amelyek a laikusok számára ugyan nehezebben befogadhatóak, viszont sokkal egyértelműbbnek tűnnek. A következő magyar terminusokat vezették be a Hohfeld tanulmányának fordításában [241].

jog	szabadság	felhatalmazottság	mentesség
kötelezettség	joghiány	beavatkozásnak kitettség	beavatkozásképtelenség

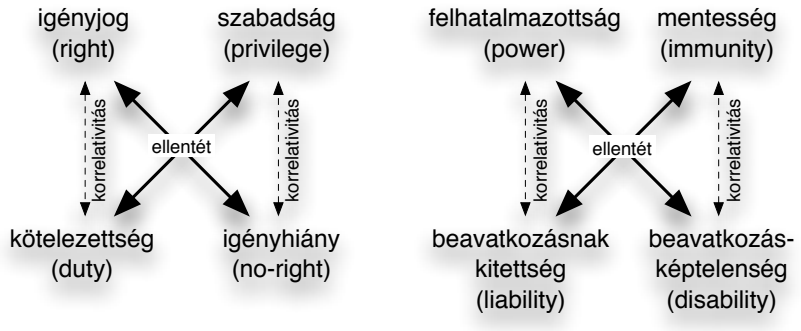
A Hohfeld után megjelenő "versengő terminusok" közül a 'claim' fordítására az 'igény' vagy 'igényjog', míg a 'liberty/freedom' terminuspárosra a 'szabadság' kifejezése került használatba. A továbbiakban igyekszem követni a hohfeldi terminológia magyar fordítását, de egy ponton tudatosan eltérek tőle, néhány más esetben pedig – stilisztikai okok miatt – olykor alkalmazok majd szinonímákat. A hohfeldi 'right' terminus magyar fordítása ('jog') helyett inkább a 'igényjog' terminust veszem igénybe, hogy ezzel is elkerüljem a 'jog' terminusának sokrétűségéből fakadó veszélyeket. Sokkal pontosabb lenne a 'joghiány' helyett az 'igényhiány' terminusát "igénybe venni", ezért alkalmazom ezt a kifejezést is, de – mivel nem akarok túlzottan eltérni a magyar fordítástól – olykor használom az eredeti magyar fordítást is. Ha nem lesz zavaró, akkor – stilisztikai okokból – olykor-olykor lecserélem a 'szabadság'-ot a 'privilégium', a 'mentesség'-et az 'immunitás' terminusára (és csak zárójelben jegyzem meg, hogy a 'mentesség' helyett a 'kivételezettség' terminusát is nyugodtan használhatnánk). Ugyan nem merült fel sehol fordítási javaslatként, mégis érdemesnek tartom megjegyezni, hogy a 'power' kategóriáját lehetne 'beavatkozásképesség'-nek, az 'immunity' kategóriáját pedig 'beavatkozásmentességnek' is fordítani. Ha így tennénk, akkor egyértelműbben látszódna, hogy a második négyes hohfeldi fogalmi csoport a beavatkozással (vagy mondhatnánk azt is, hogy a rendelkezéssel) kapcsolatos. Sajnálatos módon a hohfeldi elmélet formalizálási kísérletei során a terminológiai zavart inkább növelték, mint csökkentették azzal, hogy a kötelezettségfogalmakra is jogtípusként hivatkoztak, vagyis négy-négy jog-, illetve kötelezettségfogalom helyett nyolc jogtípus fogalomról beszéltek. Igaz, az alkalmazott formális nyelvi szinten ezt a zavart fel lehetett oldani, de a természetes nyelvi szinten ez a megoldás nehezebbé tette és teszi az értelmezést, befogadást.

[2] További terminusjavaslatokat mutat be Walter J. Kamba cikke [271].

Mielőtt sorba venném, hogyan értelmezi Hohfeld a jogosultság- és kötelezettségfogalmait, be kell mutatnom azt, hogy milyen relációkkal hogyan köti össze Hohfeld a kategóriáit. Hohfeld szerint korrelatív viszonyban vannak egymással a táblázatba foglalt elemek közül az egymás alatt levő kategóriák (azaz az igényjog és kötelezettség, a privilégium és igényhiány, a felhatalmazottság és beavatkozásnak kitettség, valamint az immunitás és beavatkozásképtelenség), és ellentét van a "keresztkapcsolatokban" (azaz az igényjog és igényhiány, a privilégium és kötelezettség, a felhatalmazottság és beavatkozásképtelenség, illetve az immunitás és beavatkozásnak kitettség között).

legal relation	jogi viszony	korrelatív párja	ellentéte
right	igényjog	kötelezettség	igényhiány
duty	kötelezettség	igényjog	szabadság
privilege	szabadság	igényhiány	kötelezettség
no-right	igényhiány	szabadság	igényjog
power	felhatalmazottság	beavatkozásnak kitettség	beavatkozásképtelenség
liability	beavatkozásnak kitettség	felhatalmazottság	mentesség
immunity	mentesség	beavatkozásképtelenség	beavatkozásnak kitettség
disability	beavatkozásképtelenség	mentesség	felhatalmazottság

A táblázatos megjelenítés mellett érdemes grafikus reprezentációt is adni a hohfeldi kategóriáknak és a köztük levő viszonyoknak, mert így jobban lehet érzékeltetni a fogalmi viszonyrendszer belső szerkezetét.



Az ábra nemcsak azt mutatja, hogy mely fogalmak között van ellentétes, illetve korrelatív viszony, de az is "leolvasható" róla, hogy a fogalmak két csoportra oszthatók. Az egyik csoport az igényjog, a másik felhatalmazottság köré szerveződik, és mindkettő csoportnak hasonló belső mikrostruktúrája van. A későbbiekben majd kiderül, hogy milyen kapcsolat van a két csoport, illetve a két csoport elemei között.

Az igényjog köré szerveződő fogalmak értelmezése

A hohfeldi jogfogalmak struktúráját már bemutattam, de se nem definiáltam, se nem adtam példákat az egyes fogalmakra, pedig az értelmezésükhöz, modellbe építésükhöz mindkettőre szükség lenne. Első körben bemutatom Hohfeld saját meghatározásait, miközben igyekszem tőle vagy a korai Hohfeld-recepcióból származó példákat hozni rájuk.

Az első fogalmi csoport középpontjában a jogigény kategóriája áll, aminek komplementere a kötelezettség és ellentéte a joghiány fogalma. Hohfeld példája így szól:

> „ . . . ha *X*-nek joga van ahhoz, hogy *Y* az ő földjén kívül tartózkodjék, ezzel korrelatív (és ennek megfelel) *Y* azon kötelessége, hogy maradjon távol attól a helytől.” [241, 73]

A fenti idézetben Hohfeld két jogi terminust nevez meg (az igényjogot és a kötelezettséget), de ezzel igazából nem két “különálló” jogviszonyt állít szembe egymással. Az igény és a kötelezettség voltaképpen ugyanannak a jogviszonynak két “nézetét” jelenti.[3] Ezért mondja rájuk Hohfeld, hogy korrelatív kapcsolatban vannak egymással. A két jogviszonyt így bonthatjuk fel összetevőkre:

jogosult	jogviszony	cselekvő	cselekvés	
Ego	igénye	Alter	nem	átjárni Ego telkén
Alter	kötezettsége	Alter	nem	átjárni Ego telkén

Az összetevőkre bontásból látszik, hogy mindkét fél (jogbirtokos) jogviszonya ugyanarra a cselekvésre (és ugyanarra a cselekvőre) vonatkozik. Ez fontos, később látni fogjuk, hogy nem minden párosítás esetében van ez így. Hohfeld következő jogfogalma a szabadság, ami a kötelezettséggel van ellentétben és a joghiánnyal korrelatív kapcsolatban. Hohfeld újra az ingatlanos példát hozza erre.

> „ . . . ha *X*-nek joga vagy igénye van arra, hogy *Y* – valaki más – maradjon távol földjétől, akkor neki magának megvan az a szabadsága, hogy belépjen e területre; másként fogalmazva, *X*-t nem terheli a távolmaradás kötelezettsége. A belépés szabadsága a kívülmaradás kötelezettségének ellentéte.” [241, 73]

Szabadság és kötelezettség tehát ellentétes fogalmak, de pontosan látni kell, mit is ért Hohfeld az ‘ellentétes’ fogalom jelentése alatt:

> „ . . . e ponton nem árt némi óvatosság; hiszen amikor azt mondjuk: egy adott szabadság egy kötelezettség puszta negációja, akkor ezt természetesen mindig úgy értjük, hogy a kötelezettség tartalma az, ami pontosan ellentétes a kérdéses szabadságéval.” [241, 74]

A telektulajdonosnak szabadsága van saját telkén átjárnia, másoknak kötelezettségük van tartózkodni az átjárástól. Utóbbi úgy is értelmezhető, hogy mások kötelesek nem átjárni a tulajdonos telkén. Ha Egora és Alterre vonatkoztatjuk mindezt, akkor a következő formában adhatjuk meg a két jogviszony összetevőit:

jogosult	jogviszony	cselekvő	cselekvés	
Ego	szabadsága	Ego		átjárni Ego telkén
Alter	kötelezettsége	Alter	nem	átjárni Ego telkén

Látható, hogy az ellentétes jogviszonyok más cselekvőre és más cselekvésre irányulnak. Fontos kérdés, hogy mi a kapcsolat ugyanazon személy, ugyanazon jogbirtokos (itt az Ego) igényjoga és szabadsága között. Joel Feinberg ezt írja:

> „Amit a jog [az igényjog] hozzátesz a szabadsághoz, az másoknak az a kötelessége, hogy tartózkodjanak a beavatkozástól.” [160, 88]

[3] Talán hasznos lehet itt egy másik példát hozni szemléltetésképpen. Amikor azt mondjuk, hogy *X* szülője *Y*-nak, akkor mondhatjuk azt is, hogy *Y* gyereke *X*-nek. Bármelyiket is mondjuk, ugyanazt a tartalmat fejezzük ki velük. A ‘szülője’ és a ‘gyereke’ reláció korrelatív párjai egymásnak (vagy másként: egymás inverzei). Ugyanez igaz az igényjog-kötelezettség párosra is.

Feinberg arra utal, hogy Hohfeld két kategóriáját szigorúan el kell választani egymástól, a szabadság nem feltételezi – sem logikailag, sem gyakorlatilag – a jogigény létezését. A szabadság önmagában még nem teremti meg, nem hozza magával azt az igényjogot, hogy az Alter nem avatkozik be, nem zavarja meg Egot abban a cselekvésben, amire Egonak szabadsága van. Hohfeld példája arra, hogy a szabadság nem keletkeztet korrelatív kötelezettséget a másik fél be nem avatkozásra, így szól:

> „*A, B, C* és *D* mint salátatulajdonos esetleg azt mondja *X*-nek: 'Edd meg a salátát, ha tudod; engedélyezzük ezt neked, de nem értünk egyet azzal, hogy ne zavarjunk meg ebben.' Ebben az esetben a szabadság létezik, s amennyiben *X*-nek sikerül megennie a salátát, akkor nem sérti meg egyik fél jogát sem. Világos azonban az is, hogy ha *A*-nak sikerülne olyan gyorsan elkapnia a tálat, hogy *X* egy falatot sem tudna abból enni, úgy *Y*-nek semmilyen joga sem szenvedne csorbát." [241, 76]

A példa szemléletesen rávilágít, hogy mi a különbség a kétfajta jogviszony, az igény és a szabadság között. De vehetek egy másik példát a hétköznapok világából. Ha egy vonatra helyjegyet veszek, akkor igényjogom keletkezik arra, hogy leüljek a székemre. Ha valaki ül azon a széken, amire a helyjegyem szól, akkor igényelhetem tőle, hogy adja át a helyét, és ezzel "párhuzamosan" neki kötelessége lesz felállni. Ha viszont nem helyjegyes vonaton utazom, akkor szabadságom van a még szabad székek bármelyikére leülnöm, de nem igényelhetem senkitől sem, hogy adja át nekem azt a széket, amin ő ül. Viszont ha én leülök egy szabad székre, utána nem leszek köteles átadni az a helyt valaki másnak, ezt senki sem igényelheti tőlem. Amíg viszont vannak szabad székek, addig bárki elfoglalhatja azok bármelyikét, és megakadályozhat engem abban, hogy a kiszemelt szabad helyre leülhessek, ha ő előbb teszi ezt meg.

Nem könnyű észrevenni a szabadság "függetlenségét" az igényjogtól, már ha valóban magában áll. A tényleges gyakorlatban a jogviszonyok az esetek nagy részében összetettek és a szabadság és igényjog együttjár. Így van ez például a tulajdonjog esetében. De ez a gyakorlati együttjárás semmiképpen sem jelent logikailag szükségszerű összetartozást köztük. Ha viszont a szabadság hol magában, hol az igényjoggal együtt fordul elő a különböző gyakorlati jogfogalmak között, akkor ezt a kétféle helyzetet, pontosabban e kétféle helyzetre jellemző szabadságfogalmat érdemes lehet jelzős szerkezetbe téve elkülöníteni egymástól. Erre tett javaslatot Jeremy Bentham, aki a **nyers szabadság** (**naked liberty**), illetve a **szerzett szabadság** (**vested liberty**) fogalompárt javasolta alkalmazni erre a kétféle helyzetre (Bentham nézeteit ismerteti: [308, 17]). Amikor a szabadság csak magában áll, akkor nyers, védtelen ahhoz képest, amikor az igényjoggal kiegészülve védve van a mások általi beavatkozástól. A nyers szabadságban megtehetünk mindent tetszésünk szerint, de mindenki másnak szabadsága van megakadályozni minket saját szabadságunk (akaratunk) érvényesítésében. A védett szabadságban a beavatkozás tiltva van a többiek részére.

Nem ejtettem még szót az utolsó hohfeldi kategóriáról, a joghiányról. Ez a fogalom az igényjog ellentéte és a szabadság korrelatív párja. A korrelativitás azt jelenti itt, hogy az egyik fél (Ego) szabadsága megtenni valamit megegyezik azzal, hogy a másik félnek (Alternek) joghiánya van (nincs igénye) arra, hogy az egyik (Ego) ne tegye meg azt:

> „. . . *X* azon jogának korrelatív párja, hogy *Y* ne lépjen a földjére, azonos *Y* azon kötelezettségével, hogy ne tegye azt; *X* azon szabadságának korrelatív párja azonban, hogy saját földjére lépjen, azonos *Y* azon 'jogának hiányával', hogy igényelhesse *X* távolmaradását." [241, 74]

Az ingatlanos példa jogviszonyait kifejező mondatok összetevőkre bontásával a következő táblázatot kaphatjuk:

jogosult	jogviszony	cselekvő	cselekvés	
Ego	szabadsága	Ego		átjárni Ego telkén
Alter	igényhiánya	Ego	nem	átjárni Ego telkén

Hohfeld ezt is korrelatív kapcsolatnak nevezi, pedig ez a viszony különbözik az igényjog és kötelezettség közti korrelativitástól. Ebben az esetben ugyanis két jogviszony nem ugyanarra a cselekvésre irányul, hanem egyszer a pozitív, másszor a negatív cselekvésre (vagyis a tevésre, illetve a tartózkodásra). A különbséget felismerve Stig Kanger, aki később formalizálta Hohfeld elméletét, két fogalmat is bevezetett a korrelativitás e két típusára, és hol az inverz, hol a konverz reláció fogalmát alkalmazta (↰344.o.)

Az igényhiány az igényjog ellentéte. Hohfeld nem mond semmit erről a kapcsolatról, pedig ez eltér a szabadság és kötelezettség közt megállapított ellentét relációtól. Amíg utóbbira Hohfeld azt mondta, hogy „a kötelezettség tartalma az, ami pontosan ellentétes a kérdéses szabadságéval”, addig az egymással ellentétes joghiány és igényjog két jogviszonya ugyanarra a tartalomra vonatkozik [308, 27]. Az ellentéte fogalmat inkább az utóbbi értelmezésben szokás használni, de akárhogy is akarjuk ezt a fogalmat értelmezni, arra mindenképpen ügyelni kell, hogy a hohfeldi elmélet kétféle ‘ellentéte’ relációját megkülönböztessük egymástól.

jogosult	jogviszony	cselekvő	cselekvés	
Ego	igénye	Alter	nem	átjárni Ego telkén
Alter	igényhiánya	Ego	nem	átjárni Ego telkén

A hohfeldi jogfogalmak első csoportja az igényjog köré szervezhető, és a fent bemutatott kapcsolatok alakítják ki a köztük levő struktúrát. Összegzésképpen egy áttekintő táblázatban – az eddigi ingatlanos példákat felhasználva – egyesítem a négy fogalom összetevőkre bontását, hogy jobban látni lehessen a fogalmak egymáshoz való viszonyát.

jogosult	jogviszony	cselekvő	cselekvés	
Ego	igénye	Alter	nem	átjárni Ego telkén
Alter	kötelezettsége	Alter	nem	átjárni Ego telkén
Ego	szabadsága	Ego		átjárni Ego telkén
Alter	igényhiánya	Ego	nem	átjárni Ego telkén

A táblázat első kettő, illetve második kettő sorában egymás korrelatív párjai szerepelnek, míg az első és negyedik, illetve a második és harmadik sorban egymás ellentétei. Bonyolódik a kép, ha figyelembe vesszük, hogy Egonak nemcsak az lehet szabadságában, hogy átjárjon saját telkén, de tartózkodhat is ettől. Ennek az a korrelatív párja, hogy Alternek igényhiánya van arra, hogy Ego átjárjon a telkén. Ezt a következő módon írhatjuk fel:

jogosult	jogviszony	cselekvő	cselekvés	
Ego	szabadsága	Ego	nem	átjárni Ego telkén
Alter	igényhiánya	Ego		átjárni Ego telkén

Ha csak az Ego telkére irányuló jogviszonyok eddig példáit nézzük, akkor is kiviláglik, hogy Hohfeld jogviszonyfogalmai nem elég pontosak. Nem mindegy adott esetben ugyanis, hogy valamely jogviszony relátumaként (paramétereként) milyen minőségű cselekvést írunk be. Hohfeld maga is figyelmeztet a szabadság és kötelezettség ellentéte kapcsán (↰333.o.), hogy ott egymás ellentéteit, a pozitív cselekvést (tevést), illetve a negatív cselekvést (tartózkodást) kell beleérteni a két ellentétes jogviszonyba. De Hohfeld nem megy túl ezen a figyelmeztetésen, nem igazán reflektál a problémára. A továbblépéshez viszont tisztáznom kell, hogy hogyan lehet pontosabban kezelnünk a hohfeldi sémán belül a cselekvés két típusát, ezért ezen a ponton egy rövid kitérőt kell tennem.

Az igényjog köré csoportosítható jogviszonyok felbontása

A hohfeldi jogviszonyfogalmak pontosítása, egyértelmű formalizálása érdekében első lépésként érdemes grafikai eszközökkel reprezentálni a kategóriák közti viszonyokat. Egy hohfeldi jogviszonyt négy változóval írhatunk le. Mindig meg kell adnunk a jog típusát, a jog birtokosát, a jog tartalmát adó cselekvés típusát és e cselekvés ágensét. Ezt az alábbi formulában összegezhetjük:

$$(i) - \textsc{right}(Agent, \pm\textsc{do}(Agent, T(\pm p, \pm p)))$$

Ezt – az egyszerűbb kezelésmód miatt, hangsúlyozottan csak ideiglenes jelleggel – írhatjuk a következő rövidebb alakban is. Azt viszont kétféleképpen is tehetjük. Először kivehetjük a formulából a teljes állapotváltozást ($T(\pm p, \pm p)$), hogy csak azt hangsúlyozzuk, a cselekvésnek van egy Ágense.

$$(ii) - \textsc{right}(Agent, \pm\textsc{do}(Agent))$$

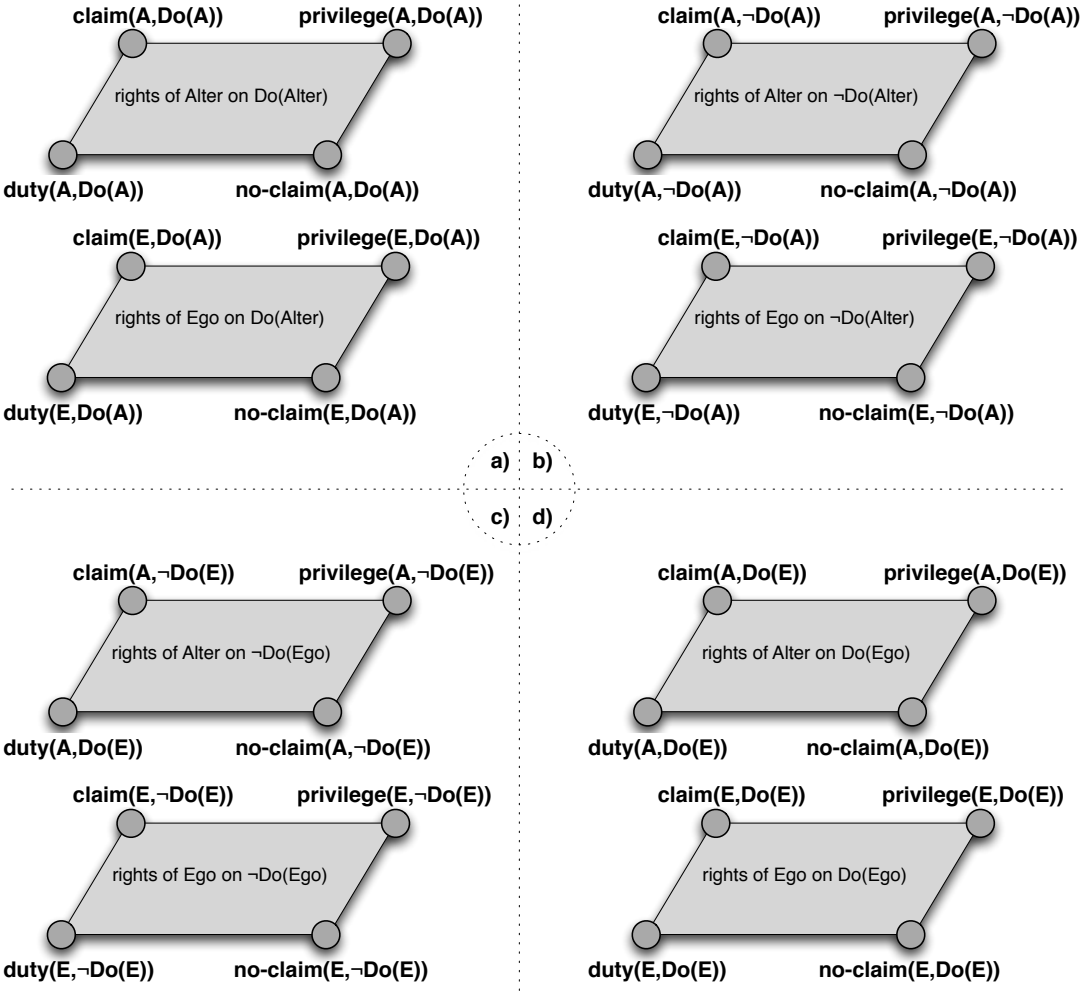
Később bemutatok olyan elméleteket, illetve azokban alkalmazott formalizmusokat, amelyekben fontos lesz valahogy jelezni azt, hogy mi a cselekvés révén "megvalósuló" állapotváltozás során előállt végállapot (ezt a későbbi elméletekhez való igazodás kedvéért az eddig alkalmazott p állapotváltozó helyett F-fel jelölöm).

$$(iii) - \textsc{right}(Agent, \pm\textsc{do}(Agent, \pm F))$$

Ez kevesebb információt képes kifejezni, mint az *(i)* formula, de többet, mint a második *(ii)* képlet. A formula a következő értékeket veheti fel az egyes változókban:

a jog (RIGHT) típusa:	CLAIM | LIBERTY | DUTY | NO-CLAIM
a jog birtokosa (*Agent*):	*Ego (E)* | *Alter (A)*
a jog tartalmát adó cselekvés (±DO):	DO | ¬DO
a cselekvés ágense (*Agent*):	*Ego (E)* | *Alter (A)*

A négy változó közül a Right-nak négy, a többinek két-két értéke lehet, így a formulának összesen 32 különböző kimenete lehet. Egy adott jogbirtokos adott ágens adott cselekvésére vonatkozó jogviszonyait egyetlen "síkban" összefogva, az alábbi nyolc jogviszonycsoportot kapjuk:

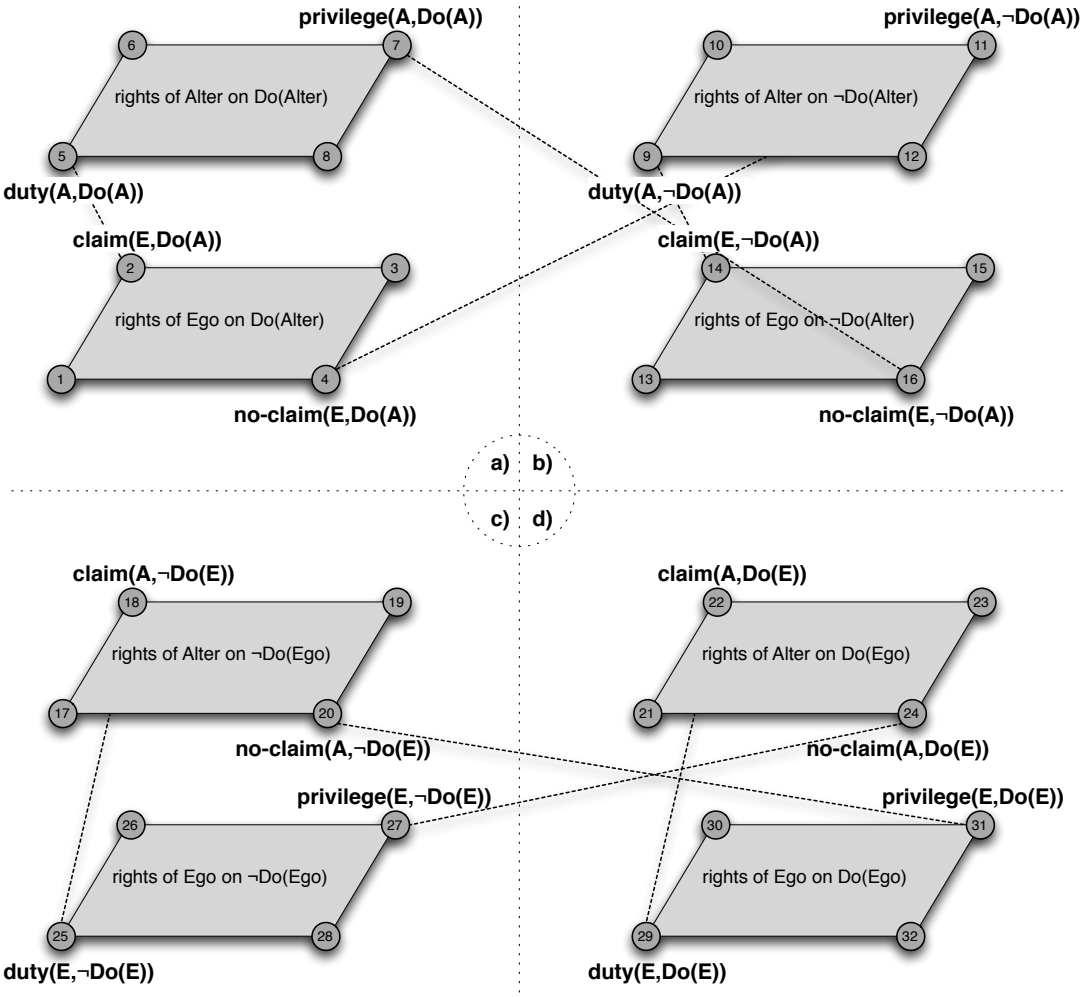


A nyolc síkot négy-négy térnegyedbe rendezhetjük azon az alapon, hogy milyen cselekvésről van szó (vagyis kinek a milyen cselekvéséről beszélünk). Az a) jelű térnegyedben Alter tevésével, a b) jelűben Alter tartózkodásával, a c) jelűben Ego tevésével, a d) jelűben Ego tartózkodásával kapcsolatos jogviszonyok láthatóak úgy, hogy egy-egy negyedbeli síkon Ego, illetve Alter jogai vannak "összekötve".

Észre kell vennünk azonban, hogy nem minden – logikailag lehetséges – jogviszony értelmes. Az igény (jogigény) például csak akkor értelmezhető, ha a jogbirtokos és a cselekvő különbözik. Nincs értelme saját cselekvésemre irányuló igényről beszélni. Ezen a ponton ugyanis feltételezhetjük, hogy egy ágens saját cselekvésére irányuló akarata egyértelmű, ha valaki akar valamit, akkor azt végre is hajtja.[4] Ugyancsak (és ugyanazért) értelmetlen az igényhiány fogalma is azonos ágensek esetében. Tartalmatlannak tarthatjuk a szabadság fogalmát is – eltérő ágensekre vonatkozóan. A szabadság csak saját cselekvésemet illetően értelmes, mással cselekvésére vonatkozóan nekem nincs, nem lehet szabadságom. Ugyanez igaz a kötelezettségre is: csak magamat illetően lehetnek kötelezettségeim, másokra nézve csak a kötelezettségre vonatkozó igényem képzelhető el.

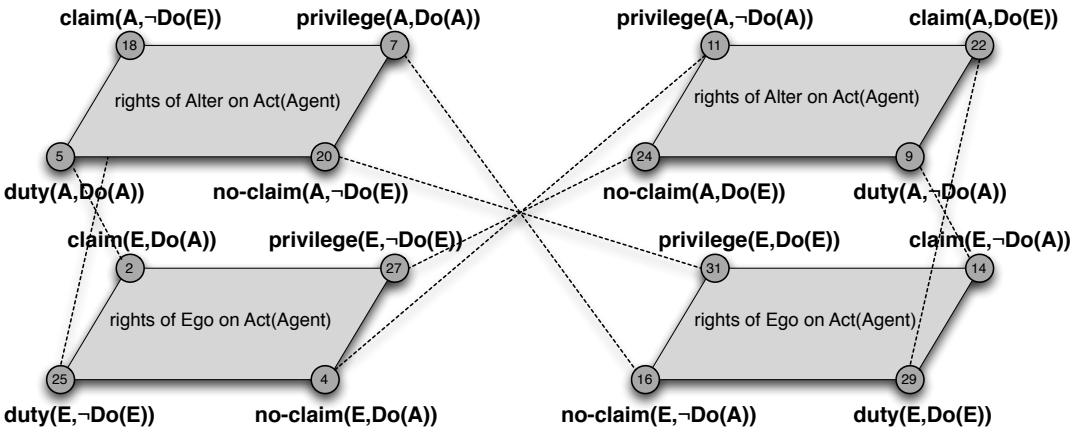
[4]Nem tagadva az akaratgyengeség problémájának cselekvésfilozófiában betöltött jelentőségét, ezt a kérdést itt, ebben a kontextusban irrelevánsnak gondolom.

Ha mindezek alapján a “nem értelmes” fogalmakat töröljük a “palettáról”, akkor már csak 16 jogviszonyhoz jutunk. Ha ezek között feltüntetem a hohfeldi korrelativitási kapcsolatokat, akkor a következő ábrát kapjuk:

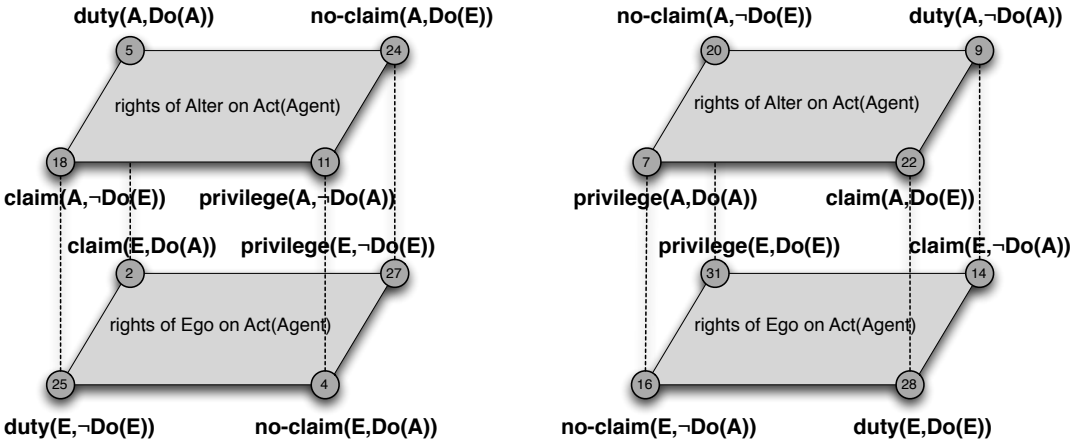


Az ábrán már nem láthatóak az “értelmetlen” jogviszonyok, csak azok szerepelnek, amelyek a hohfeldi értelmezésekben felbukkantak, felbukkanhattak. A korrelativitást jelző vonalak “átkötései” más negyedekbe szépen mutatják azt, hogy a jogfogalmak közti belső összefüggések másfajta cselekvéseket kapcsolnak össze. Az ábra tehát azt is szemlélteti egyben, hogy a fogalmak pontos értelmezéséhez mennyire fontos figyelembe venni azt, hogy milyen cselekvésre vonatkoznak a jogviszonyok. Hohfeld sajnos nem igazán reflektált erre a kérdésre.

A jogviszonyok elméletileg teljesnek mondható ábráját tovább egyszerűsíthetjük úgy, hogy két-két negyedet “összecsúsztatunk”. Ha az egymás alatt levő negyedekkel (tehát az a) és c), illetve a b) és d) jelűekkel) tesszük meg ezt, akkor olyan síkokhoz jutunk, amelyek ugyanannak a jogbirtokosnak a jogviszonyait mutatják. A két alsó síkon Ego, a két felső síkon Alter jogviszonyai láthatók. A jobb áttekinthetőség kedvéért a jobb oldali két síkot tükröztem a függőleges tengely mentén, mert így jobban láthatókká váltak a korrelativitási kapcsolatok. Ez alapján már könnyben értelmezhető ábrát kapunk, hiszen már csak feleannyi pozíciót kell értelmeznünk.



Egy utolsó átrendezést végrehajtva a Hohfeldi sémához nagyon hasonló szerkezethez juthatunk. Csak annyit kell tennünk, hogy a két felső síkot a vízszintes tengelyük körül elforgatjuk 180 fokkal (vagyis a síkok felső és alsó csúcspontjait megcseréljük egymással), majd ezután a két felső sík privilege és no-claim jogviszonyait kicseréljük egymással, végül a második kocka csúcspontjait a lapra merőleges, függőleges síkra vetítve tükrözzük. Ekkor ugyanúgy az alsó két síkon helyezkednek az Ego, a felső két síkon az Alter jogviszonyai, viszont a korrelatív kapcsolatok egymás fölé kerülnek, és az egész struktúra két részre "esik szét". A két kocka elemei csak abban térnek el egymástól, hogy mindig az ellentétes cselekvésre vonatkozó jogviszonyt írják le.



Mindkét ágens jogviszonyai között (tehát a felső és alsó két síkon) találhatunk két-két négyes csoportot. Mindkét ágens esetében négy-négy jogfogalom összetartozik azon az alapon, hogy a jogviszony az ágens saját vagy a partnere cselekvésére irányul-e (mindig a DUTY és PRIVILEGE, valamint a CLAIM és NO-CLAIM fogalmak kapcsolódnak így össze). Ezeket a fogalmi négyeseket lehet majd deontikus logikai formulák segítségével megragadni. Az ábrából az is kiderül, hogy Hohfeld négy-négy jogosultság, illetve kötelezettség fogalma kevés, valójában kétszer annyira van szükség.

Az utolsó ábra "eredményeit" táblázatos formában, szövegesen is reprezentálhatom. A pozitív és negatív cselekvést tevésnek és tartózkodásnak írva a már megszokott táblázatos formában az alábbi módon írhatjuk fel a nyolcféle jogviszonyt:

jogosult	jogviszony	ágens	cselekvés	jogosult	jogviszony	ágens	cselekvés
Ego	igénye	Alter	tartózkodni	Alter	kötelezettsége	Alter	tartózkodni
Ego	szabadsága	Ego	megtenni	Alter	igényhiánya	Ego	tartózkodni
Ego	igénye	Alter	megtenni	Alter	kötelezettsége	Alter	megtenni
Ego	szabadsága	Ego	tartózkodni	Alter	igényhiánya	Ego	megtenni
Ego	kötelezettsége	Ego	tartózkodni	Alter	igénye	Ego	tartózkodni
Ego	igényhiánya	Alter	megtenni	Alter	szabadsága	Alter	tartózkodni
Ego	kötelezettsége	Ego	megtenni	Alter	igénye	Ego	megtenni
Ego	igényhiánya	Alter	tartózkodni	Alter	szabadsága	Alter	megtenni

A táblázatban az elemi jogviszonyok mellett mindig feltüntettem a korrelatív párjaikat is. A táblázat baloldali részében van Ego, a jobboldaliban Alter nyolc-nyolc jogviszonya. Ezeket a 6.2.1 ábrán a két kocka két-két alsó, illetve felső síkján lehet megtalálni.

✠ ✠ ✠

Miután kiderült, hogy meg kell duplázni Hohfeld jogviszonyfogalmait, és nyolc alapfogalomról (illetve ezeknek nyolc korrelatív párjáról) kell beszélnünk, keresnünk kell olyan a terminusokat a jogi diskurzuson belül, amelyeket hozzárendelhetünk a kibővült fogalmi tér új elemeihez. A jogviszonyra vonatkozó tipizálási javaslatok közül a két legismertebb az **aktív jog** (ACTIVE RIGHT) és **passzív jog** (PASSIVE RIGHT), valamint a **pozitív jog** (POSITIVE RIGHT) és **negatív jog** (NEGATIVE RIGHT) megkülönböztetésén alapul.

Az aktív jog, passzív jog kettősségét David Lyons mutatja be a 'The Correlativity of Rights and Duties' című cikkében [315]. Lyons szerint az aktív jog az, amikor az ágensnek joga van arra, hogy maga cselekedjen valamit, a passzív jogról pedig akkor beszélhetünk, ha az ágensnek joga (igénye) van arra, hogy egy másik ágens cselekedjen valamit. Hohfeld eddig elemzett jogosultságfogalmai közül az igényjog passzív jognak, a szabadság aktív jognak minősíthető. A passzív jog esetében a jogosultság korrelatív párja mindig valamilyen aktív kötelezettséget jelent, az aktív jog esetében pedig a korrelativitás valamely passzív jogigény hiányáról van szó. A korrelatív kapcsolatban levő jogfogalmak tehát mindig eltérő minőségűek az aktív-passzív dimenzióban.

A jogot szokás felosztani pozitív, illetve negatív jogra is. Innentől kezdve azonban nagyon ügyelni kell arra, hogy milyen jogfogalmakat hogyan használunk. Ezen a ponton ugyanis több forrásból, más megközelítésekből származó fogalmakat kell majd egységes keretben kezelni. A jog fogalmával foglalkozó szakirodalomban az eddig elemzett jogosultságfogalmainkra (tehát mind az igényjogra, mind a szabadságra) vonatkozóan ugyanúgy alkalmazzák a pozitív/negatív jog terminusait. Nagyon vigyázni kell tehát akkor, amikor ezeket egységes keretbe foglalva akarjuk értelmezni.

Jan Narveson volt az egyik olyan szerző, aki javasolta pozitív és negatív jog fogalmainak megkülönböztetését (amit átvett tőle például a Stanford filozófiai lexikon joggal foglalkozó szócikkében a cikk szerzője, Leif Wenar is [578]) is. Narveson a hohfeldi igényjog kategóriájára gondolt, noha nem reflektált rá. Jól mutatja ezt az a mód, ahogy Narveson a jogot a kötelezettség ellenében határozta meg, de ezt erősíti az a tény is, hogy milyen példákat hoz a **pozitív jog** (POSITIVE CLAIM), illetve a **negatív jog** (NEGATIVE CLAIM) fogalmára [371, 57-58]. A negatív jogon Narveson azt érti, hogy az egyén (a jogbirtokos) "jogosult arra, hogy ne avatkozzanak bele mások a cselekvésébe". Ilyen például a "megtámadás elleni védelemhez", a "méltósághoz", a "jó hírnévhez" vagy a "magánlakás sérthetetlenségéhez" való jog. Ezzel szemben a pozitív jogon azt érthetjük, hogy az

egyén (a jogbirtokos) jogosult valamire, valamilyen jószágra, szolgáltatásra. Példa rá a "szociális segélyre" való jog.

A definíciókból, a példákból egyértelműen látszik, hogy itt a hohfeldi igényjog fogalmának alábontásáról van szó. Rögzíteni is érdemes, hogy a passzív jog fogalma alatt – a fenti értelemben – létrehozhatjuk a pozitív és negatív jog fogalmait. Amíg a pozitív passzív jog mások aktív cselekvésére (valaminek a megtevésére) való igényt, addig a negatív passzív jog mások valamitől való tartózkodásra vonatkozó igényt jelenti. Nyilvánvalóan adódik azonban itt a kérdés, vajon a többi jogosultságfogalom esetében alkalmazható-e a pozitív/negatív dimenzió a további tipizálásra. Vannak, akik nem így gondolják. A Stanford online filozófiai lexikon joggal foglalkozó cikkében például ez szerepel:

> „Since both negative and positive rights are passive rights, some rights are neither negative nor positive. Privileges and powers cannot be negative rights; and privileges, powers, and immunities cannot be positive rights. The (privilege-) right to enter a building, and the (power-) right to enter into a binding agreement, are neither negative nor positive." [578].

Nem értek teljesen egyet ezzel a gondolatmenettel. Azon a ponton mindenképpen tévesnek tartom, amikor a pozitív/negatív jog kettősségét a passzív jog fogalma alá rendeli. A pozitív és negatív jog megkülönböztetése ugyanúgy értelmes az aktív jog esetében is. Ezt mutatom meg a továbbiakban.

Adja magát, hogy azonnal hivatkozzak itt a kérdés leghíresebb szerzőjére, Isaiah Berlinre, aki a 'A szabadság két fogalma' című tanulmányában definiálta a pozitív és negatív szabadság fogalmát [47], és ezzel megalapozta, hogy e két fogalom hosszú időre bekerüljön a jogok típusairól szóló diskurzusba. Arra azonban nagyon vigyázni kell Berlin kétféle szabadságfogalma kapcsán, hogy ne tévesszen meg minket a terminusszintű azonosság. Berlin ugyanis – részben vagy teljesen – más értelemben használja a fogalmait.

Talán úgy tudom legpontosabban megmutatni, hogy hol térnek el a hohfeldi és berlini értelmezések, logikák, ha előbb megvizsgálom Jan Narveson szabadságfelfogását. Narveson is kétféle szabadságfogalmat definiál, hasonló szellemben, mint ahogy azt Berlin is teszi, csak még általánosabb formában [371, 22-23]. **Pozitív szabadság** (POSITIVE FREEDOM) fogalma alatt Narveson azt az állapotot érti, amikor minden szükséges feltétel adott ahhoz, hogy valamely cselekvés végrehajtható legyen. Ezek a feltételek lehetnek külső és belső feltételek egyaránt. Utóbbira a foci szabályainak ismeretét, megfelelő testi kondíciót, előbbire a szükséges számú társ jelenlétét, focilabda, szabad focipálya meglétét hozza fel Narveson példaként. Ezzel szemben a **negatív szabadság** (NEGATIVE FREEDOM) fogalmát Narveson olyan állapotként definiálja, amelyből hiányzik minden olyan mozzanat, amely megakadályozná valamely cselekvés végrehajtását. Példa lehet erre egy másik, ugyanazon a pályán focizni vágyó társaság hiánya. Narveson tehát a kétféle szabadságfogalom értelmezésekor bármilyen feltételt "megengedhetőnek", figyelembe vehetőnek tart.

Isaiah Berlin 'A szabadság két fogalma' című tanulmányában definiálta a pozitív és negatív szabadság fogalmát [47]. Berlin a negatív szabadság fogalmával kezdi az elemzését, és abban az értelemben Narveson definíciójához hasonlítható Berlin meghatározása is, hogy a cselekvést akadályozó feltételek hiányaként értelmezi a negatív szabadságot.

> „Kényszeríteni valakit annyi, mint megfosztani őt a szabadságától." [47, 340]
>
> „Általában azt mondják, hogy annyira vagyok szabad, amennyire más emberek vagy más

> emberekből álló csoportok nem avatkoznak be cselekedeteimbe. A politikai szabadság ebben az értelemben egyszerűen az a terület, amelyen belül az ember anélkül cselekedhet, hogy mások megakadályoznák ebben." [47, 342]
>
> „. . . a szabadság ebben az értelemben a *valamitől* való szabadságot jelenti, vagyis a változó, de mindig felismerhető határok áthágásának hiányát." [47, 352]

Narveson felfogásához képest Berlin szűkebbre fogja a negatív szabadságot biztosító, hiányzó feltételek körét – bár erre a kérdésre nem igazán reflektál tudatosan. Az mindenesetre kiderül a fenti idézetből, hogy Berlin a negatív szabadságot más emberek beavatkozásától való mentességként gondolja el. Ennyiben szűkebben definiálja a fogalmat Narvesonhoz képest, viszont ez a megközelítés gyakorlatilag megegyezik Hohfeld szabadságfelfogásával. Ahol más emberek nem avatkozhatnak be a cselekvésembe, ahol másoknak igényhiánya van velem szemben, ott vagyok szabad – Berlin negatív szabadsággal kapcsolatos értelmezésében.

Mivel sem Berlin, sem Narvenson nem különbözteti meg a pozitív és negatív cselekvéseket egymástól, ezért a negatív szabadság fogalmába sorolt példákban sem figyelnek erre a szempontra. Berlin negatív szabadságra vonatkozó példái között találhatunk a tevésre és a tartózkodásra irányuló utalásokat egyaránt. A narvesoni negatív szabadság értelmezés túl tág, a berlini felfogás már megfelelő terjedelmét tekintve, és jelentését tekintve az ő definíciója megegyezik Hohfeldével, de a két negatív szabadság fogalom egyike sem segít minket abban, hogy meg tudjuk nevezni a hohfeldi szabadság fogalom két altípusát, azt a két esetet, amikor pozitív, illetve negatív cselekvésre irányuló szabadságokról beszélünk. Más a helyzet Berlin pozitív szabadság fogalma esetében. Itt nemcsak arról van szó, hogy az új kategóriát nem tudjuk elhelyezni a hohfeldi struktúrába, nem találunk neki megfelelő párt a hohfeldi fogalmak közül, hanem arról is, hogy az értelmezéséhez más dimenzióra van szükség a hohfeldi sémához képest. Vegyük a következő néhány rövid passzust.

> „Mit ér a szabadság annak, aki nem tud mit kezdeni vele? Mi a szabadság értéke, ha nincsenek meg a felhasználásához szükséges feltételek?" [47, 347]
>
> „Egy cél elérésére való képtelenség önmagában nem jelenti a politikai szabadság hiányát. . . . Azt mondják, nagyon meggyőzően, hogy ha egy ember túl szegény ahhoz, hogy elérjen valamit, amit jogilag nem tilos – egy vekni kenyeret, egy világ körüli utat, vagy azt, hogy a bírósághoz fordulhasson jogorvoslatért –, éppoly kevéssé szabad, mint hogyha ezen dolgok elérésétől a törvény tiltaná el." [47, 344]
>
> „A »szabadság« szó »pozitív« értelme az egyénnek abból a vágyából fakad, hogy saját magának az ura legyen. Azt szeretném, ha életem és döntéseim rajtam múlnának, nem pedig holmi külső erők határoznának rólam. Saját akarati aktusomnak kívánok eszköze lenni, nem pedig másokénak. Szubjektum, nem pedig objektum akarok lenni, szubjektum, amelyet saját gondolatai, saját tudatos céljai irányítanak, s nem külső okok mozgatnak." [47, 361]

A negatív szabadságra vonatkoztatható "szabadság valamitől" felfogással szemben itt a "szabadság valamire" értelmezés lehet az érvényes Berlin szerint. A pozitív szabadság fogalmát jóval hosszabban elemzi Berlin, mint teszi azt a negatív szabadsággal kapcsolatban. Sok olyan példát hoz (vagy lehetne hozni Berlin értelmezése alapján), amely kivezet a hohfeldi jogviszonyok fogalmi köréből, mert kulturális, gazdasági teendőket "nevez meg", de sok olyan pozitív szabadság körébe tartozó példa kerül elő, amely a hohfeldi séma igényjog kategóriája alá lenne besorolható. A pozitív szabadságot tehát Berlin úgy értelmezi, hogy vagy már ismert kategóriával állítható párhuzamba (az igényjog ho-

hfeldi fogalmával), vagy nem a perszonalizált jogviszonyok fogalmi körén belül ragadható meg. Akármelyik esetről is van szó, egy a lényeg: ez a kategória nem hasznosítható a modellépítés során. Bár más terminusokat használ, mások a kérdésfeltevései, mégis érdemes felidézni itt, amit Cass R. Sunstein írt a pozitív és negatív jogok különbségeiről, mert segít értelmezni a kétféle berlini szabadságfogalom viszonyát (nyugodtan be lehetne helyettesíteni a 'jog' helyébe a 'szabadság' terminusát, ugyanúgy érvényes maradna Sunstein mondandója a berlini fogalmakra vonatkoztatva.

> A pozitív jogok „ ... nem érvényesíthetők bírósági úton: különleges szerepet töltenek be, és elsősorban arra szolgálnak, hogy az állami tisztségviselők és az állampolgárok összességének általános céljait megfogalmazzák.
> A negatív jogok ... azonban bírósági védelmet élveznek, és nem szükséges a bírói út igénybevétele előtt semmilyen más állami hatósághoz fordulni. Senki sem függhet az állami hatóságoktól, ha a jogait bíróság előtt kívánja étvényesíteni.
> A gazdasági *szabadságok* megőrzésére tett erőfeszítéseket kísérheti az a tudatos törekvés, hogy csökkentsük a gazdasági *jogok* szerepét." [513, 14]

Az idézetben nem a 'politikai üzenet' a fontos ("nem annyira kívánatos a pozitív jogokat az alkotmányba emelni"), hanem inkább az, ahogy a kétféle fogalmat szembeállítja egymással a szerző. A kétféle jogfogalom inkább a politika vagy a jog intézményrendszeréhez köthető, és e két kötődés eltérő "működési logikát" eredményez.

✠ ✠ ✠

A jogelméleti diskurzus kétféle jogtipológiája, a pozitív-negatív jog, illetve az aktív-passzív jog fogalomkettőse csak kisebb részben, de azért használható a hohfeldi elmélet pontosítására. A passzív-aktív dimenzió a későbbiekben segíthet még minket a hohfeldi jogviszonyok értelmezésében, de a modellépítésnek jelen fázisában nem alkalmazható. A pozitív-negatív jog megkülönböztetése egy az egyben ráhúzható az igényjog két altípusára, a pozitív és negatív cselekvésekre irányuló igényjogok eseteire. Amíg a pozitív igényjog esetén a partner pozitív cselekvésére (tevésére) irányuló igényről beszélhetünk, addig a negatív igényjog esetén az igény a partner negatív cselekvésére (tartózkodására) vonatkozik. A szabadság esetében is ugyanezt megtehetjük, tehát értelmezhetjük a negatív és pozitív szabadság fogalmát (tevésre és tartózkodásra irányuló szabadságokat értve alattuk), de ez a felfogás teljesen eltér a Berlin-féle értelmezéstől.

Meg kell itt jegyeznem azt is, hogy könnyen zavart okozhat az, hogy mind a pozitív és negatív, mind az aktív és passzív minőségeket vonatkoztatjuk a cselekvésekre, és a gyakorlatban nehéz a pontos különbséget tenni köztük, ami sajnos sok félreértésre, félrehasználatra, olykor túlhasználatra ad alkalmat. Ennek oka egyszerű (és érthető): mindkét dichotómia könnyen "ráhúzható" a tevés és tartózkodás kettősségére, hiszen a tartózkodás a tevéshez képest értelemezhető egyfelől negatívnak a pozitívhoz, másfelől passzívnak az aktívhoz képest. Ezen a pongyolaságon nem tudok változtatni. Csak annyit tehetek a potenciális félreértések kizárása érdekében, hogy ebben a könyvben az aktív-passzív jelzőpárost csak a jogra vonatkoztatom, míg a pozitív-negatív minőségeket mind a jog, mind a cselekvés fogalma esetében arra használom, hogy a jog, illetve a cselekvés típusait képezzem a segítségükkel. Ez a kétfajta használat azonban erőteljesen összekapcsolódik, hiszen a pozitív és negatív jog (igény vagy szabadság) altípusait (pozitív, illetve negatív jog) a pozitív vagy negatív cselekvésre (tevésre vagy tartózkodásra) irányultság alapján definiálom.

6.2.2 Az elsőrendű jogviszonyfogalmak formalizálása

A hohfeldi jogfogalmak logikai leírását a skandináv logikai iskola tagjai végezték el a hatvanas-hetvenes években. A legfontosabb szereplője ennek a történetnek Stig Kanger, svéd logikus volt. Ebben a fejezetben Kanger formalizálási kísérletét mutatom be, majd rámutatok arra, hogy bár a Kanger által teremtett formalizmus konzisztens, ezért beemelhetőnek gondolom a társadalmi cselekvés általános modelljébe is, de a kangeri javaslat nem pontosan a hohfeldi fogalmakat ragadja meg. Ezek után megpróbálom korrigálni a kangeri elmélet problémáit, pótolni a hiányosságait. Eközben természetesen többször támaszkodni fogok Kanger követőinek észrevételeire, kritikáira, javaslataira.[5]

Hohfeld-Kanger-féle elemi jogtípusok

Stig Kanger az ötvenes évek végén és a hatvanas években írta meg azokat a tanulmányait, amelyek a hohfeldi jogviszonyok formállogikai feldolgozásával foglalkoztak [274], [275], [273]. Még a kangeri formalizmus bemutatása előtt hangsúlyoznom kell, hogy egy szempontból hibásnak tartom Kanger megoldását, amit mindenképpen javítani kell. Kanger ugyanis azzal, hogy egy logikai szintre helyezi el a két hohfeldi jogviszonycsoportot, elfogadhatatlan módon összecsúsztatja azokat, és így nem tudja a felhatalmazottság köré szerveződő jogviszonyfogalmakat jól megragadni. Ez nem azt jelenti, hogy a javasolt formalizmus nem alkalmazható. A logikai struktúra egy az egyben elfogadaható és átvehető, csak annyi változtatásra van szükség, hogy a kangeri sémát más fogalmak értelmezésére használjuk. Ebből persze az is következik, hogy mivel a felhatalmazottság körüli fogalomnégyest nem lehet figyelembe venni a kangeri séma interpretációja során, ezekre a fogalmakra új sémát kell kidolgozni. A hohfeldi jogviszonyfogalmak formalizálása és interpretációja előtt Kanger felállít egy logikai elméletet, egy öt axiómából álló sémát, amely két operátort definiál. Ezek segítségével fogja majd Kanger a hohfeldi jogfogalmakat meghatározni. Kanger a logikai sémája eredetileg ebből az öt formulából állt:

1)	$If\ F \longrightarrow G,\ then\ Shall\ F \longrightarrow Shall\ G$
2)	$(Shall\ F\ and\ Shall\ G) \longrightarrow Shall\ (F\ and\ G)$
3)	$Shall\ F \longrightarrow not\ Shall\ (not\ F)$
4)	$If\ F \longrightarrow G\ and\ G \longrightarrow F,\ then\ X\ causes\ F \longrightarrow X\ causes\ G$
5)	$X\ causes\ F \longrightarrow F$

Ez az öt axiómás elmélet erősnek mondható, amely deontikus logikai értelemben is megalapozottnak és jól-formáltnak tekinthető, valamint lehetővé teszi a hohfeldi jogviszonyfogalmak (illetve azok kiegészítéseinek) pontos formalizálását.[6] Később maga Kanger és a követői a két operátort az alábbi módon interpretálták:

$$\begin{aligned} \textsc{Shall}\ F &= \textit{it shall be that } F \\ X\ \textsc{causes}\ F &= X\ \textsc{sees to it that}\ F\ \textit{holds} \end{aligned}$$

[5] Kanger elméletét kimerítően tárgyalja Lars Lindahl könyve [308]. Kanger elméletén túl Lindahl bemutatja a Hohfed előtti legfontosabb szerzők jogfogalmakkal kapcsolatos nézeteit is, és saját elméletét is kifejti a jogi pozíciókról.

[6] Erről bővebben lásd Lindahl könyvét [308, 38-54].

Az elméleti beágyazás megkönnyítése végett érdemes itt jelezni, hogy az első SHALL operátor a deontikus kötelezés (**O** – OBLIGATORY) operátorának feleltethető meg, a második 'sees to it that' pedig a cselekvés operátoraként értelmezhető (DO). Mivel a jogfogalmak formalizálásához szükség van további összetevők jelölésére, Kanger a jogviszonyok típusaira temészetesen átveszi Hohfeld kategóriáit (bár a Right helyett a Claim, a Privilege helyett a Freedom terminusait használja). A jogviszonyhordozót (bearer, holder) jelöli *X*-szel, a jogi partnert, jogi társat (counter-party) *Y*-nal, és F(*X,Y*)-nal jelöli a jogviszony tartalmát, amit állapotként értelmez (object-matter, state of affairs, condition). Fontos tény, hogy Kanger az állapotokat, nem pedig a cselekvéseket emeli be a formalizmusába. Mind a jogelmélet, mind a dentikus logika a cselekvésekre vonatkoztatja a jogviszonyokat (normatív viszonyokat). Hohfeld is is így értelmezi saját fogalmait (*X*-nek igénye van *Y*-nal szemben arra, hogy *Y* végezzen el egy cselekvést vagy tartózkodjon tőle. Kanger pedig egy jogviszonyt két személlyel és egy állapottal tart megragadhatónak. Jelöljük ezt a továbbiakban így:

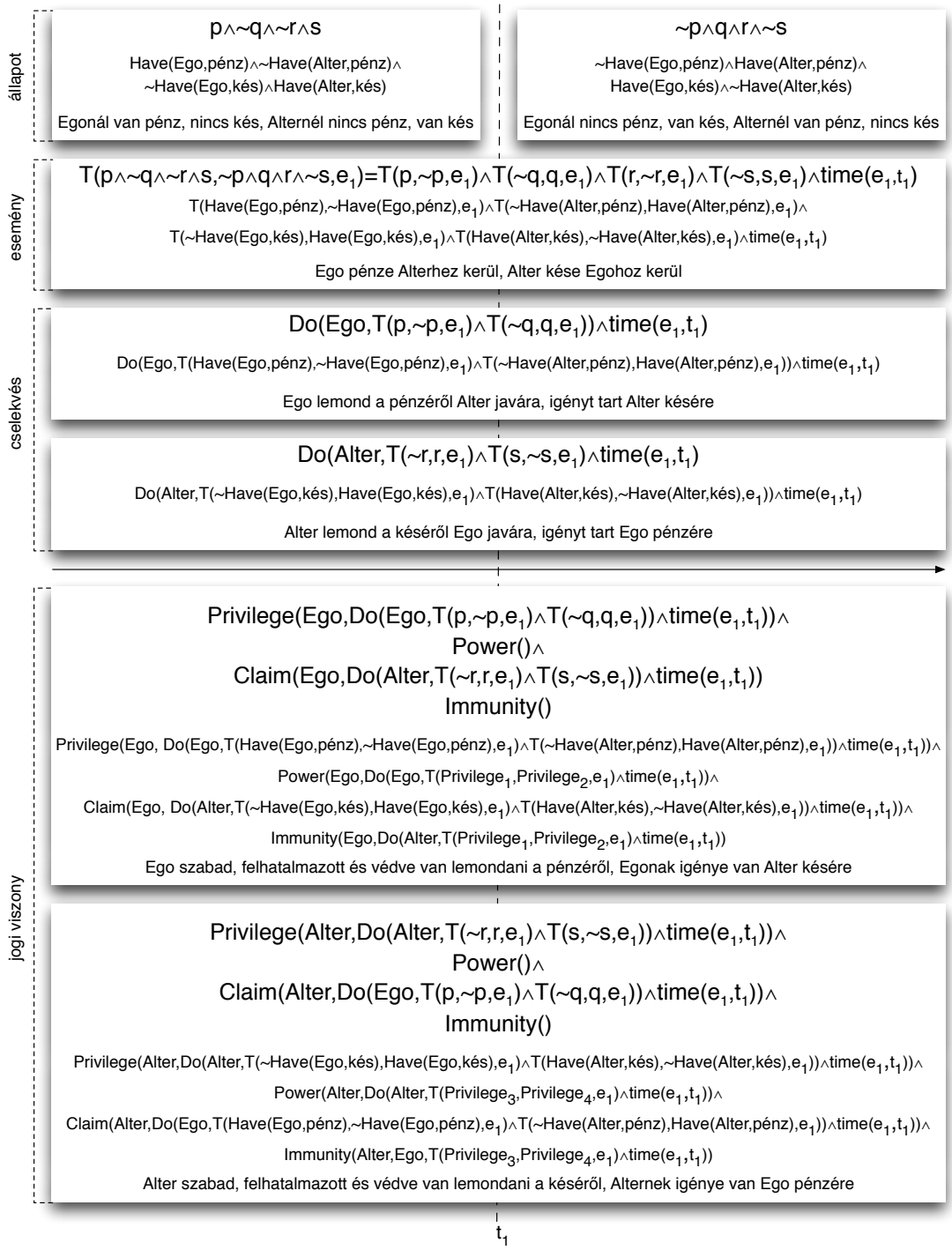
$$R(X,Y,F(X,Y))$$

Ez a megoldás nagyobb kifejezőerővel bír. Hogy miért és hogyan, azt Lindahl alaposabban bemutatja könyvében [308]. Röviden: a nagyobb kifejezőerő onnan származik, hogy ha az állapotokra vonatkoztatjuk a jogviszonyokat, akkor el tudjuk különíteni egymástól azokat az eseteket, amelyekben egy állapot vagy annak tagadása áll fent. Egy pozitív vagy negatív cselekvés (tevés vagy tartózkodás) más és más jelentéssel bírhat attól függően, hogy egy állapotnak vagy annak komplementerének fennállására irányul. Egy aktív cselekvéssel létre lehet hozni egy állapotot vagy – épp' ellenkezőleg – meg lehet akadályozni, hogy fennálljon. Ugyanez fordítva is igaz lehet: egy tartózkodás eredményeként előállhat egy állapot, de meg is szűnhet (tehát az állapot ellentéte jöhet létre). Ezeket az eseteket a kangeri formalizmussal mind meg tudjuk ragadni. Ehhez arra van szükség, hogy a cselekvés operátor mellett az állapotokat vegyük be a jogviszonyok paraméterei közé. Természetesen nem arról van itt szó, hogy Kanger a cselekvésektől eltekintene a jogviszonyok leírásakor, hanem inkább arról, hogy többet akar megragadni azáltal, hogy a cselekvések összetevőit is figyelembe veszi. Még egy lépést lehetne tenni ebben az irányban, ha a cselekvések összetevőkre bontását von Wright szellemében végeznénk el (☞137.o.). Von Wright a cselekvéseket az állapotok közti tranzíciókkal modellezte, vagyis egy cselekvés összetevői közé beemelte a kezdő és végállapotokat és a köztük fennálló átalakulás típusát. Kanger a végállapotot veszi figyelembe, ezáltal bővíti a modellje leíróerejét. Ha a von wright-i megoldást követnénk, akkor még finomabb modellhez juthatnánk, de félő, hogy túl finomszemcséssé válna már a modell és nehezebben áttekinthető, kevéssé használható lenne. Ezért itt csak a további finomítás elvi lehetőséget vetem fel, de nem törekszem a javaslat kibontására. A kangeri formulát a modellembe eddig bevezetett terminusokra átírva a következő alakot kapjuk (a második képlet megegyezik a 336. oldalon található *(iii)* jelű sor második formulájával):

$$R(X,Y,F(X,Y)) \quad \rightarrowtail \quad \text{RIGHT}\big(Agent, \pm\text{DO}(Agent, \pm F)\big)$$

Fontos megjegyezni, hogy a kangeri formulához képest még van egy változás a második képletben. Kanger a cselekvés eredményeként "létrejövő" állapot leírásakor jelezni akarja azt is, hogy az adott állapot milyen ágensek között értelmezhető, ezért írja a képletébe az

$F(X,Y)$ terminust. Ezt a két paramétert elhagyhatónak érzem, mivel Kanger érdemben nem használja ki a két paraméter alkalmazásában rejlő információt. Amikor a kangeri elméletet mutatom majd be, akkor – a "történeti hűség kedvéért – én is jelezni fogom ezeket a ágenciára utaló paramétereket, de bármikor, amikor továbbviszem valamilyen módon a kangeri formalizmust, akkor már elhagyom ezeket. Szemléltetésképpen bemutatom, hogyan lehetne formálisan leírni az *árucsere* aktusában rejlő jogviszonyokat úgy, hogy von Wright-et követve (↳137.o.) az állapotváltozásokat is beemeljük a formulába.



A hohfeldi jogviszonyokat négy egymásra épülő szinten írhatjuk le. Az első szinten az állapotok vannak, a másodikon az események (változások, átmenetek), a harmadikon a cselekvések, a negyediken a jogviszonyok. Az árucserét az állapotszinten úgy írhatjuk le, hogy a kezdőállapothoz képest ($p \wedge \neg q \wedge r \wedge \neg s$) a végállapotban minden ellentétébe fordul ($\neg p \wedge q \wedge \neg r \wedge s$). A két állapot közti átmenetet az eseményszinten az alábbi formula írja le:

$$T(p \wedge \neg q, \neg p \wedge q, e_1) \wedge T(r \wedge \neg s, \neg r \wedge s, e_1) = \\ T(p, \neg p, e_1) \wedge T(\neg q, q, e_1) \wedge T(r, \neg r, e_1) \wedge T(\neg s, s, e_1)$$

A cselekvési szinten már külön kell leírni az – árucserét megvalósító – két szereplő cselekvését, hiszen két ágens cselekvése külön cselekvésnek számít. A cseréhez mindig két cselekvőre és két cselekvésre van szükség:

$$Do(Ego, T(p, \neg p, e_1) \wedge T(\neg q, q, e_1) \wedge T(r, \neg r, e_1) \wedge T(\neg s, s, e_1)) \wedge \\ Do(Alter, T(\neg p, p, e_1) \wedge T(q, \neg q, e_1) \wedge T(\neg r, r, e_1) \wedge T(s, \neg s, e_1))$$

Ezután következhet a leírás negyedik szintje, ahol a jogviszonyokat reprezentálhatjuk. A négyféle jogviszony közül egyet kiemelve, Ego szabadságára vonatkozó formula a következő:

$$Privilege(Ego, Do(Ego, T(p, \neg p, e_1) \wedge T(\neg q, q, e_1)) \wedge time(e_1, t_1))$$

Kanger jognak mond minden hohfeldi jogviszonyt (tehát a jogosultságokat és a kötelezettségeket egyaránt). A kötelezettségeket nem kezeli, legtöbb tanulmányában nem is nevezi meg őket, egyszerűen csak a jogosultságok tagadásaként hivatkozik rájuk. Viszont megduplázza a hohfeldi típusokat amiatt, hogy a jogosultság fogalmának egyik fogalmi összetevőjét, a két ágens közti – jogi – állapotot pontosabban megragadhassa, hogy ki tudja fejezni, vajon az állapot vagy annak tagadásának esete igaz-e. Az eredeti hohfeldi kategóriák közül tehát a tagadás segítségével "kiszorítja" a kötelezettségfogalmakat:

jogosultság	kötelezettség	(= jogosultság tagadása)
right	duty	(= not privilege)
privilege	no-right	(= not right)
power	liability	(= not immunity)
immunity	disability	(= not power)

A következő lépésben aztán (a right-claim, illetve privilege-freedom terminuscserék végrehajtása után) Kanger megduplázza a jogosultságok számát mondván, hogy a jogviszony más lesz akkor, ha a tartalmát az F(X,Y) állapottal vagy annak tagadásával írjuk le. Az ellentétes állapottal jellemezhető jogviszonyokat Kanger a 'counter-' előtaggal látja el, így a következő nyolc jogosultságtípust sorolja fel:

Hohfeld kategóriái	Kanger kiegészítései
claim	counter-claim
freedom	counter-freedom
power	counter-power
immunity	counter-immunity

Kanger a hohfeldi jogviszonyokat (illetve azok kiegészítéseit) **egyszerű jogtípusoknak** (**simple types of right**) nevezi. Zavaró lehet azonban az a tény, hogy amikor Kanger később a hohfeldi jogviszonyfogalmakból "összetett" jogtípusokat képez, akkor ezeket 'atomi jogtípusoknak' nevezi el. Ebből következően a kangeri elméleten belül a hohfeldi fogalmakat az atominál kisebbnek kell tételeznünk, ezért így is lenne érdemes megnevezni őket. Ha igazodni akarnánk Kanger névadási konvenciójához, akkor ezeket az egyszerű jogviszonyokat **elemi jogtípusoknak** (**elementary rights**) kellene inkább neveznünk.[7] A következőkben én így teszek.

Kanger logikai formalizmusa a hohfeldi jogviszonyfogalmak pontosabb megragadására alkalmas. Ennek egyik eredménye az lesz, hogy Kanger kétszer annyi fogalmat képes elkülöníteni, mint Hohfeld. Ehhez természetesen kétszer annyi terminusra van szükség. Ezt a bővítést a 'counter'-fogalmak felvételével valósítja meg Kanger. Kanger a hohfeldi fogalmak logikai formulákkal írja le, miközben arra az öt axiómából álló logikai sémára támaszkodik, amit korábban már bemutattam (☞344.o.). A SHALL és a DO fogalmak segítségével képezhető formulákkal akarja megfogni a hohfeldi jogviszonyfogalmakat. A passzív jog alá sorolható fogalmakat úgy lehet formulával leírni, hogy a jogviszony másik szereplőjének korrelatív kötelezettségfogalmával fejezzük ki (ezért minden ilyen viszonyt **partner kötelezettségnek** (COUNTER-PARTY OBLIGATIVE RIGHTS) vagy röviden **O-right**-nak is nevezhetünk – a deontikus kötelezés modalitása után). Kanger az alábbiakat javasolja:

claim(Ego,Alter,F(Ego,Alter))	SHALL DO$(Alter, F(Ego, Alter))$
counter-claim(Ego,Alter,F(Ego,Alter))	SHALL DO$(Alter, \neg F(Ego, Alter))$
immunity(Ego,Alter,F(Ego,Alter))	SHALL$\neg$DO$(Alter, \neg F(Ego, Alter))$
counter-immunity(Ego,Alter,F(Ego,Alter))	SHALL$\neg$DO$(Alter, F(Ego, Alter))$

Az aktív jog alá tartozó jogviszonyokat Kanger a deontikus megengedés modalitásával gondolja megragadni, ezért ezekre ráhúzhatjuk a **megengedő jogok** (PERMISSIVE RIGHTS) vagy röviden a **P-right** kifejezést is. Kanger ide a következő fogalmakat sorolja be:

power(Ego,Alter,F(Ego,Alter))	$\neg$SHALL$\neg$DO$(Ego, F(Ego, Alter))$
counter-power(Ego,Alter,F(Ego,Alter))	$\neg$SHALL$\neg$DO$(Ego, \neg F(Ego, Alter))$
freedom(Ego,Alter,F(Ego,Alter))	$\neg$SHALL DO$(Ego, \neg F(Ego, Alter))$
counter-freedom(Ego,Alter,F(Ego,Alter))	$\neg$SHALL DO$(Ego, F(Ego, Alter))$

Ez tehát a kangeri javaslat, ami a logikai formalizmust tekintve megfelelő, elfogadható, ám a formulákkal összekapcsolt fogalmak összerendelése önkényes. Kanger a második hohfeldi jogfogalom-csoportot (vagyis a felhatalmazottság köré szervezhető fogalmakat) is felhasználja, ám ezt úgy teszi, hogy ezek definiálásakor eltér Hohfeld eredeti intenciójától. A logikai formulák szintjén alkotott elmélet korrekt, tartalmas és teljesnek mondható, viszont a felhatalmazottság köré rendezhető fogalmakat beemelése annyiban elfogadhatatlan, mert Kanger és Hohfeld fogalomértelmezései e téren túl nagy mértékben térnek el egymástól. Ezzel Kanger komoly zavart okoz. Makinson szerint ezt már nem is annyira kölcsönzésnek, mint inkább lopásnak kell tartanunk („Kanger borrowed, or rather stole") [318, 408].

[7] Nevezhetnénk még szubatomi jogtípusoknak is ezeket.

Már abból a mozzanatból is látszik, hogy Kanger más interpretációt ad a Hohfeldtől átvett terminusoknak, hogy Kanger a **claim** és az **immunity** jogfogalmakat állítja párhuzamba, és ezzel Hohfelddel ellentétesen cselekszik, utóbbi ugyanis az igényjog (**claim**) "analóg párjaként" a felhatalmazottságot (**power**) említi:

> „...Hohfeld fogalmainak pontos meghatározása érdekében a két csoport fogalmai közül igazából csak egyet-egyet kell explikálnunk – az igényjog és a mentesség fogalmait." [275, 101]

Ezzel szemben Hohfeld a két csoport "vezérfogalmaként" az igényjogot és a felhatalmazottságot emeli ki. Később fogom elemezni, bemutatni, hogy Kanger miért "tér el" ezen a ponton Hohfeldtől, itt csak annyit vetítek előre, hogy Kanger nem veszi észre, hogy a második hohfeldi jogviszonycsoport tagjait másodrendű fogalmakként lehet megragadni, ezért nem ismeri fel azt sem, hogy az új fogalmaknak kettős öröklődése és így kettős tulajdonsága lesz. A magasabbrendű fogalmiság abban a mozzanatban érhető tetten, ahogy Hohfeld a második jogcsoportot meghatározza. Az első és második jogviszonycsoport abban tér el egymástól – Hohfeld intenciója és meghatározása szerint –, hogy a második esetben a jogviszonyok szűkebb cselekvési tartományra vonatkoznak. Hohfeld a **felhatalmazottság** fogalmával a jogviszonyok változtatásának lehetőségét, képességét akarta kifejezni, de a többi három új jogfogalom is más jogokra irányuló jogként értelmezendő. A **felhatalmazottság** és a **szabadság** egymáshoz való viszonyát tehát más szinten kell kezelni, és nem úgy, ahogy Kanger teszi, amikor a két fogalom közti különbséget a pozitív vagy negatív cselekvések (tevés vagy tartózkodás) hozzárendelésével írja le. Hétköznapi fogalmakra átfordítva Kanger formuláit azt kellene mondanunk, hogy a felhatalmazottság az, ami nem tilos és a szabadság az, ami nem kötelező, de ez egyáltalán nem tűnik intuitívnek. Újra hangsúlyozom: a logikai séma jó, csak az elemeihez fűzött interpretáció nem elfogadható.

A megoldás az lehet, hogy Kanger logikai sémáját megtartjuk, de a séma egyes elemeinek interpretálásakor nem alkalmazzuk a hohfeldi másodrendű fogalmakat. Ez azt jelenti, hogy a szabadság, igényjog, felhatalmazottág és mentesség helyett a szabadság és igényjog pozitív, illetve negatív altípusait használjuk a kangeri formulák értelmezéséhez. Ezzel a megoldással megmaradunk az elsőrendű logikai szintjén úgy, hogy a kangeri formalizmus erejét teljes mértékben ki tudjuk használni, így viszont a másodrendű hohfeldi fogalmak kikerülnek a formális modellből, ezért a későbbiekben még ki kell dolgozni az őket leíró formulákat.

POSITIVE-CLAIM(Ego,Alter,F(Ego,Alter))	SHALL DO$(Alter, F(Ego, Alter))$
COUNTER-POSITIVE-CLAIM(Ego,Alter,F(Ego,Alter))	SHALL DO$(Alter, \neg F(Ego, Alter))$
NEGATIVE-CLAIM(Ego,Alter,F(Ego,Alter))	SHALL$\neg$DO$(Alter, \neg F(Ego, Alter))$
COUNTER-NEGATIVE-CLAIM(Ego,Alter,F(Ego,Alter))	SHALL$\neg$DO$(Alter, F(Ego, Alter))$
NEGATIVE-FREEDOM(Ego,Alter,F(Ego,Alter))	$\neg$SHALL$\neg$DO$(Ego, F(Ego, Alter))$
COUNTER-NEGATIVE-FREEDOM(Ego,Alter,F(Ego,Alter))	$\neg$SHALL$\neg$DO$(Ego, \neg F(Ego, Alter))$
POSITIVE-FREEDOM(Ego,Alter,F(Ego,Alter))	$\neg$SHALL DO$(Ego, \neg F(Ego, Alter))$
COUNTER-POSITIVE-FREEDOM(Ego,Alter,F(Ego,Alter))	$\neg$SHALL DO$(Ego, F(Ego, Alter))$

Ez a kangeri séma annyiban féloldalas, hogy *Ego* nyolcféle jogviszonyát *Ego* és *Alter* négy-négyféle perszonális deontikus modalitással fejezi ki. Ez a megoldás a jogviszonyokat tekintve "teljesnek" mondható, de ha az ágensekre fókuszálunk (tehát Egora és Alterre), akkor mindkét esetben hiányos a fenti séma, hiszen mindkét ágens esetében csak

feleannyi perszonális deontikus modalitást tartalmaz, mint amennyi valójában létezik. Ezt Kanger is tudta, kezelte is, hiszen – ahogy azt a következő fejezetben bemutatom majd – összesen tizenhat elemi jogviszonnyal számolt. A **kötelezés** ($\underline{\mathbf{O}}$) deontikus operátor és a DO cselekvésoperátor segítségével mindegyiküket megadhatjuk:

$\underline{\mathbf{O}}(x\text{DO}F)$	$\underline{\mathbf{O}}(x\text{DO}\neg F)$	$\underline{\mathbf{O}}(y\text{DO}F)$	$\underline{\mathbf{O}}(y\text{DO}\neg F)$
$\neg\underline{\mathbf{O}}(x\text{DO}F)$	$\neg\underline{\mathbf{O}}(x\text{DO}\neg F)$	$\neg\underline{\mathbf{O}}(y\text{DO}F)$	$\neg\underline{\mathbf{O}}(y\text{DO}\neg F)$
$\underline{\mathbf{O}}\neg(x\text{DO}F)$	$\underline{\mathbf{O}}\neg(x\text{DO}\neg F)$	$\underline{\mathbf{O}}\neg(y\text{DO}F)$	$\underline{\mathbf{O}}\neg(y\text{DO}\neg F)$
$\neg\underline{\mathbf{O}}\neg(x\text{DO}F)$	$\neg\underline{\mathbf{O}}\neg(x\text{DO}\neg F)$	$\neg\underline{\mathbf{O}}\neg(y\text{DO}F)$	$\neg\underline{\mathbf{O}}\neg(y\text{DO}\neg F)$

Makinson a fenti formulák általánosításaként egy olyan nyitott formulajavaslattal élt, amely segítségévek mind a tizenhat változatot kifejezhetjük [318]. A formulát bemutatom eredeti formájában is, illetve abban az alakban is, amelyben az ágensszerepek a modellemben használatos nevekkel vannak jelölve. A formulában látható ± jel azt fejezi ki, hogy vagy pozitív vagy negatív (tehát állító vagy tagadó) értelemben használjuk-e azt az összetevőt, amely előtt a jel áll.

$$(\pm)\underline{\mathbf{O}}(\pm)\begin{pmatrix} X \\ Y \end{pmatrix}\text{DO}(\pm)F \quad \rightsquigarrow \quad (\pm)\underline{\mathbf{O}}(\pm)\begin{pmatrix} Ego \\ Alter \end{pmatrix}\text{DO}(\pm)F$$

Az a fenti javaslatom egyébként, amelyet a szabadság és igényjog pozitív és negatív altípusainak modellbe emelésére vonatkozóan teszek, megegyezik Jeremy Bentham elképzelésével is [308, 47].[8] A továbbiakban úgy mutatom be Kanger elméletének hátralevő részét, hogy az általa használt eredeti fogalmakat, terminusokat alkalmazom, de saját modellembe az itt jelölt fogalmakat fogom átemelni.

Kanger atomi jogtípusai

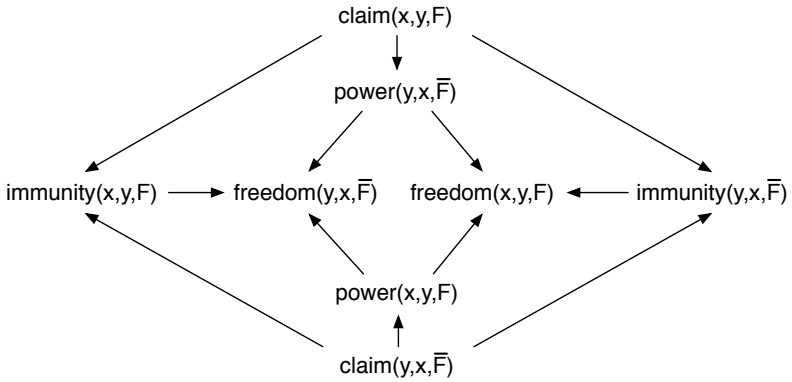
A jogok tipizálása már Jeremy Bentham számára is fontos volt [308, 63]. Bentham három jogtípust különített el, a **kötelezés**, a **tiltás** és a **nem tiltás és nem kötelezés** típusait. Ez a hármas felosztás megfelel a JEPD-elvnek, vagyis a három kategória terjedelme nem fedi át egymást, együttesen viszont kiadják a felettes fogalom (a jog) teljes terjedelmét. A hohfeldi jogviszonyfogalmak nem felelnek meg ennek az elvárásnak. Ezzel tisztában volt Kanger is, ezért a jogfogalmakra JEPD-kompatibilis felosztást dolgozott ki.

Kanger kiindulópontja az volt, hogy az általa bevezetett nyolc elemi jogtípus, illetve ezek komplementer párjai összetételével összetett jogi viszonyokat lehet képezni. A hohfeldi-kangeri elemi jogviszonyok összekapcsolásával atomi jogviszonyokat definálhatunk, mégpedig úgy, hogy a 26 egymástól különböző, tehát egymást át nem fedő atomi jogtípus összességében teljesen kitölti a jog fogalmának teljes terjedelmét. Nézzük meg, hogy jutott el Kanger a 26 atomi jogtípusáig. A nyolc elemi összetevőt (a nyolc jogosultságot és kötelezettséget), valamint ezek tagadásait logikailag 2^8 módon tehetjük össze, így 256-féle, elméletileg lehetséges elemi típusunk (possible atomic types) lehet. Ha ezeket mind feírnánk egy táblázatba, akkor 256 sort kapnánk, és valahogy így nézne ki a táblázat:

[8] Annyi különbséggel csak, hogy Bentham nem foglalkozott a Kanger által javasolt "counter"-fogalmakkal.

	1.	2.	3.	4.	5.	6.	7.	8.
1.	claim	freedom	power	immunity	counter-claim	counter-freedom	counter-power	counter-immunity
2.	claim	not-freedom	power	immunity	counter-claim	counter-freedom	counter-power	counter-immunity
3.	claim	freedom	not-power	immunity	counter-claim	counter-freedom	counter-power	counter-immunity
4.	claim	freedom	power	not-immunity	counter-claim	counter-freedom	counter-power	counter-immunity
5.	claim	freedom	power	immunity	not-counter-claim	counter-freedom	counter-power	counter-immunity
...								
255.	not-claim	not-freedom	not-power	not-immunity	not-counter-claim	not-counter-freedom	not-counter-power	counter-immunity
256.	not-claim	not-freedom	not-power	not-immunity	not-counter-claim	not-counter-freedom	not-counter-power	not-counter-immunity

Kanger szerint azonban meg lehet szűrni ezeket a sorokat, mert vannak köztük olyanok, amelyek vagy egymásnak ellentmondó elemeket tartalmaznak, így nem fordulhatnak elő, vagy egymással ekvivalensnek tekinthetők. Az ellentmondásos sorok kiesése már eleve csökkenti a 256-os számosságot, a további szűrés lehetőségét az teremti meg, hogy logikai összefüggések vannak egyes elemi jogtípusok között, és így némelyik sorban előfordulhat, hogy a nyolc összetevő közül néhányat el lehet hagyni a sorban szereplő elemi jogtípusok közötti logikai összefüggés miatt. Kanger egy ábrában mutatja meg, ezeket a logikai összefüggéseket.



Ha a fenti logikai kapcsolatok miatt mind a 256 sorból elhagyjuk a logikailag fölösleges összetevőket, akkor előfordulhat, hogy kettő vagy több sor meg fog egyezni. A 256 sor eredetileg mind különbözött egymástól, de a "ritkítás" után kevesebb egyedi sort kell kapjunk. Kanger szerint összesen 26 különböző sor marad a végén. Van olyan sor, amelyben kettő, de van olyan is, amelyben hat elemi jogtípus marad meg. Ezeket nevezi Kanger **atomi jogtípusoknak**. Az atomi jogtípusok és elemi összetevőik listája az alábbi táblázatban olvasható.

1)	Power, not immunity, counter-power, not counter-immunity
2)	Not power, immunity, not counter-power, counter-immunity
3)	Claim, not counter-freedom
4)	Not claim, power, immunity, counter-freedom, not counter-power, not counter-immunity
5)	Power, immunity, counter-power, counter-immunity
6)	Claim, power, counter-freedom
7)	Claim, not power
8)	Power, immunity, counter-freedom, not counter-power, counter-immunity
9)	Power, immunity, counter-power, not counter-immunity
10)	Power, not immunity, not counter-power, counter-immunity
11)	Not freedom, counter-claim
12)	Freedom, not power, not immunity, not counter-claim, counter-power, counter-immunity
13)	Freedom, counter-claim, counter-power
14)	Counter-claim, not counter-power
15)	Freedom, not power, immunity, counter-power, counter-immunity
16)	Power, not immunity, counter-power, counter-immunity
17)	Not power, immunity, counter-power, not counter-immunity
18)	Not power, not immunity, not counter-power, not counter-immunity
19)	Not claim, not counter-freedom, not counter-immunity
20)	Not counter-freedom, counter-immunity
21)	Not claim, not power, immunity, not counter-power, not counter-immunity
22)	Power, not immunity, not counter-power, not counter-immunity
23)	Not freedom, not immunity, not counter-claim
24)	Not freedom, immunity
25)	Not power, not immunity, not counter-claim, not counter-power, counter-immunity
26)	Not power, not immunity, counter-power, not counter-immunity

A lista meglepően hosszú. A konkrét atomi jogtípusok pedig nem igazán mondhatók intuitívnek. Nehezen értelmezhetőek a hétköznapi fogalmainkkal, tapasztalatainkkal, amit az is nehezít, hogy Kanger sajnos nem ad mindegyikre példát. Az atomi jogtípusok között azért léteznek összefüggések, és ezek felfedésével Kanger kicsit segít a 26 jogtípusból álló rendszer befogadásában.

Kanger ugyanúgy értelmezi saját atomi jogtípusai között azokat a relációkat, amelyeket Hohfeld tételezett a jogosultságok és kötelezettségek között. Saját formális elméletére támaszkodva azonban Kanger megállapítja (helyesen), hogy a hohfeldi korrelativitás relációja két altípusra osztható. Attól függően, hogy a korrelatív kapcsolatban levő jogviszonyok tartalma megegyezik-e vagy egymás tagadásai, Kanger **inverziónak**, illetve **konverziónak** nevezi a korrelatív relációt.[9] Az inverzió definícióját így adja meg Kanger:

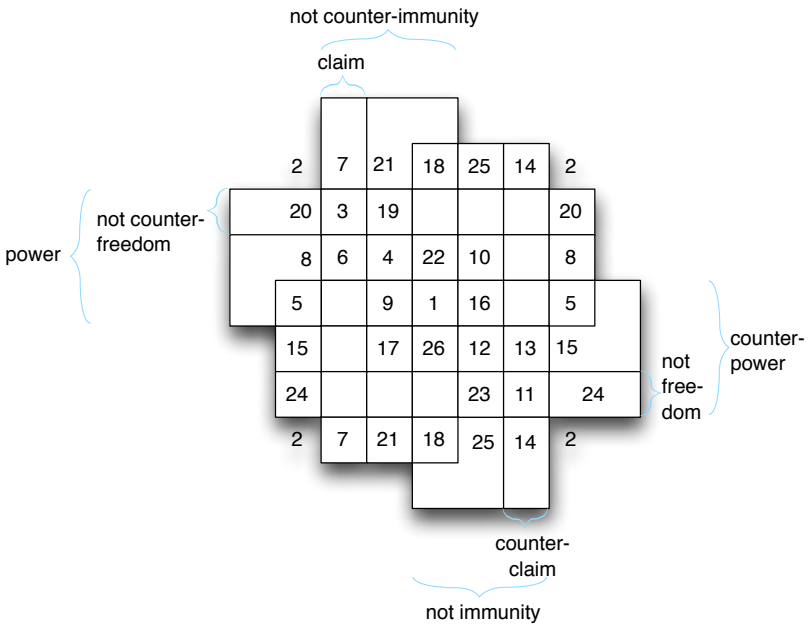
$$\text{X has versus Y a right of type P that S(X,Y)} \leftrightarrow \\ \text{Y has versus X a right of type Q that S(X,Y)}$$

míg a konverzió fogalmát az alábbi módon határozza meg:

$$\text{X has versus Y a right of type P that S(X,Y)} \leftrightarrow \\ \text{Y has versus X a right of type Q that not-S(X,Y)}$$

Kanger a fenti két "korrelativitási" reláció mellé felvesz még egy harmadik relációt, a **koordinációt**, de ezt itt nem tartom fontosnak bemutatni. Kanger a 26 atomi jogtípust az alábbi ábrába rendezte:

[9] Kanger eltérő módon definiálja az inverzió és konverzió műveletét ahhoz képest, ahogy a könyv logikai alapozó részében én bemutattam a két fogalmat (☞38.o.).



A szokatlan alakú táblázat arra jó, hogy jelezze az egyes kangeri atomi jogtípusok közti kapcsolatokat. A formalizmus alapján pontosan tudható, hogy mely típusok állnak rendre konverz vagy inverz kapcsolatban egymással.

6.2.3 Hohfeld másodrendű jogviszonyfogalmai

A hohfeldi elmélet bemutatásának talán nem "tett jót", hogy nem egyhuzamban mentem végig a hohfeldi kategóriákon, de a logikai építkezés szempontjából rákényszerültem erre. A Kanger által javasolt formalizmust én csak az elsőrendű hohfeldi fogalmakra tartom alkalmazhatónak. Ezért "mozogtam" eddig csak az elsőrendű logikai szinten. A további hohfeldi jogviszonyfogalmak tárgyalásához viszont fel kell majd lépnünk másodrendű fogalmak szintjére, és ez a tény megkívánja (és indokolja is egyben) a két szakaszra osztott tárgyalásmódot.

A felhatalmazottság köré szerveződő fogalmak értelmezése

Hohfeld a második – magasabbrendű – jogviszonycsoportba a **felhatalmazottság** (**power**), a **mentesség** (**immunity**), a **beavatkozásnak kitettség** (**liability**) és a **beavatkozásképtelenség** (**disability**) kategóriáit sorolta be. Ebben a fogalmi csoportban a felhatalmazottság az a "kulcsfogalom", ami köré a többi három fogalmat szervezve az első csoport esetében megismert struktúrához jutunk újra. A jogi felhatalmazottság a jogi beavatkozásra képtelenség ellentéte és a jogi beavatkozásnak kitettség korrelatív párja. Több részletet is idézek itt, hogy a példák segítségével jobban érthetővé váljék, mit gondol Hohfeld a felhatalmazottságról:

> „Egy meghatározott jogi viszonyban bekövetkező változás fakadhat (1) olyan új tényből, illetőleg ilyen tények együtteséből, amely kívül esik az emberi lény(ek) akaratlagos befolyásán; vagy (2) olyan új tényből, illetőleg tények együtteséből, amely egy vagy több emberi lény akaratlagos befolyása alá esik. Az esetek utóbbi csoportjában az a személy

(vagy azok a személyek), aki(k)nek akaratlagos befolyása meghatározó, (jogi) felhatalmazottsággal bír(nak) arra, hogy meghatározott változást idézzen(ek) elő a kérdéses jogviszonyban." [241, 84]

„*X* fel van hatalmazva arra, hogy hogy *Y*-ra ruházza érdekeltségét – azaz megszüntesse saját érdekeltségét, és ezzel együtt új és azonos érdekeltséget keletkeztessen *Y* javára. *X* tehát felhatalmazottsággal rendelkezik különböző fajta szerződéses kötelezettségek keletkeztetésére." [241, 85]

„A megbízási viszony keletkezése *inter alia* magában foglalja a jogi felhatalmazottság átruházását a megbízottra, s az ennek megfelelő beavatkozásnak kitettséget a megbízó oldalán. Másként fogalmazva: az egyik fél – *P* – felhatalmazottsággal rendelkezik arra, hogy megbízotti felhatalmazást hozzon létre a másik fél – *A* – oldalán, például arra, hogy átruházza *P* tulajdonát, úgynevezett szerződéses kötelezettséget keletkeztessen *P* terhére, feloldjon egy *P*-vel szemben fennálló tartozást, 'megszerezzen' egy tulajdonosi jogcímet *P* javára, és így tovább." [241, 85-86]

Bár Hohfeld a definícióban egyértelművé teszi, hogy a felhatalmazottság a jogi viszonyok változtatására irányul, igazából nem – vagy legalábbis nem eléggé – reflektál erre a mozzanatra. Pedig ez az új fogalmi négyes pontos értelmezését tekintve kulcsfontosságú. A felhatalmazottságot és a kapcsolódó másik három jogfogalmat ugyanis nem a logikai minőségét tekintve különíti el Hohfeld az igényjog köré épített fogalmaktól, hanem "csak" a fogalmak értelmezési tartományaiban tesz különbséget közöttük.

„Míg a joghelyzetek első két párja elsőrendű viszonyt, addig a második kettő másodrendűt képez. Az elsőrendű viszonyok közvetlenül az emberek viselkedésére és kapcsolataira vonatkozhatnak, a másodrendű viszonyok közvetítése nélkül; ezzel szemben a másodrendű viszonyok közvetlenül csak az emberek jogcímeire vonatkoznak, s csak áttételesen (bár meghatározó módon) az emberek viselkedésére és kapcsolataira." [292, 20]

Magasabbrendű normákról von Wright is ír a normák formális elemzésekor [564, 189-207], bár ő nem a hohfeldi fogalmak közti viszonyok esetében látja alkalmazhatónak a magasabbrendű logikát. Azt, hogy miként lehet a másodrendű fogalmakat ráhúzni a hohfeldi elméletre, rövidesen bemutatom, de előtte még meg kell ismernünk Hohfeld meghatározásait a felhatalmazottság köré építhető jogviszonyfogalmakra.

A jogi felhatalmazottságra ott lehet példákat keresni, ahol jogi viszonyok változtatásáról van szó. Ilyen egy miniszter kinevezése vagy felmentése, ilyen a házasság, az örökbefogadás, a végrendelkezés, az örökség visszautasítása, az elővételi jog gyakorlása, a megbízási érdekek képviselete, az ajánlati kötöttség, a telki szolgalom biztosítása, de minden tulajdonviszony jogszerű változtatását eredményező aktus is jogi felhatalmazottságon alapul. Lehet felhatalmazottságból eredő jogviszonyváltozás az, amikor valaki elcseréli a lakását két másikra vagy eladja a boltját, de a felhatalmazottság megnyilvánulhat abban is, ha a boltos megváltoztatja a nyitvatartási idejét, a tulajdonos bezárja a vendéglőjét vagy valaki lemond valamely tulajdonáról.

A felhatalmazottság valamiféle jogosultságot jelent, amivel ha él valaki, akkor a felhatalmazottság tartalma, valamely konkrét cselekvés megvalósul, és ezt a többiek tudomásul veszik, "jogosnak" tartják. Ez természetesen nem jelenti azt, hogy ne lehetne valamely cselekvést megtenni úgy, hogy az embernek nincs rá felhatalmazottsága. Ékes példa erre az alábbi jól ismert történet:

„Hamrák egy újságban olvasta, hogy Székesfehérvár vezetői megelégelték, hogy a piactéren vagy száll a por, vagy öles pocsolyákban tobzódnak a kofák, és elhatározták, hogy faburkolattal látják el a teret. Emberünk szépen levonatozott Fehérvárra, ahol egy

> kávéházban elhitette néhány fiatal városatyával, hogy a fa ára történelmi emelkedés előtt áll, ám neki van éppen tizenöt vagonnyi használt fatéglája, amit hajlandó 9000 aranykoronáért eladni. A hálás hivatalnokok rögtön a kezébe is nyomtak ezer aranykoronát előlegként, hősünk pedig már másnap munkásokat toborzott Pesten a Rottenbiller utca faburkolatának felszedésére.
> A módszer egyszerű, de nagyszerű volt; innen-onnan összeszedett néhány úttorlaszt, amikkel a Baross-térnél és a Lövölde-térnél lezárta az utcát, aztán a munkásokat rászabadította a faburkolatra. (Ekkor még sok helyen a macskaköves megoldásnál sokkal olcsóbb fakockás útburkolatot használták.) Mindenkinek azt mondta, hogy a városvezetésnek dolgozik, mely a faburkolat bazaltkőre cserélése mellett döntött. A munkások hat nap alatt szedték föl a majdnem egy kilométeres útszakasz burkolatát, amit a Kelenföldi pályaudvarnál vonatra raktak és elszállítottak Székesfehérvárra. A legszebb az egészben, hogy az egész művelet alig került 600 aranykoronába, így még az előlegből is maradt, a sikeres üzlet után pedig megkapta Székesfehérvártól a fennmaradó 8000 aranykoronát is." [4]

A történet értelmezésekor azt kell mondanunk, hogy Hamrák úrnak "nem volt felhatalmazottsága" az útburkolat felszedésére, értékesítésére, amit azonban úgy is mondhatnánk, hogy "nem volt megengedve neki" mindaz, amit megtett. Ez a helyettesíthetőség önmagában jelzi, hogy a felhatalmazottságot milyen deontikus logikai fogalomhoz lehet majd kötnünk, de ezt csak a későbbiekben fogom kifejteni.

A felhatalmazottság korrelatív párja a **beavatkozásnak kitettség** (LIABILITY) és ellentéte a **beavatkozásképtelenség** (DISABILITY). A beavatkozásnak kitettségre Hohfeld az alábbi példát adja:

> „A Booth v. Commonwealth esetben a bíróságnak Virginia egyik törvényi rendelkezését kellett értelmeznie, mely szerint «minden szabad fehér férfi, aki betöltötte huszonegyedik életévét és nem töltötte be a hatvanadikat, esküdtként köteles közreműködni, az alábbi kivételektől eltekintve». Világos, hogy e törvény csak beavatkozásnak kitettséget keletkeztetett, nem pedig kötelezettséget. Az beavatkozásnak kitettség, ha kötelezettséget róhatnak ránk. Ez utóbbi csak akkor következik be, amikor felhatalmazottságukkal élve a peres felek és a bíróság tagjai megtették mindazt, ami az esküdt funkciójának ellátására vonatkozó specifikus kötelezettség létrejöttéhez szükséges." [241, 91]

Az esküdtszéki munkában való részvétel mint a beavatkozásnak kitettség hohfeldi mintapéldájával párhuzamba állíthatjuk a katonai szolgálat feltételes kötelezettségét, ami számunkra talán ismerősebb lehet. Ha a beavatkozásnak kitettség a felhatalmazottság korrelatív párja, akkor minden olyan példa, amit a felhatalmazottság esetében adtam meg korábban, adekvát lehet a beavatkozásnak kitettségre is (természetesen az ágensszerepek megcserélésével). A beavatkozásnak kitettség alanyának nem feltétlen jelent hátrányt az a jogviszonyváltozás, amit a partnere okoz felhatalmazottságával élve. Ha valaki lemond a tulajdonáról (mondjuk, az órájáról), amire természetesen fel van hatalmazva, akkor ez a jogviszonyváltozás előnyös lehet mindenki más számára, hiszen ezzel az aktussal bárki szabadságot és felhatalmazottságot nyer arra, hogy eltulajdonítsa az órát [241, 92-93]. Ezt a jelenséget figyelhetjük meg a nagyvárosi lomtalanítások során.[10] Azzal, hogy a tulajdonosok kiteszik a szükségtelenné vált ingóságaikat (bútoraikat, használt ruháikat, háztartási eszközeiket, limlomjaikat) a közterületre, egyben jelzik azt is, hogy lemondanak e tulajdonaikról, és bárki viheti azokat. Ekkor potenciálisan mindenki beavatkozásnak

[10] A magyar hulladékgazdálkodási törvény 2013. január elsejével megváltozott, és ekkortól a lomtalanítás során közterületre kirakott hulladék a közszolgáltató tulajdonába kerül, de korábban nem így volt, és sok más országban is szabadon vihető az az ingóság, amiről lomtalanításkor lemondott a gazdája.

kitetté válik, ami azt jelenti, hogy mindenkinek tudomásul kell vennie, hogy a kirakott ingóságok fölötti jogosultság megváltozott, és onnantól kezdve bárki eltulajdoníthatja azokat (akár az eredeti tulajdonos is, bár ez nyilván nem valószínű opció). Az már egy másik kérdés, hogy ki lehet az új tulajdonos. Robert Sugden érdekes tanulmánya szerint az ilyen helyzetekben az elsőségnek lehet perdöntő szerepe [511].

Ha az eddig bemutatott két fogalmat, a felhatalmazottságot és a beavatkozásnak kitettséget összetevőikre bontjuk (ahogy tettük ezt az igényjoghoz kapcsolódó jogviszonyfogalmak esetében is), akkor a következő táblázathoz juthatunk:

jogosult	jogviszony	cselekvő	cselekvés
Ego	felhatalmazottsága	Ego	eladni Ego telkét
Alter	beavatkozásnak kitettsége	Ego	eladni Ego telkét

Ha felidézzük az elsőrendű hohfeldi jogfogalmak hasonló jellegű táblázatát (↪335.o.), akkor magyarázatra váró eltéréseket tapasztalhatunk. Ha ugyanis a felhatalmazottságot akár a szabadság, akár az igényjog fogalma alá akarjuk besorolni, mindenképpen eltérő mintázatot kapunk az első és másodrendű fogalmak összetevőikre bontott táblázataiban (a magyarázatot később adom meg). A **mentesség** (IMMUNITY) mint jogosultság a beavatkozásnak kitettség ellentéte és a beavatkozásképtelenség korrelatív párja. Mindenféle adómentesség, végrehajtás alóli mentesség, büntethetőséget kizáró ok fennállása, szabályok alóli kivétel e fogalom alá tartozik. Hohfeld példája a következő:

> „*X*, egy ingatlan tulajdonosa ... fel van hatalmazva, hogy elidegenítse az *Y* vagy bárki más számára. Ugyanakkor *X*-et számos mentesség is megilleti *Y*-nal vagy bárki mással szemben. *Y* ugyanis képtelen beavatkozni (azaz nincs felhatalmazottsága), amíg érdekeltséget nem szerez önmaga vagy bárki más számára. S ami igaz *Y*-ra, az hasonlóképpen igaz bárki másra is, aki sajátos keletkeztető tények folytán felhatalmazottságot nem szerez *Y* tulajdonának elidegenítésére." [241, 93]

A Hohfeld-recepció jóval kevesebbet foglalkozik a mentesség fogalmával a felhatalmazottsághoz képest. A fogalom tartalmát akkor lehet jól megérteni, ha arra fókuszálunk, hogy a mentesség a beavatkozásnak kitettség ellentéte. A diplomáciai mentesség lehet talán a legnyilvánvalóbb példája ennek. A diplomáciai rendszámmal rendelkező kocsikat nem lehet büntetni nyilvánvaló szabálytalanság esetén sem, vagyis nem lehet a használójára, tulajdonosára nézve kötelező jellegű jogi döntést hozni, nem lehet ő(ke)t jogi beavatkozásnak kitenni.

A **beavatkozás-képtelenség** (DISABILITY) a felhatalmazottság ellentéte (annak hiánya) és a mentesség korrelatív párja. Példaként lehet hozni rá a hatóságok hatáskörének vagy illetékességének hiányát valamely ügyben, a bérbeadó rendelkezési jogának szűkítését a bérbe adott dolog fölött vagy a rendőr beavatkozásképtelenségét a szabálytalankodó, ám diplomáciai rendszámmal rendelkező autóssal szemben, de idetartozik az is, hogy a tulajdonom fölött más nem rendelkezhet, nem adhatja el, nem használhatja, hacsak én nem adtam erre engedélyt. A mentesség és beavatkozásképtelenség fogalmainak összetevőkre bontása eredményeként az alábbi táblázatot kapjuk:

jogosult	jogviszony	cselekvő	cselekvés	
Ego	mentessége	Alter	nem	eladni Ego telkét
Alter	beavatkozásképtelensége	Alter		eladni Ego telkét

Ebben az esetben is ugyanazt tapasztalhatjuk a táblázat mintázatáról, mint a felhatalmazottság/beavatkozásképtelenség esetében: az elsőrendű fogalmak összetevőkre bontásához képest részben eltérő, részben egyező mintázat jellemzi a most elemzett fogalmakat. Jobban láthatóvá válik mindez, ha egyesítem a másodrendű fogalmak táblázatait.

jogosult	jogviszony	cselekvő	cselekvés	
Ego	felhatalmazottsága	Ego		eladni Ego telkét
Alter	beavatkozásnak kitettsége	Ego		eladni Ego telkét
Ego	mentessége	Alter	nem	eladni Ego telkét
Alter	beavatkozásképtelensége	Alter		eladni Ego telkét

illetve megmutatom újra, hogy nézett ki ugyanez az elsőrendű fogalmakra vonatkozóan (a sorokat némiképp átrendezve, hogy az összetartozó fogalompárok a táblázatok ugyanazon soraiban legyenek láthatóak).

jogosult	jogviszony	cselekvő	cselekvés	
Ego	szabadsága	Ego		átjárni Ego telkén
Alter	igényhiánya	Ego	nem	átjárni Ego telkén
Ego	igénye	Alter	nem	átjárni Ego telkén
Alter	kötelezettsége	Alter	nem	átjárni Ego telkén

A két táblázat összehasonlításából látszik, hogy az első-, illetve másodrendű fogalmak esetében más a korrelatív fogalmak viszonya egymáshoz. Elsőrendben az igény és kötelezettség között ugyanarra a cselekvésre irányult a jogviszony, míg a szabadság és igényhiány esetében a cselekvés és annak tagadására. A korrelativitás fogalmában megbúvó kettősséget kezelte le Stig Kanger az inverz és konverz relációk elkülönítésével (↬352.o.). Kanger szerint így elsőrendben az igény és kötelezettség közti korrelativitás inverziót jelent, míg a szabadság és igényhiány közti kapcsolat konverzió. A másodrendű fogalmak esetében a felhatalmazottság, illetve a beavatkozásnak kitettség között láthatjuk az inverz viszonyt, és a menteség, illetve beavatkozásképtelenség között a konverziót érvényesülni.

Mindez azért is problematikus, mert ha azt keressük, hogy a másodrendű fogalmakat melyik elsőrendű fogalmak alá sorolhatjuk be, akkor ellentmondásos helyzethez jutunk. Ha a felhatalmazottságot a szabadság alá rendeljük azon az alapon, hogy mindkettő **aktív jognak**[11] számít, akkor a korrelatív párjaikhoz való viszonyban lesz eltérés (a felhatalmazottság inverz, a szabadság konverz viszonyban van a párjával). A mentesség esetében is hasonlót tapasztalhatunk. A **passzív jogi** minősítésük okán a mentesség és az igényjog között kell a generikus kapcsolatot feltételeznünk, ám ha így teszünk, ez esetben is eltérő lesz a korrelativitás tartalma köztük. Meg kell említenem egy további problémát Hohfeld fogalommeghatározásaival kapcsolatban. Hohfeld ugyanis hangsúlyosan elkülöníti egymástól a kötelezettséget és a beavatkozásnak kitettséget, de nem tisztázza kellő alapossággal azt, mi a köztük lévő különbség lényege. Egy helyen azt írja, hogy:

> „A beavatkozásnak kitettség nem más, mint felelősség.” [241, 91]

de ez a mondat nem igazán segíti a tisztánlátást.[12] A felhatalmazottság és a beavatkozásnak kitettség magyarázata során kétféle értelemben is félrevezetőnek, sőt hibásnak tartom

[11] Az aktív és passzív jog meghatározását lásd 340. oldaltól.

[12] Az elemzett téma szempontjából nem tartom szükségesnek bővebben bemutatni, de jelezni mindenképpen érdemes, hogy a felelősség fogalmának elemzésekor H.L.A. Hart négy jelentésréteget is elkülönít

a hohfeldi megközelítést, amit sajnos a későbbi Hohfeld-recepció is folyamatosan életben tartott. Egyfelől ezen a ponton kerül be az nézet a hohfeldi elméletbe, miszerint a felhatalmazottság (illetve korrelatív párja a beavatkozásnak kitettség) nem jogok, hanem kapacitások, képességek létezéséről vagy azok hiányáról szól. Az angol terminusok közül a **disability** már önmagában is ezt a "sugallja". Ennek kritikájára a következő fejezetben fogok kitérni. Másfelől a felhatalmazottság és a beavatkozásnak kitettség kapacitív fogalmait Hohfeld (és utána még sok más szerző is) úgy értelmezi, mintha abban definitív szerepe lenne a feltételességnek. Megismétlem az előző Hohfeld-idézet egy részletét:

> „...beavatkozásnak kitettség, ha kötelezettséget róhatnak ránk. Ez utóbbi csak akkor következik be, amikor felhatalmazottságukkal élve a peres felek és a bíróság tagjai megtették mindazt, ami az esküdt funkciójának ellátására vonatkozó specifikus kötelezettség létrejöttéhez szükséges." [241, 91]

Az idézetből úgy tűnik, mintha Hohfeld szerint a beavatkozásnak kitettség feltételes kötelezettség lenne. Ez a feltételesség azonban nem köthető kizárólagosan a beavatkozásnak kitettséghez, előfordulhat az "egyszerű" kötelezettségek esetén is. Ha én – felhatalmazottságommal élve – kölcsönadom valakinek egy könyvemet, akkor folyamatos igényem (a partneremnek pedig folyamatos kötelezettsége) lesz arra, hogy, amikor kérem tőle, adja vissza. Ezt értelmezhetem úgy is, hogy a partnerem folyamatos beavatkozásnak kitettségben van és csak akkor keletkezik kötelezettsége arra, hogy visszaadja a könyvemet, amikor én – felhatalmazottságommal újra élve – ezt valahogy jelzem felé. Amíg azonban nem kérem vissza, addig csak feltételes, potenciális kötelezettséggel rendelkezik. Az itt jelzett problémákat csak akkor tudjuk megoldani, ha megadjuk a másodrendű fogalmak formális definícióit, és így láthatóvá tesszük egyfelől azt, hogy milyen módon kapcsolódnak, hasonlítanak az elsőrendű fogalmakhoz, illetve másfelől bemutatjuk azt, hogy milyen új vonások felvételével lehet őket saját minőségükben megragadni. A következő fejezetben ezt kísérlem meg bemutatni.

6.2.4 A másodrendű jogviszonyfogalmak formalizálása

A felhatalmazottság köré építhető hohfeldi fogalmak eddigi elemzéséből kiderült, hogy *(i)* ezek a fogalmak másodrendűek, *(ii)* ellentmondás látszik a passzív vagy aktív jog szerinti osztályozásuk, az igényjog vagy szabadság alá való besorolásuk és a korrelatív kapcsolataik eltérő minősége között, valamint *(iii)* a feltételesség, az állapot- vagy pontszerűség kérdésében eltérő minőséget lehet érezni a komplementer fogalompárok között.

A Hohfeld-recepcióban sokan úgy gondolják, hogy azon az alapon különíthetjük el az első- és másodrendű fogalmakat egymástól, hogy az első csoportot jogosultságnak, a másodikat képességnek (potenciálnak, kapacitásnak) minősítjük. Sok példát lehetne hozni arra, hogy miben különbözik a két hohfeldi jogviszonycsoport vagy másként a jogosultság és a felhatalmazottság egymástól. Makinson azt a példát említi, hogy lehet valakinek megengedett vagy tiltott meztelenül fürdeni a tengerparton, de ettől még nem rendelkezik azzal a felhatalmazottsággal, amely alapján ezen a jogosultságon változtatni tudna (megengedhetné vagy megtilthatná azt) [318, 409]. A különbségek megragadása végett egyes szerzők (például Makinson) az elsőrendű fogalmakat deontikus, a másodrendűeket kapacitív képességeknek nevezik.

---

egymástól úgy, hogy közben a hohfeldi értelmezést, a beavatkozásnak kitettséghez való viszonyát nem is igazán tisztázza ([229]).

Amikor Frederic B. Fitch egy 1967-es cikkében Hohfeld fogalmaira ajánl egy formalizmust [170], akkor, bár nem mondja ki, de a formalizmusból kiderül, hogy eltérő minőségűeknek tartja Hohfeld első és második jogcsoportjába tartozó fogalmakat. Az igényt, szabadságot, kötelességet és joghiányt deontikus operátorokkal, a felhatalmazottságot és a köré rendezett jogfogalmakat a *can* operátor segítségével definiálja.[13] Lars Lindahl is ugyanezt a különbséget tartotta fontosnak.

> „ ... a felhatalmazottságot inkább képességként, nem pedig engedélyként kell értelmezni, tehát inkább úgy, mint »képes«, nem pedig úgy, mint »szabad« (»can« rather than »may«).” [308]

Makinson figyelmeztet rá [318], hogy maga Hohfeld is erős analógiát tételezett a deontikus és kapacitív képességek egymáshoz való viszonyának tárgyalásakor, amikor így írt:

> „Ha például a közönséges tulajdonos elidegenítési felhatalmazottságára gondolunk, gondosan meg kell különböztetnünk a jogi felhatalmazottságot, a jogi felhatalmazottság »gyakorlásához« szükséges dolgok megtételének fizikai lehetőségét, s végül e dolgok megtételére irányuló szabadságot, már amennyiben ez a szabadság tényleg fennáll. Lehet, hogy fennáll, lehet, hogy nem. Így, ha X mint egy ingatlan tulajdonosa szerződést köt Y-nal, miszerint ingatlanát nem adja el Z-nek, akkor X-nek a Z javára történő értékesítési felhatalmazottság gyakorlásához szükséges magatartások gyakorlására vonatkozó szabadsága X és Y-on kívül mindenki más között fennáll. Világos azonban, hogy X és Y viszonyában az előbbi nem rendelkezik a kérdéses magatartás tanúsítására vonatkozó szabadsággal, vagy megfordítva: Y-nal szemben az a kötelezettség terheli, hogy ne tegye meg azt, ami felhatalmazottsága gyakorlásához szükséges.” [241, 90-91]

Makinson szerint ez a Hohfeld idézet is azt mutatja, hogy a felhatalmazottság és a szabadság nem mindig “jár együtt”. A tézis megerősítésére Makinson hoz egy másik példát. Amennyiben egy egyházi közösségen belül a papok csak akkor adhatnak össze két embert, akik közül az egyik nem az egyházközösség tagja, ha mindketten megígérik, hogy a gyereküket a vallási előírások szellemében nevelik fel, akkor “elválhat” egymástól a szabadság és felhatalmazottság kérdése. Ha ugyanis egy pap úgy végzi el a házassági szertartást, hogy nem kapja meg a két házasulandó fél ígéretét, attól még a szabályosan elvégzett szertartás eredményeként a házasság érvényessé válik. Az pedig már egy másik kérdés, ha a pap utána büntetést kap a feletteseitől, mert nem megengedett dolgot művelt.[14] Makinson szerint két strukturális különbség is van a deontikus és kapacitív modalitások között [318]. Az egyik különbség az, hogy *(i)* más következményekkel jár az, ha valaki olyat tesz, amire nem volt jogosultsága, illetve nem volt felhatalmazottsága. A másik különbség pedig (amit Makinson *Triponodo elvnek* nevezett el) azzal kapcsolatos, hogy *(ii)* mi történik akkor, ha a megengedés, illetve a felhatalmazottság negatív cselekvésre vagyis tartózkodásra vonatkozik. Nézzük meg egy kicsit alaposabban ezt a két elvet.

(i) Ha valaki olyan dolgot hajt végre, ami nem megengedett, akkor ez azt jelenti, hogy a cselekvés – a tiltás ellenére – megvalósult (tehát a rá vonatkozó állítás igaz). A nem megengedett cselekvés végrehajtásának következménye valamilyen szankció lesz. Ellenben ha valaki olyan dolgot tesz, amire nem volt felhatalmazottsága, akkor a cselekvés – bizonyos értelemben – nem valósul meg, érvénytelennek minősül. Felhatalmazottság

[13] Az már egy másik kérdés, hogy Fitch ezzel a megoldással túláltalánosítja a felhatalmazottság fogalmát.

[14] Makinson példája szerintem nem a két fogalom ellentmondásáról szól, hanem inkább arról, hogy két norma, kétféle jogosultság ütközésbe került egymással, tehát ez az eset sokkal inkább a normaütközés, a normakollízió problemakörébe tartozik.

hiányában is elvégezheti valaki pontosan ugyanazt a cselekvést, mint a társa, akinek felhatalmazottsága van rá, de a felhatalmazottság hiányában cselekvése érvénytelen lesz. Az útlevélhamisító pont ugyanazt a cselekvést végzi el, mint az okirat hivatalos kiadója, mégsem lesz érvényes a cselekvése, hiszen nem volt rá felhatalmazottsága.

> „Amikor végrendelkezünk, akkor a tanúk számát illetően vagy 'eleget teszünk' a Végrendeletekről szóló 1837. évi törvény 9. szakaszának, vagy nem. Ha nem teszünk eleget, akkor okiratunk nem lesz jogokat és kötelességeket teremtő 'érvényes' végrendelet; azaz 'joghatás' vagy 'jogkövetkezmény' nélküli 'érvénytelen' végrendelet lesz. De annak ellenére, hogy érvénytelenné válik, a törvényi rendelkezés be nem tartása nem valamilyen kötelezettség 'megszegése', nem is 'vétség', és zavaró lenne erről ilyen fogalmakban gondolkoznunk." [228, 41]

Ha a cselekvő nem rendelkezik felhatalmazottsággal egy adott cselekvésre, akkor nem is végezheti el azt. Formálisan kifejezve ezt (feltételezve, hogy a cselekvések tartománya a jogi viszonyok változtatására van leszűkítve):

$$\neg\textbf{power}(Agent, \textsc{do}(Agent, F)) \rightarrow \neg\textsc{do}(Agent, F)$$

Ennek a tézisnek igaz kell legyen a **kontraponáltja** is:

$$\textsc{do}(Agent, F) \rightarrow \textbf{power}(Agent, \textsc{do}(Agent, F))$$

A fenti állítás úgy interpretálható szabad szöveges formában, hogy ha egy ágens végrehajtott egy cselekvést, akkor hatalma is volt végrehajtania azt. A kérdés itt az, hogy mi az az értelem, ami szerint érvénytelennek kell minősíteni egy cselekvést még akkor is, ha sok szempontból teljesen megegyezik azzal, ami érvényes.

(ii) A deontikus és kapacitív modalitások közötti másik fontos strukturális eltérés alapja a tartózkodással szemben "mutatott különbözőség". Ha X-nek kötelessége törekedni F elérésére, akkor azt úgy is mondhatjuk, hogy számára nem megengedett nem törekedni F elérésére. Értelmes tehát arról beszélni, hogy X-nek megengedett megtenni valamit, és arról is, hogy X-nek megengedett nem megtenni valamit. Ezzel szemben azt mondhatjuk ugyan, hogy X-nek felhatalmazottsága van végrehajtani egy cselekedetet, de az értelmetlennek, jelentésnélkülinek tűnik, ha azt állítjuk, hogy X-nek felhatalmazottsága van nem megtenni ugyanazt. Makinson szerint ez az állítás nem is jelentésnélküli, mint inkább triviális, és ezt nevezi Triponodo elvnek.[15]

*

Megítélésem szerint a felhatalmazottságot nem képességként, kapacitásként kell értelmezni, és nem is a feltételesség mentén lehet igazán különbséget tenni a beavatkozásban kitettség és a kötelezettség között. Azok, akik a képeséget, a kapacitív modalitást állították szembe a deontikus modalitással, nem tudtak igazán számot adni arról, hogy milyen ontológiai szerkezetbe, hová lehetne beállítani és milyen formulákkal lehetne megragadni a képességet, kapacitív modalitást leíró operátorokat: a felhatalmazottság leírására Fitch ugyan felvette a *can* operátort [170], de nem definiálta, von Wright [564], Makinson

[15] Makinson – kis pontosítással – úgy tartja ellentmondásmentesnek ezt a tézist, hogy „. . . amikor az F jogi állapot nem egy logikai igazságot ír le, akkor mindig triviálisan igaz, hogy bárkinek felhatalmazottsága van nem megtenni F-t." [318, 413].

[318], Bulygin [87] pedig még csak kísérletet sem tett a formalizálásukra. A megoldás annak fel- és elismerésében van, hogy a **felhatalmazotság** (és a többi "társfogalom") egyfelől **másodrendű**, másfelől összetett, ún. **polihierarchikus fogalom**. Érdemes ezt az állítást alaposabban kibontani, és levonni a konzekvenciáit. A felhatalmazottság és a "társfogalmai" a jogviszonyok megváltoztatásával kapcsolatosak, míg az "igényjogi csoport" jogfogalmai bármilyen cselekvésre vonatkozhatnak. Ez nagyon fontos mozzanat, amit pontosan kell értelmeznünk. Ha az új fogalmak (a felhatalmazottság és "társai") csak annyiban térnek el a már bevezetett fogalmaktól (az igényjogcsoportba tartozóktól), hogy az egyik fogalmi összetevőjüket, a cselekvés fogalmát tekintve szűkebb terjedelműek, akkor ez az új fogalmak modellbeli helyét is meghatározza. Az új jogviszonyfogalmak nem az eddig megismert fogalmak szintjén, azokkal mellérendelt viszonyban, hanem alárendelt módon jelennek meg, tehát mind a négy új kategória (a felhatalmazottság, a beavatkozásnak kitettség, a mentesség és a beavatkozásképtelenség) valamelyik igényjogi csoportba tartozó kategória alá rendelt fogalomként (altípusként) értelmezhető. Ezt azt is jelenti egyben, hogy amikor a felhatalmazottságot és a mentességet mint a jogosultságok új típusait akarjuk majd formálisan "definiálni", akkor ezeket a már "meglevő" jogosultságfogalmak, tehát az igényjog és a szabadság fogalmai alá sorolhatjuk csak be.[16] Ha a másodrendű jogosultságfogalmakat az elsőrendű jogosultságfogalmak alá kell rendelnünk, akkor adódik a kérdés, mi hová tartozik. A válasz közel sem egyértelmű. Azt már említettem, hogy ha az aktív/passzív jogi szempontra figyelnünk, akkor a felhatalmazottságot aktív jognak minősítve a szabadság alárendeltjeként kell értelmeznünk, a mentességet mint passzív jogot az igényjog alá kell besorolnunk. Kanger tipizálási javaslata mögött is ez a megfontolás állhatott (↰348.o.). Maga Hohfeld azonban ezt a kérdést másként válaszolta meg:

> „. . . a felhatalmazottság általában ugyanúgy viszonyul a mentességhez, mint a jog [igény] a szabadsághoz. A jog [igény] mással szembeni pozitív igény, a szabadság pedig valaki más jogával vagy igényével szembeni szabadság. Hasonlóképpen: a felhatalmazottság más jogi viszonyának 'alakítására' irányuló lehetőség; míg a mentesség másvalakinek egy jogviszony 'alakítására' szóló jogi felhatalmazottságától való szabadság." [241, 93]

Az előzőekhez képest pont fordított kapcsolatot tételez Hohfeld, amikor az igényjogot és a felhatalmazottságot, illetve a szabadságot és a mentességet köti össze egymással. Igaz, Hohfeld nem beszél alá-fölérendeltségi relációról, és nem is igazán indokolja meg javaslatát. Mindkét másodrendű kategóriát valamiféle megengedésnek (egyszer lehetőségnek, másszor szabadságnak) tartja. Hart is tartja magát a Hohfeld által javasolt struktúrához (vagyis az igény a felhatalmazottsággal, a szabadságot a mentességel köti össze), de ő megpróbál valamiféle indoklást is adni az összerendelések magyarázatául:

> „A következő lehet az egyesítő mozzanat: a jogrendszer mind a négy esetben kifejezetten elismeri az individuum választását vagy negatív értelemben azáltal, hogy nem akadályozza, nem gátolja meg ebben (szabadság és mentesség), vagy megerősítő értelemben azáltal, hogy joghatást rendel hozzá (igényjog és felhatalmazottság). A negatív esetekben nincs olyan törvény, amely akadályozná az egyént abban a választásában, hogy megtegyen valamit vagy tartózkodjon valamely cselekvéstől (szabadság), illetve mindent változatlanul tartson fent saját jogi helyzetét illetően (mentesség); a megerősítő esetekben a jogrendszer joghatást rendel az egyén döntéséhez, amely alapján valaki másnak meg kell tennie

[16] Természetesen ugyanígy kell eljárnunk a kötelezettségfogalmak alá besorolandó beavatkozásnak kitettség és beavatkozásképtelenség esetében is.

valamit vagy tartózkodnia kell valamilyen cselekvéstől, illetve valaki más jogi helyzete megváltozik." [227, 35]

Hart egy lépéssel többet tesz meg, mint Hohfeld, amikor rögzíti, hogy az igényjog éa a felhatalmazottság esetében „ a jogrendszer joghatást rendel az egyén döntséhez", és ez az a mozzanat, ami összeköti e két jogviszonytípust. Az igényjog esetében egyértelmű, hogy ez a hozzárendelt joghatás a másik fél kötelezettsége valamilyen cselekvésre (valaminek a megtételére vagy valamitől való tartózkodásra),) viszont a felhatalmazottság esetében Hart nem mondja meg pontosan, hogy mi is a hozzárendelt joghatás tartalma. Legalábbis abból, hogy Hart csak annyit mond, hogy ilyen esetekben „valaki jogi helyzete megváltozik", nem lehet egyértelműen kikövetkeztetni, milyen joghatásról is van szó. Ahhoz, hogy az egymással "vitatkozó" osztályozási, alá-fölérendelési javaslatok közti ellentmondást feloldjuk, illetve a harti joghatás valós természetét feltárhassuk, szükségünk van a fogalmak formális leírására.

Eddig arról beszéltem, hogy a másodrendű fogalmakat az elsőrendű fogalmak alá kell besorolnunk. Ezt mondhatjuk másként is: Hohfeld első és második jogviszonycsoportjába tartozó fogalmak között **generikus alárendeltségi reláció** (GENERIC SUBORDINATE RELATION) létezését kell feltételeznünk.[17] A generikus alárendeltségi viszony lehet **polihierarchikus** (POLYHIERARCHICAL), illetve **monohierarchikus** (MONOHIERARCHICAL) jellegű [525]. Előbbi azt jelenti, hogy a leszármaztatott fogalomnak lehet több fölérendeltje is, utóbbi esetben csak egy fölöttes fogalom lehetséges. Amellett fogok érvelni a következőkben, hogy Hohfeld felhatalmazottság fogalma polihierarchikus módon rendelhető mindkét elsőrendű jogviszonyfogalom, a szabadság és az igényjog kategóriája alá. Abból kell kiindulnunk, hogy a felhatalmazottság (és a többi másodrendű fogalom) jogviszonyfogalom. Egy jogviszonyt a következő nyitott formulával írhatunk le:

$$\text{RIGHT}(Agent, \pm\text{DO}(Agent, T_1))$$

Amikor – Hohfeld nyomán – a felhatalmazottságot úgy értelmezzük, hogy az jogviszonyok megváltoztatására irányuló jogviszony, akkor ezt úgy ragadhatjuk meg, hogy valamely elsőrendű jogviszony (tehát kétféle szabadság vagy kétféle igényjog) megváltoztatásának jogát jelenti.

$$\text{RIGHT}(\text{Agent}, \pm\text{DO}(\text{Agent}, \text{RIGHT}(Agent, \pm\text{DO}(Agent, T_1))))$$

A fenti képletbe ekkor a RIGHT helyett POWER-t írhatunk, az *Agent*-változó helyébe *Ego*-t, a T_1 helyébe pedig valamely korábbi jogviszony megváltoztatását mint cselekvést helyettesíthetjük:

$$Change(\text{RIGHT}(Agent, \text{DO}(Agent, T_0))).$$

A felhatalmazottság definíciójában már nem kell, hogy szerepeljen a ± jel a DO cselekvésoperátor előtt. Azért nincs rá szükség, mert a felhatalmazottság esetén nincs értelme negatív cselekvésről vagyis a felhatalmazottságtól való tartózkodásról beszélni. Az új paraméterekkel a fenti nyitott és általánosított formula erre a speciálisabb alakra hozható:

$$\text{POWER}(Ego, \text{DO}(Ego, Change(\text{RIGHT}(Agent, \text{DO}(Agent, T_0)))))$$

[17] Ezt Hohfeld nem állítja, de a meghatározásaiból mégiscsak ez következik.

Ez a képlet szemléletesen mutatja a fogalom másodrendű jellegét, hiszen egy jogviszonyba beágyazva egy másik jogviszony szerepel paraméterként. Ezen a ponton viszont fel kell tennünk a kérdést, vajon milyen módon lehetne kifejezni a **power** fogalmát elsőrendű jogviszonyfogalmak segítségével, vajon a szabadság (FREEDOM) vagy az igényjog (CLAIM) alá tartozik. Az első – kézenfekvőnek tűnő – válasz az kell legyen, hogy a **power** a FREEDOM alárendeltje, hiszen mindkettő **aktív jog**, és a felhatalmazottság gyakorlásához valamilyen formában szükséges feltételnek tűnik, hogy a jogbirtokos rendelkezzen a jogviszony változtatásához szükséges szabadsággal. Vélhetőleg ez a szerkezeti hasonlóság vezethette Kangert is arra, hogy a szabadság mellé a felhatalmazottságot emelje be a deontikus logikai sémájába. Bár von Wright elsősorban kompetencianormának nevezte, de ő is permisszív normának minősítette a hatalom/felhatalmazottság fogalmát:

> „A magasabbrendű *kompetencianormákat* permisszív normáknak nevezem. Azt mondhatjuk, hogy egy kompetencianorma vagyis egy magasabbrendű permisszív norma kibocsátásának aktusában a magasabbrendű felettes hatóság *delegálja a felhatalmazottságát* egy alacsonyabbrendű *alárendelt hatóság* számára. A 'felhatalmazottság' itt nem más, mint 'normával teremtett kompetencia normák kibocsátójaként cselekedni'. Úgy is beszélhetünk erről, mint *normatív* kompetencia vagy felhatalmazottság." [564, 192]

Észre kell azonban azt is vennünk, hogy van egy másik szerkezeti hasonlóság is a hohfeldi fogalmak között. A szabadság korrelatív párja az igényhiány, aminek a hohfeldi másodrendű fogalmak között a beavatkozásképtelenség felelne meg – miközben Hohfeld a felhatalmazottság korrelatívumára, a beavatkozásnak kitettségre állandóan úgy hivatkozik, mint ami szerkezetileg, deontikus modalitását tekintve a kötelezettségnek felelne meg. A feltételes kötelezettség hohfeldi elképzelése is ezt sugallja. Ha a felhatalmazottság mint speciális szabadság aktív jog, akkor annak korrelatív párjának passzív jognak kéne lenni, viszont a beavatkozásnak kitettség szintén aktív jognak tűnik. Ellentmondásos helyzetnek tűnik, de nem az.

A látszólagos ellentmondás feloldódik, ha észrevesszük, hogy a felhatalmazottság fogalmában valójában két elsőrendű jogviszonyt érhetünk tetten. A felhatalmazottság "alapja" az a szabadság, amely "lehetővé teszi", hogy a jogbirtokos megváltoztathasson egy jogviszonyt. De ezzel még nincs vége. A jogbirtokos ezzel párhuzamosan egy igényjoggal is fellép a másik féllel szemben arra vonatkozóan, hogy a partnere ismerje el a jogviszony-változtatás tényét (érvényességét). A beavatkozásnak kitettség az ebben az értelemben vett elismerési igénnyel szembeállítható elismerésére irányuló kötelezettséget jelenti. A felhatalmazottság fogalma tehát két összetevőre bontható fel, amit így fejezhetünk ki formális módon:

(DL.22)
$$\begin{aligned}&\textsc{power}(Ego, \textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \stackrel{\text{def}}{=}\\ &\textsc{freedom}(Ego, \textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \wedge\\ &\textsc{claim}(Ego, \textsc{do}(Alter, \textsc{recogn}(\textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0)))))))\end{aligned}$$

felhatalmazottság (HUN) – power (ENG)

Ez az összetettség azt jelenti, hogy a felhatalmazottság polihierarchikus fogalom, vagyis nem egy, hanem több **generikus fölérendelt** (GENERIC SUPERORDINATE) fogalom alá lehet besorolni.[18] Ha így értelmezzük a felhatalmazottság fogalmát, akkor már magyarázhatóak a korábban jelzett "anomáliák", érthetővé válik a beavatkozásnak kitettség és a kötelezettség

[18] A generikus fölérendeltje relációt szokták még GENERIC SUPERIOR relációnak is nevezni.

oly közeli viszonya is. Ha a felhatalmazottságban van egy igényjogi összetevő, akkor annak korrelatívuma egy kötelezettség lesz, ami a beavatkozásnak kitettség fogalmába "kerül be", és ez a mozzanat alapozza meg a kötelezettség és a beavatkozásnak kitettség fogalmai között érzett hasonlóságot. A felhatalmazottság meghatározásából már könnyen levezethetjük annak inverz fogalmát, a beavatkozásnak kitettséget. A két fogalom közt fennálló inverzitás tényét kihasználva, a beavatkozásnak kitettséget is polihierarchikus fogalomként kell értelmeznünk, csak épp a felhatalmazottsághoz rendelt szabadság és elismerési igényjogi összetevők inverzeit kell a beavatkozásnak kitettség két összetevőjeként definiálni.

(DL.23) $$\begin{aligned}&\textsc{liability}(Alter, \textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \stackrel{\text{def}}{=}\\ &\textsc{no-claim}(Alter, \neg\textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \wedge\\ &\textsc{duty}(Alter, \textsc{do}(Alter, \textsc{recogn}(\textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0)))))))\end{aligned}$$

beavatkozásnak kitettség (HUN) – liability (ENG)

Kicsit meglepő, de tény, hogy a szakirodalomban a felhatalmazottsághoz képest alig foglalkoztak a mentesség fogalmával. Amennyire "egyenrangúan" kezelték, azaz ugyanakkora figyelemmel kísérték az igényjog és a szabadság kategóriáit, annyira figyelmen kívül hagyták a mentesség fogalmát. Az eddigiek alapján már megadható mind a mentesség, mind a korrelatív párja, a beavatkozásképtelenség formális definíciója, nem érzem szükségét a hosszas magyarázatoknak.

(DL.24) $$\begin{aligned}&\textsc{immunity}(Ego, \neg\textsc{do}(Alter, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \stackrel{\text{def}}{=}\\ &\textsc{claim}(Ego, \neg\textsc{do}(Alter, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \wedge\\ &\textsc{claim}(Ego, \textsc{do}(Alter, \textsc{recogn}(\textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0)))))))\end{aligned}$$

mentesség (HUN) – immunity (ENG)

(DL.25) $$\begin{aligned}&\textsc{disability}(Ego, \neg\textsc{do}(Alter, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \stackrel{\text{def}}{=}\\ &\textsc{claim}(Ego, \neg\textsc{do}(Alter, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0))))) \wedge\\ &\textsc{claim}(Ego, \textsc{do}(Alter, \textsc{recogn}(\textsc{do}(Ego, \textsc{change}(\textsc{right}(Agent, \textsc{do}(Agent, T_0)))))))\end{aligned}$$

beavatkozásképtelenség (HUN) – disability (ENG)

6.2.5 A jogviszonyfogalmak és szabályfogalmak viszonya

Az eddigiekben jogviszonytípusokról, azok értelmezési lehetőségeiről volt szó, és a deontikus operátorokat használtam a fogalmak formális leírásához. A másodrendű jogviszonyfogalmakkal egy olyan pontra értünk, amikor megpróbálhatjuk hasznosítani azokat az elméleteket, fogalmakat, amelyeket az általában vett szabályok tipizálása során alkottak meg, de ehhez az is szükséges, hogy innentől kezdve az általánosabb **szabály** fogalomról beszéljünk egy darabig az eddigi jogviszonyfogalmak helyett. A jelen kontextusban a szabály fogalmát a norma fogalmával szinonim módon fogom használni, noha a végső modellben a szabály főfogalma alá fogom majd berendelni a norma alfogalmát. Azt azonban fontos rögzíteni, hogy a szabály fogalmát az egész könyvben az emberi cselekvésekre vonatkozó szabályként használom, ebben az értelemben ez a fogalomhasználat szűkebb jelentésű a teljesen általános szabályfogalomhoz képest (a **szabály** terminus jelentése tehát ebben a könyvben megegyezik a **cselekvési** vagy **viselkedési szabály** fogalmának jelentésével). A cselekvési (viselkedési) szabály fogalmát már sokat használtam, több helyen

is elemeztem a könyvemben (↬304.o.)., illetve (↬315.o.), így itt csak egy idézettel szemléltetem, miért is fontos a szabály mind a társadalmi cselekvések végrehajtásában, mind azok magyarázatában.

> „Két jel utal arra, hogy a viselkedés szabályvezérelt, nem pedig csupán rendszeresen ismétlődő: a mintától való eltérést valamiképpen hibának vagy tökéletlenségnek tartjuk, valamint a szabályt automatikusan alkalamazzuk új esetekre is. Ha a cselekvő szembekerül egy új esettel, amellyel korábban nem találkozott, tudni fogja, mit tegyen." [470, 55]

Bár ez a szabályfogalom megengedi, hogy mind a normához, mind a szokáshoz igazodó cselekvésre ráhúzzuk, a továbbiakban csak a normativitás körébe tartozó szabályfogalmakkal foglalkozom. Az egyik fontos megválaszolandó kérdés az az, hogy milyen típusai vannak a viselkedési szabály fogalmának. Eugenio Bulygin miután tisztázza, hogy a szabály (rule) és a norma fogalma között nem tesz különbséget, valamit egymás szinonímájaként használja a kompetencia, kapacitás, potencia, hatalom, uralom, képesség, befolyás teminusokat, két szabályfogalmat különböztet meg egymástól: a **viselkedési normák** (norms of conducts) és a **kompetencianormák** (competence norms) kategóriáit [87].

Von Wright háromféle normát különít el: *(i)* a preskripciókat vagy regulációkat (prescriptions, regulations), *(ii)* a determinatív szabályokat (determinative rules) és *(iii)* a technikai normákat vagy direktívákat (technical norms, directives) [564]. A von wright-i hármas felosztás első két eleme, a preskripciók és a determinatív szabályok megfeleltethetők a harti kettős szabályfelfogásnak, a harmadik von wright-i szabályfogalom, a technikai normák fogalma viszont eltér minden eddigi tipizálási kísérlettől. Sokan mások, én magam sem tartom igazán a norma fogalma alá sorolhatónak ezt a fogalmat, mert pont a normativitáshoz szükséges legfontosabb elem hiányzik belőle.

Bulygin azt javasolja, hogy von Wright determinatív szabály kifejezése helyett inkább a fogalmi szabályok kifejezést használjuk, és tekintsünk el a harmadik von wright-i normatípustól. ezek után pedig kijelenti, hogy szerinte a kompetencianormákat nem lehet visszavezetni sem a kötelesség sem a megengedés fogalmára. A viselkedési norma megsértése esetén szankció várható (következik), ezzel szemben a kompetencianorma megsértésekor érvénytelen lesz a felhatalmazottságot közvetítő, átruházó, deklaráló cselekvés [87, 208]. Ha nem az előírt feltételekhez igazodva születik meg valami, akkor **érvénytelenség**, semmisség (NULLITY) lesz a következménye ugyanúgy, mint bármely játékban (sakk, sport) ez történik, ha megsértik a játékszabályokat. Így ír Bulygin:

> „ . . . Hart lényeges pontra tapint rá. Szankció és érvénytelenség [nullity] két radikálisan különböző fogalom és bármely kísérlet, amely beolvasztaná az érvénytelenséget a szankció fogalmába, kudarcra van ítélve. Ráadásul a szankcióra és érvénytelenségre jellemző eltérő működésmódok azt is mutatják, hogy lényegi különbség van a két szabálytípus között is. Első pillantásra a konstitutív szabály az, ami definál egy fogalmat vagy egy játékot mint tevékenységet, és ezáltal értelmessé teszi az érvénytelenségről való beszédet. Ha azonban létezik egy definíció, amely szerint ahhoz, hogy egy norma, egy cselekvés vagy egy dokumentum érvényessé váljék, bizonyos feltételeket ki kell elégíteni, akkor bármely feltétel hiánya az adott norma, cselekvés, dokumentum érvénytelenségét idézi elő. Az érvénytelenség nélkül nem létezik maga a definíció sem, így a konstitutív szabály és az érvénytelenség nem választhatóak el egymástól. E kétfajta intézmény, a büntetés és az érvénytelenség együttes létezése egyben jelzi a jogrendszer (legalább) kétféle szabályának létezését: vannak viselkedési normák és konceptuális szabályok, vagy, mindezt searle-i terminusokban kifejezve, vannak regulatív és konstitutív szabályok." [87, 214]

A **regulatív** és **konstitutív szabályok** fogalmainak bevezetését John Searle nevéhez kötik, ami a terminológiai szinten természetesen helyénvaló, de Searle-t jócskán megelőzte John Rawls, aki már 1955-ben szembeállította a **summázati** és **gyakorlati szabályokat** egymással úgy, hogy Rawls fogalomkettőse fogalmilag teljesen megfeleltethető Searle párosának [433]. Searle regulatív szabályai (és Rawls summázati szabályai) olyan cselekvésekre vonatkoznak, amelyek akkor is "léteznének", ha nem irányulna rájuk szabályozás. Ezzel szemben a konstitutív szabályok (Rawlsnál a gyakorlati szabályok) maguk "teremtik", maguk teszik lehetővé a cselekvéseket, az ilyen szabályok híján az ilyen cselekvések nem léteznének, értelmetlenekké, értelmezhetetlenekké válnának. A regulatív szabályok alá sorolható cselekvésekre deontikus operátorok irányulhatnak, és ha a szabályok által előírt cselekvések nem valósulnak meg, tehát "szabálysértés" történik, akkor ezt szankció követi. Ez utóbbi összefüggést lehet kifejezni az alábbi két formulával:

$$\textsc{Obligatory}(A) \wedge \neg A \rightarrow \textsc{s}$$
$$\textsc{Forbidden}(A) \wedge A \rightarrow \textsc{s}$$

A konstitutív szabályok által teremtett cselekvéseket máshogy lehet jellemezni. A konstitutivitás mozzanatát John Searle a **counts as** (valaminek számít) operátor segítségével véli megragadhatónak, és ezért azt javasolja, hogy az alábbi módon értelmezzük a konstitutív szabályt [470, 47]:

X counts as Y in C.

Természetes nyelvi mondatban kifejezve:

'valamely X dolog valamely C kontextuson belül Y dolognak számít'.

Y az az új dolog, entitás, amelyet a konstitutív szabály alkalmazása révén konstituálunk, X pedig az a dolog, entitás, amely a szabály alkalmazása nélkül (előtt) is létezik a világban. Utóbbira – Searle terminológiáját használva – a **nyers tények** (**brute facts**), előbbire az **intézményes tények** (**institutional facts**) vonatkoznak [470, 63]. Sokan elfogadták, helyeslően átvették Searle formuláját, és persze voltak, akik kritizálták őt ezért.[19] Joseph Ransdell egy 1971-es tanulmányában azzal "támadta" Searle-t, hogy az intézményes tényeket "hordozó" entitást kifejező Y terminusban összemos két fontos mozzanatot, amit pedig inkább szét kellene választani és érdemes lenne külön reprezentálni őket [429]. Ransdell kritikája annyiban jogos, hogy amikor Searle a fenti formuláról beszél, akkor ahhoz képest szövegszerűen maga is további összetevőkre utal, ám ezeket nem emeli be a formalizmusába. Searle szerint azzal, hogy X-re "mutatva" konstituálunk egy intézményes tényt, miközben "létrehozzuk" Y-t, ez utóbbi terminussal egyben specifikálunk egy új státust is, amihez hozzárendelünk egy vagy több funkciót [468, 46-47]. Ezek a funkciók adják a konstitúció értelmét. Ezek a státusfunkciók (status function) olyan szabályok, amelyek a konstituált új entitásra vonatkozóan rögzítik, hogy mit lehet vagy mit nem lehet tenni velük. Hiába állapítja meg ezeket Searle, a konstitutív szabály formalizmusába már nem emeli át.

Ransdell azt javasolja, hogy a konstituált új entitás két fontos összetevőjét nevezzük meg, válasszuk el egymástól [429, 388-389]. Szerinte ugyanis a konstituált entitásnak

[19] F.A. Hindriks egy tanulmánya elején rövid, de tartalmas áttekintést ad a searle-i felfogás kritikájáról [234].

(terminusnak) mindig van mellékjelentése, *konnotációja* (*connotation*) egyfelől, illetve *tartalma* (*import*) másfelől. A baseball szabályain szemléltetve Ransdell a következő példákkal világítja meg a két új fogalom jelentését. Vannak Y-terminusok (Ransdell szavaival: játékterminusok/game-terms), mint Ball (labda), Bat (baseballütő), Pitcher (dobó), Batter (ütő), Catcher (elkapó), Inning (menet) stb. Ezeknek legtöbb esetben van konnotációja és tartalma (import-ja). A Bat konnotációjába tartozik a mérete, formája, felépítése, míg a tartalmával azt lehet rögzíteni, hogy mire lehet, mire kell, mire nem lehet használni: például lehet vele ütni a labdát, de nem lehet vele elgáncsolni a futó játékost. Ransdell fontos felismerése, hogy ez utóbbiak elsőrendű játékszabályok (cselekvési szabályok), amelyek "be vannak ágyazva" a másodrendű konstitutív szabályokba. Jobban meg tudom világítani Ransdell gondolatát, ha a szemléltető példákat a baseball helyett inkább a labdarúgás területéről veszem.

*

A labdarúgás szabályait egy több, mint százharminc oldalas könyvben rögzítik [163]. Ez a szabálygyűjtemény tele van konstitutív szabályokkal. A játék játszhatósága végett mindent pontosan definiálni kell, így meghatározás olvasható arról, hogy mit jelent a 'játékos', 'cserejátékos', 'lecserélt játékos', a 'csapat', a 'labda', a 'pálya', a pályára felfestett 'vonalak' és 'pontok', a 'kapu', a 'leshelyzet', a 'szabadrúgás' stb. fogalma. A játék értelme, célja és ezért központi fogalma a 'gól'. Nézzük meg a meghatározását:

> „Gól akkor van, amikor a labda teljes terjedelmével áthalad a kapuvonalon a kapufák között és a keresztléc alatt, feltéve, hogy a gólt elérő csapat előzően nem követett el szabálysértést." [163, 30]

A definíció támaszkodik olyan, már előzőleg konstituált fogalmakra, mint a 'labda', 'csapat', 'kapuvonal', 'kapufa', 'keresztléc', valamit 'szabálysértés'. Searle formulája úgy konkretizálható, hogy a C kontextus itt a 'labdarúgást játszani', az Y a 'gól' és az X a fenti definíció szinte teljes egészében. A definíciót természetesen csak úgy lehet pontosan és egyértelműen értelmezni, ha ismerjük a meghatározásba ágyazott további konstutitív fogalmak jelentését is. De azok is mind a searle-i logika mentén határozhatóak meg. Eddig tehát tűnhetne úgy is, hogy Searle leírása megfelelő és elegendő, de ez nincs így. Vannak ugyanis olyan szabályok a labdarúgás szabályrendszerében, amelyek kilógnak a searlei formalizmusból.

Nézzük meg a büntetőrúgás (tizenegyes) végrehajtásának szabályait! A szabálykönyv korábbi pontjaiban rögzítették, hogy mi az a tíz 'szabálytalanság', amely esetén 'közvetlen szabadrúgást' kell ítélni a szabálysértő játékos csapata ellen [163, 32]. Ezek alapján lehet bevezetni a 'büntetőrúgás' terminusát. A büntetőrúgás konnotációjáként értelmezhetjük azokat a sorokat, amelyekben segítenek abban, hogyan kell a tizenegyesrúgás fogalmát értelmezni.

> „Büntetőrúgás jár az ellen a csapat ellen, amely elköveti a közvetlen szabadrúgást maga után vonó tíz szabálytalanság valamelyikét a saját büntetőterületén, amikor a labda játékban van." [163, 41]

Miután megadtuk, hogy mit jelent a büntetőrúgás fogalma, milyen esetben kell büntetőrúgást ítélni egy focimeccsen (tehát hogyan, mikor konstituálódik a fogalom), utána kell megadnunk azt, hogy milyen körülmények között és hogyan kell végrehajtani a büntetőrúgást. Ezektől az előírásoktól lesz, lehet érvényes vagy érvénytelen a tizenegyes végrehajtása. Először megadják a büntetőrúgás során előírt elhelyezkedést:

„A labdát:
- a büntetőpontra kell helyezni.

A büntetőrúgást végrehajtó játékost:
- egyértelműen azonosítani kell.

A védőcsapat kapusának:
- a kapuvonalon a kapufák közt, a rúgó játékossal szemben kell maradnia a rúgás elvégzéséig

A többi játékosnak:
- a játéktéren
- a büntetőterületen kívül
- a büntetőpont mögött
- a labdától legalább 9,15 méter távolságra kell elhelyezkedni." [163, 40]

Az elhelyezkedésre vonatkozó szabályok mind a játékosok cselekvésére irányuló, elsőrendű szabályok, melyek a szokásos módon megragadhatóak deontikus operátorokkal. A következő feltételsorban azt rögzítik, hogy milyen módon lehet magát a büntetőrúgást szabályosan végrehajtani. Ezek a szabályok nyilván a büntetőrúgást végző játékosra vonatkoznak csak.

- „Miután minden játékos a szabály által előírt helyen tartózkodik, a játékvezető jelt ad a büntetőrúgás végrehajtására.
- A büntetőrúgást végző játékosnak a labdát előre kell rúgnia.
- Nem játszhatja meg ismét a labdát, amíg az egy másik játékost nem érintett.
- A labda akkor van játékban, amikor elrúgták és előre elmozdult." [163, 41]

Ebben a definíciós részben is kötelezések és tiltások vannak, vagyis itt is alkalmazhatjuk az elsőrendű cselekvésekre irányuló deontikus operátorokat. A szabályok harmadik csoportjában a büntetőrúgással kapcsolatos szabálysértéseket és a rákövetkező büntetéseket kapcsolják össze – egymás után felsorolva azokat az eseteket, amikor a büntetőrúgást végző játékos, a kapus, a rúgást végző játékos csapattársa, a kapus csapattársa, illetve mindkét csapat játékosa vét a szabályok ellen. Ha a támadó csapat játékosa sérti meg a játékszabályokat, akkor a játékvezetőnek az alábbiak szerint kell döntenie:

„Ha a büntetőrúgást végző játékos ... [vagy] a rúgást végző játékos csapattársa sérti meg a játékszabályokat:
- Ha a labda kapuba jut, a rúgást meg kell ismételni.
- Ha a labda nem jut a kapuba, a játékvezető állítsa meg a játékot, és a mérkőzés a védő csapat javára megítélt közvetett szabadrúgással indul újra arról a helyről, ahol a szabálysértés történt." [163, 42]

Ha a védekező csapat valamelyik játékosa szabálytalankodik, akkor a bírónak a következőképpen kell eljárnia:

„Ha a kapus ... [vagy] a kapus csapattársa sérti meg a játékszabályokat:
- Ha a labda kapuba jut, gólt kell ítélni.
- Ha nem jut a kapuba, a rúgást meg kell ismételni." [163, 42]

Ezek a szabályok rögzítik azokat a szempontokat, amelyek alapján a bíró dönthet az érvényesség-érvénytelenség kérdéséről. Fontos különbség, hogy ezek az előírások a bíró tevékenységére vonatkoznak, míg a korábbiak a játékosok cselekvéseire irányultak. Az, hogy mikor lesz gól a büntetőrúgásból, két tényezőtől függ. Egyfelől attól, hogy szabályos volt-e a büntetőrúgás végrehajtása, másfelől attól, hogy a büntetőrúgás eredményeként a labda bekerült-e a kapuba. Feltételezve, hogy a büntetőrúgásra vonatkozó szabályokat betartotta a rúgó játékos, az érvényesség kérdését "leszűkíthetjük" arra a sza-

bályhalmazra, amely a büntetőterületre való belépés tilalmával kapcsolatos (illetve a gól definíciójára). Az alábbi táblázat mutatja, mi a bíró teendője a különböző esetekben.

belép-e valaki:	a labda a kapuba kerül	a labda nem kerül a kapuba
támadó játékos	a büntetőrúgás megismétlése	közvetett szabadrúgás
védő játékos	gól	a büntetőrúgás megismétlése
mindkettő játékos	a büntetőrúgás megismétlése	a büntetőrúgás megismétlése
senki	gól	nem gól, megy tovább a játék

A táblázat első oszlopa mutatja, hogy követ-e el valaki szabálysértést a belépésre vonatkozóan, és ha igen, melyik csapatba tartozó játékosról van szó. A következő két oszlopban a különböző esetekben meghozandó bírói döntést láthatjuk aszerint, hogy a büntetést végrehajtó játékos berúgta-e a labdát vagy sem a kapuba. A gól egyik feltétele az, hogy a kapuba kerüljön a labda. Ezeket a helyzeteket mutatja a táblázat második oszlopa. A gól érvényességéhez azonban arra is szükség van, hogy ne előzze meg azt a támadócsapat szabálytalansága. A szabálytalanságot itt a büntetőterületre való korai belépés jelenti. Ezért van az, hogy "csak" két esetben van gól (a lehetséges négyből): akkor, amikor a támadó csapat nem szabálytalankodott (a második és negyedik sorban). Ha a gól elsődleges feltétele nem teljesül (ha a labda nem kerül be a kapuba), akkor a a szabálytalanságok figyelése csak arra jó, hogy el lehessen dönteni, meg kell-e ismételni a büntetőrúgást vagy sem (és akkor csak kell ismételni, ha a védekezőcsapat szabálytalankodott). A játékvezető kompetenciájába tartozik, hogy megismételteti-e a büntetőrúgást vagy sem. A tapasztalatok szerint az ismételtetések elég ritkán fordulnak elő.[20]

Előfordulhat, hogy az érvényesség fogalma mellett más fogalmat is használatba kell venni. Sok esetben meg kell különböztetnünk egymástól az **érvényességi feltételt** és az **eredményességi feltételt**. A különbséget a büntetőrúgás példáján szemléltetve azt mondhatjuk, hogy ha a támadó berúgja a labdát, de szabálytalanság történt, akkor érvénytelen lesz a büntetőrúgás, míg ha a támadó nem rúgja be a labdát és nem történt szabály-

[20] A YouTube-on látni olyan videót, amelyen a bíró ötször ismételteti meg a büntetőrúgást, és csak hatodik kísérletet fogadja el érvényesnek, lásd: http://youtube.com/watch?v=VJwAE5Mtrxc.

1982. december 5-én az Ajax – Helmond Sport labdarúgó mérkőzésen az Ajax – 5:0-ás vezetésnél – büntetőhöz jutott. Johan Cruyff egy társa, Jesper Olsen segítségével különös módon hajtotta végre a tizenegyest. Cruyff nem egyetlen rúgással próbált meg gólt rúgni, ahogy előtte mindenki a labdarúgás történetében, hanem lassan oldalra gurította a labdát, amit az elrúgás után már szabályosan a tizenhatoson belülre befutó Olsen kapott meg, aki az öt és feles vonalhoz érve visszaadta a középen tartózkodó Cruyffnak, ő pedig közvetlen közelről bevágta a labdát a kapuba. Az eset mindenben megfelelt a szabályoknak, a bíró meg is adta a gólt. Az esetről több videó is látható a YouTube-on (például: http://youtube.com/watch?v=GDoqv74egBM). Később kiderült, hogy az elsőség nem Cruyffot illeti, hanem Rik Coppenst, aki már 1957-ben megcsinálta a két-lépéses büntetőt egy társával (lásd: http://youtube.com/watch?v=hFk0xqO4VkU).

Mivel „ . . . a játékvezető dönti el, mikor tekintendő a büntetőrúgás végrehajtottnak", adódik a kérdés, hogy meddig tart egy tizenegyes végrehajtása. Ez az esetek döntő részében egyértelműen megválaszolható, de rendkívül ritkán előfordul az a jelenség, amikor a sikeresnek tűnő védéssel nem "zárul le" a büntetőrúgás eseménye, mert a védés után valahogy mégis bekerül a labda a hálóba. Előfordult olyan eset, hogy miután a kapus kivédte a büntetőt, felpattant és a közönség felé fordulva ünnepelni kezdte magát, miközben a kapus testén gellert kapó labda szép lassan bepörgött a kapuba (http://youtube.com/watch?v=r4bnoMka0SE). Egy másik alkalommal, egy büntetőpárbaj során a felső kapufáról magasba pattanó labda visszahullott az öt és feles vonalára, mire a kapus örömében kirohant a pálya szélén levő társaihoz, ám eközben a visszafele pörgő labda a kapu irányába pattogva szép lassan áthaladt a gólvonalon (http://youtube.com/watch?v=s5hE6PK67bc).

talanság közben, akkor érvényes lesz a büntető, de eredménytelen. Végül, ha szabályos körülmények között rúgja be, akkor érvényes és eredményes lesz a tizenegyes. A parlamenti választások alkalmával is van érvényességi, illetve eredményességi feltétel. Az érvényességi feltételként azt lehet kikötni, hogy a szavazók x (például 25) százaléka érvényes szavazatot adjon le. Az eredményességhez az kell, hogy valamely jelölt megkapja az érvényes szavazatok valamilyen többségét (például az első fordulóban a szavazatok abszolút, az esetleges második fordulóban a szavazatok relatív többségét).

*

A labdarúgás példáján láthattuk, hogy a konstitutív szabályokat leírásához nem elégséges a searle-i 'counts as' operátor alkalmazása. A szabályok konstitutív jellegét meg lehet ragadni vele, de logikai formalizmus alapjául nem képes szolgálni. A searlei funkció (vagy a ransdelli import) kiegészítő fogalmának szükségessége pedig arra mutat rá, szükség van arra, hogy elsőrendű regulatív szabályokat kapcsoljunk a konstituált fogalmakhoz (társadalmi intézményekhez). Két kérdést is fel kell itt tenni, amelyekre a válaszokat egymással összekapcsolva lehet megadni: *(i)* milyen logikai formulákkal ragadhatjuk meg a konstitutív szabályokat; *(ii)* hogyan kapcsolhatjuk össze a konstitutív szabályokat a felhatalmazottság köré szerveződő jogviszony-kategóriákkal vagyis a másodrendű deontikus fogalmakkal.

Először is észre kell vennünk, hogy a játékvezető teendőit másodrendű szabályok írják elő, és így a játékvezető cselekvéseit (döntéseit) is másodrendű cselekvésként írhatjuk le. „Ha a kapus megsérti a játékszabályokat büntetőrúgásnál, de közben kivédi a tizenegyest, akkor a rúgást meg kell ismételni." – szól a bíróra vonakozó szabály. Egy előírásba be van ágyazva egy másik előírás (ami a kapusra vonatkozik) – ez egyértelmű jele a másodrendű szabálynak. De milyen másodrendű fogalomról van szó? A választ akkor kapjuk meg, ha a hohfeldi fogalmak segítségével próbáljuk meg jellemezni a büntetőrúgás helyzetét. Mit is jelent a büntetőrúgás? Azt a jogot a támadócsapat számára, hogy egy játékosa egyedül rúghassa kapura a labdát a kapussal szemben (és minden más játékos kötelezettségét arra vonatkozóan, hogy ezt az igényjogban kifejezett cselekvést ne akadályozza). Ha a konstitutív szabályok valamelyikét megsértik a büntetőrúgás során, akkor a bíró hozhat döntés az érvényességre vonatkozóan. Amikor a játékvezető úgy ítél, hogy megismételteti a be nem rúgott büntetőt a védőcsapat szabálytalansága miatt, akkor a támadócsapat büntetőrúgásra vonatkozó jogát teremti újra. Tehát a létező jogviszonyok valamelyikét változtatja meg. Ez azt jelenti, hogy a bíró – a hohfeldi értelemben vett – felhatalmazottsággal rendelkezik.

Mivel ez a tézis nagyon fontos mozzanat a gondolatmenetemben, ezért szeretnék még egy példát bemutatni, amikor a labdarúgáshoz képest jóval egyszerűbb versenyről szól, ezért jobban áttekinthetőek a viszonyok a játékban. 200 éves hagyomány az angliai Gloucester városában a Cooper-dombi sajtkergetés.[21] Hosszú ideig hivatalos sportversenynek számított, ám 2009-ben visszavonták a hivatalos minősítést a rendezvénytől a sok balaset miatt. Azóta "nem hivatalosan" gyűlnek össze az évente megrendezett versenyen a versenyzők, nézők. A versenyen a gloucesteri domboldal tetejéről legurítanak egy 3 és fél kilós kerek sajtot, és a résztvevők közül az lesz a nyertes, aki utoléri és el tudja kapja a sajtot. Ha ilyen nincs, akkor az nyer, aki legelőbb leér a domb aljára. A verseny

[21] http://www.cheese-rolling.co.uk/index1.htm

célja a sajt megszerzése, de ezt gyakorlatilag lehetetlen teljesíteni, tehát a győzelem igazi feltétele a leérés elsősége. A sajtkergetési versenyt mint társadalmi intézményt az alábbi – searle-i értelemben vett 'counts as' – konstitutív szabályokat révén definiálhatjuk:

Y	(is counted as) *X*
versenynap	meghirdetett dátum
startvonal	Cooper-domb teteje
célvonal	Cooper-domb alja
pálya	startvonal és célvonal közti terület
sajt	hivatalosan kijelölt sajt
kezdés	elindítják a sajtot a startvonaltól a dombon lefelé
verseny	kezdés után (elkapni a sajtot vagy végigmenni a pályán)
győzelem	versenyen elsőre (elkapni a sajtot vagy végigmenni a pályán)

Az első oszlopban láthatóak a kontituált, intézményes tényekkel leírt entitások(Y), a második oszlop tartalmazza a fizikai világban létező, nyers tényekkel leírt entitásokat (X). Utóbbi minősítés persze csak úgy igaz, ha figyelembe vesszük, hogy az (X) entitások tartalmazhatnak "beágyazott" intézményes tényeket, amelyek azonban – esetleg több lépésben – mindig visszavezethetők fizikai tényekre, entitásokra. A pálya meghatározása például tartalmaz két másik intézményes tényt (a startvonal és a célvonal kategóriáit), azonban ezek egy-egy további lépéssel már visszavezethetők nyers tényekre.

Ezek az intézményes tények, ezek a konstituált entitások még nem elegendőek ahhoz, hogy le lehessen bonyolítani a versenyt. Szükség van még olyan regulatív szabályokra, amelyek a meghatározzák, hogy mit, hogyan kell, lehet használni a konstituált fogalmak által definiált, konstituált környezetben, vagyis hogyan kell a versenyt lebonyolítani. És ha ezen szabályok közül valaki megsért egyet, akkor – nagy valószínűséggel – nem lesz érvényes "valami". Ilyesfajta szabályokról lehet szó:

előírások	szabálytalan/érvénytelen
rajtvonal mögül kell indulni	ha valaki lejjebb indul
kezdés előtt tilos indulni	ha valaki előbb indul
a hivatalos sajtot kell elkapni	ha valaki a saját sajtját kapja el
versenynapon kell versenyezni	ha valaki más napon versenyez

Hogy az ilyen szabályok nélkül nem lehetne a versenyt sem definálni, sem lebonyolítani, azt felfedhetjük azzal, ha belegondolunk, hogy milyen reakciókra számíthatna az, aki nem igazodna a regulatív szabályokhoz. Ha valaki az általa vitt sajtot maga elé ejti, majd azt felkapva felkiált, hogy 'nyertem', senki nem fogja győztesnek tartani, és ha megkérdezi, miért nem, akkor azt a választ kapja, hogy azért nem, mert nem a hivatalos sajtot kapta el vagy másként: a győzelemhez a hivatalos sajtot kell elkapni. Ha valaki előbb elindul a többiekhez képest, és ezért el tudja kapni a sajtot vagy elsőként ér le, akkor azt fogják neki mondani, 'azért nem te vagy a győztes, mert tilos volt a kezdés előtt elindulni, te pedig ezt tetted'.

Vannak, lehetnek hallgatólagos, nem reflektált szabályok is. Például van egy tényleges szabály, miszerint bárki indulhat a versenyen. Nyilván mindenki úgy gondolja ezt, hogy bárki indulhat, aki ember. Ez azonban így nincs rögzítve. A szabályzat e "hiányossága" csak akkor derülhetne ki, ha egy sajtimádó hegyikecske is pont akkor futna arra, amikor

elhangzik a kezdésjelzés, és ügyesen lefele szökdellve elkapná a guruló sajtot. Nem tűnik valószínűnek ennek bekövetkezte, de elméletileg nem lehet kizárni, és elég biztosnak mondható, hogy ha mégis megtörténne, akkor nem hirdetnék ki a kecskét győztesként, viszont beleírnák a szabályba, hogy kecskék nem versenyezhetnek.

A konstitutív szabályok működési mechanizmusa a következő. Konstituálunk bizonyos társadalmi intézményeket (amelyeket intézményes és nyers tényekkel írunk le).[22] A konstituált fogalmakat sok esetben egymásba ágyazzuk. A konstituált fogalmakhoz, társadalmi intézményekhez sok esetben elsőrendű, regulatív cselekvési szabályokat rendelünk. Ezeket a regulatív szabályokat lehet betartani és lehet megsérteni őket. Amennyiben az utóbbi valósul meg, vagyis megsértik a konstituált intézményekhez rendelt elsőrendű szabályok valamelyikét, akkor a szóban forgó társadalmi intézmény konstitív feltételerendszere nem lesz teljes, tehát a konstitutív szabály alapján létrehozni kívánt intézmény érvénytelen lesz. A kulcsmozzanat tehát az érvényesség kérdése. De mit is jelent ez? A fehatalmazottságban benne rejlő elismerési igény érvénytelenségéről van itt szó. Felhatalmazottsága a bírónak van, aki a szabálytalanságokat megállapíthatja, dönthet az elsőség kérdésében. A labdarúgás az előzetesen rögzített szabályok által konstituált játék. A felhatalmazottság először ott érhető tetten, amikor változtatnak valamit a játék szabályain. A 2010-es évek egyik nagy kérdése a videobírók bevezetése. Ha valamikor elfogadják a videobíró intézményét, akkor ehhez módosítani kell majd a szabálykönyvet, és további fogalmakat, illetve hozzájuk kapcsolódó szabályokat kell felvenni. A felhatalmazottság szinte definitív erővel jellemzi a játékvezetők tevékenységét. Ők vannak azokkal a jogosítványokkal ellátva, amelyek alapján folyamatosan változtathatják a játékot játszók jogviszonyait (jogokat adnak ide, kötelezettségeket rónak ki oda). Felmerül a kérdés, vajon van-e felhatalmazottsága a játékosnak. Nos, ritkán, kivételes alkalmak esetén még ez is lehetséges. Előfordul olykor, hogy a bíró sérülés miatt megszakítja a játékot. A megszakítás ténye azt jelenti, hogy valamelyik csapat volt a labda birtokában, és a bíró "elvette tőlük a jogot" a további birtoklás lehetőségét, hogy a sérült játékost ápolni lehessen. Az ápolás után a bíró feldobással hozza játékba a labdát, ezzel indítja újra a játékot. Ilyen esetekben, ha nem annak a csapatnak a játékosa szerzi meg a labdát, amely csapat a megszakítás előtt birtokolta a labdát, akkor az a konvencionális elvárás (ami nem a játékszabály része), hogy az ellenfél "önként" (tehát nem valamely játékszabályi előírása miatt) adja vissza a labdát. Ha ezt megteszik, akkor mondhatjuk, hogy éltek a felhatalmazottságukkal, és a labdabirtoklás jogát átadták a másik félnek.

Érdekes helyzetet produkált egy dél-koreai labdarúgó-mérkőzés.[23] Miután az egyik csapat játékosa társa sérülése miatt kirúgta a labdát a pályáról, az ápolás után a másik csapat folytathatta a játékot. A bedobás után az ellenfél támadója a konvenciónak megfelelően "visszaadta" a másik csapat kapusának a labdát, de a lövés túl erősre és "pontosra" sikerült, a kapus nem tudta megfogni a labdát, gól lett belőle. Bár a rúgó játékos mozdulatából látszódott, hogy nem így akarta, elnézést is kért érte, a bíró nem tehetett mást: gólt ítélt. A középkezdés után az "érdemtelenül", "erkölcsileg helytelenül" gólhoz jutott csapat kapusa a hozzákerült labdát direkt a saját kapujába lőtte, hogy ezzel az öngóllal

[22] Vita van arról, hogy a nyers tények (a fizikai világ X entitásai) feltétlen szükségesek-e az Y társadalmi intézmények definiálásához vagy sem. Barry Smith – John Searle-lel szemben – azt állítja, hogy nem szükséges, vannak olyan társadalmi intézmények, amelyeket nyers tényekre való hivatkozás nélkül definiálunk. A modellem felépítése szempontjából nem perdöntő e kérdésben állást foglalni, így ezt a kérdést itt nem elemzem.

[23] Az esetről beszámoló cikk: [253], a mérkőzés videója: http://www.youtube.com/watch?v=vLfNFZQdvE4

kárpótolja az ellenfelet. Nézzünk meg egy jogi példát is! Az egyéni írásbeli magánvégrendeletetre vonatkozó magyar jogszabály az alábbi alaki érvényességi feltételeket szabja:

> „Az egyoldalas „ . . . írásbeli magánvégrendelet alaki szempontból érvényes, ha készítésének ideje az okiratból kitűnik, továbbá ha a végrendelkező azt
> **a)** sajátkezűleg írt végrendelet esetén elejétől végéig maga írja és aláírja;
> **b)** más által írt végrendelet esetén két tanú együttes jelenlétében aláírja, vagy ha azt már aláírta, az aláírást két tanú előtt, azok együttes jelenlétében a magáénak ismeri el, és a végrendeletet mindkét esetben a tanúk is – e minőségük feltüntetésével – aláírják; vagy
> **c)** a sajátkezűleg írt vagy más által írt végrendeletet aláírja, és akár nyílt, akár zárt iratként a közjegyzőnél – végrendeletként feltüntetve – személyesen letétbe helyezi." (PTK 7:17. § Az írásbeli magánvégrendelet alaki érvényességi feltételei [1])

A fentiek alapján ha saját kézzel megírom a végrendeletemet, és azt személyesen zárt iratként közjegyzőnél letétbe helyezem, akkor alaki szempontok alapján érvényessé válik a végintézkedés. Ha halálom után kiderül, hogy a házamat az egyik gyermekemre hagytam, és úgy rendelkeztem, hogy a másik gyermekemé legyen a szomszéd háza, akkor tartalmilag az első rendelkezésem érvényes lehet, de a második nyilván nem lehet az. Ha a két rendelkezési aktust kettébontjuk, hogy külön kezelhessük őket, akkor lesz egy érvényes (a felhatalmazottságot példálózó) és egy érvénytelen (a felhatalmazottság hiányát példálózó) cselekményünk. A két cselekmény megegyezik alakilag. Ezt mondhatjuk úgy is, hogy mindkét esetben szabadságom volt megtenni a szóban forgó cselekvést. Amiben különböznek, az az aktushoz rendelhető igény létében van. Az első esetben érvényes elismerési igényt kapcsolhatunk a szabadsághoz, a második esetben ez hiányzik, vagyis nem rendelhető érvényes elismerési igény az – egyébként szabadon elvégezhető – tevékenységhez.

Hasonló példaként szolgálhatna az a két eset, amikor az egyikben a hivatalban levő miniszterelnök menti fel valamelyik miniszterét, míg a másikban én próbálnám meg ugyanezt. Mindkettőnknek szabadságában állna ugyanazt a mondatot kimondani, mindketten meg is tehetnénk ezt, de az én elismerési igényemet nem (vagy csak kevesen) fogadnák el, míg a miniszterelnökét igen. Az egy másik kérdés, amit még eddig nem vizsgáltam, hogy mi számít az elismerési igény társadalmi elfogadásának. Ha megnézzük az alábbi hírt a kétezres évek Magyarországáról, akkor láthatjuk, hogy különleges és konkurrens elismerési igények olykor felbukkanhatnak a történelemben.

> „Borbély József végre elkezdi a beszédét. Elmeséli, hogy ő volt a Tempó, vagyis Tiszta Energiával Magyarországért nevű párt elnöke, mígnem 2009. június 13-án hajnalban 'arra nem ébredtem, hogy Magyarország királyság.' Bármelyikünket megzavarna egy ilyen hír, de Borbély dicséretesen hamar ráeszmélt arra, hogy nemcsak királyság van, de pont ő az 'önkéntes kormányzó'. . . . 2011. november 11-én véget ért a kettős hatalom időszaka, a Kormányzó visszahelyezte jogaiba az Apostoli Magyar Királyságot, Orbán Viktor pedig azóta nem más, mint egy amúgy is illegitim párt elnöke.
> December 28-án Kecskeméten a Kormányzó felavatta az Apostoli Magyar Királyság Kormányzói Hivatalát, egyben betiltotta minden magyarországi párt működését. Ha ez nem lett volna elég, idén február 25-én partra szállt Visegrádon és bátor gerillaakcióval levonta a köztársasági zászlót." [534]

Lehet azt mondani a történet főszereplőjére, hogy "bolond", de ez nem változtat azon a tényen, hogy a hír egy olyan elismerési igény kinyilatkoztatásáról (is) szól, amelyet

kevesek azért csak elfogadtak.[24] Az eddigi fogalmainkat egy táblázaton belül ábrázolva szemléletesebbé tehetők az egyes kategóriák közti szerkezeti hasonlóságok.

kötelező	$\text{OBLIGATORY}(Ego, \text{DO}(Ego, F))$	$\wedge$	$\neg\text{DO}(Ego, F)$	$\rightarrow$	S
tiltott	$\text{FORBIDDEN}(Ego, \text{DO}(Ego, F))$	$\wedge$	$\text{DO}(Ego, F)$	$\rightarrow$	S
opcionális	$\text{OPTIONAL}(Ego, \text{DO}(Ego, F))$	$\wedge$	$\text{DO}(Ego, F) \mid \neg\text{DO}(Ego, F)$	$\rightarrow$	$\neg$S
megengedett	$\text{PERMITTED}(Ego, \text{DO}(Ego, F))$	$\wedge$	$\text{DO}(Ego, F)$	$\rightarrow$	$\neg$S
mellőzhető	$\text{OMISSIBLE}(Ego, \neg\text{DO}(Ego, F))$	$\wedge$	$\neg\text{DO}(Ego, F)$	$\rightarrow$	$\neg$S
felhat.-tt	$\text{AUTHORITATIVE}(\text{DO}(Ego, F))$			$\rightarrow$	$\neg$F
szabadság	$\text{OPTIONAL}(Ego, \text{DO}(Ego, F))$	$\wedge$	$\text{DO}(Ego, F)$	$\rightarrow$	$\neg$S
igény	$\text{OBLIGATORY}(Alter, R(F))$	$\wedge$	$\neg\text{DO}(Ego, R(F))$	$\rightarrow$	S
nem felhat.	$\neg\text{AUTHORITATIVE}(F)$			$\rightarrow$	$\neg$F
szabadság	$\text{OPTIONAL}(Ego, F)$	$\wedge$	$\text{DO}(Ego, F)$	$\rightarrow$	$\neg$S
igény	$\text{OBLIGATORY}(Alter, R(F))$	$\wedge$	$\neg\text{DO}(Ego, R(F))$	$\rightarrow$	S

A táblázat harmadik sorában az opcionális deontikus operátor szerepel, nem véletlenül. Nem lehet sem a megengedett, sem a mellőzhető deontikus operátorokat használni, mert előbbi magába foglalja a kötelező, míg az utóbbi a tiltás deontikus tartományát. Felmerülhet a kérdés, vajon a 'nem felhatalmazottságnak' lehet-e kétféle értelmezése? Az egyértelmű, hogy értelmes azt állítani, hogy valakinek nem szabad megtenni valamit és nincs is rá felhatalmazottsága. Az okirathamisító esete lehet a példa erre. De lehetséges-e, hogy valamit szabad megtenni, de nincs rá felhatalmazottság? Logikailag ez nem lehetséges, hiszen a felhatalmazottság fogalma monohierarchikusan generikus alárendeltje a szabadság fogalmának. A fogalmak jelentésének pontosítása végett nézzünk meg pár példát. A festményhamisítónak lehet szabadsága hamisítani, nekem szabadságom van leváltani az aktuális magyar pénzügyminisztert, de nem lehet igényem a változtatás elismertetésére. A pénzhamisítónak nincs szabadsága hamisítani. Vannak olyan konstituálások, amelykre lehet szabadságom, vannak, amelyekre tiltás van. De ha egyik esetben sincs felhatalmazottság, akkor nem lesz érvényesség, viszont ha volt tiltás, akkor akkor nemcsak érvénytelenség, de szankció is lesz (mivel tiltás volt). Az összehasonlíthatóság kedvéért a következő táblázatba beírtam néhány esetet, amelyben vagy van vagy nincs felhatalmazottsága a szereplőknek.

Az útlevélhamisító is pont ugyanolyan útlevelet állít elő, mint az állam, de azért nem lesz útlevél, mert a hamisítónak nem létezik felhatalmazottsága, az érvényesség feltételei nem teljesültek. Az öröklés szabályait polgári törvénykönyv egyes fejezetei és más törvények (például a családjogi törvény) ide vonatkozó részei határozzák meg. Az olyan fogalmakat, mint hogy ki számít rokonnak, házastársnak, tanúnak, ki lesz öröklésre érdemtelen, mi az öröklési sorrend stb. mindig konstitutív szabályokon keresztül definiálja a jogalkotó. A bírónak, a köztársasági elnöknek kötelező felhatalmazottsága van. A bírónak – ha megállapította a vádlott bűnösségét, akkor – nincs mérlegelési joga, vagyis köteles elítélni a vádlottat, és csak abüntetés mértékében mérlegelhet (ott – relatív – szabadsága van). A köztársasági elnöknek ki kell neveznie a beterjesztett személyeket az adott pozícióba, nem mérlegelhet, vagy alá kell írnia a felterjeszett tövényeket.[25]

[24] A hír szereplőit ezen a videón lehet látni: http://youtube.com/watch?v=avQKJJRDvtA

[25] Ezen a helyzeten nem változtat lényegileg az, hogy első körben részleges vétójoga van.

jogviszony	tartalom	jogkövetkezmény
tilos a dohányzás	dohányzom kocsmában	kötelező a szankció
megengedett a dohányzás	dohányzom otthon	tilos a szankció
kötelező büntetni	dohányzom kocsmában, rendőr büntet	érvényes a szankció
tilos büntetni	dohányzom otthon, rendőr büntet	érvénytelen a szankció
felhatalmazott később nyitni	később nyitok	érvényes
felhatalmazott pénzt nyomni	pénzt nyomtat	érvényes
nem felhat. pénzt nyomni	pénzt nyomtat	érvénytelen+szankció
felhat. lemondani	kiteszi a fotelét lomtalanításkor	érvényes
nem felhat. eladni	eladja a Rottenbiller utca burkolatát	érvénytelen+szankció
felhat. összeadni	anyakönvvez. összeadja a feleket	érvényes
felhat. végrendeletet írni	tanuk nélkül írja meg apaság elismerése	érvénytelen
nem felhat. kinevezni	kinevezem a minisztert	érvénytelen

6.2.6 A jogviszonyfogalmak teljes logikai rendszere

A jog és kötelezettség fogalmainak hohfeldi „finomszemcsés” felbontása “megkívánja” a fogalmak osztályozását, tipizálását. A logikai rend alapján már utaltam egy fontos tipizálási lehetőségre: a hohfeldi jogfogalmakat két típusba rendezhetjük aszerint, hogy első- vagy másodrendű fogalmak-e (elsőrendű jogosultságfogalom a jogigény és szabadság, másodrendű a felhatalmazottság és mentesség). A logikai rend alapján történő megkülönböztetés mellett természetesen vannak másfajta tipizálási javaslatok a jogfogalmakra vonatkozóan. A két legismertebb ezek közül az aktív és passzív jog, valamint a pozitív és negatív jog megkülönböztetésén alapul. A logikai rend, illetve az aktív-passzív dimenzió alapján a négy jogosultságtípusunkat az alábbi módon rendezhetjük el:

	aktív jog	passzív jog
elsőrendű jog	szabadság	igényjog
másodrendű jog	felhatalmazottság	mentesség

Bár a két hohfeldi jogcsoport között nagy a strukturális hasonlóság, látnunk kell a különbségeket is köztük. Többször is volt róla szó, hogy mindkét csoportnak van egy-egy központi fogalma, amelyre támaszkodva a csoport további három fogalmát is értelmezni lehet. Az első csoport vezérfogalma az igényjog, a másodiké a felhatalmazottság. Analóg szerkezeti helyzetük mellett azonban figyelnünk kell a két fogalom közti különbségre is. Egy fontos szempont szerint a két fogalom nagyon különbözik egymástól. Az igényjogot ugyanis a passzív jog, a felhatalmazottságot pedig az aktív jog típusába kell sorolnunk. Ez a különbségtétel természetesen csak akkor lesz érthető, miután már bemutattam a jogfogalmak másfajta tipizálási kísérleteit.

	Alter akarata nem teljesül	Ego akarata teljesül
elsőrendű jog	szabadság	igényjog
másodrendű jog	mentesség	felhatalmazottság

A hohfeldi elmélet tisztázatlan, vitatható pontjait rendbe lehet tenni a Kanger által javasolt formalizálási megoldások segítségével. Ettől azt remélhetjük, hogy logikailag teljesebb

	aktív jog	passzív jog
pozitív jog	pozitív szabadság	pozitív (igény)jog
negatív jog	negatív szabadság	negatív (igény)jog
pozitív jog	pozitív felhatalmazottság	pozitív mentesség
negatív jog	negatív felhatalmazottság	negatív mentesség

és konzisztensebb rendszerhez jutunk. Arra azonban a kangeri javaslatok nem kínálnak megoldást, hogy milyen nevekkel lássuk el azokat a fogalmakat, amelyeket újonnan kell bevezetni a modellünkbe. Nekem egyelőre nincsenek erre vonatkozó javaslataim, úgy vélem, időt kell hagyni arra, hogy sok példa gyűlhessen össze az egyes – még terminus nélkül álló – fogalmakra, és csak megfelelő számú alkalmazási lehetőség birtokában remélhetjük azt, hogy megfelelő kifejezéseket találunk.[26]

*

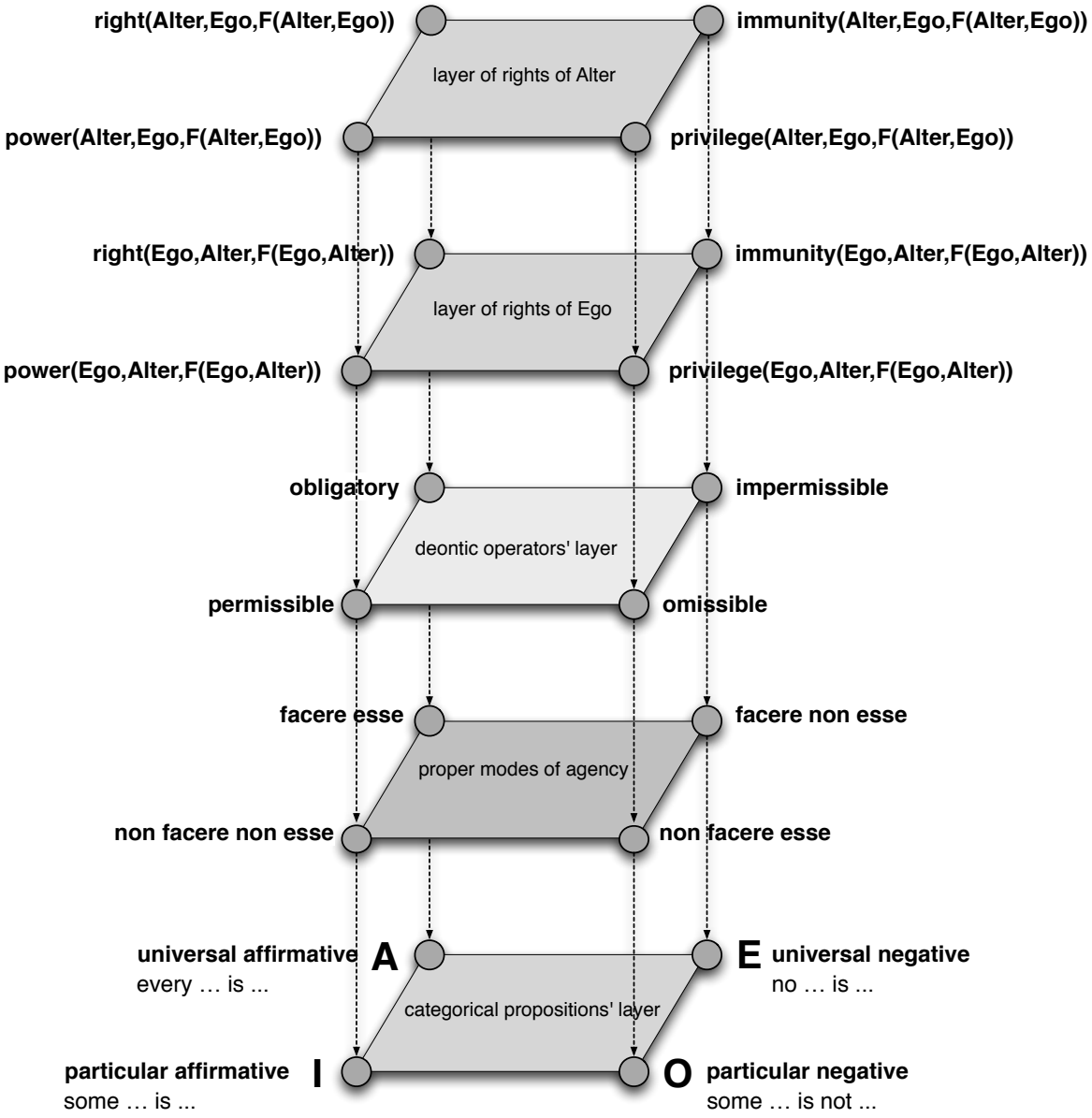
A hohfeldi jogtipológia ugyan nem a JEPD-elvnek (↷53.o.) megfelelően különíti el egymástól a négy jogtípust, hiszen nem egymást kölcsönösen kizáró jelentést rendelhetünk hozzájuk, ám ettől még jól megkülönböztethető egymástól a négyféle jogfogalom. A jog terminusában sűrűsödő négyféle jelentésréteg elkülönítése egymástól semmiképpen sem jelenti azt, hogy ezeket csak "magukban" lehetne a tényleges joggyakorlatban jogfogalmainak (tulajdonjog, birtokjog, szerzői jog, vallásszabadság, szólásszabadság stb.) meghatározásához használni. A hohfeldi jogfogalmak segítségével "összetett jogosultságokat" is definiálhatunk. A **tulajdonjog** példáján kiválóan lehet ezt szemléltetni.

Ego privilégiuma	Ego birtokolja, használja, hasznosítja saját tulajdonát, hasznot szed belőle
Ego igénye	Alter önhatalmúlag nem birtokolja, nem használja, nem hasznosítja Ego tulajdonát, nem szed hasznot belőle
Ego felhatalmazottsága	Ego rendelkezik a saját tulajdonával (eladja, elcseréli, elajándékozza, haszonbérletbe adja, végrendelkezik vele)
Ego immunitása	Alter nem rendelkezik Ego tulajdonával (nem adja el, nem cseréli el, nem ajándékozza el, nem veszi haszonbérletbe, nem végrendelkezik vele)

6.2.7 A jogviszonyfogalmak kiterjesztése

A jogviszonyok eddigi tárgyalása során jogelméleti szerzők gondolataira támaszkodtam. Ez gyakorlati szempontból talán érthető, hiszen ezen a tudományterületen született a legtöbb elemzés e kérdés kapcsán, elméletileg azonban megalapozatlan ez a leszűkítő tárgyalásmód. Az lett volna pontos (és, tán, a helyénvaló), ha a 'jogszonyok' fogalma helyett mindvégig a 'normatív viszonyok' kifejezést használtam volna, de ettől azért tekintettem el, mert nagyon nehézkesnek, kicsit elidegenítőnek, "kontextusvesztőnek" éreztem volna egy jogelméleti gondolatmenet során.

[26]Egyébként elemzési feladatokat találhatunk még a fogalmi rendszer bizonyos homályosabb pontjaival kapcsolatban is. További elemzéseket kíván például annak tisztázása, hogy a *megengedett* helyett nem az *opcionális* kategóriáját kell-e inkább használni a formuláinkban, mint ahogy azt is érdemes méregelni, hogy ha a pozitív szabadság *nem tilosként*, a negatív szabadság pedig *nem kötelezőként* értelmezzük, akkor abból az következik, hogy az előzőben benne lesz a *kötelezőség*, az utóbbiban edig a *tiltás*. Ez azonban nem igazán tűnik intuitívnak.



A továbblépéshez azonban érdemes ezt az általánosítást megtenni, mert ezzel mód nyílik arra, hogy egyrészt más tudományterületek tudását hasznosíthassuk, másrészt megpróbálkozhítunk azzal, hogy a jogviszonyok elemzése során eddig elért eredményeket máshol is "értékesíthessük". Az első lépést még egy jogász, Frederick Schauer "segítségével" tehetjük meg. Schauer ugyanis – szigorúan jogelméleti kérdéseket tárgyalva – az intézményeket az általuk hozható döntések típusai alapján különíti el egymástól.

> „Maga a hatáskör fogalma adja azt a hasznos eszközt, mely segítségével különbséget tehetünka döntéshozásra adott felhatalmazás és a meghozott döntés tartalma között. Következésképpen kétféle döntéshozó intézményt különböztethetünk meg: egyfelől azokat, melyek rendelkeznek mind a hatáskörrel felruházó, mind a döntések eredményét meghatározó tartalmi szabályokkal; másfelől pedig azokat, melyek csak az előbbi csoportba tartozó szabályokkal rendelkeznek, s így az eredmény meghatározását a döntéshozó szabályok által

nem vezérelt mérlegelésére bízzák." [461, 414]

Azért fontos Schauer megkülönböztetése, mert azáltal, hogy az intézményekhez rendelt jogviszonyok – jogelméleti – tárgyalását "áthelyezzük" egy döntéshozatali – szociológiai – kontextusba, láthatóvá, elemezhetővé válhatnak az eltérő jogosultságokkal való rendelkezés szociológiai konzekvenciái. Minden intézményt jogosultságok és kötelezettségek adott rendszerével jellemezhetünk, és minden intézményben a rá jellemző jogviszonyrendszer határozza meg azt, hogy miként cselekszenek, cselekdhetnek a szereplők. Ezek a cselekvések a szereplők döntésein keresztül is értelemezhetők, tehát a döntési kompetenciák különbségei határozzák meg az intézmények közti különbségeket.

„Azokban a döntéshozó intézményekben azonban, melyekben csak hatáskört megállapító felhatalmazó szabályokat alkalmaznak, hiányoznak a döntések kimenetelét megszabó tartalmi korlátok. A döntéshozó felhatalmazása szélesebb annál, semhogy egyszerűen szabályokat kelljen alkalmaznia a konkrét esetekre. A társadalmi és morális elvek és célkitűzések [*policy*] teljes készletére támaszkodva kell eldöntenie, hogy mi a teendő az előtte álló konkrét esetekben." [461, 415]

Schauer szemléletes példája a rendőr "intézménye", akinek a cselekvését felhatalmazó és tartalmi szabályok egyaránt meghatározzák (Schauer tartalmi szabályon a regulatív szabályokat "érti", bár erre nem igazán reflektál). A kétféle szabály közti viszonyra fény derül az alábbi idézetből:

„Az a szabály, mely felhatalmaz egy rendőrt arra, hogy mindazokat, de csakis azokat beidézze, akik átlépték az óránkénti kilencven kilométeres sebességhatárt, példaként szolgálhat arra az esetre, amikor a hatáskört megállapító felhatalmazó szabály és az eredményt meghatározó tartalmi szabály együttesen van jelen. Bár a rendőr fel van hatalmazva arra is, hogy idézést adjon ki, s arra is, hogy megállapítsa az autók sebességét, nincs felhatalmazva sem annak meghatározására, hogy mennyi legyen a megengedett legnagyobb sebesség, sem annak meghatározására, hogy mikor mentesülhet valaki a következmények alól, amennyiben átlépte a sebességhatárt." [461, 414-415]

Attól a pillanattól kezdve azonban, amikor már nem jogokról, hanem döntésekről, társadalmi intézményekről beszélünk, hirtelen "kinyílik" a fogalmi, elemzési tér, és a jogrendszeren kívüli normativitásban is ugyanúgy alkalmazhatjuk ezeket a distinkciókat. Nemcsak a döntéshozati jogosítványok alapján megkülönböztetett intézményfogalmakat, hanem minden más, a jogviszonyokról eddig bemutatott fogalmakat is alkalmazhatunk azokon a területeken, ahol jogosultságokról, kötelezettségekről általában beszélhetünk. A jog mint jogrendszer fogalma felől az általánosabb norma fogalma felé elmozdulva más dimenziókba léphetünk át. Bár Somlai Péter a családon belüli társadalmi kapcsolatok világát elemezte "csak", az alábbi meglátása általánosítható mindenféle normatív cselekvést vizsgáló tudományos tevékenység számára.

„A célracionális cselekvések modelljeivel dolgozó kutatónak tehát nemcsak vágyaktól és szokásoktól, értékvezérelt vagy irracionális döntésektől kell elvonatkoztatnia. Mindezekkel elvonatkoztat a szabályalkotás és intézményesedés folyamataitól is. Attól tehát, hogy *(a)* lehetnek olyan szabályok, amelyek valamiképp hozzáférhetők a résztvevők számára (akik formálhatják ezeket, illetve felismerhetik érvényességükben a saját kompetenciájukat, jogaikat és felelősségüket) és *(b)* a szabályok és intézmények is változhatnak a szereplők kapcsolatának és konfliktusainak eredményeként." [505]

Bármely társadalmi intézményben, ahol normák szabályozzák a résztvevők cselekvéseit, a kétféle szabály- vagy jogfogalom elemzése adhatja az eszközt az intézményen belüli

dinamikák megértéséhez. A jogrendszert elemezve tudtuk az alapvető jogviszonyfogalmakat értelmezni, pontosan definiálni, de ezek a fogalmak – kellő általánosítás után – sok más területen is használhatók. Mindenhol, ahol társadalmi intézményekben emberek saját normatív viszonyaikat próbálják meg szabályozni, fenntartani, változtatni. Vagyis mindenhol, ahol emberek cselekszenek. Pamela Oliver is tett utalást arra, hogy a társadalmi ösztönzők – többek között – arra is alkalmasak, hogy az emberek saját cselekvéseit befolyásoló, jogulstásokkal, kötelezettségekkel teli társadalmi környezetüket megváltoztathassák.

> „Az a lehetőség teszi a szelektív ösztönzőket a kollektív cselekvés dinamikus folyamatainak oly' fontos részévé, hogy megváltoztathatjuk azokat a paramétereket, amelyek meghatározzák mások (és akár sajátmagunk) viselkedését. Az emberek számára nem az az egyetlen lehetőség, hogy reagáljanak saját helyzetükre. Tehetnek is azért, hogy maga a helyzet változzon meg, amely addig korlátozta saját választásaikat." [392]

Az nyilván nem véletlen, hogy Hart híres másodrendű szabályai között is megtalálhatjuk a **változtatási szabályokat** (az elismerési szabályok és ítélkezési szabályok mellett) [228], hiszen ezek teszik lehetővé egyáltalán a jogszabályok változtatását. De ha a jog egészén mint a normatív rendszerek egy típusán belül szükséges valamely mechanizmus, valamilyen eszköz magának a jogrendnek a változtatásához, akkor ezt nyilván minden más normatív rendszeren (a konvencionális vagy erkölcsi rendszerken) belül is ugyanúgy meg kell találnunk. Nem is kockáztatunk túl sokat, ha azt mondjuk, hogy a magasabbrendű jogosultságokat és kötelezettségeket a társadalmi kapcsolatok dinamikáját biztosító eszközök közé kell sorolnunk. Talán mint a legfontosabbakat.

6.3 Társadalmi szerep

A hohfeldi jogosultságelmélet logikai formalizmusának kidolgozása után és alapján elég nyilvánvaló volt, hogy ki kellene, ki lehetne dolgozni a társadalmi pozíciók formális elméletét. Mivel azonban nem társadalomtudósok, hanem logikusok, mérnökök álltak neki ennek a feladatnak, ezek a kísérletek nem kapcsolódtak, nem tudtak kapcsolódni a társadalomtudományokban sokféle szempont szerint elemzett, kidolgozott társadalmi pozíció-, státus-, szerep- és hatalomelméletekhez.[27] Pedig lett volna hová. Arra nyilván nem vállalkozhatok itt, hogy bemutassam mindazokat az elméleteket, amelyeket kidolgoztak a társadalomtudomány jeles képviselői a jelzett fogalmakkal kapcsolatban. Egyetlen gondolati ívet szeretnék csak itt felvázolni, mégpedig azt, amin keresztül össze lehet kötni Hohfeld jogi pozíciókra vonatkozó elméletét a cselekvéselmélet számára különösen fontos uralomelmélettel.

Úgy látom ugyanis, amit a könyvem egyik fontos tézisének is tartok egyben, hogy az ontológiai leszármaztathatóság, levezethetőség értelmében, a formális modellépítés szempontjából a jogosultságokra és kötelezettségekre alapozott pozíció fogalma megelőzi a cselekvéselmélet két kiváltképp fontos fogalmát, a norma és az uralom kategóriáját. Hohfeld azt írta le az elméletében, ahogy az által elkülönített jogosultság és kötelezettségtípusok egymással korrelatív kapcsolatban állnak. Itt azonban megállt, és nem vizsgálta tovább, hogy lehet-e még mondani valamit az egymással korrelatív viszonyban álló jogi

[27] Alan Ross Anderson [7], Ingmar Pörn [420], Christen Krogh és Henning Herrestad [295], [296], Marek J. Sergot és Andrew J.I. Jones [486], [267] tanulmányait kell itt elsősorban megemlítenem.

pozíciókról. Elsőként Sir David Ross [447], majd ő követve Alvin W. Gouldner [195] volt az, aki felfedezte és a társadalomelméleten belül a megfelelő helyére tette a komplementaritás fogalmát.[28] Gouldner számára az volt az igazi kérdés, hogy a reciprocitás fogalmát hogyan kell értelmezni, és azon belül is különösen az érdekelte, hogy mi a különbség a reciprocitás és a komplementaritás fogalmai között [517]. A reciprocitás itt nem annyira fontos, így nyilván a két fogalom viszonya sem érdekes, viszont a gouldneri **komplementaritás** fogalma annál inkább. Ez a fogalom jelentését tekintve megfeleltethető a hohfeldi korrelativitás kategóriájának vagy a kangeri inverzitás-konverzitás fogalomkettősnek. Ettől persze még nem lenne fontos. A változást és az újdonságot Gouldnernél az jelentette, hogy a komplementaritás fogalmát, illetve a terminussal megragadható jelenséget – tulajdonképpen öntudatlanul – belehelyezte a cselekvéselmélet kellős közepébe azzal, hogy úgy kezelte, mint amivel sok esetben magyarázatot lehet adni a társadalmi cselekvéseink szabályszerűségére. Amikor ugyanis a reciprocitást a komplementaritás kategóriájával állította szembe, akkor a komplementaritás fogalmát vizsgálva a **társadalmi szerep** fogalmát is elemezte egyben. A társadalmi szerep viszont nem más, mint jogosultságok és kötelezettségek összekapcsolása úgy, hogy egy másik szerephez, *a* komplementer társadalmi szerephez a komplementer kötelezettségek és jogosultságok vannak hozzárendelve. Vegyünk egy egyszerű példát! Az egyetemen a tanári szerepet egy sor jogosultság és kötelezetség hozzárendelésével határozhatjuk meg. A tanárnak kötelesége órát tartani, tanévkezdéskor be kell jelentenie, mikor lesz zárthelyi a félévben, joga van zárthelyit megíratni a bejelentett napon, joga van rászólni a fegyelmezetlenkedő hallgatóra stb. Ezzel szemben a hallgatói szerephez az tartozik hozzá, hogy joga van órákra járni, joga van tanévkezdéskor rákérdezni, mikor lesznek a zárthelyik, viszont kötelessége beszámolót írnia az adott napon, és szót kell fogadnia a tanár felszólításának. A két társadalmi szerep között igazi komplementaritás létezik. Ez az igazán lényeges itt: nem pusztán az egyes jogosultság- és kötelezettségpárok vannak korrelatív, komplementer viszonyban egymással, de a szereppárok is. Nincs hallgatói szerep a tanári szerep nélkül és fordítva. Tanárként nem állhatok meg a villamoson diákkorú emberek előtt, hogy vegyenek elő papírt, mert lediktálom a zárthelyi dolgozatuk kérdéseit. Ha ezt tenném, értetlenség, kiröhögés várna rám. De nem teszem, mert tudom, hogy a villamoson nem lesznek ott azok, akik a hallgatói szerepben lehetnének, vagy ha netán ott lennének, akkor sem akarnának a hallgatói szerephez igazodni.

A cselekvéselmélet számára ez nagyon fontos előrelépés, mert amíg a hohfeldi fókusz a jogi pozíciókon volt, addig nem lehetett az elméletét hatékonyan alkalmazni a tényleges társadalmi cselekvések leírására. A jogi pozíciók túlzottan elemi szintjét jelentik a cselekvéseinknek. Szükség volt és szükség van a hohfeldi alapokra, de szükség volt egy szinttel feljebb lépni, szükség volt arra az általánosításra, amit a társadalmi szerep fogalma hoz(hat)ott el. Ezen a szinten már olyan fogalmakat képezhetünk, amelyek a mindennapi társadalmi gyakorlatban is kezelhető, beazonosítható, kommunikálható szabályszerűségeket ragadnak meg. Azt mondani, hogy tanár, diák, rendőr, eladó, sofőr, gyalogos, anya, gyerek stb., ezek mind olyan szabályszerűségekre mutató fogalmak, amelyeket könnyen és gyakorta felismerhetünk az életünkben. De ennek az a feltétele, hogy minél több, egy adott helyzetekre jellemző jogosultságot és kötelezettséget tudjunk egy komplementer szereppárhoz hozzárendelni.

[28] Sem Ross, sem Gouldner nem reflektált a hohfeldi elméletre, tehát nem beszélhetünk itt kapcsolódásról, inkább párhuzamosan kidolgozott elméletekről.

Van még egy további nagyon fontos érv arra vonatkozóan, hogy miért is fontos a társadalmi szerepeknek ez az összekapcsolódása. Talcott Parsons egyik észrevétele érdemes itt felidézni [405]. Parsons úgy érvelt, hogy a komplementer szerepek egyik legfontosabb funkciója az, hogy békés úton old meg egyébként eredetileg, önmagában véve konfliktusos társadalmi kapcsolatokat. Mindegyik társadalmi szerepet betöltő fél számára sokszor rendkívül gazdaságos lehet, hogy a komplementer szerepekbe komplementer módon "bedolgozott" kötelezetségeket önként, békében elfogadják az érintettek, és nem érzik úgy, hogy hatalmi helyzetben, kényszer hatására kellene teljesíteniük a rájuk háruló kötelezettségeket.

Tudom, kicsit furcsa, talán sokaknak elfogadhatatlan is, mégis azt állítom, hogy sem a jogi pozíciók, sem a társadalmi szerepek nincsenek feltétlen összekötve a normativitás jelenségével. A gyakorlatban persze tényleg nem nagyon lehet elválasztani őket egymástól, fogalmilag azonban a kapcsolat nem szükségszerű köztük. A komplementer szerepek kialakulhatnak és fennmaradhatnak normatív támogatottság, tehát az elvárások megsértése esetén alkalmazott szankció hiányában is. Elképzelhető, hogy kizárólag az érdekviszonyokra támaszkodó szerepmegosztás alakuljon ki a felek között, és bár mindig sérülékeny lehet az ilyen kapcsolat, de attól még az érdek összetarthatja a résztvevőket.

6.4 Norma

Mivel a társadalmi norma fogalmával valamilyen cselekvési szabályt írhatunk le, a normához igazodó vagy normatív cselekvések a szabályszerű társadalmi cselekvések egyik típusát jelentik. Bár nem lehetett teljes mértékben sikeresen megvalósítani, mégis arra törekedtem eddig, hogy a könyv korábbi fejezeteiben mellőzzem a normativitás fogalmát. Ez nagyjából sikerült is, ami azt mutatja, hogy az eddigi elemzett társadalmi jelenségeket durván le lehet írni a normativitás figyelembevétele nélkül is. Az is igaz viszont, hogy az eddigi elemzésekből több ponton már nagyon hiányzott a norma kategóriája. Ez nyilván nem véletlen, hiszen a társadalmi cselekvések nagyon fontos típusát adják a normatív cselekvések. Fontossága, mindenhol jelenlevősége miatt egyrészt nehéz észrevenni azt, hogy nagyon sok jelentésű fogalomról van szó, másrészt könnyű összekeverni a felszínen hasonló látszó, ám a mélyben máslényegűnek bizonyuló jelenségekkel (például az érdekkövetés nyomán kialakuló szabályszerűségekkel). Ebben a fejezetben megmutatom, hogy milyen módon lehet a norma fogalmát formális eszközökkel megragadni, ami segíthet majd abban, hogy el tudjuk választani a kategóriához tapadt, de fogalmilag nem hozzá tartozó jelentésrétegeket róla. Annyiban viszont kicsit bonyolítom a képet, hogy ugyanebbena fejezetben tárgyalom az uralom fogalmát is mégpedig úgy, hogy közben tételezem és bemutatom a norma és az uralom kategóriái közti nagyon szoros logikai-fogalmi kapcsolatot. A **norma** fogalma alatt a **társadalmi normát** értek, de amikor nem okoz zavart, akkor a két kifejezést szinonímnak tekintem.

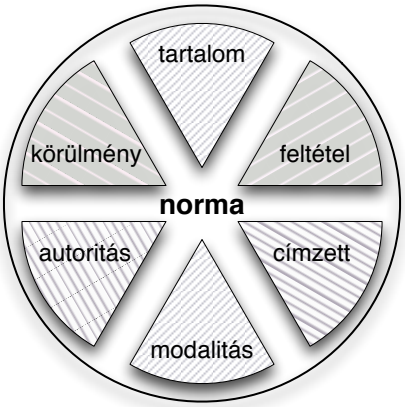
A következőkben bemutatom azt, miként lehet a norma jelenségét leírni. Először áttekintem, hogy a norma fogalmát milyen összetevőkre bonthatjuk fel. Ennek során a von Wright javaslatából indulok ki [564], de néhány új elemmel kiegészítem az eredeti von wright-i elképzelést. A kiegészítéseket viszont nem sorolom a norma összetevői közé, ezért a kibővített készletre a **norma jellemzőiként** hivatkozom. Ezek a normajellemzők arra is alkalmasak lesznek, hogy segítségükkel a fejezet második részében egy sok szempontú, kiterjedt nomatipizálást tudjak elvégezni.

A norma jellemzői

A norma leírásához összetevőire kell bontanunk a fogalmat, hogy ezáltal láthatóvá váljék, milyen elemi adatok szükségesek ahhoz, hogy egyértelműen megadhassunk egy adott normát. Von Wright hat összetevőt tartotta elkülönítendőnek [564, 70]:

- norma tartalma (content of norm)
- norma modalitása (character of norm)
- norma feltétele (condition of norm)
- norma alkalma (occasion of norm)
- norma autoritása (authority of norm)
- norma címzettje (subject of norm)

Nem vitás, hogy ezek az összetevők szükségesek a normák értelmezéséhez. A kérdés csak az, hogy elégségesek-e, nincs-e szükség újabb összetevők felvételére. Azt állítom, hogy szükség van további információ "begyűjtésére", és létezik még két olyan szempont, ami mentén további adatokat rendelhetünk a norma jelenségéhez. Mielőtt azonban ezekre kitérnék, egy ábra segítségével is rögzítsük az eredeti, von wrighti összetevők körét.



Ezek után nézzük meg, mivel is lehetne kiegészíteni a fentieket. Von Wright és a logikusok mindig is érzéketlenek voltak bizonyos társadalmi mozzanatok iránt a normatív jelenségek vizsgálata során. Az egyik ilyen vakfolt annak fel nem ismerése, hogy a normativitás csak egy részét jelenti a cselekvésbefolyásolás jelenségének, és a fenti von wrighti keretrendszer nemcsak a normákra húzható rá, de alkalmas lehet másfajta cselekvésbefolyásolási helyzetek leírására is. Ha viszont ez így van, akkor valahogyan jelezni kell tudnunk azt, hogy a normativitás és más cselekvésbefolyásolási jelenségek, mint például a hatalmi helyzetek miben térnek el egymástól. Ennek "kezelésére" valamilyen módon kezelnünk kell a legitimitás, a **legitim érvényesség** kérdését (az érvényesség létét vagy nem létét, illetve annak típusait).

A másik hiányossága a fenti hat-elemű összetevőre bontásnak az, hogy nem veszi figyelembe a normativitáshoz elengedhetetlenül hozzátartozó szankcionálás jelenségét.[29] A normativitásban tetten érhető, jövőre irányuló cselekvésbefolyásolási szándék önmagában

[29] Von Wright ugyan mint fontos jelenséget megemlíti a **szankció** jelenségét, ám azzal a kikötéssel, hogy az nem tartozik a norma összetevői közé [564, 70].

nem elégséges ahhoz, hogy a szándékolt cselekvés megvalósuljon. Ehhez szükség van arra, hogy amennyiben az elvárt cselekvés nem valósul meg, akkor valamilyen szankció kövesse azt. Korábban bemutattam, hogy egy teljes deontikus kalkulust lehet definiálni a szankció fogalmának felvételével, ami már önmagában is elég okot szolgáltat arra, hogy ezt a kategóriát is figyelembe vegyük a norma jelenségének leírásakor. De azért is szükség lehet erre a fogalomra, mert a normasértéshez kötött szankció jelensége biztosítja a talán legfontosabb normatipizálási lehetőséget is egyben. És ha már a norma fogalmi modelljébe beemeljük a szankcionálás mint másodrendű cselekvés fogalmát, akkor meg kell tennünk ezt a normákhoz szükségszerűen kapcsolódó másik cselekvéstípussal, a **kibocsátás** fogalmával is. A norma kibocsátása, illetve a normasértés szankcionálása másodrendű cselekvésnek számít, míg a normához való igazodás, a norma követése, illetve a normasértés cselekvései elsőrendű fogalmakkal ragadhatók meg. Utóbbiakra érdemes egy átfogó fogalmat alkalmazni, hogy könnyebben, egységesen hivatkozni tudjunk rájuk. Erre a célra fogom használni a **normavezérelt cselekvés** kifejezést.

A fenti hat összetevő közül von Wright a modalitást, a tartalmat és a feltételt együttesen minősíti a **normamagnak** (norm-kernel) azzal a tézissel együtt, hogy ez minden előírásnak (preskripciónak) része. A további három összetevőről tehát azt állítja, hogy nem feltétlen találjuk meg őket bizonyos normatípusokban.

A norma tartalma

A **normatartalom** fogalma alatt azt a cselekvést kell értenünk, amire a normatív befolyásolási szándék irányul. Ebben a felfogásban a tartalom valamilyen cselekvés, nem más. Nem világállapot, nem folyamat, nem változás, nem puszta esemény. Cselekvés, mint például a dohányzás, az autóvezetés, a szobába belépés, a köszönés stb. Azt, hogy a cselekvést hogyan lehet leírni a világállapotok, az átmetek, valamint a cselekvésoperátorok segítségével, a könyv egy korábbi fejezetében bemutattam (↪**??**.o.). A normák leírásakor nincs szükség többre, mint, hogy az ott bemutatottak szerint írjuk le a cselekvéseket, és ha így teszünk, egyúttal megadjuk a norma tartalmát is. Fontos figyelni arra, hogy a normatartalom megragadásakor nem vehetjük bele a norma deontikus értékét. A ‘Ebben a helységben tilos a dohányzás!’ norma esetében a norma tartalma “csak” a dohányzás lesz, a tiltás modalitását a norma másik összetevőjével írjuk le.

A norma modalitása

Amit von Wright a norma karaktereként (character of norm) ad meg, azt én a **norma modalitásaként** nevezem. Modalitáson pedig azt a deontikus operátorral kifejezhető normatív szándékot ragadhatjuk meg, amellyel a normát kibocsátó (ellenőrző, számonkérő) személy a norma címzettjét cselekvésre akarja bírni. A deontikus modalitás kérdését alaposan elemeztem a könyvben, bővebb áttekintésekért ezekre a fejezetekre hivatkozom (116, illetve 327. o.).

A norma feltétele

Amilyen fontos fogalom a norma felétele, annyira könnyű félreértelmezni a jelentését. A ‘feltételesség’ sokjelentésű fogalma miatt könnyen zavar alakulhat ki. A **norma feltételének** – legalábbis a von wright-i interpretáció szerint – nincs köze a **norma körülményei-**

hez, ennek ellenére könnyű "összekeverni" őket a két fogalomban "megbúvó" feltételesség mozzanata miatt. Von Wright szerint minden normának van valamilyen **alkalmazási feltétele** (condition of application), ami lényege szerint a norma tartalmától függ. A sokszor idézett példa szerint ha egy normál ajtóval szemben állunk, akkor kiadhatják nekünk azt a parancsot, hogy 'Csukd be az ajtót!', de ennek az a feltétele, hogy *(i)* az ajtó nyitva legyen, és *(ii)* ne csukódjon be magától. Ellenben ha egy magától bezáródó ajtóról van szó, akkor értelmetlen kiadni az előző parancsot, hiszen az világban végbemenő állapotváltozás eredményeként becsukódik az magától. Egy ilyen világállapotban más parancsnak lehet csak értelme, például annak, hogy 'Ne engedd becsukódni az ajtót!'. Fontos hangsúlyozni, hogy minden normához tartozik a fenti értelemben vett alkalmazási feltétel, sőt, pontosabb lenne úgy fogalmazni, hogy tartoznia kell. De természetesen az alkalmazási feltételen kívül lehet még további feltételeket is a normákhoz rendelni, amelyek már nem szükségszerűen kapcsolódnak a normákhoz. Von Wright példája erre a 'Ha esik az eső, csukd be az ablakot!' felszólítás. Egy ilyen feltétel esetlegesen van a normához rendelve – szemben a norma alkalmazási feltételével.

A kategorikus normák esetében az alkalmazási feltétel a norma cselekvéstartalma ismeretében is megállapíthatjuk, ezért mondja azt von Wright, hogy a kategorikus normák esetében elég az alkalmazási feltétel a tartalommal együtt adott [564, 74], ezért nincs szükség annak külön jelzésére a normát leíró logikai formalizmusban. Ezzel szemben a norma további feltételeit – ha vannak – meg kell tudnunk adni, ezért van szükség külön paraméter felvételére a normák összetevőkre bontása során. A hipotetikus normák feltételei tetszőlegesek lehetnek, elméletileg bármely világállapot lehet egy norma feltétele. A cselekvéselmélet számára azonban kiemelten fontosak azok a feltételek, amelyek egy másik cselekvésre, még pontosabban: egy másik ágens cselekvésére hivatkoznak (ilyen például a 'szemet szemért' parancs). Ezek a normák a reciprocitás és komplementaritás társadalmi helyzeteiben kiváltképp gyakoriak és fontosak.

A norma alkalma

Vigyázni kell, hogy a norma alkalmazási körülményét, **alkalmát** (occasion of norm) ne keverjük össze a norma feltételével. Persze mondhatjuk azt, hogy a norma körülménye az a feltételrendszer, amely mellett a normát érvényesnek és követendőnek tartják az érintettek (vagy elgalább azok egy része), de ilyenkor is speciális feltételekre kell gondolnunk. A dolgokat kicsit leegyszerűsítve mondhatjuk, hogy a norma tér-idő meghatározottságáról van itt szó. A körülmény azért is fontos összetevő, mert a fogalmi terjedelmébe tartozó adatok alapján tudjuk tipizálni a normákat a partikularitás/univerzalitás dimenziója mentén (legalábbis részben).

A norma autoritása

A norma mint a jövőbeni cselekvésekre vonatkozó szándékkinyilvánítás logikailag megköveteli, hogy legyen egy olyan ágens, aki a normát kibocsátja. Ezt a társadalmi szereplőt nevezzük a **norma autoritásának**. Vannak normák, amelyek esetében könnyen és egyértelműen meghatározhatjuk a norma kibocsátóját, de előfordulhatnak olyan esetek is, amikor igencsak nehéz megmondani, ki volt az adott norma autoritása. Az erkölcsi parancsok, a szokáserkölcs előírásai szinte mind ilyenek: a kibocsátójuk személye a történelem homályába vész. Részben a norma autoritása alapján válik lehetővé, hogy az

autonómia kérdését meg tudjuk fogalmilag ragadni, és így egy újabb normatipogizálási lehetőséghez jussunk.

A norma címzettje

A **norma címzettje** (subject of norm) az az ágens, akire vonatkozik a normában megtestesülő cselekvésbefolyásolási szándék, akinek a normában előírt cselekvástartalomhoz a norma modalitásértéke szerint kell (vagy lehet) viszonyulnia. Von Wright másfajta megfogalmazásában: a címzett hajtja végre azt a cselekvést, amelyre vonatkozóan az autoritás **toleráns** vagy **intoleráns** viszonyt nyilvánít ki. A címzett lehet partikuláris, egyedi (particular), illetve univerzális, általános (general). Ez a megkülönböztetés később – részben – alapul szolgálhat a normák egyfajta tipizálásához.

A normasértés szankciója

Von Wright, bár megemlítette, mint fontos, a normához kapcsolható jelenséget, de nem vette fel külön a szankcionálás jelenségét, a **társadalmi szankció** fogalmát a norma összetevői közé. Mégha el is fogadom, hogy a szankcióra nem a norma összetevőjeként kell tekinteni, ez nem jelenti, nem jelentheti azt, hogy ne kellene mindenképpen beemelnünk a fogalmat a norma modelljébe. Szankció hiányában nem beszélhetünk normáról, ezért szükség van rá a norma jellemzőinek tárgyalásakor.

Amikor von Wright a normák főbb típusait mutatta be, akkor normaként értelmezte a technikai normákat is, de én ezt az interpretációt nem fogadom el. Azon az alapon, hogy a technikai normák, a főzési receptek, a kezelési, szerelési útmutatók (például az IKEA-bútorok összeszerelését támogató vizuális segédletek), a használati leírások, szoftverkezelési kézikönyvek jelenségében nincs ott a társadalmi szankció mozzanata, ezért nem nevezhetjük ezeket társadalmi normának. Ezt Émile Durkheimhez is így gondolta, ahogy az kiderül a következő idézetből:

> „Ha ügyeimet nem a mesterség szabályainak megfelelően viszem, azt kockáztatom, hogy nem érek el sikert; de a társadalom nem emel szót az ellen, hogy így cselekedjek. Hagyja, hogy cselekvésem teljesen szabadon végbemenjen. Ez a cselekvés nem juthat el azokhoz a célokhoz, amelyek felé törekszik, de ezért még nem tiltatik meg.” (Durkheim, idézi: [376])

Hiába van egyfajta “előírás”, valamiféle mesterségbeli útmutató (technikai norma), és hiába nem követem az abban leírtakat, nem lesz rá társadalmi reakció. Ez nem jelenti azt, hogy ne lesz semmi következménye az útmutatótól való eltérésnek. Lehet, hiszen a technikai norma követése által remélt célt esetleg így nem érem el. De az eltérést ilyen esetekben nem követi szankció. Ha nem a recept szerint főzök, akkor ehetetlen lesz az étel, és ez következményként lehet igen kellemetlen, de társadalmi szankcióra nem kell számítanom. Nagyon fontos tehát, hogy a technikai előírások megsértéséből származó következményeket ne azonosítsuk, ne keverjük össze azzal, amikor a normasértést társadalmi szankció követi.

> „Ha megsértem a higiénia szabályait, amelyek azt parancsolják nekem, hogy tartózkodjam a gyanús érintkezésektől, e cselekedetet automatikusan követi a káros következmény, ... a betegség. Az elkövetett cselekmény magától váltja ki a következményt.” [138, 172]

Ezzel szemben a normasértés mindig valamilyen társadalmi helytelenítéssel találkozik, és a társadalom (közösség) vagy igyekszik megakadályozni a normasértőnek ígérkező cse-

lekvést, vagy - ha ez már nem sikerült, akkor – valamilyen szankcióval sújtja a normától való eltérést.

> „Ha egy cselekvés, amelynek természeténél fogva egy morális szabályhoz kell alkalmazkodnia, attól eltér, a társadalom, ha erről tájékoztatva van, közbelép, hogy megakadályozza ezt az eltérést. ...Ez a társadalmi reakció igazi szükségszerűséggel követi a normasértést; gyakran még módjában is előre meghatározott." (Durkheimet idézi: [376])

A normasértés helytelenítése és szankcionálása elválaszthatatlan a norma jelenségétől, de persze mindkét jellemzőt (a helytelenítést, illetve a szankcionálást) megfelelő módon kell értelmeznünk. Nem követelhetjük meg azt, hogy akár a helytelenítés, akár a szankcionálás konszenzuális legyen, vagyis mindenki egyetértsen, illetve részt vegyen ezekben. Azt sem állíthatjuk, hogy minden normasértést szankció követne, sőt, ki kell mondanunk egy fontos tételt: amilyen mértékben előzetesen várható a szankcionálás tényleges bekövetkezése, olyan mértékben fognak az emberek igazodni a normákba foglalt előírásokhoz. Az is fontos kérdés, hogy a szankcionálásnak milyen eszközei, lehetőségei vannak, de ezt a normák tipizálásakor fogom megválaszolni (és a típusokat bemutatni).

A norma érvényessége

A norma érvényessége alatt legitim érvényességet értek. Ezt a kérdéskörét az uralomról szóló fejezetben tárgyalom (↰400.o.).

A normához való igazodás

A normavezérelt magatartásnak két fő típusa van: a normasértő vagy a normához igazodó cselekvés. A normasértésről már szóltam, most a normakonform magatartás jellemzésén van a sor. Előtte azonban tisztáznom kell valamit. A normához való igazodás nem feltétlen vonzza maga után a normakövető magatartást. Mind elméletileg, mind gyakorlatilag lényegi és szoros kapcsolatot tételezhetünk a normakövetés és a normához való igazodás között, de ez nem jelent kizárólagosságot. Csontos László a következő tényezőket sorolja fel a normakövetés magyarázata során:

> „A normakövető egyéni cselekvés magyarázata során - durván - négy lehetséges mechanizmusra hivatkozhatunk: (1) a normától eltérő vagy normasértő (deviáns) viselkedés szankcionálása; (2) az egyéni stratégiaválasztástól függő nyeremények pozitiv interdependenciája (a normakövető magatartás - pozitív - externális hatása); (3) az egyéni preferenciákba épített konformizmus; és végül (4) ex ante kommunikáció az ex post normakövető csoport tagjai között." [113]

A pontosabb értelmezés érdekében bemutatok egy hétköznapi példát arra vonatkozóan, hogy milyen cselekvési stratégiákat lehet követni a nagyvárosi tömegközlekedési eszközök igénybevétele során. Az alapnorma úgy szól, hogy minden használatért fizetni kell (bérlet- vagy jegyvásárlással), a normasértés esetén pedig büntetést kell fizetni. Ebben a helyzetben az alábbi stratégiákra könnyedén találhatnánk valós példákat az utazóközönség tagjai közt:

Az 1. és 2. esetben az a közös, hogy az ágens érvényesnek tartja a normát, és igazodik hozzá, ezért fizet a szolgáltatás igénybevételéért. A két eset közti kölönbség pedig abban van, hogy az utazási szokások ismerete alapján vagy a bérlet- vagy a jegyvásárlás kerül kevesebbe. Egyszerű kalkulációval kiszámolható, hogy hány utazás után éri meg jobban a bérletet megvásárolni. Ha valaki általában autóval jár, és nagy ritkán száll csak

1.) valaki havi bérletet vált
2.) valaki minden utazási alkalommal előre jegyet vesz és érvényesíti
3.) valaki minden utazási alkalommal előre jegyet vesz, de csak akkor érvényesíti, ha ellenőrökkel találkozik
4.) valaki nem vesz sem jegyet, sem bérletet, de ha a járművön találkozik a kalauzzal, tőle vesz jegyet (ilyen rendszer működik a Szentendrei HÉV-en)
5.) valaki nem vesz sem jegyet, sem bérletet, ha ellenőrök elkapják, büntetést fizet

át tömegközlekedési eszközre, akkor megérheti az egyedi jegyvásárlás, különben nem nagyon. Más a helyzet a 3. eset ágensével, aki igazodik az alapnormához, de belsőleg nem tartja azt érvényesnek, csak a büntetéssel való külső fenyegetés tényleges létezése esetén cselekszik normakonform módon. Azzal, hogy az ellenőrök felbukkanása esetén érvényesíti a jegyét, öntudatlanul is "mutatja", hogy igazodik az alapnormához, de a normakonformitása nyilván csak olyan mértékű, amilyen mértékű az ellenőrzés gyakorisága. Amennyiben az ellenőrzések gyakorisága annyira alacsony, hogy az ilyenkor érvényesített jegyekért összesen kevesebbet kell fizetni havonta, mint amennyi a havi bérlet ára, akkor ez a "módszer" megérheti – csak racionális kalkuláció alaján. A 4. és 5. esetben a normakonformitás nagyon gyenge, ha van egyáltalán. A 4. eset annyiban különbözik a 3.-tól, hogy az előbbiben az utazási szabályzat jogszerűvé teszi azt a magatartásmintát, ami az utóbbiban jogszerűtlen (hogy ti. valaki úgy szálljon fel a járműre, hogy előzetesen jegyet vegyen). A körülményekben bekövetkezett változás lehetővé teszi a teljes racionalizálást, vagy más oldalról tekintve a teljes "normátlanítást", mert ebben az esetben a normatív vonatkozásoktól teljesen eltekintve csak anyagi mérlegelés kérdése, hogy valakinek a bérlet vásárlása vagy a jegyek időnkénti kifizetése kerül-e többe. Az 5. eset annyiban megegyezik a 4.-kel, hogy mindkettőt a "csak racionális" mérlegelés dominálja. Az 5. eset ágense kalkulálhat úgy magában, hogy ha statisztikai értelemben kifizetődőbb a ritkán előforduló büntetéseket kifizetni, mint folyamatosan bérleteket, jegyeket venni, akkor megteszi ezt a lépést akkor is, ha ennek a normasértő magatartást övező társadalmi helytelenítés az ára (esetleg ezeket a pszichológiai költségeket is beárazhatja). Láthatjuk tehát, hogy a racionális mérlegelésen alapuló érdekkövetés és a – norma érvényességének belső elfogadása esetén – nem mérlegelésen alapuló normakövetés, mint két szélső pólus között számtalan átmeneti, kevert eset is elképzelhető.

Amíg az ember nem gondol bele, talán paradoxonnak tűnhet azt állítani, hogy a norma társadalmi elfogadottsága, követése pozitív kapcsolatban állhat a normasértés jelenségével. Pedig Durkheim szerint a normasértés indirekt módon igenis hozzájárulhat a normakövetés erősödéséhez és ezáltal a társadalmi rend – legalább részleges – megszilárdulásához:

> „A kollektív érzelmek ugyanis, amelyeket a történelem adott időszakában egy nép büntetőjoga védelmez, csak akkor hatolhatnak be azokba az emlékekbe, amelyek eladdig zárva voltak előttük, s csak akkor tehetnek szert akkora hatalomra, amellyel addig nem rendelkeztek, ha a korábbinál nagyobb lesz az intenzitásuk. Az egész közösségnek erősebben kell éreznie ezeket az érzelmeket. Csakis innen meríthetnek ugyanis ezek az érzelmek olyan, a korábbbinál nagyobb erőt, amelynek révén rákényszeríthetik magukat azokra az egyénekre is, akik ezt megelőzően leginkább ellenük szegültek. A gyilkosok csak akkor tűnhetnek el, ha a vérontástól való irtózás a korábbinál jóval erősebbé válik azokban a társadalmi rétegekben is, ahonnét a gyilkosok rendszerint kikerülnek." [139, 87]

Rögtön ezek után persze Durkheim hozzáteszi azt is, hogy a kollektív érzelmek megerősödése, amely – jó esetben – valamely bűntípus visszaszorulását, esetleg eltűnését eredményezi, "hozzájárul" ahhoz, hogy más területeken újfajta bűncselekmények "jelenjenek meg". Durkheim persze ennek is megadja a magyarázatát:

> „. . . a közös tudat ilyen intenzív állapotai nem erősödhetnek meg anélkül, hogy ne erősítenék fel azokat a gyengébb állapotokat, amelyek áthágása addig csupán erkölcsi vétségnek számított. . . . [a kollektív] tudat sokkal érzékenyebben reagál majd azokra a sérelmekre, amelyek korábban alig érintették: erőteljesebben lép fel ellenük, jobban elítéli őket, s némelyiküket egyszerű erkölcsi fogyatékosságból bűncselekménnyé változtatja." [139, 88]

Ezen a ponton kiváltképp "tetten érhető" a normavezérelt magatartás mélyén kitapintható, elemi erővel hatni képes értékalapja, érzelmi töltete. Eddig erre mutattam példát, hogy a normakövetés és normához való igazodás nem feltétlen esik egybe. Arra is figyelnünk érdemes azonban, hogy nem minden normális, nem minden szabályszerűnek tűnő cselekvést magyarázhatunk valamilyen mögöttes normához való igazodással. Rorbert B. Cialdini és szerzőtársai megkülönböztetnek kétféle kategóriát, a **leíró norma** (descriptive norm) és **rendelkező norma** (injunctive norm) fogalmát [99]. A leíró norma fogalma alatt ők elsősorban arra gondolnak, hogy ha valamilyen mintát sokan követnek, azt mindenkinek érdemes lesz követni, ezért az normává válik. Az én értelmezésemben Cialdiniék leíró normája nem igazán norma, hiszen nincs mögötte szankcionáló erő. Ezt nevezhetjüük hálózati hatásnak, közös tudásnak, szabályszerűségnek, de normának nem. Szankció nélkül nincs norma.

A norma másodrendű kontextusa

A von wright-i normaösszetevők kiegészítését akkor érthetjük meg igazán, ha látjuk (és a modellünkben feltüntetjük), hogy a **normavezérelt cselekvések** sosem magukban állnak, így nem is értelmezhetők magukban. Az elemzés központjába természetesen a normavezérelt cselekvést kell helyeznünk, de emellett figyelembe kell vennünk még két másik cselekvéstípust is. A normavezérelt cselekvést a norma címzettje hajtja végre. Ennek során vagy igazodik a normába foglalt szándékhoz vagy megsérti a normatív elvárást.

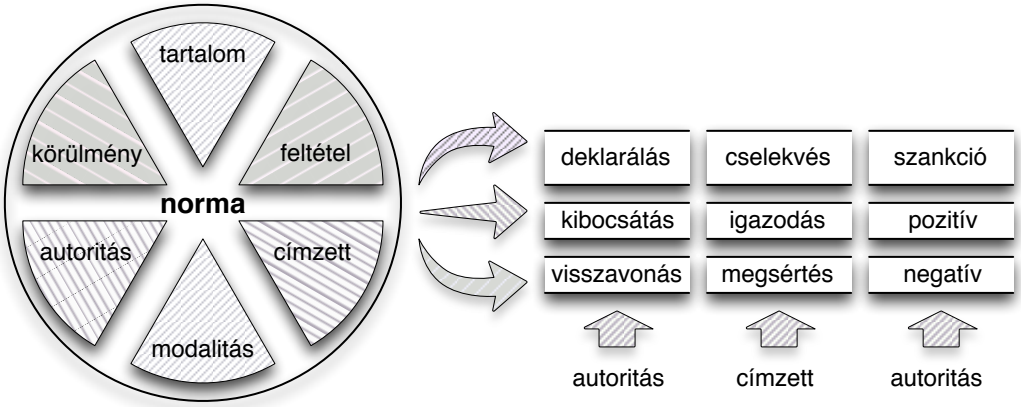
A normához való igazodáshoz azonban meg kell előzze az a cselekvés, amelynek során a normát magát kibocsátja valaki.[30] Ha viszont azt a szempontot is figyelembe vesszük, hogy a normákat nemcsak **kibocsátani** lehet, de megszüntetni vagy **visszavonni** is, akkor szükség van egy olyan általánosabb fogalomra, amely mind a kibocsátás, mind a visszavonás aktusát magába foglalja. Erre alkalmas lehet a **deklarálás** fogalma. A deklarálás tevékenységét az autoritáshoz köthetjük, és fontos kényszerként rögzíthetjük, hogy ennek a cselekvésnek időben meg kell előznie a normavezérelt cselekvést magát. Mivel pedig a norma deklarálása olyan cselekvés, amely más cselekvést magukba foglaló normákra vonatkozik, így logikai minőségét tekintve a deklarálást másodrendű cselekvésfogalomnak kell tekintenünk.

A normavezérelt cselekvéshez tartozik még a szankcionálás aktusa is, ami a cselekvést követően, tehát ahhoz képest időben később valósul meg, de fogalmilag ugyanúgy elválaszthatatlanul kapcsolódik hozzá, mint a a kibocsátás cselekvése. A szankcionálás is másodrendű, hiszen mindig csak valamilyen elsőrendű (alacsonyabbrendű) cselekvésre hivatkozva értelmezhető, önmagában vett szankcióról sosem beszélhetünk. A szankció

[30] Itt most újra eltekintek attól a problémás kérdéstől, hogy ki bocsátja ki a morális normákat.

lehet pozitív, valamilyen jutalom, illetve negatív, valamilyen büntetés. A szankcionálást az autoritás végzi el, de fontos látni, hogy az autoritás szerepfogalom, és nem feltétlen egyezik meg az a két ágens, amely egy adott normával kapcsolatos kibocsátó, illetve szankcionáló szerepeket viszi.

A normavezérelt cselekvés másodrendű kontextusán a fent bemutatott másodrendű cselekvéseket értem. A normatív jelenségek értelmezéséhez a norma összetevői mellé fel kell venni ezeket a jelemzőket is.



A norma összetevőinek bemutatása után belevághatunk a norma típusainak tárgyalásába. Ezt többszörös tipizálást lényegleg a normaösszetevőkre támaszkodva végzem majd el.

6.5 Normatípusok

A norma összetevői és egyéb jellemzői alapján már módunk nyílik arra, hogy a normákat különböző szempontok szerint tipologizáljuk. Bemutatom, hogy majdnem az összes normakompenens szerint lehet értelmesen normatípusokat képezni, de előfordul majd olyan felosztás is, amelyben egyszerre több összetevőt figyelembe véve lehet majd alábontani a norma fogalmát.

Normatípusok a szankció típusa szerint

A normavezérelt cselekvés következménye valamilyen joghatás lesz. Joghatás alatt érthetjük a szankció bekövetkeztét is, nem részletezve azt, mit is érünk szankció alatt. Még azt is mondhatnánk, hogy a szankció lehet pozitív vagy negatív, tehát beszélhetünk jutalomról és büntetésről. A tényleges társadalmi gyakorlatok ismeretében azonban nem tévedünk nagyot, ha a szankció fogalma alatt itt csak a büntetést értjük, és a talán legfontosabb három normatípust a rájuk jellemző szankciók különbsége alapján határozzuk meg. Ezt a hármas tipológiát Max Webertől vehetjük át, akitől most egy kicsit hosszabban idézek:

„6.§. Valamely rend legitimitása biztosítható:
I. tisztán belülről, mégpedig
1. tisztán indulati-érzelmi alapon: érzelmi átéléssel, belső odaadással;
2. értékracionálisan: a végső és kötelező értékek (erkölcsi, esztétikai vagy bármely más értékek) kifejeződéseként fölfogott rend abszolút érvényességébe vetett hittel;
3. vallásos hittel: azzal a meggyőződéssel, hogy a rend betartásán keresztül el lehet jutni az üdvözüléshez;
II. továbbá úgy is (vagy csak úgy), hogy sajátos külső következményekhez kapcsolódnak bizonyos várakozások.
A rendet
a) konvenciónak nevezzük, ha érvényessége külsődlegesen, azáltal van biztosítva, hogy az emberek megadható körén belül a rendtől való eltérés várhatóan (viszonylag) általános és gyakorlatilag érezhető helytelenítésbe ütközik;
b) jognak nevezzük, ha ugyancsak külsődlegesen, de azáltal van biztosítva, hogy az emberek egy külön erre fölkészült csoportja várhatóan (fizikai vagy pszichikai) kényszert alkalmaz a rend betartásának kikényszerítésére, illetve megszegésének megtorlására." [575, 60-61]

Ahogy az idézetből kiderül, Weber először elkülöníti egymástól azokat a normatív cselekvéseket, amelyek során a cselekvők belső, illetve külső tényezők miatt igazodnak a normához. A belső normakövetéshez rendeli hozzá az erkölcs kategóriáját. A külsődleges hatásokra történő normakövetés során a cselekvő társadalmi környezete, a közösség valamilyen befolyása érvényesül. Az erkölcsi normát megsértő cselekvés belső folyamatokat indít el az emberben (önmarcangolás, lelkiismeret-furdalás). A konvenció és a jog megsértése külsődleges büntetéssel jár. A konvenció esetében normasértéskor a helytelenítésen keresztül közösségi nyomás nehezedik az egyénre (aminek legszélsőségesebb esetében a közösség kizárja soraiból a normát megsértő egyént), a jogsértés pedig az adott közösség lelki vagy fizikai erőszakon alapuló szankcióit vonja maga után. Utóbbi általában úgy valósul meg a mindennapokban, hogy a közösség vezetése a börtönbe záráshoz szükséges fizikai erőszak-alkalmazás lehetőségére hivatkozik egy kifejezetten erre szakosodott apparátus segítségével – amely egyébként a folyamat végén el is végzi az olykor valóban szükséges erőszak-alkalmazást.

A normasértésekhez rendelhető szankciók két szempont figyelembevételével írhatók le. Egyfelől tudni kell, hogy ki szabja ki a szankciót a normasértőre, másfelől azt kell számításba venni, hogy milyen szankciót lehet kiszabni. Az első szempont szerint külső vagy belső szankcionálásról beszélhetünk. A második szempont szerint normasértés esetén háromféle lehet a büntetés: az egyénben ébredő lelkiismeret-furdalás, vagy a normafenntartó közösség által alkalmazott lelki, illetve fizikai kényszer (vagy az azzal való fenyegetés). Mindezek alapján megkapjuk Weber három alapvető normatípusát:

- **erkölcs** (belső szankció: az egyén lelkiismeret-furdalása)
- **konvenció** (külső szankció: a közösség helytelenítése, kritikája)
- **jog** (külső szankció: az erőszak-alkalmazással való fenyegetés)

A három szankciótípus és persze a három normatípus mindegyike vonatkozhat ugyanarra a cselekvésre, ezért minden egyes konkrét társadalmi kapcsolat elemzésekor meg kell mondani, hogy éppen akkor melyik normatípus érvényesül a leginkább. Előfordulhat ugyanis, hogy egy társadalmi helyzeten belül egyidejűleg többféle normatípus áll fenn vagy érvényesül. Betarthatom én a sebességkorlátot azért, mert *(i)* látom az út szélén a rendőrt (tehát a joghoz igazodom), *(ii)* a feleségem szúrós szemekkel tekinget felém

(konvencionális nyomást fejt ki rám), illetve *(iii)* sajátmagam belső késztetésére (rosszul érezném magam, ha normát sértenék).

Szükségesnek tartok itt pár további megjegyzést tenni. Először is fontos rögzíteni, hogy Weber konvenciófogalma jelentős mértékben eltér attól az értelmezéstől, ami David Lewis 1969-ben megjelent könyve óta elterjedt a játékelmélettel foglakozó kutatók körében [305]. Lewis a koordinációs helyzetekben jelentkező teljesen vagy közel egyenrangú döntési alternatívák közti kollektív választás egyik eszközének tekinti a konvenciót. A koordinációs helyzetek döntetlen jellegét oldja fel az, hogy valamilyen módon (szaliencia-, fókuszpont- vagy precedenstechnikákra támaszkodva) konvenciók alakulnak ki egy közösségen belül.[31]

A másik megjegyzésem Weber erkölcsfogalmára vonatkozik. Először is rögzítem, hogy a könyvemben egymás szinonimájaként kezelem az erkölcs, az etika és a morál terminusokat. Akármelyik terminust is használjuk, azt is látni kell, hogy a weberi definíció se nem igazán reflektált, se nem igazán hasonlít a megszokott, széles körben elterjedt erkölcsfelfogáshoz. Reflektálatlanságon azt értem, hogy ritkán azonosítják erkölcsként a fenti weberi idézet I. pontjában leírtakat, de szokatlan az a definíció is, amely az erkölcsöt a belső lelkiismerettel támogatott normával azonosítja.[32] Pedig maga Weber erről a hármasságról ír könyvének más oldalán.

> „A »kívülről« biztosított rendnek azért még »belső« garanciái lehetnek. A szociológia számára nem probléma a jog, a konvenció és az »etika« kapcsolata. Az »etikai« mérce itt azt jelenti, hogy az emberek bizonyos speciális értékracionális hite normaként szolgál az olyan emberi cselekvés számára, amely az »erkölcsi jó« predikátumára tart igényt." [575, 62]

Ebben a gondolatmenetben már egy szinten, egyenrangúan (és kizárólagos módon) szerepel a rend három típusa, a jog, a konvenció és az etika, csak az nincs megfogalmazva, hogy az etikához (erkölcshöz) kapcsolható speciális szankció a lelkiismeretfurdalás. Amikor az alapvető cselekvéstípusok bemutatás végén felrajzoltam egy ábrát (↪231.o.), amivel az alapbeállítódások (cél- és értékracionalitás, indulat, habitualitás) egymáshoz való viszonyát próbáltam érzékeltetni, akkor a normavezérelt cselekvést úgy jelöltem, mint ami egyszerre "támaszkodik" az indulati-érzelmi, illetve az értékracionális cselekvéstípusokat jellemző beállítódásra, belső mechanizmusokra. A normák itt megadott – a szankcionálási mechanizmus különbségeire épített – hármas felosztása során már bővebben értelmeztem a norma fogalmát a korábbiakhoz képest. Amikor a társadalmi cselekvések típusait elemeztem, akkor még csak a másoktól független egyénre, egyéni cselekvésre figyeltem. Az egyes cselekvéstípusok elkülönítésekor is csak a belső motivációkra, belső hatásokra támaszkodnak. Most viszont már a külső befolyásolás lehetőségt is figyelembe vettem a normák tipizálása során. A konvenció és a jog fogalmaival már olyan normativitást ragadhatunk meg, amelyben a normához való igazodást külső szankciók figyelembe vételével magyarázhatjuk meg. Mit is jelent ez a cselekvéstípusokra nézve? Azt, hogy ezek a "sajátos külső következményekhez kapcsolódó várakozások" racionális, mérlegelő magatartáson keresztül érvényesülnek. Vagyis amilyen mértékben beszélhetünk a konvencióhoz vagy joghoz igazodó cselekvésről, olyan mértékben válik meghatározóvá a racionális beállítódás, vagyis a racionális megfontolásokkal magyarázható normakövetés. Természetesen mindig elképzelhető, hogy valakiben mindhárom szankcionálási mechanizmus egyszerre van jelen, amikoris azt mondhatjuk, hogy az illető voltaképpen mind-

[31] A david lewis-i értelemben vett konveciófelfogásról ír magyarul Pete Krisztián[411].

[32] Az erkölcsről magyarul olvasható Orthmayr Imre tanulmánya [397].

három normatípushoz igazodik a cselekvésében (vagyis egyszerre erkölcsi, jogkövető vagy konvencionális cselekvést hajt végre). De az is könnyen elképzelhető, hogy egy normához igazodó cselekvést csak hideg mérlegelés alapján hajt végre, és semmi belső motiváció nincs benne a norma követése iránt (például az élvezetből gyorshajtó ember csak akkor tartja be a sebességkorlátokat, amikor komoly esélyét látja a lebukásnak – ez a színtiszta racionális alapon álló normakövetés példája).

Normatípusok a logikai rend szerint

A norma tartalma és szankciója alapján **elsőrendű** vagy **magasabbrendű normákat** különíthetünk el egymástól. Elméletileg tetszőlegesen magas rendű normát el lehet képzelni, de a gyakorlatban vélhetőleg harmadrendű normáknál nem nagyon találunk magasabbrendűeket. A szakirodalomban az első- és másodrendű normákkal sokat foglalkoznak, harmadrendűekkel már jóval kevesebbet, és arról nem tudok, hogy a harmadrend fölött bármilyen elemzés született volna. Pedig azt gyanítom, hogy egy alaposabb elemzés kideríthetné, hogy – főleg, mióta az európai uniós jogkörnyezetbe kerültünk – előfordulnak negyed-, netán ötödrendű normák tényleges joggyakorlatban, csak ez – már, ha tényleg létezik – azért nem okoz gondot, mert az adott kontextusban, a tényleges jogi környezetben "közelről" nem lehet érzékelni a magasabb rendbe való kerülést. De nem is az az igazán fontos kérdés, hogy milyen másodrendnél magasabb normák vannak (egyáltalán vannak-e), hanem sokkal inkább azt kell látnunk, hogy hol, milyen másodrendű normákat találhatunk.

Von Wright a normáról szóló könyve egyik fejezetét a magasabbrendű normáknak szenteli [564, 189-207]. Itt vezeti be a kompetencianorma fogalmát, amire már utaltam a hohfeldi elmélet elemzésekor (↪363.o.). A fogalom definíciójából következik, hogy a kompetencianormát (a hohfeldi felhatalmazottságot) delegáló normáknak is tekinthetjük, ami egyfajta normahierarchia kialakítására és kezelésére lehet alkalmas. Jog esetében az állam az alárendelt hatóságok számára átengedhet felhatalmazottságot, hogy így a közjót jobban ki tudja szolgálni.

> „Előfordul, hogy a legfelsőbb hatóság bizonyos dolgokat szabályozni akar, mondjuk úgy, a közjó érdekében. Ilyenkor a jogszabályi részletek kidolgozását olyan alacsonyabb szintű hatóságra hagyhatja, amely megfelelő ismeretekkel rendelkezik a cél, az adott közjó megvalósításához szükséges speciális követelményekről." [564, 193]

Persze, von Wright nem volt se jogász, se szociológus, se antropológus, így a társadalmi cselekvések elemzésével kapcsolatban nem tudott, nem is akart továbblépni. A magasabbrendű normatív fogalmakat a jogviszonyok, jogrendszer elemzése kapcsán először jogászok bontották ki alaposabban. Már Hans Kelsen is beszélt első- és másodrendű jogszabályokról, de ő nem logikai értelmezés szerint sorolta ilyen típusokba a jogszabályokat, hanem a jogrendszer működési logikája szerinti fontosságot tekintette meghatározónak [276]. Ezért ő a logikai interpretációhoz képest pont fordítva minősítette a jogszabályokat, és a bírókat az ítélkezési folyamat során kötelező szabályokat nevezte elsőrendűnek, míg az egyszerű állampolgároknak, a jog alanyainak tetszőleges, mindennapi tevékenységét szabályozó normákat tekintette másodrendűeknek. Ez a megközelítés pont az ellentéte a logikai-fogalmi gondolkodásnak. Természetesen itt csak eltérő terminológiai választásról van szó, fogalmi hibáról nem beszélhetünk.

Amikor H.L.A. Hart elsődleges és másodlagos szabályokról írt, akkor ő már a szokásos logikai interpretációnak megfelelő szemléletben osztotta fel a jogszabályokat [228]. A

három híres másodrendű szabálya közül az **elismerési szabály** abban az értelemben különbözik a másik kettőtől, az **ítélkezési** és a **változtatási szabályoktól**, hogy az előbbi a jog egészének érvényességével kapcsolatos, míg az utóbbi kettő a jogrendszer fontos szereplőinek tevékenységére vonatkozik. A bírónak, rendőrnek másodrendű kötelessége van a szankciókat kiosztani a normasértések esetén, az ítélkezési szabályok a rájuk vonatkozó jogviszonyokat szabályozzák. A törvényhozónak, a jogalkotónak magasabbrendű joga (felhatalmazottsága) van az elsőrendű (olykor a másodrendű) jogviszonyokat megváltoztatni.

A felhatalmazottságról szóló fejezetben már jeleztem, hogy a jog elemzése során kialakított jogviszonyfogalmak, az azokkal kapcsolatos tanulságok általánosíthatók és kiterjeszthetők az általánosabb normafogalmakra is. A legkülönfélébb társadalmi intézményekben, a családban, a munkahelyeken, az iskolákban, a társadalmi, kulturális, politikai szervezetekben mindenhol találhatunk elsőrendű szabályokat, normákat, amelyek a résztvevők mindennapi cselekvéseit szabályozzák, és vannak másodrendű normák, szabályok, amelyek az éppen aktuális normarend megváltoztatásának módját határozzák meg. Ezek lehetnek konvenciók, lehetnek szervezeti szabályzatok mint gyenge jogszabályok, de lehetnek akár morális szabályok is. A másodrendű szabályoknak pedig minden esetben, minden területen ugyanazok a funkciói, mint amit már a jog esetében korábban megismertünk: elismerni (eredeztetni), változatni, ítélkezni (fenntartani).

Azt jeleztem ugyan, hogy léteznek a másodrendű szabályoknál magasabbrendűek is, de eddig igazából csak a másodrendű normákról beszéltem. Bár a másodrendű szint alapvető fontosságát (elsődlegességét) nem akarom kétségbe vonni, szükségesnek tartom jelezni, hogy minden normatív rendszerben vannak, kellenek, hogy legyenek harmadrendű szabályok. Ezért bemutatok egy egyszerű, hétköznapi példát a harmadrendű norma létezésére, működésére. Egy családnál gyerekzsúrt tartanak, és a kisgyerekeket egyedül hagyják egész délután a nagyszobában játszani. Este, miután minden vendég elment, az anyuka észreveszi, hogy a csokitorta egy részét valaki belepaskolta a drága perzsaszőnyegbe. Amikor kérdőre vonja ezért a kisfiát, a gyerek azzal vág vissza, hogy nem ő volt, aki ezt csinálta, hanem Zsoltika. Az első megszólalásával az anyuka másodrendű cselekvést hajt végre, hiszen elsőre szóbeli feddésben részesíti vagyis szankcionálja a fiát. Amikor azonban kiderül, hogy az elsőrendű cselekvés szintjén nem a fia volt a rendetlen (a normasértő), akkor az anyuka visszakérdez: 'Miért nem szóltál rá Zsoltikára, hogy nem szabad ilyet csinálni?' Ez a feddő kérdés már harmadrendű cselekvésnek számít, hiszen az anya feddése annak a mulasztásnak szól, hogy a fia nem szankcionálta (másodrendű cselekvés) a vendégét, aki rosszalkodott (elsőrendű cselekvés).

Normatípusok a deontikus modalitás szerint

A norma modalitása alapján is adott egy természetes tipizálási lehetősége, hiszen a deontikus logikai kategóriák alapján a normákat élesen elkülöníthetjük egymástól. Attól függően, hogy milyen deontikus logikai operátorokat veszünk fel a deontikus logikai modellünkbe, beszélhetünk három, négy vagy hat modalitás szerinti normatípusról (úgymint: **kötelező, megengedő, tiltó, mellőzhető, opcionális vagy preskriptív normákról**).

Ahhoz képest, hogy eldöntjük, hányféle deontikus modalitással akarjuk kezelni a normatív teret, illetve melyek ezek, az már egy másik kérdés, hogy ezeket a deontikus operátorokat fogalmilag, logikailag hogyan írjuk le, és egy formális ontológiába miket

hogyan veszünk fel a természetes nyelvi szinten használt kifejezésekből. Von Wright a kötelezést és a tiltást egymásból levezethetőnek tartja, ezért elegendőnek tartja csak az egyiket felvenni a modelljébe [564, 71]. Ezt a kötelező normával (O, O-NORM) teszi meg. A megengedés aktusát viszont – jelezve annak vitatott státusát – önállóan kezelendőnek ítéli, és ezért külön felveszi saját ontológiájába a permisszív normát (P, P-NORM). Mások szerint a logikai szinten elegendő egyetlen deontikus operátor, és minden más egyéb levezethető belőle (bővebben erről (↪116.o.).

A normák deontikus modalitás szerinti tipizálásában érdemes Hohfeld elméletét, fogalmait is figyelembe venni. Bár a hohfeldi másodrendű jogfogalmak deontikus logikai minőségüket nem jelentenek új típusokat, gyakorlati jelentőségük miatt érdemes felvenni őket a normák deontikus modalitás szerinti osztályozása során. Eközben persze azt is figyelembe kell venünk, hogy a két másodrendű hohfeldi fogalmat, a felhatalmazottságot és a mentességet nemcsak itt használhatjuk önálló altípusként, de a normák logikai rend szerint csoportosításakor is.

A klasszifikációhoz ugyan nem járul hozzá a Hohfeld által megállapított jogok és kötelezettségek közti komplementaritás figyelembe vétele, hiszen a hohfeldi korrelativitás csak a normatípus közti viszonyokat ragadja meg, nem ad új alosztályt. De a korrelatív kapcsolatban levő normaviszonyok mégiscsak a deontikus modalitás szerinti normatípusokat kapcsolják össze egymással, így a korrelativitás "követése" révén további tudást lehetséges a normatív viszonyok

Normatípusok a tartalom szerint

A logikai elemzésekben nemigen találhatunk olyan kísérleteket, amelyek a norma tartalma alapján tipizálná a normákat. Vélhetőleg azért nem, mert az ilyen vizsgálat kivezet a logika területéről. Ebből a szempontból tekintve nyilván sokféle tipizálás elképzelhető. A legtermékenyebbnek a játékelmélet játékosztályai szerint elvégezhető tipizálási lehetőségeket érzem, amit nagy vonalakban be is mutattam a játékelméleti fejezetben (↪269.o.). Robinson és Goforth mind a 144 kétszemélyes alapjátékot nagyobb játékosztályokba sorolta azok "geometriai jellemzői" alapján [443]. Ugyancsak játékelméleti alapokon állva Edna Ullmann-Margalit az alapjátszmákat három nagy osztályba sorolva be elkülönítette a **koordinációs normákat** (coordination norms), a **fogolydilemma normákat** (Prisoners' Dilemma norms) és a **részrehajló normákat** (inequality or partiality norms) egymástól [550]. A dawid lewisi konvencióértelmezés [305] nyomán kialakuló diskurzusban a koordinációs problémák két fő osztályát a tiszta koordinációs és az antikoordinációs helyzeteket feloldó konvenciók mentén lehet egyfajta normatipizálást elvégezni. Robert Sugden három konveciótpípust határolt el, amikor elkülönítette egymástól a koordinációs helyzetekkel, a tulajdonnal, valamint a reciprocitással kapcsolatos konvenciókat [512]. A normák egyik fontos tartalmi altípusának kell tekintenünk a méltányossági normákat, melyek kialakulásáról Ken Binmore cikke ad érdekes áttekintést [55].

Normatípusok az érvényesség alapján

A normákat – a jogi normákat biztosan, de a konvencionális és morális normákat is, mégha kicsit nehezebben – tipizálhatjuk az érvényesség, pontosabban a hatályba lépés szempontja szerint. Egy norma lehet **hatályos**, illetve **hatályon kívüli**. Hatályos norma esetén az autoritás (a kibocsátói vagy a szankcionáló szerepben) a legitim érvényességre

hivatkozva igényli a normához való igazodást a címzettektől, míg a hatályon kívüli norma esetében nincs ilyen igény.

Normatípusok az autoritás alapján

Fogalmilag már kivezet a módszertani individualizmus világából, ha az autoritástípusok alapján próbáljuk meg tipizálni a normákat, ám a komoly szociológiai relevanciája miatt mindenképpen be kell mutatnom ezt a klasszifikációs sémát. Ha a széles értelemben vett jogra gondolunk, arra a tételes normarendre, amit emberek hoznak saját életük szabályozására ideértve mind az egységes és centralizált állami jogrendszer, mind az osztottabb, korlátozottabb hatókörű, ám nagyobb számosságú szervezeti világ működési szabályait, akkor a normák kibocsátói szerint elkülöníthetjük az alábbi normatípusokat: **törvény, rendelet, ítélet, szabályzat**. Az elkülönítés alapja az, hogy milyen típusú szerv a norma kibocsátója. Ha a törvényhozó szerv, akkor a norma **törvény**, ha államigazgatási hatóság, valamilyen végrehajtó szerv az autoritás:, akkor a norma **rendelet**, ha bírósági szerv az autoritás, akkor a norma bírósági **ítélet**, és ha az autoritás nem egy államigazgatási szerv, hanem valamilyen gazdasági, társadalmi, politikai, de államtól független szervezet, akkor a norma szervezeti és működési **szabályzat**.

Normatípusok a címzett és autoritás viszonya alapján

A norma címzettje és autoritása az a két társadalmi szerep, amely mentén az ágenseket a normákhoz kapcsolhatjuk. E két szerep egymáshoz való viszonya alapján egy olyan felosztáshoz juthatunk, amely egy jól ismert és nagyon széles körben használt kettősséget eredményez, jelesül az **autonóm** (autonomous norm), illetve a **heteronóm norma** (heteronomous norm) dichotómiáját [564, 76]. Ha a norma kibocsátója, tehát az autoritás és a norma címzettje megegyezik egymással, akkor beszélhetünk autonómiáról, autonóm normáról, ellenkező esetben pedig a norma heteronóm.

Ez a felosztás jól alkalmazható a jog, a racionális uralom világában, ahol sok helyen megtalálhatjuk mind az autonómia, mind a heteronómia jelenségét (ehhez presze a jogot általánosabb értelemben kell interpretálnunk, mint az állam által teremtett normatív rendszert). Nem ilyen egyértelmű a helyzet a konvenció és a morál területén, hiszen ilyen esetekben nem mindig lehet (vagy szinte sosem lehet) megmondani, hogy ki a norma kibocsátója. Sokszor nem is beszélhetünk kibocsátóról, kibocsátásról. Von Wright ezt pont ellentétesen látja. Szerinte a morál lehet autonóm, a preskripció (durván: a jog) nem igazán. Ezt azon az alapon állítja, hogy a morálnak (és a konvenciónak, bár ezt ő nem említi) nem ismert a kibocsátója, és így az autonómia azt jelenti, hogy nem valaki mástól (nem egy konkrét ágenstől) származik a norma. Ezzel szemben von Wright nem tartja lehetségesnek, hogy valaki önmagának bocsásson ki egy normát, vagyis a preskripció szerinte nem lehet autonóm – legfeljebb csak valami másodlagos, analógiás, átvitt értelemben [564, 77]. Ez a von wright-i ellenvetés hamar súlyát veszti, ha az egyéni cselekvők szintjéről egy kicsit elmozdulunk a kollektív cselekvések irányába. Mégha egyet is értenénk von Wright nézetével, a szervezetek világában már egészen biztosan értelmesnek, használhatónak kellene gondolnunk az autonóm és heteronóm minősítéseket.

Normatípusok a címzett és az alkalom szerint

A norma összetevői között a címzett, az alkalom és a feltétel az a három elem, amelyben egyedi, partikuláris dolgokat is kifejezhetünk. A másik három öszetevőben általános, generális információt adhatunk meg. A normák egyedi, partikuláris minőségét értelemszerűen csak az ennek kifejezésére képes összetevők segítségével ragadhatjuk meg, míg a normák általános, univerzális minőségéhez reprezentálására bármely normaösszetevőt használhatjuk. Mivel nem minden normához tartozik feltétel, ezért egy norma partikularitását annak címzettjén, illetve alkalmán keresztül jelezhetjük. Ezt teszi von Wright is, amikor a normákat a partikuláris-univerzális dimenzió mentén tipizálja [564, 77-81]. Igazodva az általa javasolt terminológiához elkülöníthetjük a partikuláris, **egyedi normákat** (PARTICULAR NORM) és az univerzális, **általános normákat** (GENERAL NORM) egymástól.

Egy norma esetében mind a címzettre, mind az alkalomra teljesen egyedi, teljesen konkrét adatokat megadhatunk. Mondhatjuk azt, hogy egy vagy több konkrét emberre vonatkozik a norma, de egyedi tér-idő adatok megadásával megszabhatjuk a különös vagyis egyedi alkalmazási körülményeket is. Eszerint a norma címzettje és alkalma lehet partikuláris, egyedi, illetve generális, általános. Ezek alapján definiálhatunk új normatípusokat. Az **egyedi normákhoz** (PARTICULAR NORM) akkor jutunk, ha mind a norma címzettje, mind az alkalma partikuláris, egyedi. Ezzel szemben egy **általános normát** (GENERAL NORM) úgy adhatunk meg, hogy vagy a címzett vagy az alkalom (vagy mindkettő) általános [564, 81]. Ehhez hozzáteszi még von Wright azt is, hogyha mindkét összetevő általános, akkor a normát **teljesen általánosnak** (EMINENTLY GENERAL NORM) lehet nevezni.

Normatípusok a feltételesség alapján

A norma feltételének fogalma a normák természetes tipizálásra is módot ad [564, 74-75]. Von Wright elkülöníti egymástól a **kategorikus** (CATEGORICAL NORM) és a **hipotetikus normákat** (HYPOTHETICAL NORM). Ezt azon az alapon teszi meg, hogy van-e a természetes alkalmazási feltételeken kívül további feltétel a normához rendelve. Ha nincs, akkor beszélhetünk kategorikus normákról, ha van, akkor hipotetikus normákról van szó.

A kategorikus és hipotetikus norma bemutatásakor már jeleztem, nagyon vigyázni kell arra, hogy ne keverjük össze egymással *(i)* a norma kategorikus feltételeit, *(ii)* a hipotetikus norma "feltételességébe tartozó" feltételeket, valamint *(iii)* a norma alkalma, alkalmazási körülményei fogalma alá sorolható feltételeket. További nehézséget okozhat az a további körülmény, hogy *(iv)* a normához tartozó szankciót is kondicionális logikai kapcsolóval, azaz egy feltétellel kötjük össze a normasértés tényét rögzítő propozícióval. A 'feltétel' terminust tehát legalább négyféle értelemben alkalmazhatjuk, és ebben a helyzetben könnyen lehet tévedni. Jakab András a norma szerkezetének vizsgálatáról szóló tanulmányában azt nyilván helyesen állapítja meg [255], hogy a tényállás és jogkövetkezmény között van egy kondicionális kapcsoló, amit lehet többféleképpen is ki lehet fejezni (és ezáltal kategorikus vagy hipotetikus formában reprezentálni), de a von wright-i értelemben vett hipotetikusság-kategorikusság csak a tényállásra, még pontosabban arra a diszpozícióra vonatkozik, amely a cselekvésre vonatkozó előírást tartalmazza. Jakab állításával ellentétben tehát továbbra van értelme hipotetikus normáról beszélni. A hipotetikusság értelmességére (és szükségességére) vonatkozóan bemutatok egy ismert példát, hogy rajta keresztül érzékeltethessem a hipotetikus és kategorikus normák közti különbséget. Vegyük az alábbi bibliai példamondatokat (normákat):

1.) Ha megdobnak kővel, dobd vissza kenyérrel!

2.) Szemet szemért, fogat fogért!

Akárhogy is akarnánk átformálni ezeket a példamondatokat, csak hipotetikus normákként értelmezhetnénk csak őket. A bennük megtalálható hipotetikusság megszüntethetetlensége pedig akkor válik nyilvánvaló, ha átalakítjuk őket kategorikus normákká. Ekkor, tehát kategórikus normaként, így néznének ki:

1'.) Dobd meg kenyérrel!

2'.) Szúrd ki a szemét, üsd ki a fogát!

Mindenki látnia kell a különbséget 1.) és 1'.). illetve 2.) és 2'.) mondatok között, nem szorul különösebb magyarázatra a kétféle megnyilatkozás szembeállítása. A hipotetikus normáknak igen nagy szerepe van a társadalmi érintkezésben. Említettem már, hogy a reciprocitás jelenségéhez kapcsolható cselekvési stratégiák, viselkedési szabályok a legtöbb esetben hipotetikus normaként (vagy logikai struktúráját tekintve annak megfelelő érdekvezérelt szabályként) vannak megfogalmazva.

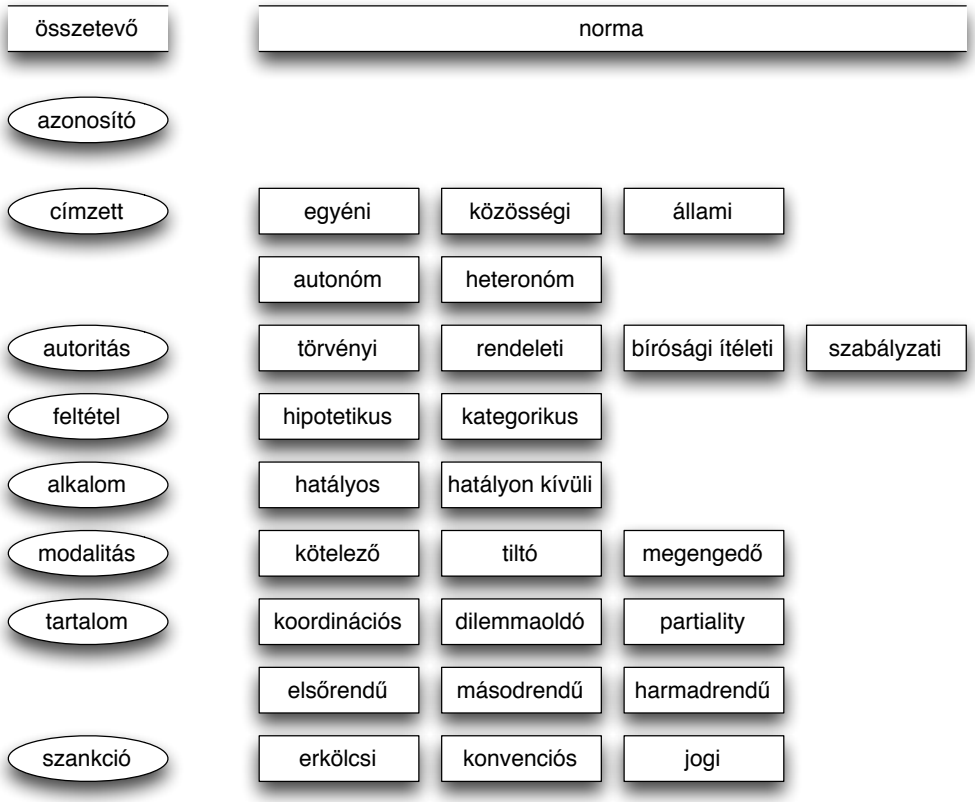
A normatípusok áttekintése

A különböző tulajdonságok, összetevők mentén megalkotott normatípusokat egyetlen áttekintő ábrában bemutatva láthatóvá válik, hogy az egyes szempontok, dimenziók szerint milyen típusokat különíthetünk el egymástól. Ha pedig arra is gondolunk, hogy ezeket a különböző típusképző szempontokat egyszerre is érvényesíthetjük, tehát további altípusokat kaphatunk, akkor látszik, hogy egy igencsak kiterjedt, sokszempontú normatipológia áll rendelkezésünkre. A normák tipizálási lehetőségeinek bemutatása után lehetne még a normák kialakulásának és fennmaradásának mechanizmusairól írni, de ezt a kérdéskört most nem tárgyalom (magyar nyelven ajánlom Orthmayr Imre áttekintését a társadalmi normák döntéselméleti és evolúciós magyarázatairól [398]).

6.6 Uralom

Egyáltalán nem szokványos dolog az uralom és hatalom jelenségét egymástól elválasztva tárgyalni. Maga Weber is szinte egymás mellett definiálta a két fogalmat, és a későbbi Weber-recepció egyszerre kezelte őket. Érdekes módon viszont sem Weber, sem követői nem törekedtek igazán arra, hogy megmondják, mi is a két fogalom közti különbség. Ezt a kérdést e fejezet végén próbálom majd a lehető legpontosabban megválaszolni, de addig is érdemes megindokolnom azt, hogy miért itt bontom ki az uralom fogalmát. Már csak azért is, mert az indokolásul felhozandó tételnek köze lesz a hatalom és uralom közti különbség magyarázatához is, és ennyiben megelőlegezi a később kifejtendő választ is. Az uralom fogalmának tárgyalását azért ágyaztam be a normát taglaló fejezetbe, mert a hatalomhoz képest az uralom sajátosságát az utóbbi normatív minősége biztosítja. Illetve még erősebbet lehet mondani: a norma és az uralom fogalma lényegében elválaszthatatlanok egymástól – nemcsak gyakorlati, de fogalmi értelemben is.[33] Mielőtt ezt a tézist

[33]Halkan jegyzem csak meg, hogy mind a fogalmi építkezés szempontjából, mind keletkezéstörténeti szempontból előbbrevalónak tartom az uralom fogalmát a normához képest. Az első utalmi mozzanat



bővebben kifejteném és megvédeném, előtte tisztáznom kell még pár dolgot. Először idézzük fel, hogyan definiálta Weber az uralom fogalmát.

> „Uralomról beszélünk, ha van rá esély, hogy egy meghatározott tartalmú parancsnak megadható személyek engedelmeskedni fognak... az »uralom« szociológiai szempontból csakis azt jelentheti, hogy egy parancs várhatóan engedelmességgel fog találkozni.” [575, 77]

Fontos rögzítenünk, hogy itt elsődlegesen cselekvésbefolyásolásról beszélünk, aminek a központi mozzanata az egyén cselekvésére irányuló **akarata**. A cselekvésbefolyásolás lényege abban van, hogy valakinek a saját cselekvésére irányuló akarata megváltozik egy másik ember kinyilvánított akaratának hatására. Lehet valakit vitán keresztül befolyásolni, aminek eredményeként megváltoztatja a véleményét. Bármennyire is fontos ez a helyzet a társadalmi cselekvések világában, nem ezen a ponton helyénvaló foglalkozni vele. A véleménybefolyásolás kérdésköre lényegi módon áll távol az uralom (és hatalom) jelenségétől. A véleménybefolyásolásnak lehetnek különböző technikái, módjai, de abban mind közösek, hogy olyan hiedelemvilágot alakítanak ki a másik emberben, ami alapján azt gondolhatja magáról, hogy külső akarati befolyástól mentesen, szabad akarata szerint cselekszik. A hatalom és uralom szempontjából ez az opció érdektelen, legfőképpen azért, mert ilyenkor csak véleménybefolyásolásról, de nem cselekvésbefolyásolásról

---

akkor született meg az ember történelmében, amikor valaki valamilyen engedelmeskedési igénnyel lépett fel egy másik emberrel szemben, és ezt az igényt utóbbi elfogadta. Ezzel persze ők ketten azon nyomban “megalkották” a norma jelenségét is egyben, tehát a “történelmi elsőség” viszonylagos. Ebből következően a fogalmi építkezés logikája szempontjából tekintve előbb kellett volna tárgyalnom az uralom, majd csak utána a norma fogalmát, de didaktikai megfontolások miatt végülis ezzel ellentétesen döntöttem.

beszélhetünk. Ez utóbbi lényege (az előbbihez képest) abban van, hogy a cselekvés-befolyásolási szándék a másik ember cselekvését meghatározó akarat befolyásolására irányul. Valamely cselekvést lehet akarni vagy nem akarni, de valamilyen vélekedést nem lehet sem akarni sem nem akarni, egy vélekedés vagy van, vagy nincs. Természetesen lehet arra törekedni, hogy valaki rendelkezzen valamilyen tudást, ebben az értelemben lehet akarni azt, hogy tudjon valaki valamit, de a tudást, vélekedést akarni nem lehet.[34]

Persze fel lehetne tenni a kérdést, hogy hol is látjuk, láthatjuk-e egyáltalán az akarat fogalmát magát a weberi definícióban. Nos, ott van a **parancs** fogalmában. Ha meg akarjuk határozni az akarat fogalmát, akkor több összetevőt kell egymás mellé helyezni. Weber egy helyütt **kinyilvánított akaratként** (manifasted will) hivatkozik rá [574, 946]. Ez fontos komponens, de önmagában még kevés. Hozzá kell tenni azt is, hogy ez a kinyilvánított akarat mások cselekvésére irányul. Önmagunknak nem adunk parancsokat, csak másoknak. Annyit tehát rögzíthetünk, hogy a parancs mások cselekvésére vonatkozó, kinyilvánított akarat. Ehhez azonban hozzá kell még tenni valamit, de azt csak később tudom indokolni, hogy mit és miért.

A második észrevételem az uralom meghatározásával kapcsolatban az, hogy az uralom jelenségét egy pillanatig sem szabad úgy felfogni, mint aminek létezését, előfordulását a jog, a politika, az állam világára lehetne leszűkíteni. Az uralom ott van szinte minden társadalmi viszonyunkban, ott van minden társadalmi intézményben, a családban, a párkapcsolatban, az iskolában, a munkahelyen, társadalmi szervezetekben stb. Mivel már összekapcsoltam az uralom és a norma fogalmát, mondhatom, hogy az uralom minden normatív viszonyban megtalálható. És abból bizony "nagyon sok van".

Harmadjára azt emelném ki a weberi definícióból, hogy ami parancsadás az egyik oldalról, az engedelmeskedés a másik oldalról nézve. Az uralom fogalmának talán legfontosabb összetevője a mindenféle érdektől független **engedelmeskedni akarás** mozzanata. A meghatározásban ugyan "csak" várható engedelmeskedésről beszél Weber, de a definícióhoz kapcsolt értelmező szövegrészben már lényegi elemként hangsúlyozza az engedelmeskedni akarás mozzanatát. Ezt azt jelenti, hogy az uralom önkéntes, a parancsnak nem valamilyen kényszer hatására engedelmeskednek, hanem önszántukból, szabadon. Akik végrehajtják mások uralmi parancsait, azok valamilyen módon, valamilyen szinten akarják az engedelmeskedést magát. Ez az engedelmeskedni akarás jelenti, ez jelzi, ez konstituálja az uralom érvényességét, elfogadottságát. És ez adja a döntő különbséget a hatalommal szemben is, ami mindig a kényszeren alapul. De erről majd később.

Fe lehet, sőt fel is kell tenni a kérdést, hogy miért engedelmeskednek egyáltalán az emberek az uralmi kapcsolatokban. Erre megint csak később tudok pontosan válaszolni, de azt – a weberi meghatározás negyedik elemeként – már itt is rögzíthetem, hogy amikor az "alávetettek", vagyis a parancs végrehajtására felszólítottak valóban teljesítik a parancsot, akkor azt úgy kell értelmeznünk, hogy a kiadott parancsot az alávetettek elismerésre méltónak (példaszerűnek, mintaszerűnek), vagyis belsőleg érvényesnek tartják. Azt is mondhatjuk erre, hogy elismerik az uralmat magát. Meg azt, hogy az uralomban mindig van valami autoritás, valami tekintély. Ebből fakad az, amit Joseph Raz a jog **preemtív jellegének** nevez [434], de amit nyugodtan ráhúzhatunk minden uralmi jelenségre, hogy ti. minden uralmi (és tegyük most már hozzá: minden normatív) jelenségben ott van az a sajátosság ("képessége"), hogy a cselekvések gyakorlati indokainak keresése, mérlegelése során képes mindent kizárni, az érvelési teret teljesen kiüresíteni, oda egyedül sajátmagát

[34] Ezen a ponton a kétfajta searle-i intencionális irányultság különbsége a perdöntő.

belehelyezni, minden más gyakorlati indok helyébe lépni.

Fel kell tárnom még egy további fontos mozzanatot az uralom jelenségéből, ami a weberi definícióban nincs ugyan explicite jelen, de az értelmező megjegyzések között már felbukkan, fogalmilag pedig nélkülözhetetlen. Ha belegondolunk egy egyszerű uralmi szituációba, akkor észrevehetjük, hogy a parancsot kiadó fél egy igénnyel lép fel a másikkal szemben, mégpedig egy **engedelmeskedési igénnyel**. Nem könnyű észrevenni, hogy a működő, már "beállt" uralmi viszonyok mennyire magától értetődően támaszkodnak erre az engedelmeskedési igényre, ami az uralmat gyakorló fél fejében lapul. Könnyen felfedezhetjük azonban ennek az összetevőnek a fontosságát (mi több, definitív módon szükségszerű létezését), ha elképzelünk egy kísérletet magunkkal. Ha azt a feladatot kapnánk, hogy "menjünk oda az utcán ismeretlenekhez, és mondjuk nekik azt, hogy meg kell írniuk a második zárthelyi dolgozatukat, vegyenek elő papírt, hogy lediktálhassuk a kérdéseket", akkor ezt a – tanárként egyébként havonta rutinosan elvégzett – műveletet képtelenek lennénk megcsinálni. A legfőbb indokunk pedig az lenne a végrehajtás ellen, hogy nem tudunk semmi érvet felhozni annak védelmében, hogy miért kéne az ismeretleneknek végrehajtaniuk a kiadott parancsainkat. Az engedelmeskedési igény hiányozna bennünk a leginkább (amit a tantermi szituációkban annyira természetesnek érezzük, hogy már észre sem vesszük). Tovább bonyolódik azonban a kép, ha felismerjük, hogy a résztvevők még egy másik és másfajta igényt is "kezelnek" ebben a helyzetben. Weber ezt **legitmitásigénynek** nevezi [575, 222]. A legitimitásigény fogalmának értelmezéséhez H.L.A. Hart egy szembeállítását hívhatjuk segítségül. Hart a jog és a szokás fogalmát úgy állította egymással szembe, hogy a joghoz a **belső szemlélet** (internal aspect), míg a szokáshoz a **külső szemlélet** (external aspect) hozzárendelését javasolta.

> „. . . ahhoz, hogy egy társadalmi szabály létezzen, legalább néhány embernek úgy kell tekintenie a kérdéses magatartást, mint a csoport egésze részéről követendő általános mintát. Egy társadalmi szabálynak létezik tehát egy »belső szemlélete« is, azon a külső szemléleten túl, melyben nem különbözik egy társadalmi szokástól, és ami nem más, mint az ismétlődő viselkedés, amit egy megfigyelő észlelhet." [228, 73]

Ez a közös belső szemlélet az, ami képes összetartani – legalább bizonyos emberek számára – a normák, és ezen belül az uralmi normák vagy másként: az uralmi (engedelmeskedési) igények érvényességére vonatkozó hiteket. Az uralom lényegéhez tartozik a parancsot kiadók engedelmeskedési igénye egyfelől, illetve az alávetettek engedelmeskedni akarása másfelől, és ezekre irányuló belső hiteket valamilyen gyakorlati érvekkel meg kell tudni támogatni. Az (uralmi) normák érvényességére (legitimitására) vonatkozóan azonban nemcsak versengő hitek, de a hiteket megelőző, megalapozó versengő világszemléletek is vonatkoznak. Magára a normatív érvényességre, a legitimitásra többféle és – ami fontosabb – többfajta érvet lehet szolgáltatni. A legitimitásigények eltérésének szemléltetésére idézek egy politikai nyilatkozatot, amelyen belül egyértelműen tetten érhető az eltérő legitimitáshitek léte és konfliktusa.

> „Az iráni kormány nem ismer el más tekintélyt vagy hatalmat, csak a Mindenható Istenét, és semmiféle jogi hagyományt, csak az iszlám törvényét. Ilyen feltételek mellett az iráni delegáció újra megerősíti, hogy a nemzetközi szervezetek iszlámmal ellentétes nyilatkozatainak, határozatainak vagy döntéseinek semmiféle érvényessége nincs az Iszlám Köztársaságban. Az »Emberi Jogok Egyetemes Nyilatkozata«, amely a zsidó-keresztény hagyomány világi koncepcióját illusztrálja, nem alkalmazható a muzulmánok által, és egyáltalán nem kapcsolódik az Iráni Iszlám Köztársaság által elismert értékrendhez; az Iráni Iszlám Köztársaság ezért nem habozhat e rendelkezések megsértésekor, ha az ország

> isteni törvényének megsértése vagy a világi egyezségokmányok megsértése között kell választania." [452]

A fenti idézet önmagában elég ahhoz, hogy lássuk, miért is kell többféle legtimitásigényről (és legitimitástípusról) beszélni, illetve a konkrét példában az is szépen látszik, hogy ezek a különböző legitimitásigények, legitimitáshitek valóban versengő viszonyban vannak egymással. Összegzésként tehát azt kell rögzítenünk, hogy minden uralmi kapcsolatban kétféle igény, egy engedelmeskedési igény, illetve egy legitmitásigény bújik meg, és az uralom tényleges megvalósulásáról akkor beszélhetünk, ha ezt a két igényt az uralmat gyakorolni kívánók sikeresen el tudják fogadtatni az uralomnak alávetettekkel. A legitimitásigény fogalma kapcsán nyilván fel kell tenni a kérdést, hogy milyen alapon, milyen típusokat lehet elkülöníteni egymástól. Weber ezt a kérdést nem elemezte, ezért szükség lenne a fogalom explikálására. Ezt az elemzést azonban csak a következő fejezetben fogom elvégezni, miután már bemutattam az uralom legfontosabb típusait.

6.7 Uralomtípusok

Amikor a *Gazdaság és társadalom*-ban Max Weber kifejti az uralomtipológiáját, akkor három típust mutat be, a tradicionális, a karizmatikus és a racionális uralmat [575, 224]. Korábbi munkáimban én négyelemű utalomtipológiát használtam [528], [517], de nem igazán támasztottam alá, miért gondolom kiegészítendőnek a weberi tipológiát. Most megpróbálkozom vele. Először röviden vegyük a három klasszikus, weberi uralomtípust. Itt még nem foglalkozom azzal, melyik típus értékelhető korábbinak, alapvetőbbnek a másikhoz képest, és nem reflektálok egyelőre az egymáshoz való viszonyukra sem. Az első típus a **tradicionális uralom** kategóriája. Íme a jól ismert meghatározás.

> „ . . . elsődlegesen az emberemlékezet óta érvényes tradíciók szentségébe és a tradíciók által autoritással felruházott személyek uralmának legitim voltába vetett mindennapi hiten alapul." [575, 224]

A definíció kulcsmozzanata a **szent tradíció**. A tradíciók által autoritással felruházott személyek parancsadási jogosultsága, legitimitása nyilván származtatott jelenség, amit nem érthetünk meg a szent tradíció minősége nélkül, viszont az előbbit már levezethetjük az utóbbiból – mind a fogalom jelentését, mind a jelenség létezését. A tradicionális uralom esetében tehát a tisztázásra váró fogalom a (szent) tradíció kategóriája. A következő uralomtípusként nézzük a **racionális uralom** fogalmát, amire Weber a következő definíciót húzta rá.

> „ . . . elsődlegesen a legalitásba – a tételes rendnek és a tételes rend által az uralom gyakorlására kijelölt személyek utasítási jogának legális voltába – vetett hiten alapul." [575, 224]

Ebben az esetben is származtatott kategóriának tarthatjuk az uralom gyakorlására kijelölt személyek parancsadási jogának legitimitását (legalitását), és a definíció központi elemeként a **tételes rend** fogalmát kell megragadnunk.[35] Mivel Weber a rend kategóriáján durván a norma fogalmát értette, így az igazi minőséget itt a tételes jelző adja, úgyhogy ezt kell jól megmagyarázni. Azt már érdemes jelezni, hogy az eddigi két uralomtípus középpontjában két rendfogalom áll, hiszen a tradícionális uralomhoz rendelt tradíció is egyfajta rendként (normaként) értelmezhető. A harmadiknak maradt klasszikus weberi uralomtípus a **karizmatikus uralom**.

[35] Később természetesen tisztázni kell még a legitimitás és a legalitás fogalmai közöti kapcsolatot is.

„ . . . elsődlegesen nem mindennapi odaadáson alapul egy szent, hősies vagy példamutató személy, illetve az általa kinyilatkoztatott vagy megteremtett rend iránt.” [575, 224]

A definíció kapcsán először arra hívnám fel a figyelmet, hogy Weber itt sorrendet vált abban az értelemben, hogy előbb beszél az uralmat gyakorló személyekről, és csak azután az uralmi rendről. Ez eddig pont fordítva volt, sőt, az előző két típus esetén azt mondtam, hogy a személyek uralmi jogosítványai levezethetőek az adott uralmi rend érvényességéből. A karizmatikus uralom meghatározásának középpontjába viszont a személy került, és akár tűnhetne úgy is – mégha ez tételesen nincs is kimondva –, hogy itt a személyből lehet a származtatni az általa kinyilatkoztatott rendet. Ezt a váltást, illetve az ebből kibontható fogalmi dimenziót nemsokára elemezni fogom, de előtte még lenne pár megjegyzésem a karizmatikus uralom kapcsán. Van egy fontos definíciós összetevő, a ‘nem mindennapi odaadáson alapuló’ hit eleme, amit Weber itt alkalmaz először, és csak ekkorlehet észre venni igazán, hogy ez szembeállítható a tradícionális uralomban alkalmazott, ellentétes minősítéssel, a ‘mindennapi hitre’ való alapozás momentumával. Ez a különbségtétel azért fontos, mert a mindennapiság kontra nem mindennapiság szempontrendszere olyan elemzési dimenziót biztosít, amely mentén később oszályozni tudjuk az uralomtípusokat. Értelmeznünk kell a magának a karizmának a fogalmát is. Ezzel kapcsolatban azt szokás idézni Webertől, hogy a karizmát egyfajta “kegyelmi ajándékként” foghatjuk fel – amit nyilván valamilyen transzcendenciától származhat csak. Amikor Weber meghatározást próbál adni erre a fogalomra, akkor a felsorolásszerű indoklásai közé bevon olyan példákat is, amelyek voltaképp kilógnak a meghatározás hatálya alól, és ezzel kicsit kétértelművé teszi Weber ezt a kategóriát. A karizmáról a következőket írja.

„»*Karizmának*« nevezzük egy személyiség nem mindennapinak számító képességeit . . . , amelyek miatt ezeket a személyeket természetfölötti vagy emberfölötti vagy legalábbis sajátos, nem mindennapi, nem mindenki számára hozzáférhető erőkkel vagy tulajdonságokkal [megáldott] embereknek vagy isten küldöttének vagy példaképnek, s ezért »*vezérnek*« tekintették.” [575, 248]

Weber itt a karizmatikus uralmat jellemző nem mindennapiságot akarja megragadni, amit – több helyen, itt is, máshol is – érezhető szándéka szerint kirárólagos módon szeretne összekötni a természetfölöttivel, emberfölöttivel. Amikor azonban a fenti felsorolásba beemeli a kivételes (ám – és itt ez a fontos – nem transzcendes minőségű) tulajdonságokra való hivatkozást is, akkor ezzel megszünteti az éles határvonalat a kivételes emberi és az emberfölötti között. Az, hogy mennyire fontos Weber számára a karizma transzcendes eredet, szolgáljon újabb bizonyítékul a következő rövidebb passzus.

„A karizma akkor érvényes, ha valamiképpen – eredetileg mindig csoda folytán – *beigazolódik*, s ennek következtében az alávetettek szabadon – csupán a megnyilatkozás iránti odaadásból, a hős iránti tiszteletből vagy a vezér iránti bizalomból – *elismerik*. De (a valódi karizmánál) ez az elismerés nem a legitimitás *alapja*, hanem *kötelesség* . . .” [575, 248]

A természetfölötti minőség adja a karizma lényegét, ami mindig csak valami csoda folytán igazolódhat be. De fontos nekünk a fenti idézet egy másik momentuma is, mert segítségével rendbe tehetjük a fent jelzett kétértelműséget. Javaslatom szerint úgy explikálhatjuk a karizmatikus uralomra vonatkozó, eredeti weberi intenciót, hogy Weber meghatározását a következő tézisekre bontjuk: *(i)* az eredeti karizmatikus uralom transzcendens világképen alapul, *(ii)* az engedelmeskedni akarás a transzcendenciának való alárendelődési hitből és az annak való megfelelési szándékból fakad, *(iii)* a karizma a transzcendencia felől érkező kegyelmi ajándék, ami a vezér mint szükségképpen földi személy kiválasz-

tottságát is jelzi egyben a többiek számára, végül *(iv)* a vezér személyét és utasításait nem annak kivételes képességei miatt ismerik el az emberek, hanem az a hit kötelezi őket a vezér követésre, hogy ő a transzcendencia kiválaszottja, és őt követve voltaképp a transzcendencia akarata szerint járnak el. Ebbe az interpretációba nem férnek bele Weber azon példái, amelyek a vezérek nem mindennapi, de földi eredetű, kivételes képességeire vantkoznak. Ha jobban megfigyeljük a következő idézetet, akkor kiderül, hogy Pierre Bourdieu is másként értelmezte a karizmatikus uralom fogalmát az eddigiekhez képest.

> „A karizmatikus vezető, ahelyett, hogy – ahogy az a szimbolikus küzdelemben alulmaradottak esetében fennáll – saját maga számára is az lenne, aminek a többiek tartják, eléri azt, hogy az őt vezetővé választó csoport számára is azzá váljék, aminek saját maga véli magát. Ő teremti meg az őt magát megteremtő közvéleményt." [66, 156]

Weber példái a nem mindannapi, de földi képességekkel rendelkező vezérekről, Bourdieu nézete a sajátmagát megteremtő karizmatikus vezérről felvetik annak a lehetőségét, hogy esetleg képezhetünk egy új, egy negyedik uralomtípust. Bármennyire is kimondja szószerint és többször is Weber maga, hogy három uralomtípust definál, bármennyire is csak három uralomtípust mutat be az uralom szociológiájában, bármennyire is ugyanezt visszhangozzák a követők Weber óta, maga Weber az, aki tesz egy olyan gesztust a *Gazdság és társadalom* egy pontján, amit értelmezhetünk úgy, mint egy tétova kísérletet a negyedik uralomtípus elkülönítésére vonatkozóan. Az uralom szociológiájáról szóló fejezet hetedik pontjának címe *A karizma uralomellenes átértelmezése*, és erről a következőeket Weber a fejezet kezdő paragrafusában.

> „Az elsődeges értelme szerint a tekintélyuralmat megalapozó karizmatikus legitimitáselv olyan értelmezést is kaphat, amelyben már nem a tekintélyuralmat támasztja alá. A karizmatikus tekintély tényleges érvényessége ugyanis valójában teljes egészében azon múlik, hogy karizmája »beigazolódik-e«, mert ez a feltétele annak, hogy az alávetettek *elismerjék*, ami persze a karizmatikus képességekkel rendelkező *és ebből következően* legitim úrral szemben *kötelességük*. A szervezeti kapcsolatok növekvő racionalizálása esetén azonban kézenfekvő, hogy ezt az elismerést már nem a legitimitás következményének, hanem e helyett a legitimitás *alapjának* fogják tekinteni ... s végül a közösség részéről történő elismerést úgy fogják föl, mint »választást«." [575, 271-272]

Mivel mindez a karizmatikus uralomról szóló fejezetbe van ágyazva, mondhatnánk azt, hogy itt továbbra is csak a karizmatikus uralomról van szó. Weber nem mondta ki, hogy ő úgy értelmezi a fenti részt, mint amelyben egy új, önálló uralomtípust vezetne be. Ebből következően nyilván nem mondhatom, hogy Weber négy típusról beszélt (vagy akart volna beszélni). Ennek ellenére érvelni fogok az elkülönítés mellett, és javaslom felvenni az újabb kategóriát, mert ha így teszünk, konzisztensebb és hézagmentesebb uralomtipológiához jutunk az eredeti hármas felosztáshoz képest. Weber ugyanis úgy jellemzi le a **átértelmezett karizmatikus uralom** kategóriáját, ami egyrészt szöges ellentétben áll egy adott dimenzióban az eredeti karizmatikus uralom jelentésével, másrészt viszont egy kétdimenziós fogalmi térben értelmezve látványosan ki is egészíti azt, ami nélkül hiányos lenne, amivel együtt viszont teljessé válna az uralomtípusok felosztása, vagyis pótolni lehetne vele a "hiányzó láncszemet". A megoldás felé elindulva elsőként idézzük fel azt a tényt, hogy Weber az uralomtípusokat a hozzájuk kapcsolható legitimitásigény alapján tartotta elkülöníthetőnek egymástól.

> „ ... célszerű az uralom fajtáit a rájuk jellemző legitimitásigény szerint megkülönböztetni. ... Itt az a mérvadó, hogy maga az uralom milyen fajta legitimitásra tart igényt, és saját legitimitásigénye releváns mértékben érvényesül-e ... " [575, 222]

Ez a weberi tipizálási javaslat teljesen érthető és elfogadható, ha belegondolunk abba, mi az uralom lényege. Valaki kiáll egy másik ember elé, és megfogalmaz egy parancsot a másiknak. A parancsot kiadó félben ott van az igény arra, hogy a másik ember teljesítse a kiadott parancsot, és a parancsolónak ezzel párhuzamosan fel kell készülnie arra is, hogy milyen érveket sorakoztathatna fel abban az esetben, ha a másik oldalon vitatnák az ő parancsadási igényét. Az, hogy milyen típusú érvekkel tudná, akarná megvédeni saját uralmi igényét, leginkább attól függ, hogy milyen érvényességfogalommal rendelkezik saját felhatalmazottságára vonatkozóan. Weber ezt nevezi legitimitásigénynek. Ha a legitimitásigény alapján határozhatjuk meg az uralomtípusokat, akkor nyilván meg kell tudnunk mondani azt, hogy milyen szempontok szerint lehet a legitimitásigény fogalmának típusait megalkotni. Két dimenziót vezetek be ezzel kapcsolatban, amelyek segítségével konzisztens módon értelmezhetjük az itt javasolt négyes felosztást az uralom fogalmára vonatkozóan. Émile Durkheim a társadalmi jelenségek magyarázata során elkülönített két fontos kategóriát, a **szent** és a **profán** fogalmát. Ez a fogalomkettős segíthet minket a normativitás és az uralom értelmezésében, valamint az uralom tipizálásában egyaránt. Durkheim így ír a két fogalomról.

> „ . . . a szentet a profánhoz viszonyítva már csakis különneműségük alapján határozhatjuk meg. E különneműség azért elegendő a dolgok ezen osztályozásának jellemzéséhez és az összes többitől való megkülönböztetéséhez, mert igen sajátos: abszolút. Az emberi gondolkodás történetében nincs még egy példa a dolgok ennyire mélyen különböző, ennyire ellentétes két kategóriájára.” [141, 46]

Durkheim ezt a kettősséget tartja a leginkább egymást kioltó fogalmi minőségnek, amihez képest például a jó és rossz vagy az egészség és betegség közti viszony jóval enyhébb szembenállást jelent. Utóbbiakban ugyanis van valami közös. A jó és rossz fogalma az erkölcsiség két ellentétes minősége ugyan, de mindkettő az erkölcs kategóriája alá sorolható be, ennyiben valamilyen minimális közös vonásnak léteznie kell mindkettejükben. Ezzel szemben ami szent vagy ami profán, abban nincs semmi közös.

> „Amikor azt állítjuk róluk, hogy szentek, ezen azt értjük, hogy értékük nem mérhető össze a többi értékkel. Szent az, amit elkülönítettek, aminek nem lehet közös mértéke a profánnal.” [138, 190]

Van tehát két fogalmunk, amelyekkel jelezhetünk, összeköthetünk egy-egy világot, egy-egy világnézetet is. Mindazon társadalmak, amelyeknek a tagjai domináns módon transzcendens világképpel rendelkeztek, a szent, a szentség minőségével kapcsolták össze az uralmi kapcsolataikat (mint minden más egyebet is). Ezzel szemben azokban a társadalmakban, amelyekben a transzcendenciában való hit legalább abból a szempontból megszűnt, visszaszorult, hogy az emberek közti uralmi igények indoklásai során nem hivatkoz(hat)tak a földöntúli akaratokra, ott az uralmi kapcsolatokat a – szentséggel élesen szembenálló – profán minőség jellemez(het)te. Ez a dichotómia adja az uralmi osztályozáshoz szükséges egyik fogalmi dimenziónkat. A tradicionális és karizmatikus uralom érvényességi igényei esetében lehet hivatkozni **emberfeletti akaratokra** vagy földöntúli akaratokra is, a racionális és átértelmezett karizmatikus uralom esetében pedig nem. Utóbbiakban csak az **emberi akaratokra** való hivatkozást fogadják el érvényesnek. Ennek nagyon fontos következménye az, hogy amíg az emberi akaratokra hivatkozó szabályokat mindig meg lehet támadni, mindig lehet vitatkozni fölöttük, addig a emberfeletti akaratokra épített, pontosabban azokra hivatkozó szabályokat sosem lehet nyíltan támadni.

A másik fogalmi dimenziót úgy találhatjuk meg, ha felidézzük, hogy milyen módon szólt Weber az uralom rend vagy uralmi szabályok, illetve az uralmat gyakorló személyek viszonyáról, "sorrendiségéről". Felhívtam rá a figyelmet korábban, hogy a tradícionális és a racionális uralom meghatározásakor Weber az uralmat gyakorló személyekhez képest előbbre vette, bizonyos értelemben fontosabbnak tartotta a parancsadási lehetőségeket rögzítő, azokat folyamatosan megteremtő szabályokat, rendtípusokat, a tradíciót, illetve a tételes rendet. Ezzel szemben a karizmatikus uralom definíciójában megcserélte ezt a sorrendet, és a vezér személyét tette a fogalom központjába mint olyan fogalmi összetevőt, amelynek ismeretében már levezethető az általa kinyilatkoztatott rend, szabály is. Ugyanezt természetesen elmondhatjuk az átértelmezett karizmatikus uralomra vonatkozóan is. A nagy kérdés itt az, hogy milyen címkékkel, milyen leírással láthatjuk el ezt a dimenziót, amely tehát egyik oldalra teszi a tradicionális és racinális uralmakat, amelyeket a központba tett szabállyal, rendtípussal azonosíthatunk, míg a másik oldalra sorolja a személyeket előtérbe helyező kétféle karizmatikus uralmat. Értelmezésem szerint az a legáltalánosabb közös része ennek a kétféle besorolásnak, hogy valamely uralomban milyen módon lehetséges a parancsadási lehetőségeket, az uralmi szabályokat, különösen azok tartalmát megváltoztatni. A tradícionális uralomban nincs vagy alig van vagy csak trükkösen van mód megváltoztatni a tradíciókba lefektetett szabályokat, de amilyen mértékben ez mégis lehetséges, úgy mindig a tradíciókra, tehát a szabályokra magára hivatkozva lehet ezt megtenni. Ugyanez igaz a racionális uralomra is, csak ott az érintettek által korábban lefektetett, tételes rend (általában a jogszabályok) újbóli tételezésén keresztül lehet változtatásokat elérni. Mindkét esetben az a közös, hogy a szabályváltoztatás a szabályok (újra)értelemzésén keresztül történik. Ezzel szemben a karizmatikus uralmak esetében a személyt kell változtatni ahhoz, hogy az uralom jellege megváltozhasson. A karizmatikus uralomban mindig van egyfajta "önkény" abban az értelemben, hogy elválaszthatatlan tőle a vezér személyéhez való kötöttség, önkényesség. Jézus paradigmatikus példáját idézve: "*Hallottátok, hogy a régiek ezt a parancsot kapták:... Én pedig azt mondom nektek ...*" [230]. Amíg a vezért követik az önkéntesek, addig mindig beszélhetünk önkényességről. Ebben a viszonyrendszerben az uralom jellegének, az uralmi szabályok megváltoztatásának nincs más módja, mint a vezér személyének lecserélése. A két most bevezetett dimenziót egyszerre figyelembe véve konzisztens osztályozási lehetőséghez jutunk, amit az alábbi táblázatos formában írhatunk fel [529, 219].

	az uralmi szabályok változtatásnak módja:	
a legitimititásigény indoklásakor:	szabályértelmező	személycserélő
emberfeletti akaratra is hivatkoz(hat)nak	tradicionális	karizmatikus
csak emberi akaratra hivatkoz(hat)nak	racionális	átértelmezett karizmatikus

Ebben a felosztásban nagyon is helye van az újonnan bevezetett átértelmezett karizmatikus uralomnak, oly'annyira, hogy nélküle hiányos lenne a tipológiánk. Az átértelmezett karizmatikus uralomban is vezérről beszélhetünk, de itt megfordul az a sorrend, amit Weber az eredeti karizmatikus uralom kapcsán tételezett. Itt az alávetettek, a követők elismerése a legitimitás alapjává válik, míg korábban annak következménye volt. Ez azt is jelenti egyben, hogy az átértelmezett karizmával rendelkező vezér (vagy mondhatjuk rövidebben és kevésbé nyakatekert módon, hogy a modern karizmatikus vezér) uralmát a többiek akaratára vezethetjük valahogy vissza – csakúgy, mint a racionális uralom esetében.

A kandidátusi értekezésemben bemutattam, hogy a kommunista-bolsevik uralom nagyon sok hasonlóságot mutat a tradicionális uralommal [524]. Éppen ezért azt javasoltam, hogy – bármennyire is ellenétetesnek tűnik az intuícióinkkal – a kommunista-bolsevik legitimitáselven alapuló történelmi formációt besorolhatjuk a tradicionális uralom fogalma alá. Ez sokaknak elfogadhatatlan volt, nekem sem tetszett, de nem tudtam jobbat javasolni helyette. Ameddig a tradicionális uralom fogalmát vagy inkább terminusát meg akarjuk tartani, addig gyakorlatilag elfogadhatatlannak tűnt és tűnik a javaslatom. Ám ha kicseréljük a tradicionális uralom terminusát valami másra úgy, hogy közben érdemben nem alakítjuk át a hozzákapcsolt fogalom jelentését, akkor talán már vállalható osztályozási rendszerhez és történelmi besoroláshoz juthatunk. A tradicionális uralom magjában a szent tradíció áll, ami – szentségéből adódóan – sérthetetlen, támadhatatlan. De ehhez valami nagyon hasonlót tapasztalhatunk a kommunista-bolsevik uralom működésével, legitimitásigényével kapcsolatban is. Mondhatjuk, hogy ez a történelmi formáció voltaképpen kiterjesztette a szentség elvét a bolsevik-kommunista (marxista-leninista) világnézetre vonatkozóan arra hivatkozva, hogy az az egyetlen tudományos világnézeten alapuló világnézet.[36] És mivel – felfogásuk szerint, természetesen téves felfogásuk szerint – a tudomány mindig igazságot, méghozzá egyetlen igazságot szolgáltat, ezért az "új" társadalmi (uralmi) rendet úgy kell felépíteni, hogy az csak egyetlen lehet. Ahogy azt a marxizmus-leninizmus leírta. Ami persze támadhatatlan. És ami ugyanolyan mechanizmusokon keresztül védte magát, mint tette és teszi azt minden más tradicionális uralom. Ebből a szemszögből nézve a kommunista-bolsevik uralom egylényegű volt a tradicionális uralommal. Ettől természetesen anakronizmusnak tűnne tradicionálisnak nevezni, de ezt a terminológiai ellenérzést felszámolhatjuk azzal, hogy új terminust választunk a 'tradicionális uralom' kifejezés helyett. Javaslatom erre a **zártrendű uralom** kategóriája, ami – intuícióim szerint – egyszerre képes kifejezni a tradicionális uralomnak a tradíciók szentségébe vetett hitét, és a kommunista-bolsevik uralomnak a marxista tanok kizárólagosságára irányuló vélekedését. Ennek alapján a következő módosított táblázathoz jutunk.

	az uralmi szabályok változtatásnak módja:	
a legitimititásigény indoklásakor:	szabályértelmező	személycserélő
emberfeletti akaratokra (is) hivatkoznak	zártrendű	karizmatikus
csak emberi akaratokra hivatkoznak	racionális	átértelmezett karizmatikus

Az uralomtípusok kapcsán szeretnék még egy további kérdéskört érinteni. Weber is fontosnak tartotta, hogy szembeállítsa az egyik oldaon a tradicionális, amásik oldalon karizmatikus és racionális uralmat egymással azon az alapon, hogy melyik uralomtípust lehet a status quo fenntartójaként, és melyiket a változásthozó forradalmi erőként minősíteni. Egyáltalán nem akarnám vitatni most Weber megállapításainak érvényességét, csupán csak szeretném levonni azt a következtést, ami az uralomtipológia kiegészítéséből meg kell tennünk. Weber a következőket írta az uralomtípusok egymáshoz, illetve változáshoz való viszonyáról.

> „A tradíciókhoz köthető korszakokban a karizma jelenti *a* nagy forradalmi erőt. Eltérően a másik, ugyancsak forradalmasító erőtől, a »ratio«-tól amely vagy egyenest kívülről hat – úgy, hogy megváltoztatja az életkörülményeket és az életproblémákat, és ezzel közvetve

[36] Ebben természetesen hatalmasat tévedtek, de ez a dolgok menete szempontjából irreleváns.

> a hozzájuk való viszonyulást is – vagy azon keresztül hat, hogy intellektualizálja a problémákat, a karizma kínból vagy lelkesedésből fakadó belső átalakulást *tud* előidézni, amely különféle viszonyulásokban – minden egyes életformához és általában a »világhoz« való viszonyulásban – kialakuló tökéletesen új beállítottság hatása alatt teljesen megváltoztatja a leglényegesebb érzületi és cselekvési irányokat." [575, 251-252]

Webernek természetesen igaza van akkor, amikor azt állítja, hogy a tradíciókhoz képest a ráció is lehet forradalmasító erő abban a közvetett értelemben, ahogy azt Weber leírja, de a karizmatikus uralom érzelmi alapjait, a kínből és lelkesedésből fakadó belső átalakulás mozgosító, forradalmi erejét ugyanúgy érdemes kiterjeszteni az átértelmezett karizmatikus uralomra vonatkozóan is. Ha viszont az átértelmezett karizmatikus uralmat is figyelembe vesszük ebben a közös értékelésben, akkor mondhatjuk azt, hogy a modern vezérek megjelenése, létezése ugyanolyan forradalmasító potenciállal bír a racionális uralommal, mint a hagyományos karizmatikus uralom a tradicionálissal uralommal szemben – és az okok természetesen ugyanazok mindkét esetben. E tézis további alátámasztásra érdemes felidézni azt, amit a racionalitás és az érzelmek, vágyak, késztetések egymáshoz való viszonyáról (arról, hogy a racionalitás önmagában nem rendelkezik mozgósító erővel) írtam korábban (↬155.o.).

A gondolatmenet lezárásaként annyit jegyeznék meg, hogy történelmileg is, fogalmilag is az a legvalószínűbb, hogy a karizmatikus uralom jelent meg legelsőként az emberek között, majd idővel valamelyik karizmatikus uralom emlékeire már úgy hivatkoztak, hogy a tényleges uralomgyakorlás formái egyre inkább a tradicionális uralomra kezdtek hasonlítani. A racionális uralom a világ varázstalanításával párhuzamosan terjedt el, majd kezdett idővel dominássá válni. Ahogy megjelent a racionális uralom és elkezdte a kinevezés elvén álló, hierarchikus uralmi szerkezetekbe rendezni az uralmi viszonyokat, úgy jelent meg – szükségszerűen – az az igény, hogy a hierarchikus szerkezetek csúcspontjain levő pozíciókba, amelyekket nem lehet(ett) a kinevezés elve alapján betölteni, a (személy)választás elvét alkalmazva keressenek megfelelő – karizmatikus képességekkel rendelkező – személyeket.

Hatalom és uralom viszonya

Korábban már explikáltam a hatalom fogalmát (↬266.o.), most bemutattam a norma és az uralom kategóriáit, így minden adott, hogy megpróbáljam tisztázni a köztük levő kapcsolatot. Emlékezzünk rá, hogy a hatalom definiálásakor a két szembenálló fél ellentétes akaratáról volt szó, és ez a két konfliktusos akarat ugyanarra a cselekvésre irányult. Ez utóbbi kötötte őket össze, és ezen az alapon lehet egyáltalán a köztük levő konfliktusról beszélni. Az *A* személy áll az ajtóban, mire egy másik ember odalép, *B* rászól, hogy álljon félre az útból, mert szeretne bemenni. Ha *A* nem akarja *B*-t beengedni mondván, hogy ő a bank biztonsági őre és a bank még nem nyitott ki, akkor megállapíthatjuk akaratkonfliktus van köztük. Eddig mesélve még nem tudhatjuk, hogy a hatalom vagy az uralom jelenségéről van-e szó. Amennyiben *B* előránt egy pisztolyt, és *A*-nak szegezve követeli a banki trezorban levő pénzt, akkor már megállapíthatjuk, hogy hatalomról beszélhetünk – már amennyiben a rablás sikeres lesz. A lényeg az, hogy a hatalom megvalósulásához kényszer alkalmazására volt/van szükség. Ehhez képest nagyon más logika mentén működik az uralom jelensége. Weber alapján azt mondtuk, hogy az uralomhoz mindig kell engedelmeskedés, amit a hatalom esetében nem találtunk. De mit is jelent az engedelmesség?

> „Az »engedelmesség« azt jelenti, hogy az engedelmeskedő személy cselekvése úgy megy végbe, mintha az illető a parancs tartalmát pusztán a parancsért tette volna a viselkedése maximájává, s *csakis* a a formális engedelmességi viszony miatt tett volna így, függetlenül attól, hogy mi a saját véleménye a parancsról mint olyanról, értékesnek tartja-e vagy sem." [575, 223]

Az uralom vagyis a más által kiadott parancsnak való engedelmeskedés azért valósul meg, mert az alávetett fél elfogadja a másik parancsadási jogosultságát. Ez azt jelenti, azt kell, hogy jelentse, hogy az alávetett félnek adott esetben önként, kényszer nélkül is fel kell tudnia adni saját akaratát azért, hogy végrehajtsa a kiaott parancsban foglaltakat. Könnyen elképzelhető, hogy a parancskiadás pillanatában a parancs tartalmával ellentétes akarattal rendelkezik, ezt figyelmen kívül kell hagynia. Az előző banki példára utalva lehet, hogy *B* személy azért akar bemenni a bank ajtaján, mert ő az új főnök, és ezt még a biztonsági őr nem tudja. Ebben az esetben *B*-nek nem kényszer alkalmazására van szükség, hogy bejusson, hanem nyilvánvalóvá kell tennie azt az *A* felé, hogy ő az új fiókvezető, és abban a pillanatban, ahogy ez sikerül neki, az őr már feladja saját korábbi – a beengedést tiltó – akaratát. Kétféle módon is lehetne interpretálni azt a tényt, hogy a biztonsági őr megváltoztatja az akaratát. Mondhatnánk azt, hogy a cselekvést megelőző mérlegelése során erősebb érvnek bizonyult a főnök státusának felismerése az illetéktelenek beengedésére vonatkozó tiltás, vagyis az akaratokat kezelhetnénk úgy, mint amik ugyanazon a szinten helyezkednek el. Ezzel szemben pontosabbnak tartom azt az értelmezést, amely szerint az uralmi helyzet felismerése egy másodrendű akarat megjelenését okozta, ami a személy eredeti, elsőrendű akaratát mintegy zárójelbe tette, felülértelmezte. Ez a jelenség teljesen analóg a jogosultságok és kötelezettségek rendszerében belül elkülönített első- és másodrendű fogalmak egymáshoz viszonyulásával (↪330.o.). A helyzet uralomként való (át)értelmezése azért következett be, mert a biztonsági őr a belépni kívánkozó emberben a főnökét felismerve felhatalmazottság tartalmú jogosultságot tulajdonított neki, amivel szemben benne egy beavatkozásnak kitettség tartalmú kötelezettség keletkezett, és e másodrendű késztetés hatására megváltoztatta saját korábbi elsőrendű akaratát. Az analógia csak annyiban nem mondható száz százalékosan megalapozottnak, hogy az uralom esetén akaratokról, a perszonális pozícióknál pedig jogosultságokról és kötelezettségekről beszélünk. De ezen az eltérésen túl a szerkezeti hasonlóságot teljesnek ítélhetjük.

A hatalom és uralom kategóriáját – bármennyire is hasonló jelenséget írhatunk le velük – szigorúan el kell választanunk egymástól. Így tette ezt egyébként maga Weber is. Erre természetesen a legfontosabb bizonyíték az, hogy eleve különböző fogalomként definiálta őket, és az uralom szociológiáját messze külön kezelte a gazdasági hatalom kérdésétől. De az alábbi idézetben explicit módon figyelmeztetett arra, hogy a két jelenséget nem szabad összemosnunk.

> „Másfelől viszont a monopolhelyzetből származó gazdasági »hatalmat«, vagyis azt a lehetőséget, hogy az egyik cserepartner »diktálja« a csere feltételeit a másiknak, önmagában véve még nem nevezzük »uralomnak«, mint bármilyen más – teszem azt az erotikában, a sportban, a vitában, vagy egyebütt élvezett – fölényből eredő »befolyást«." [575, 222]

A hatalom fogalmának tárgyalásakor bemutattam az a folyamatábrát (↪267.o.), ami az uralom értelmezésében is segíthet. Amikor ugyanis az elsőrendű akaratok különbsége, konfliktusa nyilvánvaló válik a kapcsolat résztvevői számára, akkor az lesz a nagy kérdés, hogy milyen gyakorlati érveket szolgáltat a kapcsolat domináns szereplője. Amennyi-

ben valamilyen létező, legitimnek tartott normára mutat rá, amely az ő parancsadási igényét alátámasztja, és az alávetett fél ezt elfogadja, akkor uralmi helyzetről beszélhetünk. Ez jelenti a döntő különbséget a hatalom és uralom között. Az uralmi igény alátámasztására mindig kell valamilyen norma, a hatalomhoz ez sosem szükséges. Ennek következményeként a hatalomgyakorlást mindig méltánytalannak érzi az azt elszenvedő fél, míg a normakövetésből eredő engedelmeskedni akarás nem jár együtt a méltánytalanságérzet kialakulásával.

Chapter 7

Közösségi döntések

A kő kicsorbítja az ollót.
Az olló elvágja a papírt.
A papír becsomagolja a követ.

Az egyéni döntések jelenségét a döntéselmélet, a racionalitáselmélet, a kilátáselmélet és még sok minden más diszciplina is vizsgálja, sőt, tulajdonképpen mondhatjuk azt, hogy a mikrogazdaságtan egésze is – legalább részben – ezzel a problémával foglalkozik. A társadalmi választások elmélete fókuszában a közösségi döntések vizsgálata áll. Az elmélet egyik fő kérdés az, hogy olyan helyzetekben, amelyekben nem az egyéni döntéseken van a hangsúly, hanem egyének által alkotott közösség döntésén, akkor hogyan lehet az egyéni döntéshozók egyéni döntéseiből egyetlen közösségi döntést előállítani, levezetni. Erre a talán legelterjedtebb módszer a szavazás technikájának alkalmazása, és nyilván ez magyarázza meg azt is, miért beszélnek, beszélünk társadalmi választások elméletéről (és mondjuk nem társadalmi sorshúzások elméletéről). A szavazás vagy választás technikáját mondhatjuk aggregálási módszernek is, ami megintcsak azt jelenti, hogy az egyéni döntésekből valamilyen összegzési, aggregálási eljárásra táamszkodva ki kell hoznunk egy egyértelmű közösségi végeredményt. Ebben a fejezetben az aggregálási technikák elméletét (és gyakorlatát) fogom vázlatszerűen bemutatni.

7.1 Aggregálás mint közösségi döntési technika

Amikor a közösségi döntésre gondolunk, elsőre valóban a választás/szavazás technikája ugrik talán mindenkinek, de az induláskor nem árt leszögezni, hogy azért nem ez az egyetlen módszer arra, hogy közösség kitermeljen magából egy közösségi döntést a közösség tagjainak egyéni teljesítményeire támaszkodva. Csak jelzésszerűen, anélkül, hogy ismertetném az egyes lehetőségeket, felsorolok pár aggregálási, közösségi döntési tech-

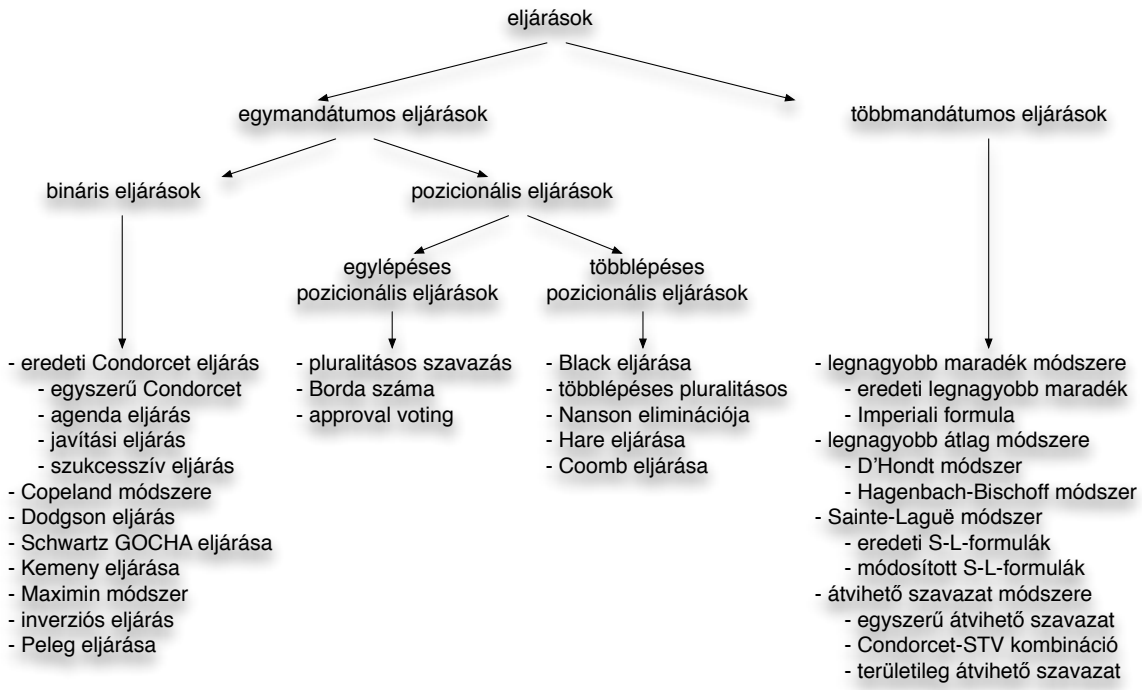
nikát. A közösségi döntési módszer alapulhat az egyének közti harcon (combat), versengésen, versenyen (contest) vagy a tagok közti alku(dozás) (bargaining) folyamatán. Dönthet egy közösség úgy, hogy a tagok közti eszmecserét, tárgyalást (discussion on merits) addig tartja fent, amíg egységes vélemény ki nem alakul közöttük, de gyorsabban elintézheti a döntést, ha a szavazás (voting) vagy a véletlen sorshúzás (chance) technikáját alkalmazza. Végül igazodhat egy közösség valamilyen autoritatív eljárás (authoritative determination) során született döntéshez is.

A társadalmi választások elméletének fókuszában az a folyamat áll, amelynek keretében (és eredményeként) az egyéni választásokból (döntésekből) közösségi választás (döntés) lesz valahogy. Közösségi döntést egyéni döntésekből többféleképpen lehet "csinálni". A történelmileg leggyakoribb megoldás a **döntési privilégium** kialakítása és fenntartása, amikor egy társadalmi norma kijelöli az emberek valamely körét, akiknek a kezébe teszi a közösség egészét érintő döntés jogát. Ezt a lehetőséget - bármennyire is fontos a történelemben - itt nem vizsgáljuk. A másik közösségi döntési mód az érintettek közötti **konszenzus** megteremtése, vagyis olyan mechanizmus kidolgozása és működtetése, amelynek révén mindenki ugyanazt a döntést fogadja el egységesen. Mivel ez a döntési mechanizmus nagyon erős feltételen nyugszik, hisz a konszenzus megteremtése sokszor nagyon idő- és energiaigényes folyamat, a módszer fontos eszköze a **vétó** technikájának alkalmazása. A teljes pozitív konszenzus eléréséhez képest gyengébb a negatív konszenzus, amikor ha nem is teljes mértékben értenek egyet a felek, de legalább nem ellenzik a vita során javasolt dolgokat. Hogy ilyen alkalmak során ne lehessen eltérni a konszenzus követelményétől, a vétójog az utolsó eszköz az egyet nem értők kezében. A konszenzusos döntések fontos típusát jelentik a **kompromisszum** alapú döntési eljárások, amelyekről Max Weber mint kompromisszumtestületekről ír [575]. A kompromisszum elve "gyengített" konszenzus elvként minősíthető, és akkor alkalmazzák, amikor azonos társadami szerepek mentén csoportosuló emberek meg kell, hogy egyezzenek egymással úgy, hogy a csoportok (vagy azok képviselői) között fenntartják a konszenzus elvét, de a csoport tagjai egyénileg valamennyien már nem nem vesznek részt a vitában. A kompromisszumtestületek működésére az érdekképviseleti szervek (állami szervek, munkaadók, illetve munkavállalók képviseleti szervei) közti tárgyalások szolgáltatják a legjob példát. Ilyenkor a döntéshozó testületnek olyan konszenzusos döntést kell hoznia, ami minden oldal számára elfogadható, ami szinte mindig valamilyen kompromisszum eredménye. A kompromisszum jelenségét találjuk meg a stratégiai szavazás, azon belül is a szavazatcsere intézménye mögött (↪449.o.). Amikor mindenki egyenrangú félként beleszól(hat) a közös ügyek megvitatásába, akkor már szavazótestületekről, **szavazás** alapú döntéshozatalról beszélhetünk. Itt mindig valamilyen többségi elv alapján születik a döntés úgy, hogy valamilyen szavazási eljárással összegzik az egyéni véleményeket egyetlen közösségi döntéssé. A továbbiakban erről a közösségi döntési módszerről lesz szó.

A szavazás elve

A szavazás tehát a közösségi döntéshozatal egyik fontos, bár nem kizárólagos módszere. A szavazásra jogosult szavazók leadják a szavazataikat és egy összegző eljárás segítségével közös döntést hoznak ki – ezt jelenti a szavazás fogalma. Az egész döntési folyamatban sok elemzendő kérdés van, amelyek közül párat később még tárgyalni fogok. Bármennyire is fontos kérdés, hogy pontosan milyen is a szavazási eljárás menete, hogyan

bonthatjuk fel összetevőkre a szavazás folyamatát, itt nem foglalkozom ezzel. Létezik egy "bonyolult", sok összetevős és sok lépésből álló folyamat, amely az emberek fejében levő sok-sok egyéni véleményt valahogyan egyetlen közösségi véleménnyé aggregálja. Ehhez először a véleményt szavazattá kell alakítani, aztán – olykor, ha szükséges – a véleményt polarizálni kell (eközben preferencia esetén azt kell figyelni, hogy az egyén mit preferál, faktuális vélemény esetén hogyan vélekedik). Előfordulhat az is, hogy a (binárisan polarizált) véleményeket súlyozni szükséges. A folyamat végén valamilyen algoritmus szerint összegezni kell az egyéni szintű adatokat. Sok technikai kérdést lehetne vizsgálni, de nem teszem. A teljes szavazási folyamatot le lehet írni pontos képletekkel, formulákkal. A választások elméletéről matematikailag nagyon jól megalapozott szakkönyvek születtek, de én kihagyom ezt a kérdéskört. Ami egy cselekvéselmélet számára fontos lehet, azt megemlítem itt-ott, a részletek pedig már nem az emberről szólnak, hanem a technikáról. Ebben a könyvben ez másodlagos. Azzal sem foglalkozom, hogy milyen szavazási eljárások, szabályok, algoritmusok léteznek. Alan D. Taylor vagy húsz szavazási szabályt mutat be alaposan a könyvében [539]. Korábban Mészáros József kollégámmal együtt mi is összeállítottunk egy könyvecskét erről, ott sok részlet olvasható e tárgyban [343]. E könyv alapján mutatok be egy összefoglaló ábrát a legismertebb választási eljárásokról.



Nem állítom, hogy az ábra kimerítően bemutatná az összes fontos algoritmust. Vannak, akik másként csoportosítanák az eljárásokat. Az egészből csak azt tartom lényegesnek, hogy az ábra azt mutatja, nincs is olyan sok szavazási eljárás.

Arra a kérdésre, hogy mit is kell aggregálni a közösségi döntések során, a legáltalánosabban adható válasz az, hogy **véleményeket**, amiből fakadóan a társadalmi választások elméletét tarthatjuk a véleményaggregálások elméletének is. A vélemények aggregálásának gyakorlati szempontból tekintve két ága, két fajtája létezik. Előfordulhat, hogy:

- egy személy több, különböző értékelési szempontok szerint képzett véleményhalmaza alapján kell egyetlen aggregált döntést hozni, vagy az, hogy
- több személy véleményhalmaza alapján kell egyetlen aggregált döntést hozni.

Természtesen az is lehetséges, hogy a két probléma egyszerre jelentkezik, amikor több személynek, több szempont szerint kell egyetlen döntést hoznia. Ezzel a lehetőséggel a továbbiakban nem foglalkozom. A két aggregálási feladat egységes modellben kezelhető. A különböző elméletek, modellkomponensek általában valamelyik problémakörhöz kapcsolódnak, és értelemszerűen magunkon viselik a terület terminológiáját. Ettől nem lehet teljesen elszakadni, ezért előfordulhat, hogy egyes fejezetekben egyik vagy a másik terület terminológiája, példái kerülnek túlsúlyba. Ez azonban sosem változtat az egységes modell lehetőségén, létezésén és az egységes kezelésmód iránti elköteleződésen.

Nem könnyű megmondani, mi is a különbség az **egyéni döntés** és a **közösségi döntés** között. Az azonosságot könnyebb megállapítani köztük. Mindkét esetben a döntés a lehetséges világállapotokból való választást jelenti. A döntés feltételezi, hogy van miből választani (vannak alternatívák), a döntés pedig ezt az alternativitást szünteti meg. Ha nem foglalkozunk azzal a kérdéssel, hogyan keletkezik (illetve hogyan változik) maga a döntés az emberben, akkor a legnyilvánvalóbb eltérés az egyéni és közösségi döntés között az, hogy az előbbiben egy, az utóbbi során több, mint egy ember hoz döntést.[1] Ha viszont több embernek kell döntést hoznia, akkor felmerül a kérdés, hogy milyen módon lehet a több egyéni döntésből egyetlen közösségi döntést előállítani, vagyis hogy lehet az egyéni döntéseket összegezni. Sokféle kifejezést használnak erre az összegzési folyamatra: **aggregálás**, kombináció, összehasonlítás (aggregation, combination, comparison). Az 'összegzés' teljesen megfelelne, de jóval elterjedtebbnek érzem az 'aggregáció' terminusát, ezért inkább ez utóbbit fogom használni. A közösségi döntés definitív mozzanata az egyéni döntések vagy **komponensdöntések** aggregálása.

Az egyéni döntésekhez képest a közösségi döntések elemzése nemcsak azért nehezebb, mert több ember döntésének vizsgálata értelemszerűen több feladatot kíván. A közösségi döntések vizsgálata teljesen újfajta problémákat hoz a felszínre. Később megpróbálom majd tisztázni, mi a különbség a vélemény- és a preferenciakinyilvánítás között. De amíg el nem jutok oda, addig is szükségesnek tartom jelezni, hogy a közösségi döntések világában el kell különíteni egymástól, hogy:

- milyen problémák, paradoxonok keletkez(het)nek a **véleményaggregálás** (judgement aggregation) folyamata során
- milyen társadalmi elvárásoknak, feltételeknek lehet (illetve nem lehet) egyszerre megfelelni a **preferenciaaggregálás** (preference aggregation) folyamata során
- milyen hatása van az **aggregálási eljárásoknak** a folyamat végeredményre
- milyen módon lehet biztosítani az egyéni megnyilatkozások őszinteségét (sincerely voting) vagy másként a csalásbiztosságot, illetve hogyan lehet kezelni (ha kell, elkerülni) a stratégiai vagy taktikai szavazást (strategic or tactical voting)

Kétféle véleményt érdemes megkülönböztetni. Vannak a preferenciakinyilvánítások, amelyek **összehasonlító ítéletek** (comparative judgements).[2] Ezek valamilyen értékelési szempont (evaluative point of view) szerint a lehetséges világállapotok mint alternatívák közül választanak ki egyet [361]. Ezt a problémahelyzetet lehet értelmezni úgy is, hogy az

[1] Ezen értelmezés mentén gondok adódhatnak a diktatórikus döntés jelenségének kezelésekor, de ezzel a kérdéssel itt most nem foglalkozom.

[2] Az összehasonlító ítéletek az értékítéletek alá sorolhatók, de ezt a tényt itt nem használjuk ki.

összehasonlítandó világállapotok mint alternatívák helyett a világállapotokat leíró propozíciók az alternatívák, így az ítélet ilyen esetekben is propozíciókból való választást jelent. A preferenciakifejezésekkel szembeállíthatóak a tényszerű ítéletek vagy **tényítéletek** (factual judgement), amelyek világra vonatkozó propozíciók közül való választást jelent. Utóbbi esetben a propozíciók igazsága vagy hamissága a kérdés, a választás az igaznak tartott állítások választását jelenti, a preferenciakinyilvánítás során pedig a jobb/rosszabb alternatívát leíró propozíciók közül választunk. Mindkét helyzetben közös viszont az, hogy az egyéni ítéleteken mellett gyakran igény merül fel közösségi vagy kollektív ítéletek (collective judgement) meghozatalára is. Ebben a fejezetben ennek problémáit tárgyaljuk. Mit jelent az egyéni vélemény, pontosabban: egyéni kinyilvánított vélemény? Annak kinyilvánítását, hogy az adott propozícióval (alternatívával) egyetért, nem ért egyet vagy tartózkodik vele szemben (mert nem akar vagy nem tud véleményt mondani). A kollektív vagy közösségi ítélet (vélemény) a közösség tagjainak egyéni véleménye alapján előállított, aggregált származtatott vélemény. Az aggregálás folyamatát nevezhetjük szavazásnak is. Nevezzük a véleményt (ítéletet, szavazatot) nyilvánítókat szavazóknak (döntéshozóknak, megnyilatkozóknak), akik kifejezik a véleményüket (ítélkeznek, szavaznak) az alternatívák halmazával kapcsolatban.

A társadalmi választások elméletét támogató logikát aggregációlogikának neveztem el (AL), aminek definiálásához az alábbiakra van szükség.

(AL.0) $\mathscr{L}_{AL}(\mathscr{L}_{PL}, \mathbb{S}, \mathbb{A}, \Phi, \mathbb{C}, \{x, y, z\}, \{\alpha, \beta, \gamma\}, \mathbf{F})$, ahol

$\mathscr{L}_{PL}$	propozícionális logika
$\mathbb{S}$	világállapotok halmaza
$\mathbb{N}$	ágensek halmaza
Φ	vélemények halmaza
$\mathbb{C}$	választások halmaza
$\mathbf{F}$	választási függvény

Az új nyelvünkbe több új predikátumot fel kell vennünk, hogy meg tudjuk ragadni az aggregálási folyamat legfontosabb komponenseit, úgymint a világállapotokat, amelyekre az ágenesek véleményei vonatkoznak, a választások halmazát vagy magát a választási függvényt. Vezessük be a következő jelöléseket.

(AL.1) $\mathbb{N} = \{1, \ldots, i, \ldots, |\mathbb{N}|\}$, ahol $|\mathbb{N}|$ a szavazók száma

szavazók halmaza (HUN) – set of voters (ENG) – set of individuals (ENG)

(AL.2) $\mathbb{X} = \{m, n, \ldots, o\}$

választható (világ)állapotok (HUN) – conceivable (social) state (ENG)

(AL.3) $\phi_i \equiv \mathbf{R}_i \equiv \mathbf{P}_i \equiv \mathbf{I}_i$

egyéni preferenciarendezés (HUN) – egyéni vélemény (HUN) – voter's preference (ENG) – individual's opinion (ENG)

(AL.4) $\{\mathbf{R}_i\}_{i\in\mathbb{N}} \stackrel{\text{def}}{=} \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{|\mathbb{N}|}\} \equiv \{\phi_i\}_{i\in\mathbb{N}} \equiv \{\phi_1, \phi_2, \ldots, \phi_{|\mathbb{N}|}\}$

egyéni preferenciarendezések halmaza (HUN) – véleményhalmaz (HUN) – set of voters' preferences (ENG) – set of individuals' opinions (ENG)

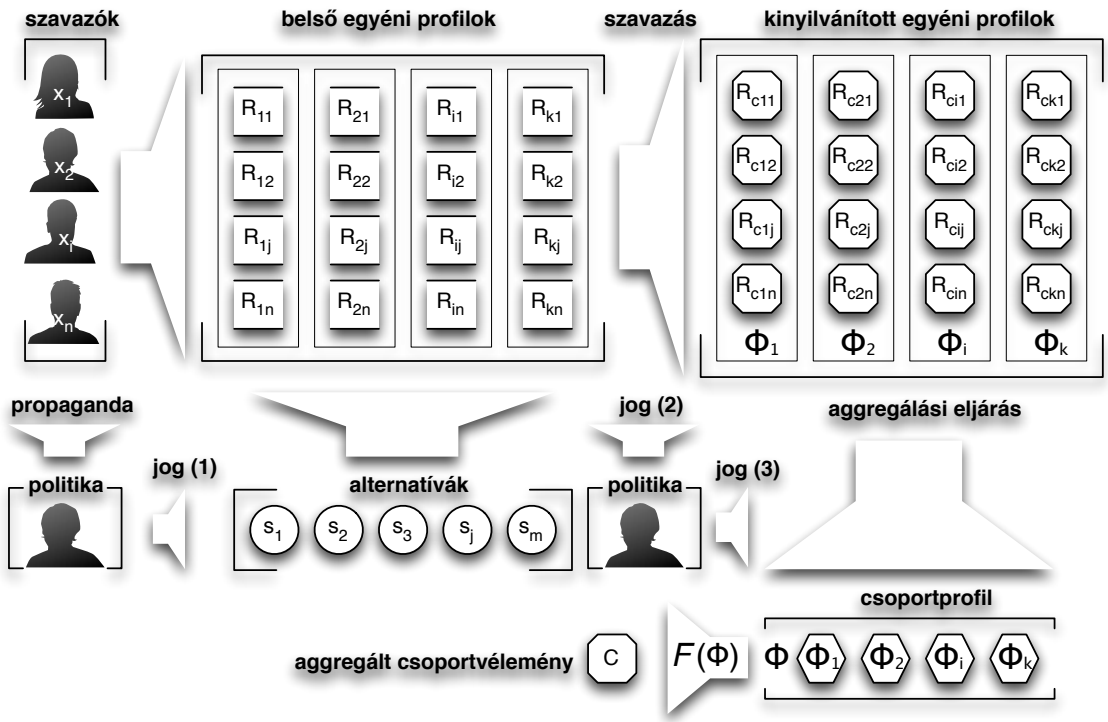
(AL.5) $a \stackrel{\text{def}}{=} \langle \mathbf{R}_1^a, \ldots, \mathbf{R}_i^a, \ldots, \mathbf{R}_{|\mathbb{N}|}^a \rangle \equiv \langle \mathbf{R}_i^a \rangle_{i\in\mathbb{N}} \equiv \Phi \equiv \langle \Phi_i \rangle_{i\in\mathbb{N}} \equiv \langle \phi_k, \ldots, \phi_m \rangle$

profil (HUN) – preferenciaprofil (HUN) – véleményprofil (HUN) – profile (ENG)

(AL.6) $\mathbf{F}(a) \equiv \mathbf{F}(\langle \Phi_i \rangle_{i \in \mathbb{N}}) \equiv \mathbf{F}(\langle \mathbf{R}_i^a \rangle_{i \in \mathbb{N}}) \stackrel{\text{def}}{=} C(\langle \mathbf{R}_i^a \rangle_{i \in \mathbb{N}}) \equiv C^a \equiv \mathbf{F} : \mathbb{D}_F \mapsto \mathbb{X}$, ahol $\mathbb{D}_F \subseteq \mathbb{X}^{|\mathbb{N}|}$

közösségi vélemény függvény (HUN) – csoportdöntési fügvény (HUN) – group decision function (ENG) – társadalmi jólét függvény (HUN) – social welfare function (ENG) – preferenciaaggregálási függvény (HUN) – preference aggregation function (ENG) – constitution (ENG) – közösségi preferencia (HUN) – social preference (ENG) – kollektív választás(i szabály) (HUN) – collective choice (rule) (ENG) – kollektív döntés(i szabály) (HUN) – collective decision (rule) (ENG) – allokációs mechanizmus (HUN) – allocation mechanism (ENG) – (társadalmi) választási függvény (HUN) – (social) choice function (ENG) – társadalmi döntési módszer/függvény (HUN) – social decision method/function (ENG) – szavazási rendszer (HUN) – voting system (ENG) – voting scheme (ENG) – voting mechanism (ENG) – social choice procedure (ENG)

Bár a közösségi döntések elméletében mindig abból indulunk ki, hogy a közösségi döntést az egyéni döntések, preferenciarendezések határozzzák meg, az elmélet középpontjában nem annyira az egyéni rendezések, hanem inkább az egyéni rendezések, egyéni véleménynyilvánítások halmaza, de még inkább rendezett struktúrái állnak. Ezeket nevezzük **profiloknak**, és a későbbiekben ezekre hivatkozva tudjuk megadni a legfontosabb fogalmak definícióját. A bemutatandó elmélet javarésze a preferenciaaggregálások kérdéseivel foglalkozik, ezért a legtöbbször a preferenciarendezésekre utaló jelöléseket fogom használni, de a végefelé szóba kerülnek a véleményaggregálás problémái, és akkor majd váltani fogok egy általánosabb jelölésmódra. Az alapfogalmak bevezetése után érdemes egy közös ábrában is feltüntetni a most jellemezni kívánt aggregálási folyamat legfontosabb összetevőit, az ezek közti kapcsolatokat.



A helyzet legfontosabb kapcsolata az egyének és a világállapotok közti kapcsolat, amelyről itt semmit sem tudunk mondani. Tényként vesszük itt tudomásul, hogy egy adott

csoport tagjai kifejezik a véleményüket a világállapotok valamely halmazáról. Attól függően, hogy melyik véleményaggregálási helyzetben vannak, más lesz az egyéni vélemények típusa. Az egyéneknek véleményük lehet a világállapotok kívánatosságáról, amit preferenciarendezés formájában fejezhetnek ki, de véleményük lehet a világállapotokra vonatkozó faktuális kijelentés is. Ezek az egyéni vélemények – alapértelmezés szerint – még őszinték, viszont nem nyilvánosak abban az értelemben, hogy az ágensek nem nyilvánították ki őket valamilyen szavazási aktus keretében. Ha ez a szavazás megtörténik, akkor beszélhetünk az ágensek kinyilvánított véleményeiről, és ezeket az ágensekhez rendelve megkapjuk a külső egyéni profilokat. A szavazás aktusában lehet megragadni a stratégiai szavazás jelenségét. A külső egyéni profilok alapján lehet létrehozni a csoportprofilt, majd ez alapján a közösségi döntés végeredményét. Az aggregálási eljárás – szűkebb értelemben vett – fogalma ezt a véleményössszesítési folyamatot ragadja meg. Ez az ábra az aggregálási folyamat legáltalánosabb értelmezését jelenti. Magába foglalja az aggregálási feladatok mindkét ágát, a vélemény- és a preferenciaaggregálás jelenségét. A következő két nagyobb fejezetben bemutatom mindkét területet, előbb a preferenciák, majd a vélemények aggregálásának mikéntjét.

Preferenciaprofilok tipológiája

Ha az egyéni preferenciák helyett társas vagy közösségi preferenciákat vizsgálunk, akkor az egyéni preferenciák valamely összességét preferenciaprofilnak hívjuk. Ezek a preferenciaprofilok rendelkezhetnek olyan tulajdonságokkal, amelyekre támaszkodva fontos és érdekes tételek vezethetők le. Ebben a fejezetben ezeket a speciális mintázattal rendelkező profilokat mutatom be, de itt nem foglalkozom a belőlük levezethető tételekkel. Amikor preferenciaprofilokat jellemzünk, akkor nemcsak az összahasonlítandó alternatívákra vonatkoztatjuk az állításainkat, de a minősítésbe bevonjuk a preferenciarendezéseket végrehajtó személyeket is. Ezt persze interpretálhatjuk úgyis, hogy ilyenkor magukra az egyéni preferenciarendezésekre mondunk ki állításokat, ami logikai szempontból azt jelenti, hogy másodrendű logikai konstrukciókat kell kezelnünk (↬55.o.), hiszen relációk (predikátumok) tulajdonságaira, kapcsolataira kell a kvantorainkat vonatkoztatni.

Az első elemzett preferenciaprofil az egycsúcsú preferencia volt, amivel kapcsolatban fontos, ún. lehetségességi tételeket mondtak ki a társadalmi választások elméletén belül (↬439.o.). Ken-ichu Inada volt az, aki először tett kísérletet egy – bizonyos szemponból – kimerítőnek minősíthető preferenciaprofil-tipológia kidolgozására [251], [252]. Inada 1969-es cikkében véglegesített tipológiáját befolyásolta Amartya K. Sen javaslata, amely eredményeként egy általánosabb kategória alá be lehetett sorolni Inada legfontosabb típusait, az ún. értékkorlátozott preferenciákat [474], [484]. Inada profiltipológiáját később – egy új szempont figyelembe vételével – egy további típussal kibővítették, viszont az inadai típusok közül többeket elhagytak idővel, és inkább csak a Sen által is fókuszált értékkorlátozott profilokkal foglalkoztak alaposabban. Az egyes preferenciaprofilok ismertetésekor kezdetben igazodom Inada logikájához, de később eltérek az általa követett gondolatmenettől, és a legfontosabb, a mai napig elemzett típusok bemutatásakor már a későbbi szerzők megközelítését veszem át.

Inada preferenciaprofiljai

Ken-ichu Inada először három olyan preferenciaprofilt mutatott be, amelyek tetszőleges számú szavazó esetén tranzitív közösségi preferenciát képesek adni az egyszerű többségi döntési eljárással. A preferenciaprofil-típusok leírásához Inada abból a kérdésből indult ki, hogy elméletileg milyen egyéni preferenciarendezések lehetségesek. Ha adva van három alternatívánk (m, n, o), akkor azokat tizenháromféleképpen lehet preferenciasorba állítani. Hatféleképpen rangsorolhatjuk őket úgy, hogy szigorú rendezést feltételezünk mindhárom alternatíva között, három-háromféle sorrend lehetséges akkor, ha egy alternatívát jobbnak, illetve rosszabbnak tartunk a másik kettő, egymással közömbös alternatívához képest, valamint egyféleképpen fordulhat elő az a helyzet, amikor mindhárom alternatíva közömbös egymással szemben. A szemléletesebb megjelenítés miatt kivételesen eltérve az eddigi gyakorlattól, a preferenciarelációk argumentumait infix-jelölésmódban kezelve, az alábbi táblázatban mutatom be a fent kifejtett lehetőségeket (csak a példa kedvéért jelzem, hogy az infix jelölésmód alkalmazásával a következő két formula ekvivalens egymással: $\mathbf{P}(m, n) \wedge \mathbf{P}(n, o) \equiv m\mathbf{P}n\mathbf{P}o$.

$m\mathbf{P}n\mathbf{P}o$	$m\mathbf{P}n\mathbf{I}o$	$m\mathbf{I}n\mathbf{P}o$	$m\mathbf{I}n\mathbf{I}o$
$m\mathbf{P}o\mathbf{P}n$	$n\mathbf{P}o\mathbf{I}m$	$n\mathbf{I}o\mathbf{P}m$	
$n\mathbf{P}m\mathbf{P}o$	$o\mathbf{P}m\mathbf{I}n$	$o\mathbf{I}m\mathbf{P}n$	
$n\mathbf{P}o\mathbf{P}m$			
$o\mathbf{P}m\mathbf{P}n$			
$o\mathbf{P}n\mathbf{P}m$			

Mivel az indifferencia reláció szimmetrikus, így az argumantumai felcserélhetőek egymással, a második és harmadik oszlopban szereplő konfigurációkban az indifferenciareláció két oldalán levő argumentumok szerepelhetnének fordítottt sorrendben is. A táblázat az egyéni rendezések lehetséges mintázatait mutatja. A szavazás során a szavazók ezen rendezésminták bármelyikét választhatják, így előfordulhat az az eset, amikor mind a tizenhárom különböző rendezésminta szerint szavazott legalább egy szavazó. Amikor tehát az egyéni preferenciák aggregálását kell modelleznünk, akkor azt kell feltételeznünk, hogy a fenti tizenhárom mintázat mindegyike szerint szavaznak valamennyien, és ezek alapján kell az összesítést elvégeznünk. A kérdés az, hogy vannak-e olyan konstellációk a rendezésminták összessége között, amelyekre képesek lehetnek a többségi szavazat konzisztenciáját, tranzitivitását boztosítani. Ken-ichi Inada ezen lehetőségeket vizsgálva különített el különböző típusú preferenciákat [252]. Az inadai preferenciaprofil-típusok mind olyanok, amelyek bizonyos feltételek mentén biztosítani képesek a többségi eljárás konzisztenciáját. Fontos hangsúlyozni, hogy itt nem egyéni preferenciákról van szó, hanem egyéni preferenciatípusok adott konfigurációjáról, vagyis preferenciaprofilokról.

Amennyiben feltételezzük, hogy a szavazói populáció olyan, hogy az egyéni preferenciák között nem fordulnak elő szigorú rendezések (vagyis a fenti táblázat első oszlopában látható rendezésminták szerint senki sem szavaz), akkor Inada szerint **dichotóm preferenciáról** beszélhetünk. Ilyenkor a szavazók mindegyikére igaz, hogy az alternatívákat két csoportba osztva rendezik el úgy, hogy az egyes csoportokon belüli alternatívák között közömbösek. Az alternatívák így két csoportra oszlanak el és a szavazók így valóban dichotóm módon választ(hat)anak egyik vagy másik csoportból maguknak. Mielőtt megmutatnám, hogy milyen formula segítségével lehet leírni a dichotóm preferenciaprofil

fogalmát, szükséges egy megjegyzést tenni Inada jelölési megoldására vonatkozóan. A következő formulákban át fogok térni az alternatívákat jelző m, n, o változók helyett az x, y, z változók használatára, hogy egyértelműen jelezhessem, a szavazási minták eltérhetnek egymástól a konkrét értékeiket tekintve, miközben a mintázatukat tekintve azonos rendezési logikát mutatnak. Az $m\mathbf{P}n\mathbf{P}o$ formula így a fenti táblázat első oszlopának mind a hat elemét reprezentálhatja. Ezek alapján a dichotóm preferencia formuláját a következőképpen írhatjuk fel.

(AL.7) $(\text{DIP}) \stackrel{\text{def}}{=} \forall x \forall y \forall i(\neg(\mathbf{P}_i(x,y) \wedge \mathbf{P}_i(y,z)))$

dichotóm (preferencia) (HUN) – dichotomous preference (ENG)

Inada kimondott egy tételt az ilyen preferenciaprofil esetére vonatkozóan, de ezt később mutatom majd be (↰439.o.). A következő preferenciaprofilokat már úgy kapjuk meg az inada építkezés során, hogy feltételezzük azt, hogy előfordulhatnak szigorú egyéni preferenciarendezések (vagyis a preferenciaminták között szerepelnek a fenti táblázat első oszlopából). Ez a feltétel önmagában azonban még nem szolgáltat érdekes mintázatot számunkra, tehát további feltételeket kell megfogalmaznunk. Az egyik ilyen pluszfeltétel lehet, hogy kikötjük azt, hogy amikor fennáll egy adott szigorú hármas preferenciarendezés, akkor kizárjuk a rendezés két végén álló alternatíva közti fordított irányú szigorú rendezés létezését. A fenti táblázat alapján ez azt jelenti, hogy az első oszlop mintái közül megengedjük azokat, amelyek az egyik alternatíva (x) szigorúan jobb, mint a másik (z), de fordítva nem áll fent a kapcsolat köztük, vagyis az első három sorban levő típusok létezését engedjük csak meg. Inada ezt a típust nevezte el **visszhangzó preferenciának**. A kifejezéssel arra akart utalni, hogy ha van egy szavazó, aki tranzitív preferenciákkal rendelkezik és a többiek is követik (visszhangozzák) őt ebben, akkor nyilvánvalóan teljesen a közösségi szavazat tranzitivitása is. Ezt a triviális esetet lehet általánosítani a most megfogalmazott feltétellel, amit a következő módon formalizálhatunk.

(AL.8) $(\text{ECP}) \stackrel{\text{def}}{=} \forall x \forall y \forall z \forall i(\mathbf{P}(x,y) \wedge \mathbf{P}_i(y,z) \rightarrow \neg \mathbf{P}_i(z,x))$

visszhangzó (preferencia) (HUN) – echoic preference (ENG)

A következő preferenciaprofil felvezetéséhez Inada újracsak egy triviális példát hozott értelmezési mintául. Ha a szavazó közösség két diszjunkt csoportra osztható a szavazók "ízlése" szerint, azaz amíg az egyik csoport tagjai $x\mathbf{P}y\mathbf{P}z$ sorrend szerint szavaznak, addig a másik csoportba tartozók a $z\mathbf{P}y\mathbf{P}x$ mintát követik, akkor a többségi eljárás konzisztens közösségi végeredményt képes kihozni. Ezt a triviális esetet úgy lehet általánosítani, hogy amennyiben vannak, aki az $x\mathbf{P}y\mathbf{R}z$ sorrend szerint preferálják az alternatívákat, akkor vagy vannak olyanok is, akik a fordított rendezési sort preferálják ($z\mathbf{P}y\mathbf{R}x$), vagy vannak, akik a két szélső értéket tekintve közömbösek ($x\mathbf{I}z$). Ha ez így van, akkor – Inada javaslata alapján – **antagonisztikus preferenciáról** beszélhetünk, ami a következő formulával fejezhetünk ki.

(AL.9) $(\text{ANP}) \stackrel{\text{def}}{=} \forall x \forall y \forall z \forall i((\mathbf{P}_i(x,y) \wedge \mathbf{P}_i(y,z)) \rightarrow (\mathbf{P}_i(z,y) \wedge \mathbf{P}_i(y,x)) \vee \mathbf{I}_i(x,z))$

antagonisztikus (preferencia) (HUN) – antagonistic preference (ENG)

A dichotóm, a visszhangzó és az antagonisztikus preferenciaprofilok fogalmával szemben Inada úgy vezette be a **tabu preferencia** fogalmát, hogy jelezte, hogy utóbbi létezése

esetén csak páratlan számú szavazó esetén remélhetünk konzisztens végeredményt a többségi eljárás alkalmazása során. A tabu preferencia lényege az, hogy egyfajta "preferenciaminta" nem fordulhat elő. Ezt azzal a kettős feltétellel vélte megragadhatónak Inada, hogy kizárta az alternatívák közti teljes közömbösség, valamint a kizárandó preferenciarendezés lehetőségét az egyéni preferenciarendezések köréből. Ezt az alábbi módon írhatjuk le.

(AL.10) $(\text{TAP}) \stackrel{\text{def}}{=} \forall x \forall y \forall z \forall i (\neg(\mathbf{I}_i(x,y) \wedge \mathbf{I}_i(y,z)) \wedge \neg \mathbf{P}_i(y,x))$

tabu (preferencia) (HUN) – taboo preference (ENG)

Az első három bemutatott inadai preferenciaprofil tetszőleges számú, a negyedjére elemzett mintázat csak páratlan számú szavazó esetén képes konzisztens végeredményt produkálni. Az első három, illetve a negyedik típus között elemezte Inada azokat a preferenciamintákat, amelykre későbbb Sen az értékkorlátozott minősítést alkalmazta. Mivel az idők során ezek a preferenciarendezések kerültek igazán az elemzői figyelem fókuszába, ezeket önállóan, a következő fejezetben mutatom be.

Értékkorlátozott preferenciaprofilok

A legelőször vizsgálatba vont és azóta is a legtöbbet hivatkozott preferenciaprofil-típus az egycsúcsú preferencia. A fogalmat Duncan Black vezette be 1948-ban [60], Arrow is hivatkozott rá 1951-es könyvében [25], majd utánuk még sokan mások is használták a kategóriát. Ha a szigorúan rendezett alternatívák halmazában van olyan elem, amely minden más elemhez képest jobbnak minősül, és a többi alternatíva értékelése (rendezése) speciális mintázat szerint, montonon módon csökkenve változik a maximális elemtől két irányban távolodva adott skálán mérve, akkor beszélünk **egycsúcsú preferenciáról**. Ekkor a lineárisan rendezett alternatívákra vonatkozóan a szavazók preferenciái vagy mindig szigorúan növekvő vagy szigorúan csökkenő vagy egy pontig először szigorúan növekvő, majd a ponttól szigorúan csökkenő függvényképet mutatnak.

A fogalom értelmezéséhez segítséget jelenthet, ha bemutatok néhány példát az egycsúcsú preferenciákkal kapcsolatban. Az egyik példa hétköznapi. Ha egy nagy teremben több ágy van sorban egymás mellé rakva (mint egy laktanya körletében), és csak egy melegítőtest van, akkor minden egyes ágy birtokosának egycsúcsú a preferenciája, vagyis mindenki meg tudja mondani, hogy hol lenne számára a hőforrás ideális helye. Ez a hely természetesen mindenki esetében más és más pont lehet a teremben, ami részben a terem hossztengelye mentén elfoglalt pozíciótól, részben az ágytulajdonos "hőigényétől", fázósságától függ. Akik jobban szeretik a meleget, azok közelebb szeretnének a kályhához lenni, de lehet egy olyan pont, aminél közelebb már nem lenne annyira jó nekeik, mert túlzottan sok meleg jönne rájuk. Ilyen helyzetben mindenki meg tudja valahogyan határozni azt, hogy neki mi lenne a hőforrás ideális elhelyezése, és az mindenki számára egycsúcsú preferencia elrendezést jelentene a lehetséges helyek halmazán.

Gyakran hivatkoznak az egycsúcsú preferenciákra a közgazdaságtanban, illetve a politikai szociológia világában. Az érdeklődést Harold Hotelling híres cikke váltotta ki először [244], amiben a szerző egy tengerparti sétányon vagy egy kisváros főutcáján működő boltok elhelyezkedéséről állított fel egy meglepő és igencsak "erős" tézist. Hotelling modellezte az a jólismert tényt, hogy amikor két konkurrens boltos helyet keres magának a város főutcáján (vagy az üdülőövezet sétányán), akkor mindketten a főutca (sétány)

közepén fogják megnyitni a boltjukat. A magyarázat egyszerű. Ha feltételezzük, hogy a verseny miatt a boltosok minden minőségükben (árban, kínálatban, kiszolgálásban stb.) megegyeznek, akkor a potenciális vevőknek csak az marad meg egyetlen döntési szempontként, hogy melyik bolt van közelebb hozzájuk (a lakásukhoz, nyaralójukhoz). A Hotelling tétel szerint a boltosok akkor remélhetik a legnagyobb forgalmat, ha a lehető legtöbb vevőnek a lehető legkisebb sétálnivalója lesz, ezt pedig akkor érhetik el, ha valamennyien az utca, a sétány közepére települnek. Az tétel azért érdekes számunkra, mert a magyarázat erősen támaszkodik az egycsúcsú preferencia létezésére (illetve annak tulajdonságaira). A szavazói preferenciák egycsúcsúságával magyarázzák a politikai pártok medián szavazókra optimalizáló viselkedését, és hasonló okok miatt kap fontos szerepet ez a fogalom a társadalmi választások elméletén belül a lehetségességi tételek között (↬439.o.). Az egycsúcsú preferenciák jellegzetessége az, hogy van egy maximális elem és a többi alternatíva ehhez képest egyre rosszabbnak minősül, bármelyik irányban is távolodunk a maximális elemtől. Ennek a speciális mintázatnak elképzelhetjük egyfajta tükörképét is, amikor van egy minimális elem, és minden más elem ehhez képes egyre jobbra értékelt, minél jobban távolodunk a minimális elemtől mindkét irányban. Az ilyen preferenciát Ken-ichi Inada **egymélypontú preferenciának** (vagy egyvölgyű preferenciának) nevezte el [251].

Az egymélypontú preferencia fogalmával kifejezhetjük például a "közrosszhoz" való viszonyunkat. Amikor olyan döntést kell hozni, amit nem nagyon támogatnak az érintettek (hol épüljön börtön, menekülttábor, hulladéktároló telep, szemétfeldolgozó üzem stb.), akkor a preferenciáinkat nem a jobb, hanem inkább a rosszabb alternatívák jelzésével szerint fejezzük ki. William Vickrey már Inada cikke előtt is utalt erre a preferenciatípusra, bár ő egymélypontú (single troughed) preferenciának nevezte el [561]. Úgy tűnik, hogy az inadai 'single-caved' kifejezés helyett végül inkább a 'single-dipped' nyelvi szerkezet terjedt el a szakirodalomban [279], [280], [35].

Inada egy harmadik profiltípust is definiált, amit ugyanazzal a módszerrel határozott meg, mint amivel az egymélypontú preferencia fogalmát is "elválasztotta" az egycsúcsú preferencia fogalmától. Ekkor azonban a legnagyobb (legjobb) elem helyett nem a legkisebb (legrosszabb) elem cseréjét javasolta, hanem ehelyett inkább a hármas összehasonlítás harmadik elemére fókuszált az új fogalom definiálásakor. A **szeparált preferenciának** elnevezett új típus lényege az, hogy az alternatívák közül a szavazók mindig a két szélsőértéket hasonlítják össze vagyis minden esetben figyelmen kívül hagyják a "középső értéket". A három preferenciaprofilt a következőképpen definálhatjuk.

(AL.11) $$(\text{SPP}) \stackrel{\text{def}}{=} \forall m \forall n \exists! o \forall i \big(((m < n \wedge n \leq o) \rightarrow \mathbf{P}_i(n,m)) \wedge ((m > n \wedge n \geq o) \rightarrow \mathbf{P}_i(n,m)) \big)$$

egycsúcsú preferencia(profil) (HUN) – egycsúcspontú preferencia(profil) (HUN) – single-peaked preference (profile) (ENG)

(AL.12) $$(\text{SDP}) \stackrel{\text{def}}{=} \forall m \forall n \exists! o \forall i \big(((m < n \wedge n \leq o) \rightarrow \mathbf{P}_i(m,n)) \wedge ((m > n \wedge n \geq o) \rightarrow \mathbf{P}_i(m,n)) \big)$$

egymélypontú preferencia(profil) (HUN) – egyvölgyű (HUN) – egyvermű (HUN) – single-caved (ENG) – single-dipped preference (profile) (ENG) – single troughed (ENG)

(AL.13) $$(\text{GSP}) \stackrel{\text{def}}{=} \forall m \forall n \exists \mathbb{M}_1 \exists \mathbb{M}_2 \forall i \big(m \in \mathbb{M}_1 \wedge n \in \mathbb{M}_2 \wedge \mathbb{M}_1 \cap \mathbb{M}_2 = \varnothing \wedge \mathbb{M}_1 \cup \mathbb{M}_2 = \mathbb{X} \wedge \mathbf{R}_i(m,n) \vee \mathbf{R}_i(n,m) \big)$$

(szeparált preferencia) (HUN) – (two-)group-separated preference (ENG) – group-separable preference (ENG)

Inada preferenciafogalmainak vagyis az egycsúcsú, az egymélypontú és a szeparált preferencia általánosításaként Amartya Sen javasolta használatba venni az **értékkorlátozott preferencia** (value-restricted preference) fogalmát [474], [484]. Sen abból indult ki, amit már Inada is észrevett (sőt, a tételeiben ki is használt), hogy Arrow a lehetségességi tételeinek bizonyításakor (↰439.o.) valójában csak hármas összehasonlításokkal operált, vagyis nem használta ki azt, hogy a tétel feltételei az összes alternatívára "egyszerre" vonatkozzanak. Sen az inadai három preferenciatípust arra a közös vonásra hivatkozva vonta egy felettes kategória alá, hogy azok az alternatívák hármas összehasonlításakor mindig kizárják valamelyik elemet. Három elem (m, n, o) viszonylatában ki lehet hagyni a legjobb, a legrosszabb és a középső elemet, és ennek megfelelően beszélhetünk **felső-kizáró** (best-restricted), **alsó-kizáró** (worst-restricted) és **közép-kizáró** (medium-restricted) preferenciáról.

Az egycsúcsú preferenciák Sen jellemzésében az alsó-kizáró típusba tartoznak, hiszen ebben az esetben nincsen legrosszabb elem egyik szavazó összehasonlításában sem. Az egyvölgyű preferenciáknak a felső-kizáró preferenciaprofil felel meg, hiszen senki preferenciarendezésében sem szerepel a legjobb elem. A szeparált preferencia esetében pedig közép-kizáróságról beszélhetünk, amikoris a középső alternatívát hagyja ki az összes szavazó a rendezéséből. Sen generikus alá-fölérendeltséget állapított meg az inadai típusok és saját értékkorlátozott preferenciaprofiljai között, de nem szabad elfeledkeznünk arról, hogy ez a kapcsolat "egyirányú". Ez azt jelenti, hogy minden egyvölgyű (egymélypontú, illetve szeparált) preferencia felső-kizáró (alsó-kizáró, illetve közép-kizáró) profilú, de a fordított kapcsolat nem áll fent. A kérdés az, hogy milyen további feltételekkel lehet egyértelműen meghatározni az inadai típusokat a Sen-féle értékkorlátozott mintázatokból. Miguel A. Ballester és Guillaume Haeringer erre a kérdésre adott választ további feltételek kidolgozásával [32]. Ballester és Haeringer egyfelől fölöslegesnek tartotta az alternatívák lineáris rendezését megkövetelő feltételt mondván, sok esetben nem lehet értelmesen interpretálni az alternatívák szigorú rendezettségét, másfelől bebizonyította, hogy két tulajdonság feltételezésével egyértelműen – és az alternatívák rendezettségére való hivatkozás nélkül – meghatározható az egycsúcsú (egymélypontú, szeparált) preferenciaprofil fogalma. Ballesterék szerint a Sen által bevezetett értékkorlátozott típusokra úgy tekinthetünk, mint amelyek szükséges feltételt jelentenek a három inadai profil definiálásához. Az elégséges feltételhez a szerzők új tulajdonságokat $(\alpha, \alpha^{-1}, \beta)$ definiáltak. Az α-tulajdonság akkor érvényesül, ha két szavazónak három alternatívára vonatkozóan teljesen ellentétes a preferenciarendezése, de egy negyedik alternatívát mindketten jobbnak tartanak a hármas alternatívasor középső eleménél. Az inverz α-tulajdonság csak annyiban különbözik ettől, hogy a negyedik alternatívát rosszabbnak tartja mindkettő szavazó a hármas sor középső eleménél. A β-tulajdonsággal azt az helyzetet írhatjuk le, amelyben egyrészt két tetszőleges szavazó által legrosszabbnak, illetve legjobbnak tartott alternatívák különböznek egymástól, másrészt amíg az egyik szavazó két szélsőértékének sorrendjében megegyeznek a szavazók, addig a másik szavazó két szélsőértékének sorrendjében eltérően értékelik. A tulajdonságok formulái a következők.[3]

(AL.14) $\alpha\text{-RESTRICTED} \stackrel{\text{def}}{=} \exists m \exists n \exists o \exists r \exists i \exists j \big((m \mathbf{P}_i n \mathbf{P}_i o \wedge n \mathbf{P}_i r) \wedge (o \mathbf{P}_j n \mathbf{P}_j m \wedge n \mathbf{P}_j r) \big)$

α-tulajdonság (HUN) – α-property (ENG)

[3] A formulák egyszerűsítése végett a preferenciarelációt itt is infix-módban és a lineáris rendezéseket "egybe" írva jelzem. Az itt bemutatott formulákat részben Robert Bredereck, Jiehua Chen és Gerhard J. Woeginger cikke alapján adom meg [77].

(AL.15) α^{-1}-RESTRICTED $\stackrel{\text{def}}{=} \exists m \exists n \exists o \exists r \exists i \exists j \big((m\mathbf{P}_i n \mathbf{P}_i o \wedge r\mathbf{P}_i n) \wedge (o\mathbf{P}_j n \mathbf{P}_j m \wedge r\mathbf{P}_j n) \big)$

inverz α-tulajdonság (HUN) – converse α-property (ENG)

(AL.16) β-RESTRICTED $\stackrel{\text{def}}{=} \exists m \exists n \exists o \exists r \exists i \exists j \big((m\mathbf{P}_i n \mathbf{P}_i o \mathbf{P}_i r) \wedge (n\mathbf{P}_j r \mathbf{P}_j m \mathbf{P}_j o) \big)$

β-tulajdonság (HUN) – β-property (ENG)

Mindhárom fenti tulajdonságot összekapcsolhatjuk valamelyik érték-korlátozott preferenciaprofillal, de ezt tagadó módon tehetjük meg. Ha ugyanis azt tapasztaljuk, hogy valamelyik tulajdonság igaz egy "ismeretlen" preferenciaprofilra nézve, akkor azt mondhatjuk, hogy ez a profil nem tartozhat abba a típusba, amit hozzárendeltük a szóban forgó tulajdonságot. Ezt persze azt is jelenti egyben, hogy az inadai három preferenciaprofil-típusát definiálhatjuk úgy is, hogy a meghatározásaikba belevesszük a velük összekapcsolt tulajdonságok tagadását.

(AL.17) (SSP) ≡ SINGLE-PEAKED ↔ WORST-RESTRICTED ∧ ¬α-RESTRICTED

Ballester-Haeringer egycsúcsú preferencia tétele (HUN) – Ballester-Haeringer's theorem on single-peaked preference profile (ENG)

(AL.18) (SDP) ≡ SINGLE-DIPPED ↔ BEST-RESTRICTED ∧ ¬α^{-1}-RESTRICTED

Ballester-Haeringer egymélypontú preferencia tétele (HUN) – Ballester-Haeringer's theorem on single-dipped preference profile (ENG)

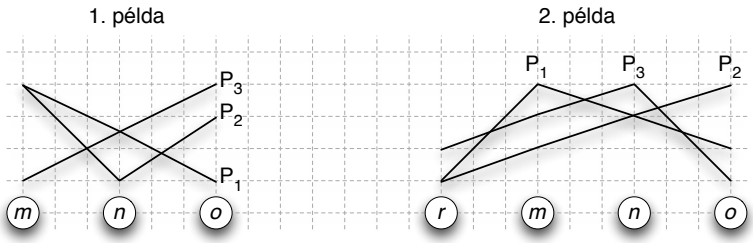
(AL.19) (GSP) ≡ GROUP-SEPARATED ↔ MEDIUM-RESTRICTED ∧ ¬β-RESTRICTED

Ballester-Haeringer szeparált preferencia tétele (HUN) – Ballester-Haeringer's theorem on group-separated preference profile (ENG)

Ezzel az újraértelmezéssel könnyebben felismerhetjük az egycsúcsúság (egymélypontúság, szeparáltság) tulajdonságát, amikor egy "ismeretlen" preferenciaprofilt reprezentáló adatokra tekintünk [459]. Az alábbi két példára nézve "könnyen" megállapíthatjuk, hogy az első táblázat nem elégíti ki az egycsúcsúság feltételét, hiszen mindhárom alternatíva szerepel a legrosszabb helyen valamelyik egyéni preferenciarendezésben. Ezzel szemben a második tábla adatai alapján ráismerhetünk az egycsúcsú preferenciaprofilra, ha észrevesszük, hogy van olyan elem (n), amely sosincs a legrosszabb pozícióban a hármas összehasonlítások során (praktikusan ez azt jelenti, hogy a 2. példa adatait "oszlopszerűen" nézve az alternatívák utolsó két oszlopában nem fordul elő az n alternatíva).

1. példa	2. példa
$m\mathbf{P}_1 n\mathbf{P}_1 o$	$m\mathbf{P}_1 n\mathbf{P}_1 o\mathbf{P}_1 r$
$m\mathbf{P}_2 o\mathbf{P}_2 n$	$o\mathbf{P}_2 n\mathbf{P}_2 m\mathbf{P}_2 r$
$o\mathbf{P}_3 n\mathbf{P}_3 m$	$n\mathbf{P}_3 m\mathbf{P}_3 r\mathbf{P}_3 o$

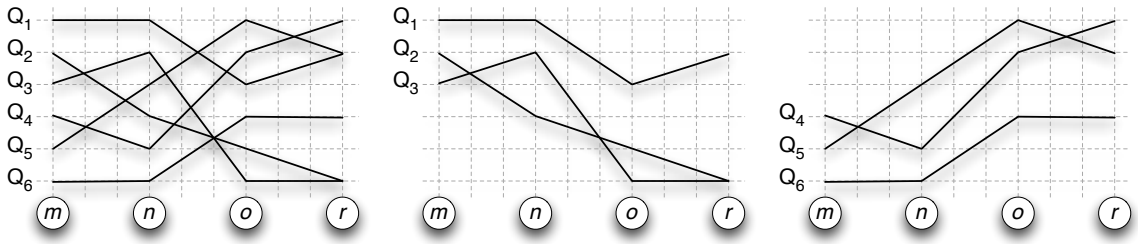
Az adatok alapján persze azt is felismerhetjük, hogy az első táblázat egyben példát mutat az egymélypontú preferenciára. Praktikusan ezt az alapján dönthetjük el, hogy az alternatívák első "oszlopában" nem szerepel az egyik elem (n), vagyis ennek a profilnak ez a elem sosem a legjobb eleme. Ha az adatokat grafikonon ábrázoljuk, akkor ezek a tulajdonságok még szemléletesebben előtűnnek.



Az ábrákra tekintve egyből látható a kétféle preferencia különbsége, az, ahogy a csúcsosság felfelé vagy lefelé érvényesül, az előbbi az egycsúcsú, az utóbbi az egymélypontú preferenciák esetében. Ha a harmadik inadai típust is ábrázolni akarjuk, akkor ahhoz az alábbi adatokból érdemes kiindulnunk.

$\mathbf{Q}_1 = m\mathbf{I}n\mathbf{P}r\mathbf{P}o$	$\mathbf{Q}_2 = m\mathbf{P}n\mathbf{P}o\mathbf{P}r$	$\mathbf{Q}_3 = n\mathbf{P}m\mathbf{P}o\mathbf{I}r$
$\mathbf{Q}_4 = r\mathbf{P}o\mathbf{P}m\mathbf{P}n$	$\mathbf{Q}_5 = o\mathbf{P}r\mathbf{P}n\mathbf{P}m$	$\mathbf{Q}_6 = o\mathbf{I}r\mathbf{P}m\mathbf{I}n$

Az egyéni preferenciarendezések itt olyanok, hogy az alternatívák minden esetben két csoportba tartoznak. Ebben az értelemben az $A = \{m, n\}$ és az $B = \{o, r\}$ alternatívák vannak elkülönítve egymástól, ami tehát azt jelenti, hogy a a két csoport (A és B) elemei közti viszony olyan, hogy az A-ba tartozó alternatívák minden egyéni preferencia esetén vagy jobbak vagy rosszabbak a B csoprotba tartozó elemekhez képest. Természetesen ezeket az adatokat is érdemes ábrával szemléltetni, de a táblázatos formából a grafikus megjelenítésre váltva csak akkor látszik igazán jól, hogy az egyéni preferenciákat valóban két csoportba sorolhatjuk, ha nem egyszerre jelenítjük meg mind a hat grafikont, hanem az azonos csoportba tartozó preferenciákat külön-külön tüntetjük fel.



A második és harmadik grafikonon jól látszik, hogy az ott látható három-három görbe hasonló egyéni preferenciákat takar. A Q_1, Q_2, Q_3 esetben az m és n alternatívák mindig többre vannak értékelve a másik kettőnél, a Q_4, Q_5, Q_6 esetben ez pont fordítva van, és az o és r alternatívákat mindig jobbnak tartják az m-nél vagy n-nél. A szeparált preferenciaprofil további fontos vonása még (ami persze itt nem látszik), hogy bináris gráffal lehet reprezentálni [251].

Egyszerváltó preferencia

A preferenciaprofilokról szóló diskurzusba az inadai típusok mellé a hetvenes-nyolcvanas évektől kezdve bekerült még egy további mintázat, az **egyszerváltó preferencia** (single-

crossing preference) profiltípusa is. A fogalommal kapcsolatos példaként gyakran hivatkoznak az adózási kulcsokra vonatkozó szavazások jelenségére. Kevin W.S. Roberts volt az, aki elsőként leírta ezt a jelenséget, bár ő még más terminust használt rá, amikor bevezette a **hierarchikus tapadás feltétel** (hierarchical adherence condition) fogalmát [442]. Az egyszerváltó preferenciáról akkor beszélhetünk, amikor nemcsak az alternatívák halmaza, de a szavazók halmaza is lineáris sorba rendezhető. Gondolhatunk itt arra a helyzetre, amikor az adókulcsok mértékről kell szavazni, és eközben maguk a szavazók egyértelműen sorba rendezhetők a jövedelmi-vagyoni helyzetük szerint. Ilyen esetben megtapasztalhatjuk az egyszerváltó preferencia jelenségét, ami azt jelenti, hogy az embereket a jövedelem (vagyon) alapján szigorú sorba rendezve észrevehetjük, hogy amíg a gazdagabbak kisebb adókulcsokat szeretnének, addig a szegényebbek magasabb (és progresszív) adókulcsokra szavaznának. Ez érthető is, hiszen az adózás értelme – ebből a szempontból tekintve – éppen az, hogy a magasabb jövedelműek rétege felől áramoltassanak át erőforrásokat az alacsonyabb jövedelműek rétegei felé. Ebben a helyzetben az egyszerváltó tulajdonság azt jelenti, hogy az egyéni preferenciákra vonatkozóan – elég egyértelműen – kijelenthetjük, hogy a "szegényebb" pólus felől indulva sokáig a magasabb kulcsokat támogatják a szavazók, a másik irány felől indulva visszafelé sokáig az alacsonyabb kulcsokat preferálják, és az is elképzelhető, hogy valahol a két pólus között van egy olyan átmeneti réteg, amely indifferens ebben a kérdésben, de az is előfordulhat, hogy nincs ilyen halmaz, és ilyenkor létezik egy pont, ahol az egyéni preferenciák "előjelet", értéket váltanak. Ez az egyszeri trendváltás az egyszerváltó preferenciaprofil lényege. A pár oldallal korábban bemutatott kettős példára hivatkozva azt mondhatjuk, hogy az első példa nemcsak az egymélypontú, de az egyszerváltó preferenciára is vonatkoztatható lenne.

Ahogy Miguel A. Ballester és Guillaume Haeringe megmutatta azt, hogy milyen feltételekkel lehet Inada és Sen profiltípusai között egyértelmű megfeleltetést találni, úgy Robert Bredereck, Jiehua Chen és Gerhard J. Woeginger elvégezte azt a munkát az egyszerváltó preferenciák esetében is [78].

(PL.20) $(\text{SCP}) \stackrel{\text{def}}{=} \forall m \forall n\ \forall i \forall j((n > m \wedge \mathbf{P}_j > \mathbf{P}_i \wedge \mathbf{P}_i(n,m)) \rightarrow \mathbf{P}_j(n,m)) \wedge \forall m \forall n\ \forall i \forall j((n > m \wedge \mathbf{P}_j > \mathbf{P}_i \wedge \mathbf{P}_j(m,n)) \rightarrow \mathbf{P}_i(m,n))$

egyszerváltó preferencia (HUN) – single-crossing preference (ENG)

Ennek a preferenciatípusnak komoly szerepe van a reprezentatív szavazó elméletében (↪457.o.), és lehetségességi tételek is megfogalmazhatók vele kapcsolatban. A most tárgyalt kategóriák alapján meg lehetne próbálkozni a fogalmak valamiféle általánosításával, de erre a feladatra én nem vállalkozom. Ilyen kísérletek természetesen vannak, Salvador Barberà és Bernardo Moreno például a topmonotonitás fogalmát bevezetve az egycsúcsú és az egyszerváltó preferencia [36], de egyrészt hiányzik a kellő tudás, és talán a kellő motiváció is. A továbbbi absztrakciónak egyelőre nem látom az igazi hozadékát.

7.2 Preferenciaaggregálás

A preferenciáról szóló fejezetben bemutattam, hogy e jelenségről szólva mindig négyféle preferencia-vonatkozású relációt kell figyelembe vennünk (↪78.o.). A társadalmi választások elméletében - csakúgy, mint más preferencia-vonatkozűsú elméletekben – elsősor-

ban a gyenge preferenciarendezés (↪84.o.) és az erős preferenciarendezés (↪83.o.) fogalmait használják, de olykor-olykor felbukkan a preferenciarelációk körébe sorolható közömbösség (↪84.o.), illetve az összehasonlíthatatlanság (↪83.o.) relációja is. Mielőtt a preferenciaggregálás sajátosságainak tárgyalásába kezdenék, röviden felvázolok egy olyan értelmezési keretet, amely végig segítségemre lehet a jelenségek bemutatásában azáltal, hogy összeköti az egyes elemzendő kérdésköröket.

7.2.1 Manipuláció

A társadalmi választások elméletén belül a szavazás jelenségének talán legfontosabb vonása a manipuláció lehetősége. Az egyéni preferenciák aggregálása egyetlen közösségi véleménnyé nem kizárólag az egyéni véleményektől függ, hanem sok minden mástól, és ezeket mind valamilyen szinten meg lehet változtatni, amivel a végeredményt is lehet befolyásolni. A manipulációnak azonban sok értelme, sok formája, sokféle "színtere" lehetséges. A társadalmi választások elméletében legelterjedtebb manipulációértelmezés az egyéni preferenciák kinyilvánításával, a leadott szavazatok őszinteségével kapcsolatos (↪447.o.). A stratégiai szavazásnak nevezett jelenség a szavazók által direkt módon megvalósított szavazatmanipulációt jelenti, amikor a saját egyéni preferenciarendezésüket megváltoztatják a szavazatgazdák, vagyis a szavazók. Ezzel szemben áll az az indirekt szavazatmanipuláció, amikor ugyanúgy az egyéni preferenciarendezések változnak meg, de itt nem a szavazók maguk változtatják meg tudatosan saját véleményüket, hanem mások hatására, a külső véleménybefolyásolások eredményeként lesz más a vélemény. Ekkor a szavazók őszintén szavaznak, szemben a stratégiai szavazás őszintétlenségével. A külsődleges véleménybefolyásolás hatására bekövetkező vélményváltoztatást azonban, bár fontossága nem kérdőjelezhető meg, itt mégsem vizsgálom, mivel ez más tudományterület elemzési körébe tartozik.

Másfajta manipulációról beszélhetünk akkor, amikor az alternatívák halmaza változik meg valamiért. Az nagyon fontos kérdés, hogy miért is változ(hat)nak meg az értékelendő alternatívák. A dolgokat itt is leegyszerűsítve azt mondhatjuk, hogy szavazási kontextusban az alternatívák maguktól, a szavazók felől tekintve nem változnak meg. Ha változik valami e téren, akkor mindez azoknak a hatására történik, akik felelnek a szavazásért, akik kontroll alatt tartják a választási folyamatot. Ha ezeket elnevezzük aggregációgazdáknak, akkor azt mondhatjuk, hogy az **alternatívamanipuláció**, amikor az alternatívák, a jelöltek listáját megváltoztatják az az aggregációgazdák, kívül esik az aggregációs folyamatok logikáján. Ez a kérdés a választás normatív kontextusában elemezhető, ezért ennek tárgyalását is elhagyom itt. Fontos kérdés, de nem az aggregálási folyamat jelenségébe tartozik.

Hasonlókat mondhatunk a **szavazójog-manipuláció** kérdéséről is. A szavazások, választások végeredményét befolyásolja az a tény, hogy kik a szavazók, kiknek a véleménye számít. Ez is fontos történelmi, társadalmi kérdés, gondoljunk csak a nők választójogának megjelenésére, a cenzus alapú választójog eltűnésére, a 'ki a szabad', 'ki a felnőtt' kérdésekre adott válaszok történetére, a szavazati jog megvonása mint egyéni vagy kollektív büntetés jelenségére, vagy – még konkrétabban – a Magyarország határain kívül élő magyarok, magyar állampolgárok szavazati jogának kezelésére. Ebben a jelenségben is van manipulációs potenciál, nyilván ez is a választáshoz tartozik, de ezt sem elemzem itt. Ugyanazokért nem, mint ahogy a korábbi témaköröket is elhagytam: más tu-

dományterület relevanciájába tartozik.

Van azonban két olyan manipulációs lehetőség a választásokkal kapcsolatban, amit végre már "elemezhetek". Az egyik az **algoritmusmanipuláció**,[4] amikor az aggregálási eljárást megváltoztatják az aggregációgazdák, hogy ezzel befolyásolják a végeredményt, a másik a **körzethatár-manipuláció**, amikor a körzethatárokat változatják meg politikai, gazdasági előnyök reményében. Az első kérdéskört alaposabban körbe járom (↫427.o.), a második problémát csak röviden említem (↫463.o.).[5]

7.2.2 Algoritmusmanipuláció

Azt, hogy a szavazási eljárás választása önmagában befolyásolhatja a csoportszavazás végeredményét, már legalább a rómaiak óta tudjuk. Nem lehet nem megidézni azt a mesét, amiről ifjabb Plinius egyik levelében számolt be [415, 8. könyv, 14. levél], és ami a társadalmi választások elméletének talán leggyakrabban idézett példájává vált [154], [441], [395]. A történetet Plinius eredeti meséjéhez képest kicsit megváltoztatva mutatom be, hogy szemléletesebb, karakteresebb legyen a tanulsága. Ifjabb Plinius beszámolója szerint Afranius Dexter szenátor holtan találják a villájában, és a gyanú a felszabadított rabszolgáira terelődik. Tegyük fel, hogy a szenátorok egy 12 tagú csoportjának kell meghoznia kell a végső döntést, akik három alternatíva közül választhatnak:

F	felmentés
K	kivégzés
S	száműzés

Tegyük fel még azt is, hogy a szenátorok három csoportokba sorolhatók. Az 1. csoport tagjai hisznek a vádlott ártatlanságában, tehát szeretnék felmenteni, vagyis az F alternatívát teszik saját preferenciasoruk elejére. Továbbá úgy gondolkodnak még, hogy ha valamilyen oknál fogva nem lehet felmenteni a vádlottat, akkor a száműzés (S alternatíva) mellett voksolnának, s utolsó helyre tennék a kivégzés lehetőségét (a K alternatívát). A referenciarendezésük tehát az alábbi formában írható le röviden: FSK. A második csoport rangsora legyen KSF (vagyis: kivégzés-börtön-felmentés), míg a harmadik csoportot jellemezzük a SFK száműzés-felmentés-kivégzés) rendezéssel (e csoport tagjai olyan érvekre támaszkodhatnának, hogy bár bűnösnek tartják a vádlottat, az első helyen ezért szerepel a börtönbüntetés kiszabása, de a halálbüntetést ellenezve második helyre inkább a felmentési alternatívát tennék). Tegyük fel, hogy a három csoport létszáma a következő:

1. csoport	FSK	5
2. csoport	KSF	4
3. csoport	SFK	3

[4] Direkt módon nem így nevezi meg, de tartalmilag mint napirend-manipulációról (agenda manipulation) beszél erről a jelenségről Peter C. Ordeshook. Az eljárási algoritmusokat a szavazások lebonyolításának ügyrendjeként felfogva lehetne érvelni a napirend, de még inkább az ügyrend fogalmának használata mellett, de a politikai elemzések világában később teret nyert napirendelemzési technikák miatti terminológiai ütközés miatt én nem használom Ordeshook kifejezését.

[5] A manipuláció bizonyos lehetőségeiről, területeiről magyar nyelven érdekes áttekintést ad Meszerics Tamács tanulmánya [352], a stratégiai szavazásról, a szavazói tanulás jelenségéről pedig [351].

A szavazási összesítést kétféle módon is elvégezzük, nézzük meg, hogyan. Vegyük előbb a plurális eljárást!

Plurális eljárás. Vizsgáljuk meg, melyek alternatívára hányan szavaznak külön-külön, és a legtöbb szavazatot kapott alternatíva legyen a nyertes. Ebben az esetben az eredmény a következő:

F	5 szavazat	az 1. csoport mindegyik tagja ide szavaz
B	3 szavazat	a 3. csoport mindegyik tagja ide szavaz
K	4 szavazat	a 2. csoport mindegyik tagja ide szavaz

Tehát a nyertes az F alternatíva, vagyis ebben az esetben a vádlottat felmentenék. Most nézzük meg a másik, az ún. bináris eljárás eredményét!

Bináris eljárás. Az eljárás két lépésből áll. *(i)* Először vizsgáljuk meg, a szenátorok bűnösnek tartják-e a vádlottat vagy nem? Ekkor az eredmény a következő:

bűnös:	4+3 szavazat	a 2. és 3. csoport tagjai ide szavaznak
nem bűnös:	5 szavazat	az 1. csoport mindegyik tagja ide szavaz

Tehát a szenátorok szerint a vádlott bűnös! *(ii)* Másodjára vizsgáljuk meg, hogy milyen büntetést kell kiszabni a vádlottra. Ekkor már csak két döntési alternatívánk marad, mert a felmentés lehetőségét ki kell zárnunk. A döntési menet tehát:

K	4 szavazat	a 2. csoport mindegyik tagja ide szavaz
B	3+5 szavazat	a 3. csoport az elsődleges preferenciája, míg az 1. csoport a másodlagos preferenciája szerint szavaz ide

A nyertes a B alternatíva, a vádlottat börtönbe csuknák. A bemutatott eset azt szemlélteti, hogy a szavazások során nemcsak a leadott szavazatok tényleges eloszlása, de az alkalmazott aggregálási eljárás is befolyásolhatja a végeredményt. Sokféle paradoxont beazonosítottak, megneveztek e jelenségkörön belül, párat röviden bemutatok közülük. Az egyiket Ordeshook dominált győztes paradoxonának (dominated-winner paradox) nevezte el [395], és a példában mutatott eljárás azt az elvárásunkat sérti meg, hogy ha van egy alternatíva, amely egy másikkal szembeni összehasonlításban mindig alulmarad (vagyis mindenki preferálja x-t y-nal szemben), akkor a közösségi döntésben is megnyilvánuljon ez a sorrend. Hogy lássuk ezt, feltételezzünk három szavazót, négy alternatívát és a következő preferenciarendezéseket:

1. szavazó	$x\,y\,b\,a$
2. szavazó	$a\,x\,y\,b$
3. szavazó	$b\,a\,x\,y$

Az aggregálási eljárás pedig legyen olyan, hogy először az x és a alternatívák közötti páros összehasonlítások alapján hoz ki részgyőztes, majd ezt követően a győztest előbb a b-vel, majd az újabb győztest az y-nal veti össze, és ezzel véget is ér az algoritmus. Ebben az összehasonlítási folyamatban x és a versenyéből utóbbi jön ki győztesen (hiszen a 2. és 3. szavazónál is előbbre van a sorban). A második menetben a-t és b-t kell

összevetni, aminek b lesz a nyertese (az 1. és 3. szavazó miatt). Harmadjára b és y kerül egymással szembe, és ezt a fordulót y nyeri (mert az 1. és 2. szavazónál előbb van b-nél). Ez az algoritmus tehát a kollektív döntés győzteseként az y alternatívát hozza ki, ami szembeállítható azzal a ténnyel, hogy egyénileg mindhárom szavazó jobbnak tartotta x-t y-nál.

Másféle paradoxon adódik akkor, ha a Borda-pontok módszerével aggregálunk közösségi döntést. Ordeshook két paradoxont is megnevezett [395], itt csak azt mutatom be, aminek az "elsőkből utolsók" (winner-turns-loser) paradoxon nevet adta. Tételezzük fel az alábbi preferenciaeloszlásokat hét szavazó és négy alternatíva között.

1.	2.	3.	4.	5.	6.	7.	alternatíva	Borda-pont
a	b	c	a	b	c	a	a	$18 = (3\times4)+(2\times2)+(2\times1)$
b	c	x	b	c	x	b	b	$19 = (2\times4)+(3\times3)+(2\times1)$
c	x	a	c	x	a	c	c	$20 = (2\times4)+(2\times3)+(3\times2)$
x	a	b	x	a	b	x	x	$13 = (2\times3)+(2\times2)+(3\times1)$

A Borda-módszer szerint a legjobb helytől indulva a gyengébbek felé egyre kevesebb (itt: 4-3-2-1) pontot kell adni minden egyéni rangsorban elért helyezésért, és a közösségi döntéshez ezeket kell összegezni. Így eljárva a fenti példában a c alternatíva jön ki győztesnek. A paradoxon akkor keletkezik (válik láthatóvá), amikor ugyanezen szavazók és alternatívák közül kiveszünk egyet, az x alternatívát (mondjuk azért, mert – ha az alternatívák politikai jelölteket reprezentálnak – az x azonosítójú jelölt meghal rögtön a szavazás után). Ekkor újra kell számolni a Borda-pontokat úgy, hogy a preferenciarendezésekből kivesszük az x alternatívát.

1.	2.	3.	4.	5.	6.	7.	alternatíva	Borda-pont
a	b	c	a	b	c	a	a	$15 = (3\times3)+(2\times2)+(2\times1)$
b	c	a	b	c	a	b	b	$14 = (2\times3)+(3\times2)+(2\times1)$
c	a	b	c	a	b	c	c	$13 = (2\times3)+(2\times2)+(3\times1)$

A tanulság egyből levonható a két táblázat összevetéséből (most az a nyert a b-vel és a korábbi nyertes c-vel szemben). Az a tény, hogy a vesztesek közül valaki kiesik a versenyből, megváltoztathatja az elsőség kérdését. Ez a paradoxon egyben fontos példaként tekinthető arra vonatkozóan, hogy nem mindig érvényesül az irreleváns alternatívák függetlenségének elve.[6] A Borda-pontozás további paradoxonokat is okozhat. Ha a pontozási módszeren változtatunk, és az egyéni győzteseknek nem annyi pontot adunk, mint az alternatívák száma, hanem kevesebbet (és ekkor értelemszerűen az egyéni listák végefelé már nem fognak pontot kapni az alternatívák), akkor más és más győztest hozhatunk ki [395]. Ennek ismerete és alkalmazása olykor komoly manipulációs potenciált jelenthet a Borda-pontozást alkalmazó tendereztetések, versenyeztetések lebonyolítói számára.

A különböző szavazási, választási eljárások leírásával, következményeik feltárásával több tanulmányunkban is foglalkoztunk Mészáros József kollégámmal közösen [343], [349]. Ezek tanulságait nem mutatom be itt részletesebben, csak egyetlen példát idézek fel arra vonatkozóan, hogy milyen hatása lehet a különböző többmandátumos eljárásoknak

[6] Ezt az elvet a későbbiekben még részletesebben tárgyalni fogom (↰433.o.).

mandátumok elosztására vonatkozóan. Az 1994-es parlamenti választás adatai alapján a megyei listán a parlamenti pártokra leadott tényleges szavazatok alapján kiszámoltuk, hogy milyen eredmények születtek volna, ha más és más eljárással számolták volna ki az eredményeket [349]. A táblázatban feltüntetett eljárásokat itt nem mutatom be, most elég csak annyit tudni róluk, hogy mindegyiküket alkalmazni lehetett volna az éles számítások során (a hivatkozott korábbi tanulmányainkban bővebben lehet olvasni róluk).

	MSZP	SZDSZ	MDF	FKGP	FIDESZ	KDNP
legnagyobb maradék módszere	54,1	17,9	9,8	6,7	5,2	5,7
Imperiali módszer	55,4	17,9	9,6	6,7	4,7	5,2
d'Hondt módszer	57,5	19,4	9,1	6,0	3,3	4,1
Hagenbach-Bischoff módszer	54,9	17,1	9,6	7,0	5,2	5,7
Sainte-Laguë módszer (osztó: 1)	53,9	17,1	10,4	7,0	5,2	6,0
Sainte-Laguë módszer (osztó: 1,4)	55,7	18,4	9,8	5,2	4,4	4,9
Sainte-Laguë módszer (osztó: 1,2)	54,1	17,9	9,8	6,7	5,2	5,7

Az adatokból szépen kiolvasható, hogy ha nem is radikális, de nem is elhanyagolható változásokat okozhat az a tény, hogy kicseréljük a szavazatok összegzését végző algoritmusokat. Tudni kell azonban, hogy ennél sokkal komolyabb változást is tapasztalhatnánk, ha a választási rendszerben komolyabb átalakítást végeznénk. A magyar választási rendszer ugyanis több, egymástól jól elkülöníthető, és más logikát megtestesítő részből áll, és ha valamelyik összetevő súlyát megnövelnénk a többi rovására (például a listás rendszerkomponenshez rendelt mandátumok számát megnövelnék az egymandátumos választókerületi részrendszer rovására), akkor az sokkal jelentősebb következményekkel járna. Ennek a kérdéskörnek azonban ezen a ponton nem szentelek több figyelmet.

A bemutatott példák azt szemléltetik, hogy az aggregálási algoritmus maga befolyással lehet a közösségi döntés végeredményére, vagyis amikor közösségi döntést képezünk valamilyen eljárással, akkor az egyéni vélemények összessége mellett az összegzés technológiája is befolyással lehet a kimenetre. Ez felveti annak szükségességét, hogy megvizsgáljuk és értékeljük az összegző eljárásokat. Mielőtt ezt megtennénk, rá kell még mutatnom egy problémára, amivel a kutatók az aggregálási folyamatok nagyon korai vizsgálata során szembesültek. Ez a többszáz éve ismert paradoxon folyamatosan arra sarkallta a kutatókat, hogy keressék az ellentmondás feloldásának lehetőségét.

7.2.3 Condorcet-paradoxon

Nicolas de Condorcet francia matematikus és filozófus már a 18. században leírta azt a paradoxont, amit később róla neveztek el. Legalább Condorcet óta tudjuk tehát, hogy amikor az egyéni preferenciák alapján egy közösségi sorrendet kell kialakítani valamilyen alternatívahalmazon, akkor a többségi eljárással nem remélhetünk minden esetben végeredményt. A Condorcet-féle módszer a Condorcet-győztes elemet úgy hozza ki, hogy az egyéni rendezések alapján összehasonlítja az összes alternatívát egymással, és a páros összehasonlítások részeredményeit összesítve a többségi elv alapján hirdeti ki a győztest. A Condorcet-paradoxon lényege az, hogy könnyen preferenciakör, **tranzitív kör** alakulhat ki az ilyen eljárás során. A körkörös többség jelenségét szemléltethetjük az alábbi táblázat segítségével (szigorú preferenciarendezést feltételezve).

	feltétel	következmény
1. egyéni vélemény	$\mathbf{P}(m,n) \wedge \mathbf{P}(n,o)$	$\mathbf{P}(m,o)$
2. egyéni vélemény	$\mathbf{P}(n,o) \wedge \mathbf{P}(o,m)$	$\mathbf{P}(n,m)$
3. egyéni vélemény	$\mathbf{P}(o,m) \wedge \mathbf{P}(m,n)$	$\mathbf{P}(o,n)$
többségi vélemény	$\mathbf{P}(m,n) \wedge \mathbf{P}(n,o) \wedge \mathbf{P}(o,m)$	?

Mindegyik egyén racionálisnak mondható (mert tranzitív) preferenciákkal rendelkezik, de a csoport egészére, ha többségi szavazással kell dönteni az egyes alternatívák közti rendezésekről, akkor a tranzitivitásnak ellentmondó eredményhez jutunk. A Condorcet-paradoxon feloldására történtek mindenféle kísérletek, de ezek számunkra itt, most nem annyira érdekesek. Sokkal fontosabb az az – időben jóval későbbi – felismerés, hogy a paradoxon mélyében alapvetőbb ellentmondások rejlenek. Ennek megértéséhez meg kell néznünk, milyen tulajdonságokkal jellemezhetjük az aggregálási eljárásokat, és ezek alapján milyen összefüggéseket, tételeket állapíthatunk meg rájuk vonatkozóan.

Tartalmát tekintve ugyan nincs köze a választás problematikájához, mégis érdemes megemlíteni, hogy a hétköznapi életből jól ismert *kő-papír-olló* játék is a Condorcet-paradoxon lényegét adó "körbeverésen" alapul.

7.2.4 Arrow lehetetlenségi tétele

Nem nagyon vitatkoznék azzal, ha valaki azt állítaná, hogy a preferenciaaggregálás egész világa Kenneth J. Arrow lehetetlenségi tétele körül forog. Arrow 1950 fogalmazta meg először híres tételét [23], majd 1951-ben egy kis könyvet jelentetett meg ugyanerről a kérdésről [25].[7] Az egyéni preferenciákat aggregáló függvényt **társadalmi jóléti függvénynek** (social welfare function) nevezve, az aggregálási folyamatra vonatkozóan különböző elvárásokat megfogalmazva, Arrow kimondott egy tételt, miszerint azok az elvárások, amelyeket elég széles körű egyetértés alapján a véleményösszegzési eljárásokkal szemben támasztunk, nem elégíthetők ki egyszerre. Vegyük sorba egymás után, milyen elvárásaink lehetnek az összegző függvénnyel szemben.[8]

Nyilván meg kell követelnünk az aggregálási szabálytól azt, hogy mindenféle konstellációban képes legyen eredményt produkálni – akárhogyan is alakulnak az alternatívák és az egyéni preferenciarendezések. Ez az elvárás azonban kétféle dolgot is jelent, ami alapján a bemenetre és a kimenetre is megfogalmazhatunk egy-egy részfeltételt. A bemenetre vonatkozó feltételünk az lehet, hogy mindenféle bemenetet be tudjon fogadni a rendszer, ebben az értelemben legyen egyetemes, univerzális. A kimenetre vonatkozóan pedig azt az elvárást rögzíthetjük, hogy egyetlen, tehát egyértelmű végeredményt adjon ki a rendszer az összesítés után. Az persze előfordulhat – szinte bármelyik aggregáló szabály esetében –, hogy a szavazatok összegzése döntetlent eredményez. Ilyenkor szükség lehet egy másodlagos **döntetlentörő szabályra** (tie-break rule), ami segít a döntetlent megszüntetni. Ez azonban a főszabályra vonatkozó elvárásunkat nem érinti lényegi módon. A két tulajdonság, amit a fentiek alapján definiálhatunk a következő.

[7] Az 1951-es könyvet kisebb módosításokkal úra kiadták 1963-ban, én is erre hivatkozom itt.

[8] Megintcsak sokféle terminust használhatunk akkor, amikor elvárásokról beszélünk. Az összegzőfüggvény **tulajdonságaira** (property) vonatkozóan **feltételeket** (condition), **kényszereket** (constraint), **kritériumokat** (criterion) írhatunk elő. A későbbiekben ezeket a terminusokat mind használni fogom – ugyanabban az értelemben.

(AL.21) $(\text{UNI}) \equiv \text{UNIVERSAL}(\mathbf{F}) \stackrel{\text{def}}{=} \mathbb{D}_F = \mathbb{X}^n$

univerzális (értelmezési tartomány) (HUN) – korlátozatlan értelmezési tartomány (HUN) – universal (domain) (ENG) – unrestricted domain (ENG)

(AL.22) $(\text{DEC}) \equiv \text{DECISIVE}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a(|\mathbf{F}(a)| = 1)$

döntésképes (HUN) – ügydöntő (HUN) – decisive (ENG)

Az első követelmény az **univerzalitás** tulajdonsága, ami tehát azt jelenti, hogy az összegző függvény értelmezve van minden lehetséges bemenetre. Matematikailag ezt úgy ragadhatjuk meg, hogy az $\mathbf{F}$ aggregáló függvény $\mathbb{D}_F$ értelmezési tartományát egyenlővé tesszük a teljes állapotteret jelző $\mathbb{X}^n$ halmazzal. Az univerzalitás felétele azt jelenti tartalmilag, hogy az aggregáló függvénytől elvárjuk azt, hogy bármilyen (bárki által megfogalmazott) lehetséges preferenciarendezést figyelembe vegyen a közösségi vélemény kialakításakor, ami egyet jelent azzal, hogy az egyének szabadon alakíthassák ki saját egyéni sorrendjeiket a világ dolgait illetően. Ez az elv tehát az szabad véleménykinyilvánítás egy fontos esetét írja le. Ha elhagyjuk az univerzalitás követelményét, akkor valamilyen módon korlátozott tartományról kell beszélnünk. Erre példa az egycsúcsú preferenciák esete (↳421.o.), amelyre vonatkozóan Duncan Black fontos tétel fogalmazott meg, amit később fogok bemutatni (↳439.o.). Bár a jelenség lényegét tekintve nem ide tartozik, de azért érdemes megjegyeznem, hogy amikor a diktatórikus rendszerek megakadályozzák, hogy a szavazáson tetszőleges jelöltek indulhassanak, akkor voltaképpen az itt tárgyalt univerzalizási kritériumnak nem tesznek eleget.

Amíg az első elvárás a függvény bemenetére rögzített egy kritériumot, addig az aggregáló függvény **döntésképes** jellegére irányuló feltétel a függvény kimenetére irányul. A fentebb már jelzett elvárás, miszerint az összegző eljárásnak minden lehetséges bemenetre egyértelmű végeredményt kell adnia, azt a szimpla és nagyon praktikus elvet akarja érvényesíteni, hogy értelme legyen az aggregálási eljárás lefolytatásának. Arra ugyanis pont azért van szükség, mert a sok egyéni vélemény nem egyezik meg egymással, tehát nincs konszenzus az egyéni szinten, az érintettek nem tudnak kollektíve dönteni. Ilyen helyzetekben azért kell az aggregálási eljárást elindítani, hogy legyen kollektív döntés. De csak akkor beszélhetünk döntésről, ha egyetlen kimenetünk van. Ez biztosítja az eljárás döntésképes tulajdonsága.

A lehetetlenségi tétel első megfogalmazásakor Arrow még öt tulajdonságot különített el egymástól [23], és csak később alakította át úgy az elméletét, hogy négy tulajdonságról legyen szó. A tétel első megfogalmazásában még szerepelt a **monotonitás** fogalma is, amit a relációelméleti fejezetben mi már definiáltunk, mint fontos függvénytulajdonságot (↳37.o.). A monotonitás kritériumát itt nyilván az $\mathbf{F}$ összegzőfüggvényre kell vonatkoztatnunk.

(AL.23) $(\text{MON}) \equiv \text{MONOTONE}(\mathbf{F}) \stackrel{\text{def}}{=} (\mathbf{F}(\langle \mathbf{R}_i^a \rangle_{i \in \mathbb{N}}) = m \wedge (\mathbf{R}_i^a = m \rightarrow \mathbf{R}_i^b = m)) \rightarrow \mathbf{F}(\langle \mathbf{R}_i^b \rangle_{i \in \mathbb{N}}) = m$

monoton (HUN) – monotone (ENG)

Arrow ezt az egyéni és csoportos értékek közti pozitív kapcsolat (positive association of social and individual values) elvárásaként is értelmezi [23]. Ha az egyéni rendezésekben az egyik alternatíva mindig egy másik előtt szerepel, akkor ugyanennek a rendezésnek kell megjelennie a kollektív preferencia szintjén is. A monotinitás helyett azonban nagyon hamar egy másik tulajdonság került be a lehetetlenségi tétel(ek)ről folytatott diskurzusba,

az **irreleváns alternatíváktól való függetlenség** elvárása.[9] A fogalom értelmezését kezdhetjük azzal, amit Arrow maga mondott erről.[10]

> „Az ordinális hasznossághoz való újkeletű ragaszkodást a megkülönböztethetetlenek azonosságának leibniz-i elvével magyarázhatjuk. Csak a megfigyelhető különbségeket tehetjük meg a magyarázataink alapjául. ... Az Irreleváns Alternatívák Függetlenségének Feltétele egy lépéssel tovább megy a megfigyelhetőség követelményén. Ha adott a társadalom számára választható alternatívák halmaza, akkor – elméletileg – elvárható, hogy valaki képes lehet az összes alternatíva mérlegelésére, de nem várhatjuk el azt, hogy figyelembe vegyen a társadalom számára nem elérhető (nem létező) alternatívák közti preferenciákat." [25, 109-110]

Azt az elvárást kell itt megfogalmazni valahogyan, hogy a közösségi preferenciarendezések eredményei csak az összehasonlításba bevont alternatívákra vonatkozó egyéni preferenciáktól függjenek, semmi mástól. Ha az egyéni preferenciák két alternatíva között nem változtak, de – esetleg – más alternatívák között igen, akkor a közösségi preferencia sem változhat a szóbanforgó alternatívákra.

(AL.24) $(\text{IIA}) \equiv \text{IIA}(\mathbf{F}) \stackrel{\text{def}}{=} \forall i \forall a \forall b \forall m \forall n ((\mathbf{R}_i^a(m,n) \leftrightarrow \mathbf{R}_i^b(m,n)) \rightarrow (\mathbf{F}(a)(m,n) \leftrightarrow \mathbf{F}(b)(m,n)))$

független az irreleváns altenatíváktól (HUN) – independent from the irrelevant alternatives (ENG) – binary independence (ENG)

Az elv szemléltetésére nézzünk meg egy egyszerű példát. Tegyük fel, hogy van három szavazónk, v_1, v_2 és v_3, valmint négy alternatívánk, m, n, o és p. A következő táblázatban tüntessük fel azt, hogy a szavazók hogyan rangsorolják az alternatívákat, illetve milyen sorrendet hoz ki az $\mathbf{F}$ szavazási függvény két alternatíva, m és n között. Az irreleváns alternatívák függetlensége azt jelenti, hogy ha az első sorban feltüntetett egyéni rendezésekből az összegző függvény a m alternatívát hozza ki a n elé, akkor minden más profilegyüttesre is tartania kell ezt a sorrendet.

v_1 szavazó	v_2 szavazó	v_3 szavazó	$\mathbf{F}$	
(m,n,o,p)	(n,p,o,m)	(o,m,p,n)	$m > n)$	
(m,o,p,n)	(o,n,p,m)	(m,n,o,p)	$m > n)$	jó
(p,m,n,o)	(n,m,o,p)	(o,p,m,n)	$m > n)$	jó
(n,p,o,m)	(o,p,n,m)	(p,m,o,n)	$m > n)$	rossz
(m,n,o,p)	(n,p,o,m)	(o,m,p,n)	$m < n)$	rossz

A táblázatban két-két példát mutatok be az elvnek megfelelő (2-3. sor), illetve az elvet megsértő (4-5. sor) konstellációra. A negyedik sor példája azzal sérti meg a vizsgált elvet, hogy a v_1 szavazó esetében nem áll fent két alternatíva közti sorrend változatlansága, az ötödik sor pedig azért mutat "rossz eloszlást", mert az összegzőfüggvény megváltoztatja az alternatívák viszonyát egymáshoz. Azok a választási szabályok tehát, amelyek az első három sor alapján működnek, kielégítik az irreleváns alternatívák függetlenségének kritétiumát, a két utolsó sor pedig olyan választási függvényeket reprezentál, amelyek

[9] Az irreleváns alternatívák függetlensége kifejezés korábban már előkerült, az egyéni racionalitás jellemzésekor (↰172.o.). A két fogalom különbözik egymástól. Itt a kollektív preferencia kialakítását végző eljárás egyik fontos feltételét, korábban a racionális egyén egyik tulajdonságát jelenti.

[10] Az idézet első mondatának értelmezéséhez fontos tudni, hogy itt ordinális rendezésről beszélünk.

megsértik az (IIA) elvét. Korábban egyébként már bemutattam egy algoritmust, ami ugyancsak nem felelt meg az irreleváns alternatívák függetlensége elvének (↪429.o.).

A következő szempont arra vonatkozik, hogy az összegző függvény hogyan "viselkedjen" akkor, ha az egyéni profilok nagyon hasonlítanak egymásra, netán ugyanolyanok, vagyis mi legyen a végeredmény az egyéni vélemények egyhangúsága esetén. Ezt a közösségi véleményt kifejező függvény Pareto-hatékonyságával fejezhetjük ki. A Pareto-hatékonyságot azonban háromféle értelemben is megadhatjuk, és a háromféle értelmezést más és más tételek során lehet felhasználni. Nézzük meg először a definíciókat!

(AL.25) $(\text{WPE}) \equiv \text{WEAKLY_PARETO_EFFICIENT}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall m \forall n((a = (\mathbf{R}_1^a, \mathbf{R}_2^a, \ldots, \mathbf{R}_{|\mathbb{N}|}^a) \wedge (m,n) \in \mathbb{X} \wedge C^a = \mathbf{F}(a)) \rightarrow ((m,n) \in \bigcap_{i \in \mathbb{N}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow (m \in C^a(\{m,n\}))))$

gyengén Pareto hatékony (HUN) – weakly Pareto efficient (ENG) – gyengén Pareto optimális (HUN) – weakly Pareto optimal (ENG) – WPO (ENG)

(AL.26) $(\text{PE}) \equiv \text{PARETO_EFFICIENT}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall m \forall n((a = (\mathbf{R}_1^a, \mathbf{R}_2^a, \ldots, \mathbf{R}_{|\mathbb{N}|}^a) \wedge (m,n) \in \mathbb{X}) \rightarrow ((m,n) \in \bigcap_{i \in \mathbb{N}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow (m \in S \rightarrow n \notin C^a(S))))$

Pareto hatékony (HUN) – Pareto efficient ((ENG)) – Pareto optimális (HUN) – Pareto optimal (ENG) – PO (ENG)

(AL.27) $(\text{SPE}) \equiv \text{STRONGLY_PARETO_EFFICIENT}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall m \forall n((a = (\mathbf{R}_1^a, \mathbf{R}_2^a, \ldots, \mathbf{R}_{|\mathbb{N}|}^a) \wedge (m,n) \in \mathbb{X} \wedge C^a = \mathbf{F}(a)) \rightarrow ((m,n) \in \bigcap_{i \in \mathbb{N}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow (m \in S \rightarrow n \notin C^a(S))))$

erősen Pareto hatékony (HUN) – strongly Pareto efficient (ENG) – erősen Pareto optimális (HUN) – strongly Pareto optimal (ENG) – SPO (ENG)

A Pareto-feltételeket egyfajta egyhangúságnak, egyöntetűségnek minősíthetjük, aminek az a lényege, hogy ha az m alternatívát minden szavazó prefereálja az n-nel szemben, akkor n nem lehet győztes. Fontos azonban jelezni, hogy ha az egyöntetűséget mint a konszenzus elvének megvalósítását értelmezzük, vagyis a szavazás egyhangúságával a teljes konszenzust várjuk el, akkor a szavazók a szavazás során könnyen stratégiai szavazókká válhatnak, és elkezdhetik "működtetni" a szavazatcsere intézményét (↪449.o.).

Az is fontos elvárás lehet a választási függvénnyel szemben, hogy ne egyetlen személy (vagy csoport) véleménye érvényesüljön az aggregálás során, vagyis azt várjuk el, hogy mindenki véleménye számítson az összegzés során, és ne fordulhassanak elő oligarchikus vagy diktatórikus megoldások. Ha arra a szempontra figyelünk, hogy milyen az egyéni és csoportos vélemények közti viszony, kétféle tulajdonságot is definiálhatunk, a **kikényszerítettséget** és a **diktatórikusságot** (illetve ezek komplementerét). Ezeket azonban akkor érthetjük meg jól, ha előtte még más fogalmakat is bevezetünk a modellünkbe. Először is képeznünk kell a szavazók tetszőleges csoportját, koalícióját, amelyek közül kiemeljük majd azokat, amelyeknek valamilyen módon meghatározó befolyása van a szavazás kimenetére.

(AL.28) $(\text{DEC}) \equiv \text{DECISIVE}(\mathbf{V}) \stackrel{\text{def}}{=} \exists \mathbb{V} \forall a \exists m \exists n(((m,n) \in \mathbb{X} \wedge \mathbb{V} \subset \mathbb{N} \wedge \mathbb{V} \neq \varnothing) \rightarrow ((m,n) \in \bigcap_{i \in \mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow \{m\} = C^a(\{m,n\})))$

(lokálisan) döntő (HUN) – (locally) decisive (ENG)

(AL.29) $(\text{ADE}) \equiv \text{ALMOST_DECISIVE}(\mathbf{V}) \stackrel{\text{def}}{=} \exists\mathbb{V}\forall a\exists m\exists n(((m,n)\in\mathbb{X} \wedge \mathbb{V}\subset\mathbb{N}\wedge\mathbb{V}\neq\emptyset) \rightarrow (((m,n)\in \bigcap_{i\in\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \wedge (n,m)\in \bigcap_{i\in\mathbb{N}\setminus\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a)) \rightarrow \{m\} = C^a(\{m,n\}))$

(lokálisan) majdnem döntő (HUN) – (locally) almost decisive (ENG)

(AL.30) $(\text{BLO}) \equiv \text{BLOCKING}(\mathbf{V}) \stackrel{\text{def}}{=} \exists\mathbb{V}\forall a\exists m\exists n(((m,n)\in\mathbb{X} \wedge \mathbb{V}\subset\mathbb{N}\wedge\mathbb{V}\neq\emptyset) \rightarrow ((m,n)\in \bigcap_{i\in\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow m\in C^a(\{m,n\})))$

(lokálisan) blokkoló (HUN) – (locally) blocking (ENG)

(AL.31) $(\text{ABL}) \equiv \text{ALMOST_BLOCKING}(\mathbf{V}) \stackrel{\text{def}}{=} \exists\mathbb{V}\forall a\exists m\exists n(((m,n)\in\mathbb{X} \wedge \mathbb{V}\subset\mathbb{N}\wedge\mathbb{V}\neq\emptyset) \rightarrow (((m,n)\in \bigcap_{i\in\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \wedge (n,m)\in \bigcap_{i\in\mathbb{N}\setminus\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a)) \rightarrow m\in C^a(\{m,n\}))$

(lokálisan) majdnem blokkoló (HUN) – (locally) almost blocking (ENG)

Az itt megadott, a szavazók koalíciójára vonatkozó ‘döntő’ (decisive) értelmezés eltér a korábban definiált ‘ügydöntő’ (decisive) tulajdonságtól (↬434.o.), amit majd később, a May-tételhez kapcsolódva használunk fel (↬442.o.). A majdnem döntő minőség formulája itt azt fejezi ki, hogy a majdnem döntő koalíció tagjai egyhangú véleménye határozza meg a közösségi döntés végeredményét úgy, hogy eközben mindenki más preferenciája pont a koalíció tagjainak preferenciájával ellentétes. Ha tehát a majdnem döntő csoport tagjai mind m-t preferálják n-nel szemben, és rajtuk kívül mindenki más az n-t az m ellenében, legyenek utóbbiak akármennyien, akkor is a koalíció véleménye számít, tehát az m alternatíva lesz a győztes. A majdnem döntő koalíció definíciójában szerepel a koalíción kívüliek ellenkező véleménye, míg a döntő csoport meghatározásából ez a mozzanat hiányzik. Éppen ezért a döntő csoport fogalma általánosabb, mert a definíciójában nincs semmilyen további feltétel kikötve. Kicsit félrevezetőnek is tűnhet a ‘majdnem döntő’ kifejezés, mert ha a meghatározása szerint a majdnem döntő csoport tagjai a csoporton kívüli összes szavazó egyhangú ellenvéleménye ellenére képesek meghatározni a közösségi végeredményt, akkor felmerülhet a kérdés, hogy a teljes ellenkezés dacára elért győzelemnél mi lehet még erősebb (hogy ne legyen szükség a ‘majdnem’ jelző alkalmazására). A válasz pedig az, hogy logikailag általánosabb, tehát erősebb az a fogalom (a ‘döntő koalíció’ fogalma), amire semmilyen feltétel sem vonatkozik.

A (majdnem) döntő, illetve a (majdnem) blokkoló koalíció fogalmak még lokálisak abban az értelemben, hogy két adott alternatíva (m és n) összehasonlításán alapulnak, de a definíciók nem követelték meg, hogy minden alterntatívára igazak legyenek a megadott feltételek. Globálissá tehetjük a fogalmakat akkor, ha minden alternatíva összehasonlítására kiterjesztjük a feltételek érvényességét (vagyis az alternatívákra vonatkozóan egzisztenciális kvantorok helyett univerzális kvantorokat alkalmazunk).

(AL.32) $(\text{GDEC}) \equiv \text{GLOBALLY_DECISIVE}(\mathbf{F}) \stackrel{\text{def}}{=} \exists\mathbb{V}\forall a\forall m\forall n(((m,n)\in\mathbb{X} \wedge \mathbb{V}\subset\mathbb{N}\wedge\mathbb{V}\neq\emptyset) \rightarrow ((m,n)\in \bigcap_{i\in\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \rightarrow \{m\} = C^a(\{m,n\})))$

(globálisan) döntő (HUN) – (globally) decisive (ENG)

(AL.33) $(\text{GADE}) \equiv \text{GLOBALLY_ALMOST_DECISIVE}(\mathbf{F}) \stackrel{\text{def}}{=} \exists\mathbb{V}\forall a\forall m\forall n(((m,n)\in\mathbb{X} \wedge \mathbb{V}\subset\mathbb{N}\wedge\mathbb{V}\neq\emptyset) \rightarrow (((m,n)\in \bigcap_{i\in\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \wedge (n,m)\in \bigcap_{i\in\mathbb{N}\setminus\mathbb{V}} \mathbf{P}(\mathbf{R}_i^a)) \rightarrow \{m\} = C^a(\{m,n\}))$

(globálisan) majdnem döntő (HUN) – (globally) almost decisive (ENG)

(AL.34) $(\text{GBLO}) \equiv \text{GLOBALLY_BLOCKING}(\mathbf{F}) \stackrel{\text{def}}{=} \exists \mathbb{V} \forall a \forall m \forall n (((m,n) \in \mathbb{X} \ \wedge \mathbb{V} \subset \mathbb{N} \wedge \mathbb{V} \neq \emptyset) \rightarrow ((m,n) \in \bigcap_{i \in \mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \ \rightarrow m \in C^a(\{m,n\})))$

globálisan) blokkoló (HUN) – (globally) blocking (ENG)

(AL.35) $(\text{GABL}) \equiv \text{GLOBALLY_ALMOST_BLOCKING}(\mathbf{F}) \stackrel{\text{def}}{=} \exists \mathbb{V} \forall a \forall m \forall n (((m,n) \in \mathbb{X} \ \wedge \mathbb{V} \subset \mathbb{N} \wedge \mathbb{V} \neq \emptyset) \rightarrow (((m,n) \in \bigcap_{i \in \mathbb{V}} \mathbf{P}(\mathbf{R}_i^a) \wedge (n,m) \in \bigcap_{i \in \mathbb{N} \setminus \mathbb{V}} \mathbf{P}(\mathbf{R}_i^a)) \ \rightarrow m \in C^a(\{m,n\}))$

(globálisan) majdnem blokkoló (HUN) – (globally) almost blocking (ENG)

A döntő-, illetve blokkolóképesség fogalmaira támaszkodva definiálhatjuk az oligarcha (oligarchikus szavazó) és a diktátor (diktatorikus szavazó) fogalmát, amelyekre a lehetetlenségi tételek kifejtése során fogunk hivatkozni. Mind az oligarcha, mind a diktátor fogalma szavazók halmazára vonatkozik, tehát az oligarcha és a diktátor fogalma alatt alapértelmezés szerint koalíciót értünk, ami természetesen megengedi azt, hogy adott esetben a koalíció egyelemű legyen, vagyis az oligarcha vagy a diktátor fogalma valóban egy szavazóra utalhasson (ahogy a terminusok hétköznapi használata ezt megkívánná).

(AL.36) $(\text{OLI}) \equiv \text{OLIGARCHIC}(\mathbf{F}) \stackrel{\text{def}}{=} \text{GLOBALLY_DECISIVE}(\mathbf{F}) \wedge \text{GLOBALLY_BLOCKING}(\mathbf{F})$

oligarchikus (HUN) – oligarchic (ENG)

(AL.37) $(\text{NOLI}) \equiv \text{NON-OLIGARCHIC}(\mathbf{F}) \stackrel{\text{def}}{=} \neg\text{OLIGARCHIC}(\mathbf{F})$

nem oligarchikus (HUN) – non-oligarchic (ENG)

(AL.38) $(\text{DIC}) \equiv \text{DICTATORIAL}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall m \forall n \exists!\, i \in \mathbb{N}(\mathbf{F}(\mathbf{R}^a(m,n)) \leftrightarrow \mathbf{R}_i^a(m,n))$

diktatórikus (HUN) – dictatorial (ENG)

(AL.39) $(\text{NODIC}) \equiv \text{NON-DICTATORIAL}(\mathbf{F}) \stackrel{\text{def}}{=} \neg\text{DICTATORIAL}(\mathbf{F})$

diktátormentes (HUN) – nem-diktatórikus (HUN) – non-dictatorial (ENG)

Az $\mathbf{F}$ véleményösszegző függvény diktatórikus, ha az általa aggregált kollektív vélemény minden esetben megegyezik egyetlen (az i.) szavazó $\mathbf{R}_i^a(m,n)$ véleményével.[11] Ha ilyen nincs, akkor a szavazási rendszer nem diktatórikus. Meg kell jegyezni, hogy a diktátormentesség követelménye – a dolgokat a tényleges társadalmi praxis felől szemlélve – nem igazán erős. Az egyszemélyi akarat (a diktatúra) érvényesülését ugyanis ki lehet zárni már azzal is, ha két személy (egy kisebb csoport) véleménye érvényesül. A diktátormentesség megkövetelése a véleménymonopólium kizárását biztosítja, de a véleményoligopóliumét már nem.

Bár Arrow már a tétele első megfogalmazásakor definálta és használta a **kikényszerítettség** fogalmát, a témakör diskurzusában kevesebb figyelmet kapott ez a kategória, ezért én sem szentelek neki több figyelmet. A nem-kikényszerítettség tulajdonságát Arrow az állampolgári szuverenitás (citizen's sovereignty) feltételének is nevezte [23].

A társadalmi jóléti függvényre vonatkoztatható feltételek bemutatása után következhet Arrow lehetetlenségi tételének kimondása: nem létezik olyan társadalmi jóléti függvény, amely minden preferenciaprofilra helyes rendezést eredményez (tehát reflexív, tranzitív és anti-szimmetrikus relációt) és teljesíti az (UNI), (PE), (IIA) és (NOD) feltételeket.

[11] Az általam adott meghatározása tartalmazza, de általában nem szokták beletenni a definíciókba az egyetlen individuumra való szűkítést.

(AL.40) $\neg\exists\mathbf{F}\big((\text{UNI})(\mathbf{F})\wedge(\text{PE})(\mathbf{F})\wedge(\text{IIA})(\mathbf{F})\wedge(\text{NODIC})(\mathbf{F})\big)$

Arrow lehetetlenségi tétele (HUN) – Arrow's impossibility theorem (ENG) – General Impossibility Theorem (ENG) – Arrow's paradox (ENG)

Arrow saját bizonyítása a diktátor szavazó fogalmára épül.[12] A tételt az idők során sokféleképpen is bizonyították. Sen két lépésben adta meg a bizonyítást [477, Chapter 3*]. Először egy segédtételben kimondta, hogy ha egy aggregálási metódusra teljesül a (UNI), a (PE) és a (IIA) feltétel, illetve van legalább egy olyan szavazó, aki az összes alternatívapárra vonatkozóan majdnem döntő, akkor annak a szavazónak diktátornak kell lennie. Második lépésben pedig megmutatta, hogy ha az első három feltétel teljesül, akkor mindig van ilyen majdnem döntő szavazó, aki – a segédtétel értelmében – diktátor is egyben.

Másfajta bizonyítást mutatott be Salvador Barberà [33]. Bizonyítását egy új fogalomra támaszkodva adta meg, amikor bevezette a **pivotális szavazó** fogalmát. E fogalommal azt a szavazót jelölte meg, akinek a szavazata fordulópontot jelent a szavazás során, vagyis ha ő megváltoztatja a saját szavazatát, akkor a közösségi döntés maga is megváltozik.

A jól ismert, széles körben alkalmazott kollektív döntési eljárások közül, amelyek megsértik valamelyik feltételt a fenti négyből, a többségi eljárás kielégíti a (PE), a (IIA) és a (NOD) feltételeket, de az univerzalitási kritériumot már nem [477, 38]. Sen bemutatott egy konkrét példát arra, hogy a Borda-pontozási rendszer hogyan sérti meg az irreleváns alternatívák függetlensége elvét.

1. szavazás				2. szavazás			
preferenciák		Borda-pontok		preferenciák		Borda-pontok	
1. szavazó:	$m\,n\,o$	m:	$7 = 1\times3+2\times2$	1. szavazó:	$m\,o\,n$	m:	$7 = 1\times3+2\times2$
2. szavazó:	$o\,m\,n$	n:	$4 = 1\times2+2\times1$	2. szavazó:	$o\,m\,n$	n:	$3 = 3\times1$
3. szavazó:	$o\,m\,n$	o:	$7 = 2\times3+1\times1$	3. szavazó:	$o\,m\,n$	o:	$8 = 2\times3+1\times2$

Sen példája kétféle Borda-pontozást mutat be, amikor a két szavazás csak annyiban tér el egymástól, hogy az 1. szavazó másként szavaz a két fordulóban, mert menet közben átértékeli a preferenciasorrendjét, megcseréli o és n sorrendjét. Az első fordulóban m és o lesz a Borda-győztes, a második fordulót viszont o nyeri meg egyedüliként, miközben m és o egymáshoz való viszonya nem változott meg. Mivel a végeredmény mégis változott, így kimondhatjuk, hogy az eljárás megsérti az (IIA) feltételt.

Gibbard oligarcha tétele

Érdekes története van Allan Gibbard oligarcha tételének, mert a tételt megfogalmazó írását a szerző sosem publikálta hivatalos helyen, így csak kéziratban terjedt a szakmán belül [186]. Mégis sokan olvasták, sokan is hivatkoztak rá, úgyhogy az 1969-es születése után nem sokkal a szakmai diskurzus megkerülhetetlen hivatkozási alapjává vált a "cikk".[13] Gibbard később sem publikálta soha a tanulmányát, de az állandó érdeklődés

[12] Arrow bizonyítását jól leírja Amartya K. Sen [477], a tételt saját formalizmusába konvertálva elemzi Kotaro Suzumura [514]. Arrow tétele után sokféle átértelmezés, új fogalmak, új bizonyítások jelentek meg a szakmai diskurzusban. Kiderült, hogy az univerzalitás kritérium nélkül is igaz marad a tétel, tehát már három feltétel együttes fennállását is ki lehet zárni. Umberto Grandi és Ulle Endriss elsőrendű logikával írta le az Arrow-tételt (több más, kapcsolódó tétellel együtt) [196], [197].

[13] A Gibbard-tételre hivatkoznak – többek között – angolul [477], [514], magyarul [321, 16].

miatt a 2000-es években saját honlapján elérhetővé tette mindenki számára. A tétel a következő formulával írható le.

(AL.41) $|\mathbb{X}| \geq 3 \wedge \big(\text{WEAK_ORDER}(\mathbf{R}^m) \vee \text{LINEAR_ORDER}(\mathbf{R}^m)\big) \wedge (\text{PE})(\mathbf{F}) \rightarrow (\text{OLI})(\mathbf{F})$

Gibbard oligarcha tétele (HUN) – Gibbard's oligarchy theorem (ENG)

Tegyük fel, hogy az alternatívák halmaza több, mint két elemből áll ($|\mathbb{X}| \geq 3$). Legyen az $\mathbf{F}$ egy kvázi Arrow-típusú összjóléti függvény, amelyre teljesül a következő két feltétel: *(i)* az alternatívák értelmezési tartományán képzett rendezett m-esek gyengén rendezettek (↬35.o.) vagy lineárisan rendezettek (↬34.o.), illetve *(ii)* érvényesül a Pareto-hatékonyság. Ekkor az $\mathbf{F}$ függvény oligarchikus.

Sen libertárius paradoxona

Az Amartya Sen által először megfogalmazott **libertárius paradoxon** nemcsak a társadalmi választások elmélete iránt érdeklődők számára vált fontossá, de sok más irányból is felkapták ezt a témát [478], [480], [481]. Ennek elsősorban az lehetett az oka, hogy az elemzett problémának, a libertárius viszonyulásban, értékelésben benne rejlő ellentmondásnak nyilvánvalóan erős politikai, értékelvű konzekvenciái voltak.[14]

Sen példája 'Lady Chatterly szeretője' című könyvvel kapcsolatos. Ha van két szereplő, M, a prűd, merev gondolkodású és N, a nyitott, érdeklődő ember.[15] M nem szeretné, ha az általa szabadosnak vélt könyvet bárki elolvasná, mert károsnak tartja a tartalmát (legyen ez az o alternatíva). Ha mégis döntenie kell, hogy kettejük közül ki olvassa el, akkor inkább magát választaná úgy gondolván, hogy ő "ellent tudna állni" a káros hatásoknak (legyen ez az m alternatíva, és legyen n az az opció, amikor N olvasná el a könyvet). Az M számára tehát a következő preferenciarendezés adott: omn.[16] Ezzel szemben feltételezhetjük azt, hogy a nyitottabb gondolkozású N nemcsak azt szeretné legkevésbé, hogy senki se férhessen hozzá a könyvhöz, de a prüdéria csökkentése érdekében azt pártolná leginkább, hogy M olvassa azt el elsőként. A N rendezése így: mno. Ha úgy akarunk közösségi döntést hozni, hogy minimális liberalizmus elvét betartjuk, vagyis tiszteletben tartjuk az egyének döntési jogát ebben a kérdésben, akkor a következőféleképpen érvelhetünk. Mivel M nem szeretné, viszont N szeretné elolvasni a könyvet, ezért az lenne a liberális kollektív döntés, ami igazodna az om, valamint az no preferenciákhoz, amiből viszont következik a nom hármas rendezés.

M	$o\,m\,n$
N	$m\,n\,o$
liberális elv	$n\,o\,m$

Ez a helyzet azonban ellentmondást takar, hiszen az m alternatíva mindegyik egyéni rendezésben előbbre van az n-hez képest, és ennek a Pareto-elv szerint a kollektív döntésben

[14] Sen az első cikkeiben még liberális, később azonban már libertárius elvről, libertárius paradoxonról beszélt. A liberalizmus, kontra libertárianizmus szembenállításban utóbbi az egyéni, személyes szabadság elsődlegességének elvének követését jelenti.

[15] Az eredeti cikkben Sen csak a prűd embert nevezte meg, a másodikat nem. A prűd partnerére később ugyan sokan ráakasztották a 'buja' minősítést, de ezt nem érzem jó megoldásnak, ezért inkább a 'nyitott ember' kifejezéssel utalok rá.

[16] Sen egy mondatában ezt úgy fejezi ki, hogy M inkább lenne cenzor, mint cenzorált.

is így kell lennie. Ez viszont ütközik az egyéni döntési szabadság libertárius felfogásával. Julian H. Blau szerint a paradoxon abban az esetben áll fent [62], ha a szereplőknek **tolakodó preferenciái** (nosy preferences) vannak, vagyis a szóban forgó preferenciák a kapcsolatban levő emberek mindegyikének cselekvési lehetőségeire vonatkoznak. Sen példájában mindkét szereplő (a prűd és a nyitott ember) a saját, a partner és mindenki más olvasási lehetőségét "beveszi" a preferenciarendezésébe.

7.2.5 Lehetségességi tételek

Az Arrow tétel és az azt követő további lehetetlenségi tételek azt mutatták meg, hogy mi lehetetlen, vagyis mit nem tehetünk meg, de nyitva hagyták a kérdést, hogy mi lehetséges, vagyis mit tehetünk meg. Tudunk-e pozitív állításokat tenni a társadalmi választási függvényekre vonatkozóan. A válasz természetesen az, hogy igen. Ritkábban szokták idézni, de ettől még tény, hogy Arrow a lehetetlenségi tételéről szóló könyvében foglalkozott azzal a kérdéssel is, hogy vannak-e lehetségességi tételek, és rá is mutatott két esetre, amikor létezhet megoldás. A továbbiakban ezekkel a lehetőségekkel foglalkozom röviden.

Black tétel

Duncan Black 1948-ban, tehát még Arrow lehetetlenségi tételének "felbukkanása" előtt megfogalmazott egy tézist, ami egy speciális preferencia esetén az Arrow-féle kritériumok együttes kielégíthetőségét állította [60]. Black tézise nem az Arrow-tétel cáfolata, hiszen a preferenciák egy szűkebb tartományára, mégpedig az (SPP) egycsúcsú preferenciákra (↪421.o.) vonatkozóan áll csak fent az igazsága. Az egycsúcsú preferenciák létét csak úgy feltételezhetjük, ha a választási függvény értelmezési tartományára nem áll fent az univerzalitás kritériuma, vagyis korlátozott tartományról beszélünk. A Black tétel éppen ezért "létezhet" az Arrow-féle lehetetlenségi tétel mellett, mert az előbbi érvénytelenként kezeli az utóbbi egyik alapfeltételét. A Black tétel az alábbi kondicionálissal írható le.

(AL.42) $$\big(\mathbf{F} = \mathbf{F}^{AP} \wedge (\text{SPP})(\mathbf{P})\big) \rightarrow \big((\text{NODIC})(\mathbf{F}) \wedge (\text{IIA})(\mathbf{F})\big)$$

Black tétel (HUN) – Black theorem (ENG)

A Black-tétel értelmében tehát amennyiben az egyéni preferenciák egycsúcsúak, akkor a(z egyszerű) többségi döntési szabály nem diktatórikus és kielégíti az irreleváns alternatívák függetlenségének elvét. A Black-tétel úgy tudta megkerülni Arrow lehetetlenségi tételét, hogy korlátozott tartományra vonatkozva mondta ki a lehetségességi tételét (amivel megsértette az univerzális értelmezési tartomány követelményének mint egyik arrow-i kritériumnak a teljesülését). Azért beszélhetünk itt korlátozott tartományról, mert az egycsúcsú preferenciák esetén igaz a tétel, tehát nem minden preferenciatípusra igaz.

Inada tétel

Ken-ichi Inada különböző preferenciatípusokra mondta ki a konzisztens társas preferencia létezését többségi szabály mellett [251], [252]. Korábban bemutattam, hogy vannak preferenciák, amelyek jól meghatározott profillal rendelkeznek (↪417.o.). Ezek közül néhány esetben – Inafa alapján – kimondhatunk lehetségességi tételeket a társadalmi választási függvényre vonatkozóan.

- dichotóm preferencia
- visszhangszerű preferencia
- antagonisztikus preferencia
- értékkorlátos preferencia
  - egycsúcsú preferencia
  - egyvölgyű preferencia
  - szeparált preferencia
- tabu preferencia

Az első három típus esetében nincs megkötés a szavazók számosságára vonatkozóan, az utolsó kettőre annyi elvárás van, hogy a szavazók számra páratlan kell legyen. Ezeknek a tételeknek közös vonása az az elképzelés, hogy a többségi eljárás tranzitivitása nem azonos a racionális döntéssel, valamint, hogy könzisztens társas preferencia létezéséhez a tranzitivitás túl erős követelmény, elég hozzá gyengébb feltétel is [484, 179].

May tétel

Az Arrow-féle lehetetlenségi tételre azonnal reagált Kenneth O. May, aki egy 1952-es cikkében fejtette ki azt a tételét, amit később róla neveztek el [326]. May kétalternatívás helyzetet (m és n választási alternatívával), benne minden egyén számára három döntési lehetőséget ($\mathbf{P}(m,n), \mathbf{I}(m,n), \mathbf{P}(n,m)$) feltételezett, és a közösségi döntések jellemzésére bevezette a **csoportdöntési függvény** (group decision function) kifejezést, ami ugyanaz a fogalmat jelöli, mint a véleményösszegző függvény terminusa. Ezek után megfogalmazott négy feltételt, amelyek teljesülését az F csoportdöntési függvénytől "el lehet várni".

May elsőként a döntésképesség (decisive) tulajdonságát vezette be. Mivel nem adott pontos formulákat a fogalmaira, ezért a szóbeli leírásait nekünk kell interpretálni és formalizálni. A szakirodalomban ezt a may-i kritériumot leggyakrabban az univerzalitás fogalmával azonosítják [477], [111]. Ezt a fogalmat korábban már bemutattam (☞432.o.). Én kicsit másként értelmezem May felfogását. Szerintem ő azt akarta rögzíteni, hogy a csoportdöntési függvény legyen minden lehetséges bemenetre értelmezve, és minden bemenő értékre adjon egyértelmű kimenetet. Nem véletlen, hogy May maga döntésképesnek (decisive) nevezte ezt a feltételt: a tétel bizonyítása során az egyértelmű végeredmény elvárására hivatkozott. Igaz, hogy ettől még a feltételek közé kell sorolnunk a tetszőleges bemenet "kezelésének" elvárását, tehát az univerzalitást is. Ebből az következik, hogy – megítélésem szerint – May a döntésképes fogalma alatt az $\mathbf{F}$ választási függvény döntésképes és univerzális jellegét értette.[17]

May következő két tulajdonsága valamifajta szimmetriát fejez ki – egyfelől a szavazók, másfelől az alternatívák között. A szavazók közti szimmetriát az **anonimitás**, az alternatívák köztit a **semlegesség** tulajdonságaival ragadhatjuk meg. A szimmetricitást itt úgy kell érteni, hogy mind a szavazók, mind a választható alternatívák egyenrangúak egymással, nincs egyik sem kivételezett helyzetben a többihez képest, és ebben az értelemben nincs semmiféle aszimmetria közöttük. Az anonimitás definiálásához szükség van egy σ segédfüggvényre, amely a szavazók sorrendjét képes tetszőleges módon előállítani (matematikailag ez annyit jelent, hogy a $\sigma : \mathbb{N} \mapsto \mathbb{N}$ függvény a szavazók $\mathbb{N}$ halmazának tetszőleges permutációját képes létrehozni). A segédfüggvény segítségével már

[17] Az egyértelműség kedvéért azt jeleznem érdemes itt, hogy ez a may-i 'döntésképes' minősítés más, mint a korábban bevezetett, a szavazók koalíciójára vonatkozó 'döntő' tulajdonság (☞434.o.).

meghatározhatjuk az **anonimitás** tulajdonságát.

(AL.43) $(\text{ANYM}) \equiv \text{ANONYM}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall b \forall m \forall n \forall i (\mathbf{F}(\langle \mathbf{R}_i^a(m,n) \rangle_{i \in \mathbb{N}}) = \mathbf{F}(\langle \mathbf{R}_i^b(m,n) \rangle_{i \in \mathbb{N}}))$

anomin (HUN) – egalitáriánus (HUN) – egalitarian (ENG) – anonym (ENG)

Az anonimitás feltétele azt jelenti, hogy a csoportdöntés nem változik, ha a bemenő egyéni szavazatok sorrendjét tetszőleges módon megváltoztatjuk.[18] Ebből azonnal következik, hogy egyfelől mindenki szavazata egyenlőnek számít, másfelől ezáltal teljesíteni lehet az anonimitás követelményét, hiszen a végeredményt tekintve nem számít, hogy ki, mikor, melyik szavazólapon adta le szavazatát. May meg is jegyezte, hogy lehetne ezt a tulajdonságot egalitáriánusnak is nevezni, amely kategória egyértelműen utal a szavazók körében elvárt "egyenlőség" szimmetriájára.

A másik szimmetricitási tulajdonság az alternatívákra, jelöltekre vonatkozik. Ahogy a neutralitásnál azt kellett biztosítani, hogy egyik szavazó se legyen se jobb, se rosszabb helyzetben a többihez képest, úgy itt valami hasonlót kell az alternatívákra megkövetelni. Elvárjuk, hogy minden jelölt helyzete legyen ugyanolyan, vagyis az alternatívák egyenlő eséllyel rendelkezzenek, ne legyen köztük bármiféle aszimmetria. Az anonimitás kritériumával azt fejezhetjük ki, hogy a csoportdöntés érzéketlen az alternatívákra, nem kedvez egyik számára sem.

(AL.44) $(\text{NEUT}) \equiv \text{NEUTRAL}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall \forall bm \forall n \forall o \forall p ((\forall i \forall \sigma(\mathbf{R}_i^a)(m,n) \leftrightarrow \mathbf{R}_i^b)(o,p)) \wedge$
$(\forall i(\mathbf{R}_i^a)(n,m) \leftrightarrow \mathbf{R}_i^b)(p,o))) \rightarrow ((\mathbf{F}(\langle \mathbf{R}_i^a(m,n) \rangle_{i \in \mathbb{N}})) \leftrightarrow \mathbf{F}(\langle \mathbf{R}_i^b(o,p) \rangle_{i \in \mathbb{N}}))$
$\wedge ((\mathbf{F}(\langle \mathbf{R}_i^a(n,m) \rangle_{i \in \mathbb{N}})) \leftrightarrow \mathbf{F}(\langle \mathbf{R}_i^b(p,o) \rangle_{i \in \mathbb{N}})))$

semleges (HUN) – neutral (ENG)

A formula értelmezéséhez figyelembe kell venni azt a tényt, hogy May olyan helyzetet feltételezett, amelyben három döntés, tehát háromféle szavazat képzelhető csak el, amit leegyszerűsíthetünk arra, hogy adott alternatívára vagy *igen*-nel vagy *nem*-mel szavaz-(hat)tak a szavazók. Ekkor viszont az egy jelöltre leadott igenlő szavazatok komplementer viszonyban vannak a jelöltre nem adott szavazatokkal, és ha ilyenkor minden szavazat esetében megcseréljük a szavazatok minőségét (*igen*-ről *nem*-re és fordítva), majd így végezzük el az összegzést, akkor ugyanazt a végeredményt kell kapjuk. Ha ez a feltétel teljesül, akkor valóban biztosíthatjuk az alternatívák semlegességét, hiszen ebben az esetben nem tudunk különbséget tenni a jelöltek között. Az alternatívák iránti semlegességet csak a döntetlen esetén kell feladni, amikor egy másodlagos, döntetlentörő szabály segítségével kell eldönteni, ki legyen a győztes. Ez a szabály azonban kivételszabálynak minősül, nem tartozik a csoportdöntési függvény magjába.

May negyedik kritériumként bevezette a **pozitív érzékenység** tulajdonságát, amiről ő maga megjegyezte, hogy erősebb Arrow második feltételéhez képest, és amiről mi megjegyezhetjük, hogy ez az új kategória nagyon hasonlít a korábban már bemutatott monotonitás fogalmára (↪432.o.).

[18] Az anonimitás és a sorrend felcserélhetőségének összefüggőségét szépen érzékelteti az a videó, amelyen az látható, hogy egy zárthelyi dolgozatot késve leadó diák dolgozatát a tanár nem akarja elfogadni. Erre a diák megkérdezi a tanártól, ismeri-e őt, és a tagadó válasz után becsúsztatja a dolgozatát az asztal szélén álló többi dolgozat közé. Lásd: http://www.youtube.com/watch?v=wRpJM4-PoT8

(AL.45) $(\text{POSR}) \equiv \text{POSITIV_RESPONSIVENESS}(\mathbf{F}) \stackrel{\text{def}}{=} \forall a \forall b \forall m \forall n (\forall i (((\mathbf{P}_i^a(m,n) \rightarrow \mathbf{P}_i^b(m,n)) \wedge (\mathbf{I}_i^a(m,n) \rightarrow \mathbf{R}_i^b(m,n))) \wedge \exists j ((\mathbf{I}_j^a)(m,n) \wedge \mathbf{P}_j^b(m,n)) \veebar (\mathbf{P}_j^a(n,m) \wedge \mathbf{R}_j^b(m,n))) \rightarrow (\mathbf{R}^a(m,n) \rightarrow \mathbf{P}^b(m,n)))$

pozitív érzékenység (HUN) – pozitív reakciókészség (HUN) – positive responsiveness (ENG)

May azt az interpretációt adta a fogalomra, hogy ha a csoportdöntés m alternatívára nézve kedvező (vagy legalább indifferens), és az egyéni szavazók közül valaki megváltoztatja a korábbi döntését n-ról m-re, akkor a csoportdöntési függvény továbbra is kedvező kell maradjon m-re nézve.

Az aggregáló függvények fenti tulajdonságaira, valamint az **egyszerű/többségi többségi szavazás** (simple majority voting)[19] eljárásra támaszkodva May a következő tételt fogalmazta meg:

(AL.46) $(|\mathbb{X}| = 2 \rightarrow \mathbf{F} = \mathbf{F}^{AP}) \leftrightarrow \text{DEC}(\mathbf{F}) \wedge \text{ANYM}(\mathbf{F}) \wedge \text{NEUT}(\mathbf{F}) \wedge \text{POSR}(\mathbf{F})$

May tétel (HUN) – May theorem (ENG)

A két-alternatívás $\mathbf{F}$ csoportdöntési függvény akkor és csak akkor (abszolút) többségi szavazás, ha $\mathbf{F}$ ügydöntő, anonim, semleges és monoton.[20]

A May-tételben az ekvivalenciának mindkét iránya fontos. Egyfelől igaz, hogy ha egyszerű többségi szavazási eljárást alkalmazunk, akkor teljesítjük a fenti négy igazságossági feltételt, másfelől igaz az is, hogy ha a jelzett négy – a méltányosságérzetünket kielégítő – tulajdonság teljesül egy véleményaggregáló függvényre, akkor az egyszerű többségi eljárás lesz. A tétel jelentőségének és helyének érzékeltetéséhez hozzátartozik, hogy tudjuk: csak kétalternatívás közösségi döntési helyzetekben érvényes.

A többségi szavazással hozható kapcsolatba a logrolling, a szavazatcsere jelensége, ami a többségi és kisebbségi véleményeket valló emberek csoportjai közti különös viszonyból fakad, a jelenséget azonban csak azután tudom bemutatni (↪449.o.), miután már bevezettem a stratégiai szavazás fogalmát.

Young tétel

A többségi szavazás mellett gyakran alkalmazott aggregálási technika a pontozásos szavazás, a Borda-pontok módszere, úgyhogy adta magát a kérdés, hogy vajon milyen feltételek mellett alkalmazható ez a szavazási technika. J. Smith cikke volt az első 1973-ban, ami ezzel a kérdéssel foglalkozott [499], de (el)ismertebbé H.P. Young tétele vált, amit Young egy 1975-ös cikkében publikált [589]. Young a pontozásos módszer alkalmazhatósági feltételeire vonatkozóan újabb tulajdonságokat definiált, és ezekre támaszkodva mondta ki tételét. Két új tulajdonságot vezetett be, és további kettőt használt a korábban definiált tulajdonságok közül. A pontozásos módszer axiomatizálását Young egy korábbi cikkében már elvégezte [590], de ott még - részben – más tulajdonságokat használt. Mivel ekkor már közismert volt, hogy a pontozásos módszer ellentmond az irreleváns alternatívák

[19] May egyszerű többségi szavazásnak nevezte a vizsgált módszert, ami kétalternatívás helyzetekben megegyezik az abszolút többségi eljárással, de más esetekben nem feltétlenül. Az egyszerű jelzővel utalni lehet a relatív többségi eljárásra is, ezért jobbnak gondolom itt (is) az abszolút jelző használatát.

[20] A tételt leíró formula első kondiconálisában megadott feltételre igaz a következő: $|\mathbb{X}| = 2 \leftrightarrow |\mathbb{P}| = 2$.

függetlensége elvének, ezért Young javasolta elhagyni ezt a feltételt. Rögtön javasolt viszont bevezetni két új fogalmat, az egyik egy konzisztenciafeltételt, a másik egy folytonossági elvet fogalmazott meg.

(AL.47) $(\text{RIF}) \equiv \text{REINFORCED}(\mathbf{F}) \stackrel{\text{def}}{=} \exists m((\mathbb{N}_1 \cup \mathbb{N}_1 = \mathbb{N} \wedge \mathbb{N}_2 \cap \mathbb{N}_2 = \varnothing \wedge C_1(\mathbb{N}_1) \cap C_2(\mathbb{N}_2) \neq \varnothing \wedge m \in C_1(\mathbb{N}_1) \cap C_2(\mathbb{N}_2)) \rightarrow C(\mathbb{N}) = C_1(\mathbb{N}_1) \cap C_2(\mathbb{N}_2))$

megerősített (HUN) – reinforced (ENG) – konzisztens (HUN) – consistent (ENG)

Ez a konzisztenciakritérium fontos, intuitívnak mondható elvárást fogalmaz meg, amit Young a Pareto-hatékonyság kiterjesztéseként értelmezett. Ha a szavazók két körzetben ($\mathbb{N}_1$ és $\mathbb{N}_2$ csoportban), egymástól teljesen elkülönítve szavaznak ugyanazokra a jelöltekre, az $\mathbb{N}_1$ csoportban a C_1, az $\mathbb{N}_2$-ben C_2 a győztesek halmaza, és van legalább egy olyan jelölt, aki a győztesek mindkét csoportjában benne van, akkor a két szavazói csoport ($\mathbb{N}_1 \cup \mathbb{N}_2$) szavazatainak összesítése után a győztesek halmaza a két részgyőztes halmazának metszete lesz ($C_1 \cap C_2$). Peyton Young a **konzisztens** terminust használta erre a tulajdonságra, Roger B. Myerson szerint H. Moulin volt az, aki kicserélte ezt a 'reinforcement' terminussal megtartva a fogalom eredeti jelentését [370], és a javaslatot Myerson is elfogadta, én pedig a 'reinforcement' terminusát más szófaji alakkal helyettesítettem. A később bemutatandó paradoxonok közül a többkörzetes paradoxonnal (465.o.) kapcsolatban játszik szerepet ez a kritérium, mert a reinforcement tulajdonság létezése biztosíthatja azt, hogy az adott szavazási módszer esetén ne lépjen fel a többkörzetes paradoxon.

A következő tulajdonság értelmezéséhez újra azt kell feltételeznünk, hogy két teljesen elkülönült csoportban zajlik a szavazás (N_1 és N_2). A két csoportban legyen A és B jelölt a nyertes külön-külön. Ilyen helyzetben fel lehet tenni a kérdést, hogy van-e olyan feltétel, amely mentén ki lehet hozni győztesnek az egyik jelöltet a másik ellenében, ha a két körzetet (csoportot) egyesítjük. Erre az egyik igenlő válasz az lehet, hogy elvárjuk az arkhimédeszi tulajdonság érvényesülését, amit egyszer már megragadtunk a sorrendezés vagy archimédeszi rendezés fogalmával (36.o.).

(AL.48) $(\text{CONT}) \equiv \text{CONTINUOUS}(\mathbf{F}) \stackrel{\text{def}}{=} \text{ARCHIMEDEAN_ORDER}(\mathbf{F})$

folytonos (HUN) – continuous (ENG) – arkhimédeszi (HUN) – Archimedean (ENG) – ellenállhatatlan többség (HUN) – overwhelming majority (ENG)

Az újonnan definiált arkhimédeszi tulajdonság akkor áll fenn, ha az egyik csoport részgyőztesét úgy hozhatjuk ki az egyesítés utáni győztesnek, hogy egy kellően nagy számmal beszorozzuk a csoport létszámát, és ezáltal a győztesre leadott szavazatok számát, amivel már le lehet győzni a másik csoport részgyőztesét. Ugyanezt a minőséget fejezi ki az arkhimédeszi preferencia fogalma is, amit máshol elemezek a könyvben (215.o.). Az 'ellenállhatatlan többség' kifejezést Roger B. Myerson javasolta,[21] mint ami tartalmilag legpontosabb fejezi ki a szóban forgó tulajdonságot [370]. Az újonnan bevezetett tulajdonságokat kettő korábban definiált kritériummal kiegészítve Young már megfogalmazhatta saját lehetségességi tételét, amelyben a Borda-pontozási függvényre vonatkozóan tett állítást.

[21] A tulajdonságokat jelző terminusokra vonatkozóan nincs megegyezés, hogy azok milyen szófaji alakban szerepeljenek. Az egész könyvben igyekszem melléknév szófajú terminusokkal hivatkozni a tulajdonságokra, de ez nem mindig sikerül. Myerson javaslatánál is ez a helyzet: sem magyarul, sem angolul nem találtam értelmesen használható melléknévi alakot.

(AL.49) (SCR) ↔ (ANYM) ∧ (NEUT) ∧ (RIF) ∧ (CONT)

Young tétel (HUN) – Young theorem (ENG)

A tétel azt mondja ki, hogy egy társadalmi döntési módszer akkor és csak akkor lehet anoním, semleges, megerősített és folytonos, ha az a pontozásos szabály. Young összetett tulajdonságokat is képzett, és részben ezekre támaszkodva mondta ki tételét (így például az anoním és semleges függvényt Young szimmetrikusnak nevezte, amivel pusztán csak azt akarta kifejezni, hogy az adott eljárás mind a szavazókat, mind az alternatívákat egyenlő módon, elfogultság nélkül kezeli).

7.2.6 No-show paradox

A monotonitás feltétele magától értetődőnek tűnik, hiszen ha egy alternatívára több szavazat érkezik, akkor azt érezzük intuitívnek, hogy ez a többletszavazat ne rontson az alternatíva helyzetén. Mégsem mindig ez a helyzet. Vannak olyan összegzőeljárások, amelyek megsértik a monotonitási tulajdonságot. Fishburn és Brams cikkéből idézhetünk ide egy példát szemléltetésképpen [168].

Egy kisvárosban három jelölt indul a polgámesteri címért: Mrs. Bitt, Mr. Huff és dr. Wogg. A szavazás során a Hare-féle preferenciális szavazás eljárását alkalmazzák. Ennek a lényege röviden annyi, hogy ha az első preferenciaként megadott jelöltek közül valaki megszerzi a szavazatok felét, akkor ő nyer. Ha nem ez a helyzet, akkor az első helyes szavazatok alapján kihúzzák a legkevesebb szavazatot szerző jelöltet, és a többi alternatívára szűkítve folytatják tovább az eljárást – az első fordulóhoz hasonló logika mentén.[22] A szavazás során mindenkinek meg kell adnia a szavazólapon, hogy milyen sorrendben támogatja a három jelöltet. A szavazás napján a Smith házaspár kocsija lerobban útban a szavazóhelység fele, és lekésik a szavazást. Mindkettejük preferenciája megegyezett az alábbi minta szerint: Bitt > Huff > Wogg, tehát a szavazólapra a BHW sorrendet írtak volna. A másnapi újságban megjelent a hír, miszerint a választást Mr. Huff nyerte. Közölték a szavazási eredményeket is.

szavazatok	sorrendek	H > W	W > H
417	BHW	417	0
82	BWH	0	82
143	HBW	143	0
357	HWB	357	0
285	WBH	0	285
324	WHB	0	324
1608		917	691

Az újságbeszámolóból kiderült az is, hogy Mrs. Bitt (a Smith házaspár kedvence) esett ki az első forduló után, mivel az első helyes szavazatok az alábbi módon alakultak:
Mivel az első számlálás után Mrs. Bitt esett ki, a bent maradt két jelöltre már könnyen ki lehetett számolni, hogy kire mennyi szavazat jutott a második számlálási körben (ezek a számok vannak a táblázat utolsó két oszlopában), és ezek alapján jött ki dr. Wogg győzelme. Ezt látva a – korábban aktuáriusként dolgozó – Mrs. Smith kiszámolta, hogy

[22] A módszer leírása itt olvasható magyarul: [343].

Bitt	499 (417 + 82)
Huff	500 (143 + 357)
Wogg	609 (285 + 324)

mi lett volna, ha ők is időben odaérnek, és mindketten leadják a Mrs. Bitt számára kedvező szavazataikat. Ez a plusz két szavazat ugyanis megfordította volna Mrs. Bitt és Mr. Huff közti első helyes szavazatok sorrendjét, és ebben az esetben utóbbinak kellett volna kiesni az első fordulós összegzés után. Ekkor persze a megmaradt két jelöltre adott szavazatok is másként oszlottak volna el.

szavazatok	sorrendek	B > W	W > B
419	BHW	419	0
82	BWH	82	0
143	HBW	143	0
357	HWB	0	357
285	WBH	0	285
324	WHB	0	324
1610		644	966

Az újraszámolásból az derült ki, hogy a Mrs. Smithék számára "adott" plusz két szavazat nemhogy javított volna a helyzetükön, de inkább rontott, hiszen a legkevésbé kedvelt jelöltet segítette volna a szavazatuk. Ezt a paradox helyzetet Fishburn és Brams 'No-Show paradoxon'-nak nevezték el.

Az okos és kitartó Mrs. Smith azonban további felfedezéseket is tett.[23] Észrevette azt is, hogy ha abban a hipotetikus helyzetben, amikor ők is szavaztak volna, ketten másként adták volna le a szavazatukat azok közül, akik a valóságban a BWH sorrendet jelölték be (82-en), és ehelyett a WBH-re szavaznak, akkor a táblázat kicsit, a végeredmény nagyon megváltozott volna. Ekkor ugyanis az első számlálás után megint csak Mrs. Bitt esett volna ki, és a bent maradt két jelölt, Mr. Huff és dr. Wogg között nagyjából az első esethez hasonló módon lakult volna a második körös eloszlás.

szavazatok	sorrendek	H > W	W > H
419	BHW	419	0
80	BWH	0	80
143	HBW	143	0
357	HWB	357	0
287	WBH	0	287
324	WHB	0	324
1610		919	691

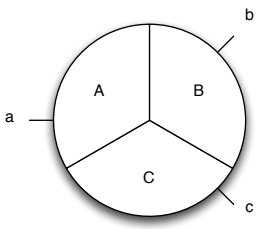
A képzelt eset lehetett volna persze teljesen reális is, így a példa azt mutatja, hogy ezen aggregáló eljárás esetében a monotonitás feltétele nem teljesült, hiszen a második esethez képest csak annyi változás történt a harmadik esetben, hogy a második eset nyertese a harmadik esetben "kapott" két plusz szavazatot, a többletszavatatok azonban nem javítottak, hanem rontottak a helyzetén. A no-show paradoxonról érdemes még megjegyezni,

[23] Fishburnék a cikkben leírnak még további két paradoxont, de ezeket itt nem mutatom be.

hogy az olyan helyzeteket jellemzi, amelyekben a szavazók bizonyos csoportja számára előnyösebb lehet nem szavazni, mint a saját preferenciáikat kinyilvánítani.

7.2.7 Reprezentációs paradoxon

Az előző fejezetben tárgyalt problémához sok szepontból hasonlít a reprezentáció paradoxona, amit Thomas Schwartz írt le először 1985-ben [467]. Schwartz azt a közkeletű nézetet támadta meg, amely szerint a közösségi döntési helyzetekben valamilyen vélemény képviseletének sikeressége mindig arányos a véleményt képviselő reprezentációs erő nagyságával. Másként, kicsit konkrétabban mondva: ha egy szavazótestületben (például parlamentben) egy véleménycsoport (például egy párt) érvényesíteni szeretné a véleményét, akkor annál biztosabban teheti ezt meg, minél több szavazattal rendelkezik. Schwartz szerint nem mindig van ez így. Előfordulhat, hogy a kevesebb szavazat, pontosabban az alacsonyabb reprezentációs szint előnyösebb kimenetet eredményezhet.



Schwartz három példát is bemutatott, amik alapvetően abban különböztek egymástól, hogy milyen magatartást (kooperatív, rafinált és naív hozzáállást) feltételezett a szavazóktól. Itt csak a naív, őszinte szavazói viselkedéshez tartozó esetet ismertetem. Ebben a történetben három körzethez (A-hoz, B-hez, C-hez) tartozó 100 képviselőnek kell dönteni abban a kérdésben, hogy építsenek-e új haditengerészeti bázist vagy ne, és ha igen, akkor hova. A döntés pillanatában már van egy bázis, az a pontban, és a lehetséges döntések között van az az eset is, hogy minden marad a régiben, nem épül új hadikikötő (status quo). Három hely közül csak az egyiken lehet támaszpont, mert túl drága lenne kettőt fenntartani, viszont egyet mindenképpen meg kell finaniszírozni. A három körzet és a három potenciális bázis helyét, ezek egymáshoz való viszonyát mutatja az ábra. A szavazócsoportokat (pártokat, lobbicsoportokat stb.) a körzetek nevével tudjuk azonosítani. Az A csoportba 23, a B-be 28, a C-be 49 mandátum (szavazási lehetőség) tartozik. A három körzet képviselői rendre a következő preferenciákkal rendelkeznek.

körzet	szavazatszám	preferencia
A	23	$a\,b\,c$
B	28	$b\,c\,a$
C	49	$c\,a\,b$

Mindenki azt preferálja leginkább, hogy a saját körzetéhez legközelebb levő helyen épüljön fel az objektum. Ebben a felállásban a képviselőcsoportok vagy beterjesztik a saját körzetükhöz közelebbi helyre vonatkozó javaslataikat vagy nem. Engedjük meg, hogy a beterjesztett javaslathoz lehessen kiegészítést (azaz a másik helyre vonatkozó javaslatot) beterjeszteni. Ha az A és B csoport nem terjeszti be javaslatot, akkor minden marad a régiben, tehát az A-nak kedvező a alternatíva, a status quo lesz a döntés. Ha B-ék javaslata b-re egyedül kerül a színre, akkor a-val versenyez, és ekkor A és C csoport együttesen leszavazza b-t, tehát újracsak marad az a alternatíva. Ha b-t is és c-t egyszerre beterjesztik, akkor előbb köztük kell a szavazást lebonyolítani, amiből 51:49 arányban a b alternatíva kerülne ki győztesen. Attól az esettől eltekinthetünk, amelyben a c-t egyedül terjesztené be a C csoport, mert ezt az ellenfeleik (akár az A, akár a B csoport) mindig

megfúrhatnák azzal, hogy kiegészítő javaslatként bedobnák a b alternatívát a c mellé, és ilyenkor a status quo maradna fent (ahogy az előbb azt bemutattam). Ez a felállás tehát minden körülmények között az a alternatívának kedvez. A B csoport tagjainak ez azt jelenti, hogy a számukra legkedvezőtlenebb alternatíva valósul meg. Tegyük fel most, hogy a szavazati erők kicsit másként oszlanak el úgy, hogy B-nek kevesebb szavazata lesz.

körzet	szavazatszám	preferencia
A	25	$a\,b\,c$
B	24	$b\,c\,a$
C	51	$c\,a\,b$

Ebben a szavazateloszlásban a C mindig meg tuja nyerni a szavazást abszolút többsége birtokában, tehát a kimenet a c alternatíva lesz. Ezzel a változással azonban a B csoport tagjai is jobban jártak, hiszen a korábbinál számukra kedvezőbb végeredmény született, noha kisebb reprenzetációs erővel rendelkeztek, mint azt megelőzően. És ezzel egyben példát is szolgáltattak a Schwartz-paradoxonra. Hunno Nurmi szerint [385] ez a paradoxon közel áll a no-show paradoxonhoz (↰444.o.).

7.2.8 Szavazatmanipuláció

A társadalmi választások kimenetét elsődlegesen az egyéni szavazók által leadott szavazatok határozzák meg. Már az is egy kérdés, hogy miért úgy szavaznak, ahogy szavaznak, és tudjuk jól, hogy a végeredményt azáltal is lehet módosítani, megváltoztatni, hogy az egyéni szavazók véleményére próbálnak meg hatást gyakorolni még a szavazás előtt, esetleg a szavazás során. Az előzetes véleménybefolyásolás, a lojalitáskialakítás és lojalitásfenntartás, a kampányolás indirekt technikái többnyire legális módokon hatnak, míg a szavazás közben alkalmazott direkt befolyásolási technikák (például nyílt szavazás során kifejtett nyomásgyakorlás) döntően illegálisak, de ettől még létez(het)nek. Bármennyire is fontosak ezek a befolyásolási jelenségek, velük itt – egyetlen kivételtől eltekintve – nem foglalkozom. Ezen a ponton azt mondom, vegyük tudomásul, hogy a szavazók valamilyen módon, korábban kialakult véleménnyel, preferenciarendezéssel érkeznek el a szavazás pillanatához. Ideális esetben ennek megfelelően adják le a szavazatukat, amikoris azt mondhatjuk, hogy őszintén szavaztak, vagyis mindenféle stratégiai megfontolás nélkül pontosan a saját preferenciáiknak megfelelően adják le a szavazataikat. Elképzelhető azonban az is, hogy mégsem ez a helyzet, vagyis vannak szavazók, akik valamiért az elsődleges/eredeti preferenciáktól eltérnek a szavazás során. Érdemes megfontolni, hogyan is nevezzük az ilyen helyzeteket, mert bár sokféle kifejezés verseng e tárgyban, ezek némelyike rossz konnotációt hozna be a diskurzusba. A stratégiai szavazás vagy taktikai szavazás még értéksemleges, a hazug szavazó (liar) vagy manipulatív szavazó (manipulation of voter) kifejezéseknek már egyértelműen negatív felhangja van. Akármelyik terminust is használjuk, a lényeg persze nem változik. Azt is látni kell azonban, hogy az a definíciós mozzanat, miszerint minden szavazás stratégiai, amennyiben nem a tényleges preferenciarendezése alapján szavaz valaki, nagyon hétköznapi, nagyon gyakori jelenségekre is ráhúzható. Eszerint ugyanis már azt is stratégiai szavazásnak kell tartanunk, ha egy politikai választáson a két nagy párt és a az egy kis párt jelöltjei közül egy szavazó

inkább valamelyik nagy párt politikusára szavaz, noha szíve szerint a kis párt politikáját követné. Ha viszont tudja, hogy a két nagy versengése mellett a kicsinek esélye sincs megnyerni a választást, és így a kis pártra adott szavazata "elveszne", nem értékesülne, és ezért inkább a neki szimpatikusabb nagy párt képviselőjére teszi le a voksát, akkor már stratégiai, hazug, manipulatív szavazónak minősülne. Lehet azon vitatkozni, ki mennyire érzi kritizálhatónak, hazugnak, manipulatívnak az ilyen magatartást, de itt és most nem az ilyen helyzetek gyakorlati értékelését, hanem csak az elméleti leírását szeretném elvégezni.[24]

A **stratégiai szavazás** során tehát a szavazó azért ad le az igazi preferenciájához képest eltérő, tehát "színlelt" szavazatot, mert azt előnyösebbnek véli az eljárás végeredményét tekintve. Nézzük meg egy egyszerű példát erre a jelenségre [152]. Legyen három alternatíva (U, B, L) és a szavazólapokon lehesen mindegyik alternatívára szavazni úgy, hogy a teljes rendezés során az első alternatíva kapjon 2 pontot, a második 1 pontot, a harmadik ne kapjon semmit (ez a Borda-szavazásnak felel meg). A táblázat baloldali része mutatja a tiszta szavazás eredményét, amikor mindenki a valósi preferenciáinak megfelelően szavazott. A jobb oldalon látható a stratégiai (hazug) szavazás eredménye, ami olyan módon állt elő, hogy mindhárom szavazótábor úgy szavazott, hogy megcserélte a második és harmadik helyen álló alternatíváit.

	igaz szavazás			hazug szavazás		
	LBU	BLU	UBL	LUB	BUL	ULB
szavazat	104	98	7	104	98	7
pontszám	306	307	14	208	203	216

A táblázatban látható eredmények magukért beszélnek. A stratégiai szavazás következményeként a legkisebb saját táborral rendelkező (U) alternatíva győzött. Nagyon életszerű magyarázatot lehet adni ennek a jelenségnek. Az első két, nagy táborral rendelkező alternatíva sokkal közelebb állt egymáshoz a harmadikhoz képest, de a győzelem reménye miatt mindkét tábor tagjai úgy gondolkodhattak a stratégiai szavazás során, hogy úgy gyöngíthetik a riválisukat, ha az esélytelennek gondolt (és mindkettőjüktől távol eső) alternatívára szavaznak második helyen, így a konkurrencia nem kapja meg a második helyekért járó egy-egy pontot. Ám amikor mindkét tábor tagjai fegyelmezetten így is tettek, akkor annak az lett a következménye, hogy a nevető harmadik – egyébként teljesen esélytelen – alternatíva győzedelmeskedett. Az is látszik a táblázatból, hogy ha csak páran lettek volna fegyelmezetlenek a nagyobb táborokból, és az igaz szavazataikat adták volna le, akkor bármelyik nagy győzhetett volna. A bemutatott példa láthatóan "mesterséges", de a stratégiai szavazásra sok valós példát is találhatnánk. Az ausztráliai parlamenti választásokon (ahol a Single Transferable Vote eljárást használják) 2004-ben egyértelműen kimutatható volt, hogy a zöld-, a demokrata- és munkáspárti szavazók a Family First konzervatív párt jelöltjére szavaztak, amit a szakértők a stratégiai szavazás

[24] A manipulálhatóság problémáját Allan Gibbard elemezte először formális módon a lehetetlenségi tételről szóló cikkében [187]. Ebben az 1973-as tanulmányában Gibbard a kérdés előzményeként Michael Dummet és Robin Farquharson 1961-es cikkében megfogalmazott sejtésére hivatkozott: „This would be to assume that every voter votes 'sincerely,' whereas it seems unlikely that there is any voting procedure in which it can never be advantageous for any voter to vote 'strategically,' i.e., non-sincerely." [137, 34]. Gibbard másik előzményként említi William Vickrey egy 1960-as cikkét is [561], amelyben Vickrey egyértelműen a preferenciák stratégiai álcázásáról (strategic misrepresentations of preferences) értekezik.

jelenségével magyaráztak [102].

A stratégiai szavazás első – eddig tárgyalt – szintjén a szavazók "magukban" szavaznak, egyedül változtatják meg eredeti preferenciáikat, nincs stratégiai kapcsolat közöttük. A szavazatok megváltoztatásának azonban vannak más lehetőségei is. A szavazók (bizonyos csoportjai) megállapodhatnak abban, hogy kölcsönös előnyt ígérő helyzetekben együttesen megváltoztatják a szavazataikat. Ezt a jelenséget nevezik szavazatcserének, amikor valamilyen "politikai koalíció" alapján szavaznak stratégiailag az emberek.

7.2.9 Szavazatcsere

A **szavazatcsere** (logrolling) vagy **szavazatkereskedelem** (voting trade) jelenségét ugyan a stratégiai szavazás körébe kell sorolnunk, mégis érdemes önálló fejezetben tárgyalni. A jelenségnek vannak ugyanis olyan jellemzői, amelyek túlmutatnak a szavazás, azon belül is a stratégiai szavazás intézményén. Minél inkább a szavazatcsere interakciós vonásaira koncentrálunk, vagyis minél inkább figyelembe vesszük a jelenségben rejlő dinamikát, annál inkább távolodunk el az egyszerű szavazás jelenségétől, és közeledünk másfajta, dinamikus módon kezelhető társadalmi kapcsolatokhoz.

Na, de nézzük meg, mit is jelent a logrolling jelensége. A megértéséhez abból a megállapításból kell kiindulnunk, hogy a (többségi) szavazás nem veszi figyelembe az egyéni preferenciák intenzitásában meglevő különbségeket [84]. Egy adott ügy iránt nagyon elkötelezett ember egyetlen támogató szavazata ugyanannyit ér, mint egy majdnem teljesen közömbös szavazóé, aki valamiért ellenző szavazatot ad le. Amikor ilyen érzelmi intenzitáskülönbségek tapasztalhatók a szavazók körében, akkor lehet értelme a szavazatcsere intézményének. Ha ugyanis az erősen elkötelezett támogató hajlandó valamilyen előnyt kínálni az addig a másik oldalon álló fél támogató szavazatának megszerzéséért, és az az egy szavazat megváltoztatása épp elég ahhoz, hogy a szavazás végeredménye megforduljon, akkor mindkét résztvevő jobban járhat a logrolling révén. Az elkötelezett ember nyereménye a szavazási eredmény megváltozása, a közömbös szavazó pedig a szavazatáért kapott előny miatt jobban járhat, mint anélkül. A logrolling alkalmazása így Pareto-optimális helyzetet teremthet. Nézzünk meg egy példát [264, 163]. Legyen két kérdés (ϕ és ψ), amiben három szavazónak (A-nak, B-nek és C-nek) kell döntenie, és a két kérdés elfogadásához rendeljük a következő nyereményeket (illetve veszteségeket). Ezekből a két kérdés elfogadása esetén származó nettó társadalmi nyereséget is ki lehet számolni.

	ϕ esetén	ψ esetén		ϕ-re	ψ-re
A nyereménye	-100	-100	A szavazata	nem	nem
B nyereménye	-200	500	B szavazata	nem	igen
C nyereménye	600	-300	C szavazata	igen	nem
nettó társadalmi nyereség	300	100	többségi szavazat	nem	nem

Ha nincs logrolling, akkor a többségi szavazat alapján mindkét kérdésben elutasító döntés születik, még akkor is, ha mindkét kérdésnek van egy-egy lelkes támogatója, akik nagy nyereményhez jutnának valamelyik kérdés elfogadásával. Mivel a többségi vélemény mindkét esetben elutasító lenne, senki sem nyerne és senki sem vesztene a szavazáson. Ha viszont a szavazás végeredményének mgváltoztatására erős motivációval rendelkező

két szavazó, *B* és *C* megegyezne egymással, hogy saját ellenző szavazataikat kölcsönösen megváltoztatják a másik kedvéért, akkor megváltozna a végeredmény, mindkét kérdés többségi véleménnyé válhatna. Egyénileg mindketten, *A* is, *B* is, jobban járnának, hiszen a két kérdés együtteséből pozitív nyereséggel jöhetnének ki, sőt, még realizálódhatna a nettó társadalmi nyereség is, vagyis a csoport egésze is nyerne az ügyön. Mindez persze azon az áron valósulhatna meg, hogy egy szavazónak (*A*-nak) így veszteséget kellene elszenvednie. A tanulság tehát az, hogy ha nem kerülnek kapcsolatba, csak saját egyéni szavazataikra "támaszkodnak" az emberek, akkor elutasítóak lesznek olyan kérdésekben is, ahol egyébként a többségnek és a közösség egészének is jobb lenne ellentétes módon dönteni. Ha mód van a szavazatcserére, akkor ez az optimálisabb helyzet elérhető.

A logrolling jelensége az egyéni preferenciák intenzitáskülönbségeiből táplálkozik. Ezt a technológiát ezért csak akkor lehet eredményesen alkalmazni, ha valamely véleménykisebbségnek jóval erősebb a véleményintenzitása a többséghez képest [84, 134]. A logrolling jelensége kis létszámú és nyilvános döntési eljárással dolgozó testületekben fordulhat elő könnyebben, ezt nevezik explicit logrollingnak. A kis létszám a szavazat értékessége miatt fontos (ha nagyon sok szavazó van, akkor egy szavazat nem sokat ér), a nyilvános szavazás pedig az ellenőrizhetőséget "biztosíthatja" (titkos szavazás esetén nem lehetne tudni, hogy megvalósult-e a szavazatcsere). Buchanan és Tullock szerint létezik szavazatcsere nagylétszámú szavazás során is, amit ők indirekt (máskor implicit) logrollingnak neveztek el. Olyan eseteket soroltak ide, amikor a kampányszervezők úgy állítják össze a szavazási napirendet (pártjelölteket), hogy legyen olyan kérdés (jelölt) benne, amely (aki) iránt erősen elkötelezett szavazókat lehet odavonzani az adott párthoz, csoporthoz még akkor is, ha a csomag más elemei esetleg már nem vonzóak az elkötelezett szavazók számára.

Az állami kiadások folyamatos emelkedését nyilván sok összetevőre bontva magyarázhatjuk csak meg, de a lehetséges magyarázó tényezők között figyelembe vehetjük azt is, hogy a logrolling jelensége sok esetben növeli a kollektív költségeket. Ezt a felismerést már 1959-ben megfogalmazta Gordon Tullock egy cikkében [547]. Ezt a tézist elmondhatjuk más formában is: a logrolling révén alkalmi többségi koalícióba lépve másokra (a kisebbségre) lehet terhelni a többség számára fontos célok költségeinek egy – akár jelentős – részét. Egyszerű példával lehet szemléltetni mindezt. Legyen egy kisvárosunk, ahol három út (A, B, C) felújítására lenne szükség, de a város költségvetésében csak kettőre van meg a fedezet (természetesen az előző évben befizetett adókból). Ekkor szavazást lehet kiírni arról, melyik kettő legyen felújítva a három közül. Az egyszerűség kedvéért tegyük fel, hogy a három út három (egyenlő lélekszámú) városrészben van, és az ott élők mind a saját városrészükben levő út rendbetételét szeretnék elérni. Ha nincs előzetes "megegyezés" az érintettek egy része között, akkor minden városrész lakói *igen*-t mondanának saját útjuk és *nem*-et a többi út felújítására, amiből a többségi vélemény mindhárom esetben az elutasítás lenne. Ha azonban két városrész lakói megegyeznének a szavazatcserében, akkor egymás útjaira is támogatólag szavazva már többségi véleményt szerezhetnének saját útjaik felújítására.

A szavazatcsere eredménye az, hogy két út megépülhet a rendelkezésre álló pénzből. Ezt az előnyt úgy nyerhették el a két városrész lakói, hogy az általuk befizetett adóhoz még hozzájárult a harmadik városrészből befolyt adók összege is, vagyis az nyertesek útfelújítását részben a vesztesek adóiból finanszírozták. Ez az időleges koalícióképzési lehetőség gyakran és könnyen egy olyan spirált indíthat be bárhol a világon, ami – a

	szavazatcsere nélkül			szavazatcserével		
	A-i út	*B-i út*	*C-i út*	*A-i út*	*B-i út*	*C-i út*
A városrész lakóinak szavazata	igen	nem	nem	igen	igen	nem
B városrész lakóinak szavazata	nem	igen	nem	igen	igen	nem
C városrész lakóinak szavazata	nem	nem	igen	nem	nem	igen
többségi szavazat	nem	nem	nem	igen	igen	nem

koalícióba lépő szereplőket állandóan cserélgetve – folyamatosan költi a mások adóját, amit a következő körben más ügyekben "kell" újabb adóelvonásokkal kompenzálni. Ez nyilvánvalóan arra készteti az ilyen koalícióképzési, szavazatkereskedelmi folyamatokat irányító politikusokat, hogy állandóan növeljék az adóbevételeket, hiszen csak ilyen módon lehet új és új időleges koalíciókat képezni.

A logrolling jelenségére mutat be egy konkrét politikai példát William H. Riker az USA történetéből [441]. Magyar példaként hivatkozhatunk az 1990-es MDF-SZDSZ paktumra, amelyben a stabil kormányozhatóság érdekében megegyezett egymással a két legnagyobb kormány-, illetve ellenzéki párt a kétharmados törvények körének kiterjesztéséről, valamint a köztársasági elnök személyéről.

A szavazatcsere, különösen az explicit logrolling intézménye annyiban speciális a stratégiai szavazás jelenségéhez képest, hogy az érintettek itt egyeztetett módon, tudatosan, kölcsönös engedményeket téve térnek el a szavazataik kinyilvánítása során a valós véleményüktől. Ezt igazán jól csak kis létszámú, a nyilvánosságtól elzárt módon egyezkedni képes szavazótestületekben lehet megtenni, másfelől ez gyakorlati eredményét tekintve már sokkal inkább a kompromisszumtestületek működésére hasonlít (↰411.o.). A szavazatcsere a kölcsönös haszon reményében tett kölcsönös engedmények eredménye. Minél inkább a kompromisszum megkeresésének, megteremtésének technikái válnak dominánssá egy adott társadalmi kapcsolatban, annál kevésbé beszélhetünk stratégiai szavazási problémáról, és annál inkább csúszunk át az alkudozáselmélet, a koalícióelmélet, a politikai hatalomtechnikákat leíró elméletek világába.

Megemlítem itt, de egyáltalán nem elemzem a korrupció kérdését is, ami – legalább részben – a stratégiai szavazás területére tartozik. A direkt szavazatvásárlás vagyis a megvesztegetés fontos esetei eredményüket tekintve azt jelentik, hogy fontos döntéshozó emberek valamely szavazás során a kapott előnyökért cserébe megváltoztatják a véleményüket. Ezen az alapon van ebben a fejezetben lenne a helye ennek a jelenségnek is, de ahogy a kompromisszumkötés világát sem elemzem itt, úgy a korrupció egész jelenségét sem tárgyalom. Ha jobban belegondolunk a logrolling és adózás kapcsolatára hozott korábbi példába, akkor észrevehetjük, hogy ide lehetne még kötni a járadékvadászat egész kérdéskörét is, de sem ennek bővebb kifejtését, sem a témakör elemzését nem vállalom itt.

7.2.10 Csalásbiztos eljárás

Miután "kiderült,", hogy egyfelől a tényleges társadalmi gyakorlatban létezik a stratégiai szavazás intézménye, másfelől az erre reflektáló elméletekben is meg lehet ragadni, jellemezni lehet a jelenséget, akkor felmerül a kérdés, vajon vannak-e olyan társadalomtechnikák, amelyek segítségével meg lehet akadályozni vagy legalább meg lehet nehezíteni azt, hogy emberek "csaljanak", hazudjanak a preferenciáinkat illetően. Van-e –

szavazatmanipulációs szempontból vett – csalásbiztos aggregálási technika? A válasz az, hogy bizonyos helyzetekben elképzelhető.

Nézzünk meg egy példát Csontos László egyik cikkéből [114], amelyben a következő történet szerepelt a Clarke-adó mint csalásbiztos technika hasznával kapcsolatban. Képzeljünk el egy szigeten elterülő kisvárost, ahol egyszercsak felmerül egy híd építésének az ötlete a sziget és a szárazföld között. Ilyenkor hihetőnek tűnik azt is feltételezni, hogy a városka lakói között lesznek, akik ellenzik, és lesznek, akik támogatják az ötletet. Ha nem akarják, hogy egy ilyen fontos kérdésben a többségi akarat alapján döntsenek, mivel az a vesztes kisebbségre túl nagy "terheket" róna, ezért más döntési technikát kell keresni. A következő megfontolásból lehet kiindulni. Minél erősebben támogatja vagy ellenzi valaki a tervet, annál inkább hajlandó lehet pénzzel is támogatni a híd felépítését, ha támogatja a tervet, vagy – épp ellenkezőleg – hajlandó lehet kárpótlást fizetni azért az ötlet támogatóinak, hogy ne épüljön meg a híd. Ezt a – pro és kontra – fizetési hajlandóságot kellene valahogyan figyelembe venni a döntés során. A probléma csak az, hogy egy egyszerű szavazás során nem könnyű biztosítani azt, hogy a szavazók megvallják a többiek számra ezeket az elköteleződéseiket. Az ellenzők igyekeznének úgy szavazni, hogy ne derüljön ki, hogy ők hajlandóak lennének kárpótlásra, míg a támogaták azt nemfednék fel szívesen, mennyire érné meg nekik az ötlet megvalósítása. Az egyik megoldása lehet ennek a problémának az ún. Clarke-adó kivetése. Ezt akkor érthetjük meg jól, ha egy táblázatban feltüntetjük, hogy az érintetteknek (háztartásoknak) milyen fizetési hajlandóságaik vannak (amikor is feltételezhetjük, hogy ezek a felajánlások nyilvánosak és egyben igazak).

	fizetési hajlandóság ($-ban)		Clark-adó
háztartás	híd mellett	híd ellen	
1. háztartás	0	9000	0
2. háztartás	0	6000	0
3. háztartás	0	1000	0
4. háztartás	4000	0	0
5. háztartás	17000	0	12000
összes háztartás	21000	16000	0

Az tényleges egyéni preferenciákat az egyének természetesen nem fedik fel egy sima szavazás során. Ennek feltárását célozza meg a Clarke-adó alkalmazása, aminek lényeg az, hogy csak azokra a szavazókra vetik ki, akik – értékben is kifejezhető – szavazata döntő a szavazás végeredménye szempontjából. Ez úgy derül ki, hogy először is mindenkinek meg kell vallania a fizetési hajlandóságát, és ezek alapján ki lehet számolni, melyik opció mellett lennének hajlandóak több anyagi áldozatra a szavazók. A jelen példában a támogatók ajánlottak fel többet (21 ezer dollárt a 16 ezer ellenében). Ebben az esetben döntő szerepe van az ötödik háztartást képviselő szavazónak, hiszen ha ő nem ajánlana fel 16 ezer dollárt a híd felépítésére, akkor az építkezés ellenzői nyernének. A híd felépítésének kérdését tehát ő dönti el. A Clarke-adó mértéke úgy állapítják meg, hogy a választási lehetőségekhez rendelt összesített hozzájárulások különbségéből levonják a döntő szavazó felajánlását. Ez a jelen példában $|(21000-16000)-17000| = 12000$ dollár. Ennyit kellene tehát fizetnie az 5. szavazónak, miközben mást nem terhelnének adóval. A fizetendő adó összege kevesebb, mint amennyit a legelkötelezettebb támogató hajlandó lenne fizetni a tényleges preferenciái alapján, de a nagy kérdés itt az, hogy miért vallaná

be ezt, ha a végén csak neki kell fizetnie adót. Nos, ha csak pár ezer dollárral kevesebbet jelölne meg (bárhol 12 és 17 ezer dollár között), akkor a Clarke-adó ugyanannyi maradna és neki kellene továbbra is megfizetnie. Ha például csak 13 ezret vallana be, akkor a következőképpen számolnák ki a Clarke-adót: $|(17000-16000)-13000| = 12000$ dollár. Ez ugyanannyi, mint korábban volt. Amennyiben 12 ezer dollárt vagy annál kevesebbet vallana meg, akkor viszont megfordulna a szavazás kimenete, és a híd ellenzői nyernének (döntetlen esetén a státus quo vagyis a híd nélküli állapot maradna fent). Ebben az esetben a döntő szavazó elesne a híd felépítéséből származó előnyöktől, ezért neki ekkor sem érné meg erre a szintre lemenni.

Belátható, hogy a másik oldalon, az ellenzők között sem éri meg a preferenciákat hamisan megadni. Amikor az ellenzők feltüntetik a fizetési hajlandóságaikat, akkor ezzel azt jelzik, hogy mennyi veszteséget jelentne számukra a híd felépítése, és ezt mekkora összeggel lennének hajlandóak megakadályozni. Ha az 1. szavazó azzal akarná megakadályozni a híd felépítését, hogy a ténylegesnél nagyobb veszteséget jelezne előre (amivel nagyobb fel-nem-építési kárpótlást vállalna magára), mondjuk 14 ezer dollárt, akkor megnyerhetnék a választást (21-21 ezer dolláros döntetlen mellett a status quo maradna fent), de neki magának ekkor 14 ezer dollárnyi Clarke-adót kéne fizetnie ($|(21000-21000)-14000|$) dollár. A fizetendő adó viszont 5 ezer dollárral több lenne, mint a tényleges fizetési hajlandósága, emiatt értelmetlen lennne hazudnia.

A Clarke-adó tehát mindkét oldal résztvevőit arra ösztönzi, hogy a valós fizetési hajlandóságaikat vallják meg, és ezáltal a tényleges egyéni preferenciák (tehervállalási hajlandóságok) alapján szülessen meg a közösségi döntés. Ez egyfajta igazságosság megvalósulását képes támogatni azzal, hogy megakadályozza a valós preferenciák eltitkolásának lehetőségét. A módszernek azonban gyakorlati szempontokból tekintve komoly hiányosságai vannak. A preferenciák feltárása fontos lehet ugyan a kollektív közérzet szempontjából, de a preferenciák, a vélt nyereségek vagy veszteségek kinyilvánításának semmi köze nincs az utánuk következő valós folyamatokhoz (itt a híd felépítéséhez). A módszer alapján nem lehet semmit sem kezdeni azzal, ha nem elég a pénz a tervek megvalósításához, de arra sem ad semmi útmutatót, hogy mit kell tenni abban az esetben, ha netán maradna a projekt végén a beszedett Clarke-adóból. Ez a megoldás nem kezeli az érintettek preferenciái, vágyai, tudásai és a megvalósult folyamatok valós költségei közti diszkrepanciát.

7.2.11 Gibbard-Satterthwaite tétel

A szavazás során lehetséges manipuláció további lehetőségét jelenti a **taktikai szavazás** (tactical voting) intézménye. Ebben az esetben a szavazó tudja befolyásolni a végeredményt azzal, hogy nem a valós preferenciái szerint, azaz nem őszintén szavaz abban a reményben, hogy ezzel a maga számára kedvezőbb végeredmény születik a szavazás során. Sok más manipulációs lehetőséghez képest ez bizonyos értelemben talán a "legdemokratikus" formája a manipulációnak, hiszen míg a manipulációs lehetősége a legtöbb esetben csak kevesek számára adottak, a taktikai szavazással elvileg mindenki élhet. A taktikai szavazást a szavazatok manipulációjának is mondhatjuk. A szavazatmanipulációval kapcsolatos leghíresebb tételt ketten is kimondták nagyjából egyidőben, Allan Gibbard [187], illetve Mark Allen Satterthwaite [460], ezért kettejük nevét összekötve hivatkoznak a "közös" Gibbard-Satterthwaite tételre.

Gibbard azzal indítja az elemzését, hogy a preferenciaaggregálási függvények (vagy ahogy Gibbard nevezi őket: a szavazási sémák) a – játékelméleti értelemben vett – játékok speciális eseteit jelentik. Ahogyan megadunk egy játékot, ugyanúgy kell leírni egy szavazási sémát is. Ami a játékokban a stratégiát jelenti, az a szavazásban megfelel egy adott preferenciarendezésnek (egy leadott szavazatnak). Ez a hasonlóság azonban még kevés ahhoz, hogy megragadhassuk a stratégiai szavazás jelenségét. Az igazi, illetve a hazug szavazatok "kezeléséhez" még tudnunk kell azt is, hogy az adott játékos (szavazó) melyik preferenciarendezése az "igazi", az őszinte és melyik nem (tehát stratégiai). Ezt megadhatjuk egy függvénnyel, ami minden szavazóhoz hozzárendeli a lehetséges preferenciarendezések halmazából az egyetlen őszinte preferenciarendezést. A manipulálhatóság tulajdonsága ezek alapján viszont már nem a játékok, hanem csak a szavazások sajátosságaként értelmezhető. Meg kell tehát adni, hogy mit jelent az, hogy egy szavazási eljárás manipulálható vagyis az szavazók számára lehetővé teszi a taktikai szavazást.

(AL.50) $(\text{MAN}) \equiv \text{MANIPULABLE}(\mathbf{F}) \stackrel{\text{def}}{=}$

manipulálható (HUN) – manipulable (ENG) – **színlelt szavazás** (HUN) – stratégiai (szavazás) (HUN) – taktikai (szavazás) (HUN) – strategic (voting) (ENG) – tactical (voting) (ENG) – sophisticated (voting) (ENG) – insincere (voting) (ENG)

(AL.51) $(\text{STRP}) \equiv \text{STRATEGY-PROOF}(\mathbf{F}) \stackrel{\text{def}}{=} \neg\text{MANIPULABLE}(\mathbf{F})$

színlelésbiztos (HUN) – nem manipulálható (HUN) – hazugságbiztos (HUN) – strategy-proof (ENG) – non-manipulable (ENG)[25]

Csak az egyértelműség kedvéért rögzítem, hogy ezek a tulajdonságok két dologra is vonatkozhatnak. Egyfelől irányulhatnak magára a szavazásra, azaz a szavazat leadásának aktusára, másfelől vonatkozhatnak a szavazási eljárás egészére. Az egyéni szavazó lehet a taktikai szavazó, a szavazási rendszer lehet színlelésbiztos, nem manipulálható vagy épp fordítva, lehet manipulálható, de nem mondhatjuk, hogy az eljárás taktikai vagy stratégiai lenne (legfeljebb csak azt, hogy az eljárás megengedi vagy épp megakadályozza a taktikai vagy stratégiai szavazást). A színlelés

(AL.52) $\big((\text{STRP}) \wedge (\text{UNI})\big) \rightarrow (\text{DIC})$

Gibbard-Satterthwaite tétel (HUN) – Gibbard-Satterthwaite theorem (ENG)

A tétel szeint egy legalább háromszemélyes helyzetben nincs olyan racionális döntési mechanizmus, amely egyszerre kielégítené az univerzalitás (UNI) és a színlelésbiztosság (STRP) feltételét, és eközben nem-diktatórikus (NODIC) maradna.[26] A tételből következik: ha azt akarjuk elérni, hogy a racionális döntés őszinte, taktikai szavazásoktól mentes legyen, azaz mindenki a legpreferáltabb alternatíváját nyilvánítsa ki, akkor szükségképpen a racionális választás elitista/diktatórikus jellegű lesz, azaz lesz egy olyan diktátor,

[25]Ezen a ponton sem könnyű feladat megfelelő fordítást találni az angol kifejezésre. A felsorolt lehetőségeken túl a 'strategy-proof' kifejezésre lehetne még a taktikázásbiztos, csalásbiztos, csalásgátló, helyezkedésgátló, taktikaérzéketlen magyar fordításokat is alkalmazni.

[26]Gibbard-Satterthwaite tétel kicsit másfajta interpretációja szerint ha az alternatívák száma több kettőnél, akkor a nem diktatórikus, racionális választás ki van téve a stratégiai manipuláció lehetőségének. Ugyanezt megfordítva mondva: kettőnél több alternatíva esetén csak a diktatórikus közösségi választási szabály képes biztosítani a színlelésmentes választási eredményt.

aki a döntési helyzetben belül 'túl nagy' hatalommal rendelkezik, és preferenciája automatikusan kollektív preferenciává válik. A Gibbard-Satterthwaite tétel széleskörű elfogadásával práhuzamosan megindult a tétel finomítása, újraértelmezése, újfajta bizonyításokat dolgoztak ki. Salvador Barberà alternatív bizonyítást adott a tételre a pivotális szavazó fogalmára támaszkodva [34]. Alan D. Taylor pedig megmutatta [539], hogy az Arrow-tétel és a GS-tétel ekvivalens egymással.

McKelvey tétel

A lehetetlenségi tételek egymást kizáró feltételeket állapítanak meg a közösségi véleményaggregálások világában. A preferenciákkal kapcsolatos lehetetlenségi tételek sorában az Arrow-tétel mondta ki a legfontosabbat: amikor az egyéni véleményeket összegezzük, nem úszhatjuk meg a diktatórikus mozzanatot, a közösségi vélemények kialakításakor valaki(k) mindig döntő szerepbe kerül(nek). A Gibbard-Satterthwaite tétel azzal pontosította mindezt, hogy ráirányította a figyelmet a stratégiai szavazás, a szavazatmanipuláció jelenségére. A tétel középpontjában két egymást kizáró feltétel került, amelyek közül választanunk kell: a szavazási rendszerünk vagy az őszinte szavazást vagy a diktátormentességet biztosíthatja, a kettőt együtt nem. A most bemutatandó, harmadik lehetetlenségi tétel szerint a szavazás során minden megtörténhet (anything might happen), mert a szavazás lebonyolítása, a **szavazási napirend** vagy menetrend (agenda setting) kialakítása olyan erőt ad annak kezébe, aki e fölött rendelkezik, amivel bármilyen végeredményt ki tud hozni. És ez a végeredmény nyilvánvalóan az az alternatíva lesz, amit ő (a napirendkezelő) legjobban preferál.

Richard Drummond McKelvey volt az, aki 1976-ban írt egy cikket, amiben megfogalmazta a – később róla elnevezett – tételét [333], amit aztán némely ponton kiegészített, módosított egy három évvel későbbi tanulmányában [334]. Érdekes kapcsolatot jelez az a tény, hogy nagyon gyakran káosz-tételként hivatkoznak McKelvey tézisére. A káosz metafora használatának azonban vannak előzményei. Harold Hotelling első tétele óta (↪420.o.) a preferenciák kezelésére, értelmezésére a téri metaforát használták. Duncan Black tétele (↪439.o.) az egycsúcsú preferenciák (↪421.o.) létezése esetén áll fent, és ez a feltétel természetesen nem mindig teljesül. Nem is teljesülhet minden esetben, hiszen az Arrow-tétel pont ennek lehetőségét tagadja. A Black-tételt így lehetségességi tételként is értelmezhetjük, és feltehetjük a kérdést, vajon vannak-e másfajta feltételek, amelyek alapján ki lehetne másfajta lehetségességi tételeket is mondani. Természetesen léteznek további ilyen feltételek és tételek. 1966-ban Plott cikkében megírta, hogy a preferenciák sugaras vagy radiális szimmetriája (radial symmetry) esetén a többségi eljárás diktátormentesen is képes racionális végeredményt produkálni. 1972-ben Otto A. Davis, Morris H. DeGroot és Melvin J. Hinich bebizonyította az euklideszi preferenciákra, hogy $W(x^*) = \varnothing$ nyerőhalmaz eleme(i) akkor és csak akkor többségi Condorcet-győztes(ek), ha megegyeznek a teljes mediánnal (total-median) [123].

A lehetségessségi tételek feltételei összefüggenek egymással: a Black-tételben kihasznált feltétel, a preferencia egycsúcsúsága speciális esete a sugaras szimmetriának, utóbbi pedig a teljes medián egyik altípusa [109]. Ezek a feltételek azonban nem mindig teljesülnek a tényleges társadalmi gyakorlatban. McKelvey meg is jegyezte egy helyen, hogy a teljes medián követelménye olyan erős, hogy gyakorlatilag lehetetlen kielégíteni. Minél több szavazó van, annál könnyebb megsérteni az elvet, hiszen elegendő egyetlen

szavazó preferenciájának miniális megváltoz(tat)ása, és borul a feltételekben megfogalmazott kényes egyensúly [334, 1085].

McKelvey az 1976-os cikkében még feltételezte a preferenciákra vonatkozóan az euklideszi tulajdonság létezését, vagyis azt, hogy az alternatívák tere euklideszi, de az 1979-es tanulmányában ezt tovább általánosította, amikor az euklideszi tér feltételezése helyett elegendőnek tartotta tetszőleges topológiai tér létezését. David Austen-Smith és Jeffrey Banks 1999-ben pedig azt bizonyította be, hogy az euklédeszi preferencia feltételezése helyett elég lehet a folytonossági feltétel rögzíteni [27]. McKelvey gondolatmenetének kiindulópontjában egy olyan lehetségességi tétel áll, amely kimondja ugyan, hogy a többségi szabály biztosíthatja a Condorcet-győztest bizonyos feltételek mellett, ám McKelvey ezek után épp azt mutatja meg, hogy ez az elvi lehetőség – az alternatívák terében tapasztalható kaotikus viszonyok miatt – gyakorlatilag sosem valósulhat meg. Emiatt húzható rá a káosz-tétel minpsítés.

(AL.53) $(|\mathbb{X}| > 2 \wedge |\mathcal{N}| > 2) \rightarrow$ (DICT)

McKelvey (káosz) tétel (HUN) – McKelvey (chaos) theorem (ENG)

A tétel szerint tehét ha több, mint három szavazó van, és legalább három politikai kérdés van napirenden, akkor a döntési folyamat nem diktátormentes. Ebből viszont már levezethető az is, hogy ilyen esetben a **napirendet** a "diktátor", azaz a napirendkezelő (agenda setter) fogja irányítani (feltételve, ha ismeri a többiek preferenciáit, illetve az ügyletnek nincsenek – komoly – tranzakciós költségek).

A napirend-manipuláció híres példájaként hivatkoznak gyakran a 'Powell Amendment' címkével arra az – Adam Clayton Powell amerikai képviselő által gyakorolt – módszerre. Powell célja az volt, hogy a szövetségi forrásokból ne támogassanak olyan állami szintű intézményeket, alapítványokat, iskolákat, amelyeken belül valamilyen szegregációs politika érvényesült, és mindezt olyan formai eszközökkel érte el, hogy az olyan törvényekhez, amelyeknek potenciálisan volt valamilyen szegregációs tartalma, relevanciája, mindig benyújtott egy kiegészítést, amely megtiltotta az olyan államok szövetségi támogatását, amelyeken belül szegregációt alkalmaztak állami intézményekben [441]. Riker utólagos elemzésekkel bemutatta, hogy amikor az iskolák szövetségi támogatásáról kellett szavazni, akkor a déli államok demokrata képviselői megszavazták volna a kiegészítés nélküli beterjesztett javaslatot, mert sok pénzt hozott volna a déli iskolák számára, de a Powell-féle "szegregációs" kiegészítéssel ellátott javaslatot már elvetették, így viszont az egész ügylet bukott.

A napirend-manipulációval kapcsolatban végeztek empirikus kísérleteket is, amelyek közül Michael E. Levine és Charles R. Plott kísérletei váltak híressé [304], [416]. Magyar kísérleteket mutat be Meszerics Tamás egy cikke [351], és magyar példát is hoz a napirend-manipulációra Körösényi András egy tanulmányában [287].

Korábban említettem, hogy a McKelvey-tétel a lehetségességi tételekkel szemben megfogalmazott lehetetlenségi tétel, ami gyakorlati szempontokra hivatkozva "utasítja el" a diktátormentesség lehetőségét. De azt is említettem, hogy mindez bizonyos típusú preferenciákra vonatkozóan áll fent. Fel lehet itt tenni azt a kérdést, hogy más preferenciatípusok esetében lehet-e még Condorcet-értelemben stabil a többségi szavazás. Stephen Ansolabehere és James M. Snyder Jr. szerint elképzelhető ilyen eset [18]. Snyderék szerint a politikai küzdőtéren nem csak olyan politikai ügyek jelenhetnek meg, amelyek megosztják a szavazókat (amelyek prefenciaprofiljait az eddigiekben elemeztük),

de előfordulhatnak, úgymond, konszenzusos témakörök is. Snyderék **egypólusú ügyek** vagy konszenzusos ügyeknek (valence issues) nevezik azokat a politikailag releváns témaköröket, amelyeket minden szavazó ugyan módon, ugyanolyan polaritás, valencia mentén ítél meg, és csak az ügyekhez való viszonyulás intenzitásában térnek el a választók nézetei. Vannak ügyek, amelyeket vagy mindenki támogat, vagy mindenki ellenez. Ilyen téma lehet a bűnözés visszaszorítása, a gazdasági növekedés gyorsítása, a nemzeti büszkeség ápolása, valamely közfelháborodást kiváltó esemény kezelése, vagy lokális szinten egy választókerület egész népessége számára előnyt jelentő szolgáltatások, erőforrások megszerzése. Ezek az egypólusú ügyek megváltoztathatják a politikai színtér "dinamikáját". Amennyiben a többdimenziós politikai térben van legalább egy konszenzusos, egypólusú ügy, akkor előállhat a McKelvey-tétel által elvi lehetőségként tételezett helyzet, amikor létezik üres nyerő halmaz, amikor diktátormentesen Condorcet-győzteshez juthatunk.

(AL.54) → ,????

Ansolabehere-Snyder tétel (HUN) – Ansolabehere-Snyder theorem (ENG)

Ebbe az irányba tett lépésként értékelhetjük még Herve Moulin bizonyítását, aki szerint az egycsúcsú preferenciák (↰421.o.) esetén a társadalmi választási függvény nemmanipulálható [369]. Ehhez hasonló Alejandro Saporiti tétele is, amit az egyszerváltó preferenciákra (↰425.o.) vonatkozóan mondta ki a nem-manipulálhatóság lehetőségét [459]. Paul Rothstein pedig a medián szavazó elméletével szemben dolgozta ki a reprezentatív szavazó elméletét [450]. Rothstein olyan új elemzési irányra, lehetőségre mutatott rá, amely azt ígéri, hogy a szavazásokkal kapcsolatos politikai elemzések régóta létező, és sokáig megingathatatlannak tűnő tradícióját lehet képes megreformálni.

7.2.12 AI-válasz

A társadalmi választások elméletének egyik kulcseleme az aggregálási, választási eljárás fogalma. A közösségi döntések meghozatalához szükséges választási eljárások algoritmusok, így nem kellene, hogy meglepő legyen az a tény, hogy az 1990-es évektől kezdve megjelentek matematikusok, mérnökök, informatikusok is ezen a kutatási terepen. A formális eszközök iránti igény korábban is vonzotta az ilyen felkészültségű szakembereket ide. Miután egyre másra jelentek meg olyan új feladatok, amikor számítástechnikai környezetben kellett szimulálni ilyen-olyan választási helyzeteket,[27] érthető, hogy a fejlesztések, fejlesztők oldalán is erősödött az igény a szavazáselméletek tanulságainak becsatlakoztatására. Az azonban mégiscsak meglepő, hogy milyen technikákat hoztak be ebbe az új diskurzusba az informatika, matematika oldaláról.

Ahhoz, hogy érteni lehessen, milyen fogalmakkal, elméletekkel operálnak a mérnökök, informatikusok ezen a terepen, tudnunk kell valamit a témánk szempontjából nagyon fontos bonyolultságelmélet világából. A számítástechnika fejlődése során a matematikusok, informatikusok sokat foglalkoztak az algoritmusok problémakörével. Az egyik kiemelten fontos kérdés az volt számukra, hogy milyen számítási feladatot jelentenek a

[27] Egyre bővül az a lista, amely a számítástechnika és a választási probléma összekapcsolódásával kapcsolatos jelenségekről szól – a multiágens alapú döntési, tervező rendszerektől a kollaboratív szűrés technikáját alkalmazó rendszereken át a weboldalakon elhelyezett linkeket szavazásokként felfogó és a keresések hatékonyságát ezáltal növelő keresőrendszerekig.

különféle algoritmusok a gépek számára, vagyis mennyi idő szükséges valamely probléma megoldására, a hozzá tartozó algoritmus lefuttatására. Anélkül, hogy a fogalmak lényegi tartalmát, s főleg a mögöttük levő elméleteket megértsük, szükségünk van arra, hogy tudjuk, milyen fogalmak mit jelentenek, mire és miért használják őket a mérnökök. P- és NP-problémákról, illetve utóbbiakon belül NP-nehéz és NP-teljes problémákról szoktak beszélni. Nekünk csak annyit kell tudni, hogy ezek a megoldandó feladatok bonyolultságát, és a megoldáshoz szükséges számításigény nagyságrendjét fejezik ki. A P osztály a polinom időben megoldható problémákból áll, azaz olyan problémákból, melyekhez létezik olyan k konstans, hogy a probléma n hosszú bemenet esetén $O(n^k)$ idő alatt megoldható. Ez annyit jelent, hogy az ide tartozó algoritmusok vállalható idő alatt lefutnak a számítógépeken. Ezzel szemben az NP-nehéz (NP-teljes) osztályba tartozó problémák olyanok, hogy nagy elemszám mellett gyakorlatilag megoldhatatlanok. A megoldhatatlanságot itt csak abban az értelemben kell érteni, hogy kiszámíthatatlanok, vagyis ha elindítanánk tetszőleges képességű és méretű számítógépen a problémát elvileg megoldani képes algoritmust, akkor az sosem jutna a futási idő végére. Elvileg tehát megoldható lenne, csak gyakorlatilag nem számítható ki.

Ezek a problémaosztályok "normál ügymenetben" kihívást jelentenek a matematikusok számára, amikor a cél az ilyen bonyolult problémák megoldása valamilyen eszköz, trükk segítségével. Kiderült azonban, hogy nagyon praktikus haszna is lehet az ilyen problémáknak. Praktikus védelmet jelenthetnek ugyanis olyan helyzetekben, amikor valamilyen informatikai erőforráshoz való hozzájutást effajta problémák megoldásának közbeiktatásával lehet megfelelő módon kezelni ("csak akkor léphetsz be ezen a ponton, ha megoldod ezt a problémát"). A NP-osztályba tartozó feladatok tehát egyfajta kapuőrként is használhatók. És ez a kifordított gondolat segített abban, hogy a mérnökök megjelenjenek a szavazáselméletek világában és használni kezdjék ezeket a bonyolultságelméleti fogalmakat. Ha ugyanis az a feladat, hogy megakadályozzuk az – egyéni vagy koalíciós – stratégiai szavazás lehetőségét, akkor erre meg kell keresni és hasznosítani kell azokat a számítástechnikai eszközöket, amelyeket más területeken már sikeresen alkalmaztak.

Az 1990-es évek kezdetén Bartholdi szerzőtársaival olyan megoldásokat dolgozott ki, amelyek a fenti elveket ültették át a gyakorlatba. Kiderült, hogy a társadalmi választási eljárásokat valóban annál kevésbé lehet manipulálni, minél bonyolultabbak ezek az eljárások algoritmuselméleti szempontból. Ez praktikusan azt jelenti, hogy az aggregálási mechanizmusok integritásának a szükséges számításigény értelmében vett komplexitás az igazi biztosítéka [38], [37]. Bartholdiék megoldási javaslatai mind a jelöltek, mind a szavazók számára nagy számokat írtak elő, így tudták biztosítani a megfelelő számítási bonyolultságot.

Vincent Conitzer, Tuomas Sandholm és Jérôme Lang azzal érveltek, hogy a jelöltekre vonatkozó számossági elvárások nem mindig életszerűek, hiszen a választási helyzetek egy részében csak kevés alternatíva közül kell választani, ezért olyan számításokat végeztek, és ezek alapján olyan új javaslatokat tettek le az asztalra, amelyek úgy biztosítottak megfelelő számítási bonyolultságot az egyes választási eljárésokkal kombinálva, hogy a jelöltek számára alacsony számú alternatívát is fent lehetett tartani [104]. Tanulmányukban először röviden bemutatták, milyen dimenziókban, milyen kérdések mentén vizsgálják a jelenséget, Kik a manipulátorok? Lehetnek egyéni vagy csoportos, koalícióba szerveződő szavazók. Mit tudnak a manipulátorok a tiszta szavazók szavazatairól, preferenciáiról? Lehet teljes vagy tökéletlen a tudásuk. Súlyozottak-e vagy sem a szavazók?

Ezáltal lehet különbséget tenni különböző szavazótípusok, szavazók között (például részvényesi szavazás során). Mi a manipuláció célja? Vagy az, hogy egy adott jelölt nyerjen (ekkor beszélhetünk konstruktív szavazásról), vagy az, hogy ne nyerjen (ez a destruktív szavazás). A vizsgálatba bevontak 9-10 szavazási módszert, és a fenti szempontok alapján sorra mondtak ki tételeket arra vonatkozóan, hogy a különböző feltételek mentén adott eljárások esetében adott (2-7 fős) jelöltszámok mellett milyen számítási problémát jelent a manipulációs szavazás végrehajtása (megakadályozása). Csak a szemléltetés kedvéért bemutatok itt egy táblázatot a cikkből, ami a súlyozott, koalíciós, konstruktív manipulációkkal kapcsolatos adatokat tartalmazza.

jelöltek száma	2	3	4, 5, 6	≥7
Borda	P	NP-teljes	NP-teljes	NP-teljes
véto	P	NP-teljes	NP-teljes	NP-teljes
STV	P	NP-teljes	NP-teljes	NP-teljes
kétfordulós többségi	P	NP-teljes	NP-teljes	NP-teljes
Copeland	P	P	NP-teljes	NP-teljes
maximin	P	P	NP-teljes	NP-teljes
randomizált kupa	P	P	P	NP-teljes
reguláris kupa	P	P	P	P
többségi	P	P	P	P

A táblázat legfontosabb "üzenete" az, hogy bizonyos eljárások esetében 3 jelöltnél is, másoknál pedig 4-7 jelöltnél már meg lehet akadályozni a manipuláció lehetőségét, mert olyan feltételek esetén a számítási bonyolultság már kellően naggyá válik.

A szavazatmanipuláció speciális esete a szavazatvásárlás vagyis a megvesztegetés (bribery). Kicsit furcsa, szokatlan, egyeseknek talán elfogadhatatlan a szavazatvásárlást megvesztegetésnek minősíteni, pedig valóban az. Valaki (itt a szavazó) pénzért (vagy más ellenszolgáltatásért, valamilyen előnyért) olyan döntést hoz, ami másoknak kedvező, és az illető ezt a döntést a pénz (előny) felajánlása nélkül nem hozta volna meg, akkor beszélünk megvesztegetésről, csalásról, megkenésről, és ez a meghatározás illik a szavazatvásárlás esetére is. Piotr Faliszewski társaival együtt amellett érvelt, hogy a megvesztegetési problémát érdemes a manipulációs problémától elkülönítve vizsgálni, mert speciális tulajdonságai vannak [151]. Nem is annyira a szűkebb értelemben vett választás jelenségén belül érhető tetten ez a speciális jelleg, hiszen az "önkéntes", külső befolyástól mentes stratégiai szavazás eljárástechnikai értelemben nem különböztethető meg a szavazatvásárlás útján megváltoztatott véleményektől. A megvesztegetés a eljárástechnikai részleteken túl válik érdekessé (és elemezhetővé). Mennyi szavazatot kell megvenni, hogy biztosítani lehessen az adott jelölt győzelmét, ugyanannyit kell-e fizetni minden megvásárolt szavazónak, vagy vannak minősített szavazók, összesen mennyibe kerül a szavazatvásárlás, lehet-e pénzen kívüli eszközökkel fizetni – ilyen kérdések merülnek fel, amelyekkel kapcsolatban ugyanúgy és ugyanolyan indokok miatt kell a bonyolultságelméletet segítségül hívni, ha meg akarjuk akadályozni az ilyen manipulációk lehetőségét. Ha el akarjuk lehetetleníteni az iylen próbálkozásokat, akkor valóban ismerünk kell minden olyan technikai lehetőséget, ami módot adhat arra, hogy a manipulációs kísérletek sikeresek lehessenek. Érdekes tanulságokat nyerhetünk az ilyen vizsgálatokból. A megvesztegetés modellezésekor például kiderült, hogy a jelenség egyik alfajaként minősített "hátul cserélős" megvesztegetés (swap bribery) során indirekt módon is lehet sikert elérni [152]. A

megvesztegetés során ugyanis lehetséges, hogy valaki vagy nagyon drágán hajlandó a saját preferencialistája első helyén levő jelöltet elmozdítani a csúcsról, vagy erre a egyáltalán nem vehető meg. A preferencialistáján hátrébb levő jelöltek között viszont már sokkal könnyebben (sokkal olcsóbban) meg változtatja a véleményét. A választás befolyásolása szempontjából azonban sok esetben sikeres taktika lehet, ha az ilyen szavazók esetében a gyenge preferenciák sávjában érik el a szavazatcseréket, és ez a sok apró rangsorváltoztatás a hátsó preferenciarégiókban egymással összeadódva mégiscsak meghozhatja a kívánt eredményt.

7.2.13 Párosításelmélet

A kollektív preferenciák kezelésének egy speciális területét jelentik azok a problémák, amelyeket a párosításelmélet vizsgál a hatvanas évek elejétől fogva. David Gale és Lloyd S. Shapley írta le először azt a problémát, amit párosítási feladatnak, majd később a problémát leíró elméletet **párosításelméletnek** neveztek el [180]. A probléma megoldásra Shapleyék egy olyan algoritmus javasoltak, ami a mai napig Gale-Shapley algoritmus néven ismert és hivatkozott szerte a világban. A stabil házasítási vagy párosítási probléma (stable matching problem) arról szól, hogy össze kell párosítani lányokat és fiúkat egymással azok preferenciarendezései alapján a lehető legoptimálisabb módon, vagyis úgy, hogy mindeki a lehető leginkább kedvelt partnerével kerülhessen egy párba. A stabil megoldást itt úgy értelmezhetjük, hogy a kapcsolatok között nincs egyetlen **blokkoló pár** sem, vagyis nem létezik olyan fiú-lány kettős, akik nem házastársak, de mindketten boldogabbak lennének egymással a jelenlegi párjukhoz képest. Ezt úgy is fogalmazhatjuk, hogy ha egy fiú és egy lány nem házas, akkor legalább az egyikük jobban szereti a jelenlegi házastársát, ezért nem lehet őt elcsábítani, vagyis az algoritmus válásbiztos (divorce-proof). Azt természetesen nem várhatjuk el, hogy minden esetben mindenkit össze lehessen kapcsolni a legjobban akart partnerével, de azt mindenképpen biztosítani kell, hogy mindenki olyan jelöltet kapjon a végén, aki rajta van a preferencialistáján (vagyis mindenki találjon magának valakit a preferencialistájáról).[28] Shapleyék erre a feladatra dolgozták ki azt az algoritmust, ami végül is képes volt minden esetben stabil megoldást szállítani.

A **késleltetett elfogadási algoritmus** (deferred acceptance algorithm) lényege az, hogy a fiúk minden körben felkérik a lányokat a preferencialistájuk alapján,[29] hogy legyenek a párjuk (táncpartnerük stb.), a lányok pedig a kérők közül a nekik legjobban tetsző fiúknak feltételes igen-t (talán-t) mondanak (ha van ilyen), míg a többieknek nem-et. Az elküldött fiúk a következő körben felkérik a preferencialistájukon szereplő következő lányt. A lányok az új kérőket elutasítják, ha azok hátrébb vannak, mint a már esetleg talán-nal talonba tett fiú vagy ha annál jobb helyen áll az új jelölt, akkor lecserélik a régit az újra. A lecserélt fiú újra beáll a kérők közé, és a következő fordulóban – a többi visszautasított társával együtt – megy a preferencialistája következő helyén szereplő lányhoz. Ekkor a lányok újra választanak. Ez így megy addig, amíg minden fiú ki nem köt valahol, vagyis amíg talán választ nem kap. Mivel a jelenség szemléltetésére Gale és Shapley nehezebben

[28] A probléma legáltalánosabb formájában meg szokták engedni azt, hogy valaki ne találjon párt magának. Az ilyen esetekre azt mondják, hogy valaki "önmagával is párba kerülhet".

[29] A preferenciarendezéseknek nem kell teljesnek lennie, vagyis megengedhető, hogy valaki ne mindenkit vegyen fel a listájára, vagy másként: bárkit ki lehet zárni a potenciális jelöltek közül.

követhető példát adott, ezért itt inkább egy Biró Péter és társai által bemutatott példát ismertetek [58]. A példában szereplő négy lány és négy fiú az alábbi sorrendben szeretne magának társat választani.

	lányok rangsorai		fiúk rangsorai
Anna:	Gábor > Henrik	Endre:	Dóra > Anna > Bea > Csilla
Bea:	Gábor > Ferenc > Endre > Henrik	Ferenc:	Dóra > Csilla > Anna > Bea
Csilla:	Gábor > Henrik > Endre > Ferenc	Gábor:	Dóra > Csilla > Bea > Anna
Dóra:	Henrik > Gábor	Henrik:	Bea > Anna > Csilla

Az első körben Endre, Ferenc és Gábor felkéri Dórát, Henrik Beát. Dóra Gábornak, Bea Henriknek mondja, hogy 'talán'. Mivel Endre és Ferenc kosarat kaptak, tovább próbálkoznak a második menetben. Ekkor Endre Annát kéri fel, de most is elutasítják. Ferenc Csillát kéri fel, aki 'talán' választ kap. A harmadik körre már csak Endre maradt pár nélkül. Ekkor megkéri a következő lányt a listáján, Beát. Beának Endre jobban tetszik, mint Henrik, ezért Henriktől elveszi és átadja Endrének a 'talán' jelzést. A negyedik fordulóban Henrik felkéri a preferencialistáján szereplő következő lányt, Annát és ezzel véget is ér az algoritmus, hiszen az utolsó fiú is párra lelt. Mivel minden fiú megállapodott, a lányok elfogadják a jelenlegi partnerüket, a párok "véglegesítődnek". Ezt a folyamatot bemutatom egy táblázat segítségével is, hogy látszódjanak az egyes fordulók eseményei. Az egyes fordulókban az adott lányhoz bejelentkező fiú vannak feltüntetve, és aláhúzással jelölöm azt, ha az adott fordulóban egy lány talonba tett vagy ott tartott valakit (a harmadik körben előfordult, hogy a talonban kicseréltek valakit).

	1. kör	2. kör	3. kör	4. kör
Anna		Endre		Henrik
Bea	Henrik	Henrik	Endre, Henrik	Endre
Csilla		Ferenc	Ferenc	Ferenc
Dóra	Endre, Ferenc, Gábor	Gábor	Gábor	Gábor

Az algoritmus mindenkit összerendelt azok közül, akiket el tudtak képzelni párjuknak a maguk számára. Természetesen nem adta ki mindenkinek a legjobb megoldást, de mindenkinek "vállalható" pár jutott, olyan, aki rajta volt az elképzelhető partnerek listáján. Az algoritmus a fiúk számára taktikázásbiztos (strategy-proof), de a lányok számára nem az. Alvin Roth egy 1984-es cikkében bebizonyította, hogy nincs olyan stabil párosítási mechanizmus, amelyben minden szereplő számára az igazmondás a domináns stratégia [449]. Ez azonban nem akkora baj, ugyanis ha a szereplők nem ismerik a többi szereplő preferenciáit, akkor nem tudják előre megmondani, hogyan kéne módosítani a sorrendjükön, hogy számukra kedvező módon befolyásolhassák a végeredményt. Ez azt jelenti, hogy az elméleti lehetőség nem igazán váltható át gyakorlati haszonná: sokszereplős piacon gyakorlatilag nem lehet megjósolni, mi lehet a kiment, és emiatt nem igazán lehet manipulálni a dolgok lefolyását.

A párválasztási probléma algoritmusa valóságidegennek tűnhet ahhoz képest, ahogy a házasság, a párválasztás ténylegesen megvalósulás a valós társadalmi gyakorlatban. Vannak azonban olyan társadalmi jelenségek, amelyek a párválasztási problémához nagyon hasonlítanak, és a megoldásuk az itt bemutatott mintát követi. Shapleyék a "programadó cikkükben" is hivatkoztak a kollégiumi felvételik során a kollégiumok és az oda

jelentkezők összekapcsolására vagy csak egyszerűen a szobatársak összerendelésének kérdésére, de ez nem nagyon tér nagyon a házassítási problémától. Talán a legfontosabb (és a legtöbbet hivatkozott) példa a párosítási feladatok közül az amerikai rezidensek központosított kórházválasztási procedúrája. A probléma lényegét tekintve itt is ugyanaz , mint a párválasztás esetében, hiszen a lehető legoptimálisabb módon kell összerendelni a kórházakat és a pályakezdő orvosgyakornokokat. Ennek a példának a gyakorlati jelentősége azonban már egy pillanatig sem kérdőjelezhető meg, és az a tény, hogy ezt a feladatot már sok évtizede a Gale-Shapley algoritmussal oldják meg folyamatosan, a megoldás erejét és jelentőségét egyértelműen bizonyítja. Kiderült, hogy ezt a módszert alkalmazták már az ötvenes évek elejétől fogva (tehát még a Gale-Shapley-algoritmus "felfedezése", leírása előtt, és a kórházak számára optimális módon). Az algoritmus módosítására csak évtizedekkel később volt szükség, akkor is csak azért, mert a hatvanas években már egyre több fiatal orvos házaspár keresett magának közös munkahelyet, vagyis nekik az a legfontosabb, hogy ugyanabban a városban kapjanak állást. Ezt feltételt az eredeti algoritmus nem tudta figyelembe venni, ezért kellett kicsit módosítani rajta [57]. Ezt a módszert természetesen nemcsak orvosok esetében lehetett alkalmazni. Az egyetemi jelentkezési rendszerekben pont ugyanez a feladat, csak ilyenkor a diákok és az iskolák, egyetemek közti párosításokat kell egyensúlyba hozni, és bizony ennek a problémának nem lehet vitatni a gyakorlati jelentőségét.[30] Ugyancsak vitathatatlan gyakorlati haszonnal jár annak azoknak a vesecsereprogramok a vezérlése, amelyet pont a párosításelmélet fenti sikerei nyomán indítottak el a világ egyre több pontján [57]. Ebben az esetben a vesére váró betegek, illetve a vesedonorok közti összerendeléseket kellett és kell minél optimálisabb módon megteremteni.

Van tehát egy algoritmus, amivel a fent bemutatott, illetve az azokhoz hasonló, további párosítási poblémákat sikeresen meg lehet oldani, a módszer bizonyított a munkaerőpiacon (rezidensek elosztása), az oktatásban (felvételik lebonyolítása), az egészségügyben (vesedonorok megszervezése). A procedúra értékelésekor fontos látnunk, hogy a módszer alkalmazása – a látszattal szemben – egyáltalán nem a központi tervezés megvalósulását jelenti. A problémakör Nobel-díjas kutatója, Alvin Roth szerint másról van itt szó.

> „Vegyük észre, hogy az itt vizsgált központosított piacok nem jelentenek központi tervezést, ahogyan a legtöbben ezt értelmezik, mivel ezek a piacok úgy lettek megalkotva, hogy érzékenyek legyenek a szereplők kinyilvánított preferenciáira, ahelyett, hogy egy tervező ezektől független céljainak elérését szolgálnák. Ami itt központosítva van, az nem az elérendő cél, hanem a piaci mechanizmus maga." idézi: [57]

Mielőtt foglalkoznék még a késleltetett elfogadási algoritmus néhány fontos következményével, bemutatom, hogy lehet a tételt formális módon leírni. Amennyiben $\mathbb{M}$-vel jelöljük a férfiak halmazát, $\mathbb{W}$-vel a nőkét, $>_a$-vel valamely a ágens preferenciarendezését és μ-vel ($\mu : \mathbb{M} \cup \mathbb{W} \mapsto \mathbb{M} \cup \mathbb{W}$) a párosítási függvényt, akkor néhány tulajdonság definiálása után formálisan is megadhatjuk a Gale-Shapley tételt: a házassági problémának mindig létezik stabil megoldása.

(AL.55) $\exists m \exists w (\mu(m) = w \leftrightarrow \mu(w) = m) \wedge \forall w \forall m ((\mu(w) \in \mathbb{M} \vee \mu(w) = w) \wedge (\mu(m) \in \mathbb{W} \vee \mu(m) = m))$

párosítás(i függvény) (HUN) – matching (function) (ENG)

(AL.56) $\text{BLOCKED_INDIVIDUALLY}(\mu) \stackrel{\text{def}}{=} \exists a \in \mathbb{M} \cup \mathbb{W} (a >_a \mu(a))$

[30] A magyar felvételei rendszer és a párosításelmélet kapcsolatáról bővebben lásd: [57], [284], [285].

egyénileg blokkolt párosítás (HUN) – matching blocked individually (ENG) – matching blocked by an individual (ENG)

(AL.57) $\text{BLOCKED_COLLECTIVELY}(\mu) \stackrel{\text{def}}{=} \exists w \in \mathbb{W} \exists m \in \mathbb{M}(w >_m \mu(m) \wedge m >_w \mu(w))$

párosan blokkolt párosítás (HUN) – matching blocked collectively (ENG) – matching blocked by a pair of agents (ENG)

(AL.58) $\text{STABLE}(\mu) \stackrel{\text{def}}{=} \neg\text{BLOCKED_INDIVIDUALLY}(\mu) \wedge \neg\text{BLOCKED_COLLECTIVELY}(\mu)$

stabil párosítás (HUN) – stable matching (ENG)

(AL.59) $\exists\mu(\text{STABLE}(\mu))$

Gale-Shapley tétel (HUN) – Gale-Shapley theorem (ENG)

A tételre Gale és Shapley konstruktív bizonyítást adtak, amikor megmutatták, hogy az algoritmust, aminek alkalmazásával mindig el lehet jutni a feladat stabil megoldásához [180]. Annak ellenére, hogy a házasítási problémának mindig van stabil megoldása, azért az is fontos tény, hogy ez az algoritmus nem mindig és nem mindenkinek biztosítja a legjobb párosítást. A Gale-Shapley algoritmus ismeretében Lester E. Dubins és David A. Freedman egy 1981-es cikkben feltette a kérdést, hogy abban az esetben, ha a játékosok között vannak olyanok, akik ismerik a többiek preferenciáit, akkor van-e esélyük arra, hogy a saját preferenciakinyilvánításaik megváltoztatásával befolyásolják az algoritmus menetét, és ezzel a végerdményt – nyilván úgy, hogy számukra kedvezőbb lehessen. Lehet-e manipulálni, érdemes-e hazudni, vagyis lehet-e itt machiavellizmusról beszélni – tette fel a kérdést Dubins és Freedman [135]. A válasz az volt, hogy az egyetemek esetében elképzelhető, hogy némelyiküknek előnyösebb lehet hazudni, azonban a hallgatók esetében ez a lehetőség már nem áll fent.[31] Minden hallgatóra (és azok minden koalíciójára) igaz, hogy az algoritmus biztosítja a legjobb stabil megoldást számukra, és ezen nem tudnak semmilyen taktikázással javítani. Dubinsék cikkére válaszul David Gale és Marilda Sotomayor példát és bizonyítást hozott arra, hogy több stabil illeszkedés esetén a másik irányban mindig lehetséges hazugsággal előnyt szerezni legalább egy szereplő számára [181].[32]

Sok más tétel és alkalmazási lehetőség lenne még bemutatható a késleltetett elfogadási algoritmussal kapcsolatban, de ezek elemzésével itt nem foglalkozom tovább. Az elmélet további tételeiről, az elméleti eredmények konkrét gyakorlati alkalmazásairól értő összefoglalást nyújt Alvin Roth 2008-as áttekintő tanulmánya [448].

7.2.14 Körzethatár-manipuláció

Az egyéni vélemények közösségi véleményekké transzformálásának következő befolyásolási lehetősége a **körzetesítés** vagyis a körzethatárok megváltoztatása (gerrymandering). Itt azokról a körzethatárokról van szó, amelyek meghatározzák az aggregálási számítások alapjául szolgáló területi egységeket. Még pontosabban: a földrajzi térben szétszórtan elhelyezkedő szavazópolgárokat egy-egy szavazataggregációs közösségbe egybefogó

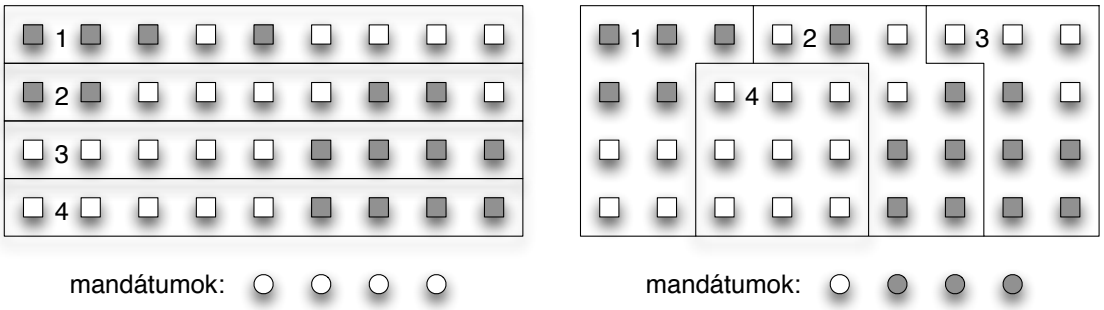
[31] Dubins és Freedman az egyetemek és a diákok összerendelésének problémáját vizsgálta.

[32] Mivel Gale és Sotomayor a házasítási problémát használták fel az érvelésük során, ezért szellemes módon utaltak Dubins és Freedman 'Machiavelli and the Gale-Shapley algorithm' című cikkére, amikor saját tanulmányuknak a 'Ms. Machiavelli and the stable matching problem' címet adták.

határokról van szó, amelyeket természetesen – mindig és szükségszerűen önkényes – politikai döntések határoznak meg. Ez a körzetesítés így mindig rendelkezik valamilyen mértékű manipulációs potenciállal.

Mielőtt bemutatnék egy egyszerű példát a körzethatárok átszabásában rejlő manipulációs lehetőségre, fel kell hívnom a figyelmet arra, hogy két fontos körzetfogalmat el kell különítenünk egymástól, nem szabad őket összekevernünk egymással. A szavazatokat szavazókörökben gyűjtik össze. Ezeket a körzeteket kicsire kell szabni, mert igazodni kell ahhoz az elváráshoz, hogy a szavazópolgárok kis fáradsággal fel tudják keresni. A szavazókörökhöz így viszont csak kevés embert lehet hozzárendelni, amiből következően a mandátumokat sem lehet egy-egy szavazókörben kiosztani. Ezért van szükség a választókerület fogalmára, ami a mandátumok kiosztásának területi egysége. Természetesen szükség van arra is, hogy kapcsolatot teremtsünk a kétféle körzetfogalom között, vagyis összerendeljük a szavazatok összegyűjtésének színtereit, a szavazóköröket és a mandátumok kiosztásának egységeit, a választókerületeket egymással. Akármilyen elveket is rögzítünk azonban, amikhez igazodni kell a szavazókörök és választókerületek párosításakor, mindig marad hely az eseti (önkényes) döntésnek, és ez adja a körzethatár-átszabások elvi lehetőségét. Ezek alapján nézzük meg, hogyan is működhet ez.

Képzeljük el, hogy van két párt, harminchat szavazókörzet. Feltételezzük fel még, hogy minden körzetben ugyanannyi ember szavaz, és az egyszerűség kedvéért tegyük fel azt is, hogy minden szavazókörzetben az egyik vagy másik pártra adják le az összes szavazatot (ez ugyan életszerűtlen feltételezés, de a lényeget ez nem érinti). Tegyük fel még fel azt is, hogy egy-egy mandátumot lehet kiosztani négy választókerületben. A kérdés már csak az, hogy a harminchat szavazókörből hogyan csináljunk négy választókerületet, ahol egy-egy mandátumot lehet nyerni. A következő ábrán a harminchat szavazókör kétféle választókerületi besorolását mutatom be.



A szürke és fehér négyzetek jelölik az egyik vagy másik pártra egységesen szavazó szavazóköröket (nevezzük a két pártot fehér és szürke pártnak). A választókerületi felosztást jelzik a szavazókörök köré rajzolt körzethatárok. Ha egyszerű többségi szavazással döntünk, akkor nagyon eltérő végeredményt ad a két felosztás. A baloldali besorolás alapján a négy választókerület mindegyikében a fehér párt nyeri el a mandátumot, míg a jobboldali felosztásban a szürke párt három, a fehér párt egy mandátumot kap.

Bár a gerrymandering valóban létező, "művelhető" manipulációs lehetőség és technika, ahhoz képest, hogy mennyire sokat hivatkoznak rá, nincs akkora potenciál benne. Sok országból, sok történelmi helyzetből lehet példákat hozni arra, hogy alkalmazták,

sőt, sokszor sikerrel alkalmazták ezt a technikát, de ettől még tudni kell a helyén kezelni a dolgokat. Mivel az 1990-ben megállapított választókerületi határok sok esetben nem feleltek meg a törvényi előírásoknak (az időközben lezajlott népmozgalmi folyamatok eredményeként), 2008-ban készítettünk egy olyan alkalmazást, ami a 2006-os parlamenti választások adataira támaszkodva, bizonyos paramétereket szabadon változtatva átszabta a választókerületek határait.[33] Bár az alkalmazás használatakor különböző paramétereket beadva egészen eltérő választókerületi határokat és ezzel eltérő szimulált végeredményeket lehetett kihozni, az is kiderült az ilyen próbálkozásokból, hogy nagyon radikálisan eltérő végeredményeket nem lehetett elérni, akárhogy is változtak a körzethatárok.[34]

7.2.15 Többkörzetes paradoxon

Sorolhatnánk a körzetmanipuláció, a gerrymandering fogalma alá azt a körzethatárokkal kapcsolatos jelenséget, amit Peter C. Fishburn és Steven J. Brams szerzőpáros **többkörzetes paradoxonnak** nevezett el, és amire egy 1983-as cikkében mutatott be egy szemléltes példát [168]. A helyzet azonban pont fordítva van. A többkörzetes paradoxon létezése biztosítja a körzethatár-manipuláció elvi lehetőségét, vagyis az előbbi jelenség logikailag megelőzi az utóbbit. A többkörzetes paradoxon ugyanis azt a helyzetet írja le, amikor a szavazatokat – a szavazók térhez való kötöttsége miatt, tehát pusztán gyakorlati okokból – körzetekbe rendezve gyűjtik össze, és a szavazási adatok összegzése során paradoxon adódik. Ha nem lennének körzetek, akkor nem adódna semmi gond, de ha vannak körzetek, akkor paradoxon keletkezhet. És ha már van paradoxon, akkor az ilyen helyzeteket már lehet manipulálni, vagyis akkor már helye van a gerrymanderingnek is. A többkörzetes paradoxon thát az alap, de még nem tudjuk, miről is szól pontosan ez a paradoxon. A no-show paradoxon kapcsán korábban már bemutatott kisváros (↰444.o.) szavazási anomáliái közé tartozik ez az új eset is. Bramsék cikkének főszereplője, Mrs. Smith csinált egy újabb kimutatást a kisvárosi választás adatai alapján. A szavazást két szavazókerületre bontva, egy keleti és egy nyugati körzeten belül végezték el a városka lakói. Az adatok így a két körzetre bontva is rendelkezésre álltak Mrs. Smith számára. Az alábbi táblázat mutatja a különböző preferenciasorrendekre leadott szavazatokat a két körzet szerinti bontásban.

összes szavazat	sorrendek	keleti körzet	nyugati körzet
417	BHW	160	257
82	BWH	0	82
143	HBW	143	0
357	HWB	0	357
285	WBH	0	285
324	WHB	285	39
1608		588	1020

Az adatok elemzéséből meglepő következtetést lehet levonnni. Korábban már bemutat-

[33] Az EOVK-határoló alkalmazásról beszámoló cikk: Pártos térképészek manipulálhatják a választásokat, at: http://index.hu/belfold/eovk4876/

[34] Nem kifejezetten a gerrymandering szempontjából, hanem inkább a szavazatok területi megoszlásainak vizsgálata miatt kollegáimmal korábban több tanulmányban, könyvben is foglalkoztunk a választások térbeliségével [346], [345], [348], [344], [347], [350].

tam, hogy ezen a választáson a három jelölt, Mrs. Bitt, Mr. Huff és dr. Wogg közül Mr. Huff lett a nyertes (☞444.o.). Ha azonban a kerületekre bontható adatok alapján kiszámolnánk, hogy ugyanazon eljárást alkalmazva ki nyerte volna meg a választást a két körzeten belül külön-külön, meglepő eredményhez jutnánk. A két körzetben az alábbi szavazatösszesítést kapnánk.

jelöltek	keleti körzet	nyugati körzet
Bitt	160 (160+0)	339 (257+82)
Huff	143 (143+0)	357 (0+357)
Wogg	285 (0+285)	324 (285+39)

Az első körben kiesne a legkevesebb szavazatot kapott jelölt, azaz a keleti körzetben Mr. Huff, a nyugatiban dr. Wogg. A második körös számítás során tehát Mrs. Bittnek a keleti körzetben dr. Wogg, a nyugatiban Mr. Huff lenne az ellenfele, és az alábbi adatok alapján számolnák ki a körzetnyerteseket.

jelöltek	keleti körzet	nyugati körzet
Bitt	303 (160+143)	624 (257+82 +285)
Huff		396 (357+39)
Wogg	285 (0+285)	

A két körzetben külön-külön számítás esetén tehát mindkét esetben Mrs. Bitt nyerte volna meg a versenyt, miközben tudjuk, hogy az összesített eredmények alapján más lett a győztes. A paradoxon nyilvánvaló.[35]

7.2.16 Többesetes paradoxon

Bizonyos értelemben a választási paradoxonok vég nélkül szaporíthatóak, ezért fontos kérdés, hogy ezek között vannak-e olyanok, amelyek visszavezethetők valamely más paradoxonra. A többkérdéses vagy **többesetes választás paradoxona** (multiple-elections paradox) jelenségét egy 1998-ban megjelent cikkben Steven J. Brams, D. Marc Kilgour és William S. Zwicker nevezte meg először [73], és bár a cikk végén megmutatták azt is, hogy a jelenség értékelhető a szavazási paradoxon egyik változataként, mégis önálló névvel, újfajta minőséggel hivatkoztak erre a problematikára. A paradoxon olyan választási helyzetekben jelentkezhet, amelyekben egyfelől egyszerre több témában is szavazni kell, másfelől mindegyik témában *igen* vagy *nem* lehet a választási lehetőség. Bramsék az egyes témákat, szavazási kérdéseket propozícióknak, míg az összes szavazási kérdés együttesét kombinációnak nevezték el. Az ilyen szavazások során a szavazókat annyiféle szavazói profilba sorolhatjuk be, ahány szavazatkombináció lehetséges elméletileg. Ha valóban csak *igen*/*nem* értékek közül lehet választani, akkor k szavazási kérdés esetén a szavazatkombinációk száma összesen k^2 lehet. Az egyes szavazók által leadott tényleges szavazatkombinációkat mindig besorolhatjuk valamelyik szavazatkombináció-profilba, és összeszámolhatjuk, mennyi szavazat érkezett összesen ezekre. A szavazatkombinációkat az adott sorszámú kérdésre adható igen vagy nem szavazatok szeint I-vel vagy N-nel jelölhetjük, tehát egy szavazatkombináció mindig annyi I vagy N betűből áll,

[35] Donald G. Saari szerint ez a jelenség a Simpson-paradoxon speciális esetének tekinthető [456].

ahány kérdést tesznek fel a választás során. Hogy miként lesz ebből paradoxon, jobban érthetővé válik, ha megnézünk egy konkrét példát. Tegyük fel, hogy három kérdésről kell szavazni, csak *igen*/*nem* válasz lehetséges, 13 szavazónk van, és az alábbi szavazatok érkeztek az egyes szavazatkombinációkra.

profilok	III	IIN	INI	INN	NII	NIN	NNI	NNN
szavazatok	1	1	1	3	1	3	3	0

Az adatokat úgy értelmezhetjük, hogy négy szavazatkombinációra 1-1, háromra 3-3, egyre pedig 0 szavazat érkezet. A legtöbb támogatás tehát azok a "kimenetek" kapták, amelyekben két *nem* és egy *igen* szavazat érkezett. Ezekre a leggyakoribb szavazatkombinációkra mondhatjuk azt is, hogy ezek a leggyakoribb választótípusok is egyben, de ez az eloszlás nem a szavazásra feltett kérdésekre vonatkozik, hanem a kérdésekre adott válaszok adott kombinációjára. Ha az egyes szavazási kérdésekre akarunk végeredmény kihozni, akkor ha egyszerű többségi szavazással döntünk, akkor mindhárom kérdésben (az elsőben, másodikban és harmadikban) 7:6 arányban a nem szavazatok kerülnének többségbe, vagyis a közösségi döntés a három 'nem' (NNN) lenne. És pont ez a mozzanat, amit paradox jellegűnek tarthatunk. Az NNN szavazatkombinációra ugyanis senki sem szavazott, vagyis senki sem érezheti úgy, hogy az ő kedvére való végeredmény született a szavazás során. A közösségi döntés áll a legmesszebb az egyéni szavazói profiloktól, vagy másként: a győztes (szavazat)kombináció kapta a legkevesebb szavazatot. Kétféle aggregálási mód létezik tehát, amelyek eltérő végeredményeket hozhatnak ki olykor. Amikor az összes kérdésre leadott szavazatprofilok gyakoriságát hasonlítjuk össze egymással, akkor kombinációaggregálásról (combination aggregation) beszélhetünk. Amikor az egyes kérdésekre (propozíciókra) adott – igenlő – szavazatokat egyenként összegezzük, tehát először kérdésenként döntjük el, hogy a támogató vagy ellenző vélemények lettek-e a győztesek, majd ez alapján képezhetjük a győztes véleményprofilt az összes kérdésre együttesen, akkor Bramsék szerint propozícióaggregálásról (proposition aggregation) van szó. A szembeállítás tartalma nyilvánvaló, a javasolt kifejezéspárt azonban nem tartom szerencsésnek, mert nem szemantikailag egynemű fogalmakat használ, amikor a kétféle aggregálási lehetőséget a propozíció és kombináció eltérő minőségeivel akarja megragadni. Ezen a ponton jobbnak tartom az **esetenkénti aggregálás** (vagy: egyenkénti aggregálás) és a **kombinált aggregálás** terminusait alkalmazni.

A Brams-Kilgour-Zwicker szerzőtrió még sokféle formában bemutatta a paradoxont, elemezték kettő- vagy négy-kérdéses változatban, a tartózkodási lehetőség megengedésével vagy olyan paraméterek mentén, amikor a győztes kombinációra adott legkisebb szavazatszámban döntetlen adódik, vagyis más kombinációk is ugyanannyi szavazatot kapnak, mint a végső győztes kimenet. Bemutattak egy olyan konstellációt, amikor négy kérdésre adott 31 szavazat az alábbi megoszlást mutatta.

profilok	IIII	IIIN	IINI	INII	NIII	IINN	ININ	INNI
szavazatok	0	4	4	4	4	1	1	1
profilok	NIIN	NINI	NNII	INNN	NINN	NNIN	NNNI	NNNN
szavazatok	1	1	1	1	1	1	1	5

Ez az eset azért különösen érdekes paradoxon, mert az egyenkénti aggregálás során 16:15 arányban minden kérdésben az *igen* szavazatok győznek, vagyis a IIII profil a győztes,

miközben egyetlen szavazó sincs, aki a négy kérdésben ilyen mintázat szerint szavazott volna. Ezzel szemben a kombinált aggregálás végeredményeként az NNNN kombináció jönne ki győztesként, hiszen a legtöbben (öten) erre a kombinációra szavaztak. A kétféle aggregálási technika tehát a két szélsőséges kombinációt, az IIII- és az NNNN-profilt hozná ki győztesnek, vagyis az eljárás felcserélése teljesen megfordítaná a szavazás végeredményét. Ezért nevezték el Bramsék ezt az esetet teljes megfordítás paradoxonának (complete-reversal paradox).

A paradoxon lehetőségét szemléltető elméleti példák bemutatása után a szerzőtrió bemutatott pár konkrét példát az amerikai politikatörténetből, amikor a többkérdéses választás paradoxona valósult meg. 1990-ben California államban 28 kérdésből álló népszavazást tartottak. A Los Angeles városában és a környékén leadott szavazatokat elemezve kiderült, hogy az egyenkénti aggregálás szerinti nyerő kombinációra egyetlen szavazat sem érkezett. Ez persze volt nem akkora meglepetés, hiszen 28 kérdés esetén az összes lehetséges szavazatkombináció száma 2^28, azaz majdnem 270 millió, ami a leadott durván két milliós szavazathoz képest nagyságrendekkel nagyobb, vagyis a lehetséges kombinációk döntő részére nem érkezett egyetlen szavazat sem. Ennek ellenére azért voltak olyan kombinációk, amelyek az egyenletes eloszláshoz képest jóval nagyobb szavazatszámot kaptak. A legtöbb szavazatot (1,75%-ot) a minden kérdésre tartózkodást adó profil volt,[36] a második helyen végzett a *teljes-nem* kombináció (1,72%), ötödik lett a Los Angeles Times újság által sugallt szavazási javaslat (0,29%), kilencedik pedig a *teljes-igen* mintázat (0,20%).

A történelmi példák között Brams és társai olyan jelenségre is hivatkoztak, ami ugyan nem a népszavazás témakörébe tartozik, mert nem *igen*/*nem*-mel megválaszolható kérdésekre kell válaszolni, mégis "produkálja" a többesetes választási paradoxont. Az elnökválasztás évében háromfajta politikai posztra is szavaznak az emberek az USA-ban, az elnöki pozícióra, valamint a szenátusi és a képviselőházi tagságra. Ha a szavazási lehetőségként csak a demokrata (D), illetve republikánus párti (R) jelöltekre adott szavazást tekintjük (ez Amerikában teljesen helytálló feltételezés), valamint a szenátusi és képviselőházi tagságot egy-egy pozíciónak tekintjük (ez is elfogadható leegyszerűsítés, mert politikai szempontból sokszor az az igazán lényeges, hogy melyik pártnak van többsége az adott testületben), akkor három leadható szavazat (elnök, képviselőház, szentáus) és kétféle szavazási lehetőség (R vagy D) létezik. Ilyen körülmények között a többkérdéses paradoxon ugyanúgy előállhat, csak ekkor – Bramsék szerint – ekkor nem propozícióaggregálásról, hanem pozícióaggregálásról (office aggregation) beszélhetünk. Ekkor a három pozícióról külön-külön szavaznak, és esetenkénti aggregálással döntenek a poszt betöltéséről, ám a politikai életet leginkább a három különálló kérdésben kialakult eredmény kombinációja határozza meg. Ennek a kérdésnek, vagyis az egységes vagy osztott kormányzás (unified or divided government) helyzetnek az értékelése az amerikai politikai elemzések fontos témája régóta. 1976-ban egységes, 1980-ban osztott kormányzás jött létre, és a szavazatoknak az egyes kombinációkra vonatkozó eloszlását a Bramsék által feldolgozott formában (a szavazatokat 315-316 körzetre vetítve) mutatja a következő táblázat.

Az adatokból könnyen kiolvasható, hogy 1976-ban nagy volt a nyomás az egységes kormányzásra, hiszen a DDD és az RRR kombináció együtt megkapta a szavazatok majdnem kétharmadát. Ekkor a teljes demokrata párti győzelem született, tehát nem alakult ki para-

[36] A szavazás során három válasz volt lehetséges, a támogatáson és ellenzésen túl lehetett tartózkodni is.

kombináció	DDD	DDR	DRD	DRR	RDD	RDR	RRD	RRR	
1976	*40,8*	5,7	2,6	1,9	11,7	8,2	8,5	20,6	100% (n=316)
1980	22,2	2,5	2,2	0,6	16,8	12,7	*14,3*	28,6	100% (n=315)

doxon. 1980-ra viszont nagyot változott a szavazók véleménye. Ekkorra már a szavazók fele elfogadta az osztott kormányzás lehetőségét, sőt, az esetenkénti aggregálás eredményeként ténylegesen is létrejött az osztott kormányzás. Viszont a kialakult végeredmény, az RRD állapot csak a negyedik legnépszerűbb kombináció lett, tehát az egyenkénti és a kombinált aggregálás által kihozott eredmények elég messze kerültek egymástól. Mivel nem a legrosszabb kombinált eredmény lett az egyéni győztes, ezért a szerzők szemiparadoxonnak nevezték el ezt a konkrét eredményt. A valóságos történelmi példa tehát nem az elmélet által előírt, “tiszta"" módon valósult meg, de a dolgok paradox jellegét azért még így is felismerhettük a gyakorlatban. Bramsék hoztak még egy harmadik példát a törvénykezési gyakorlatból, de ezt már nem mutatom itt be.

A többszörös választási paradoxonhoz hasonlít az Ascombe-paradoxon, de mivel a szakirodalomban önálló néven hivatkoznak rá, én is külön fejezetben mutatom be.

7.2.17 Anscombe-paradoxon

A G.E.M. Anscombe nevével jelzett paradoxon a többségi elv egy furcsa deficitjére mutat rá [17]. A jelenség szemléltetésére Anscombe adott egy példát, amelyben tizenegy szavazó (A,B, . . . , K) tizenegy kérdésről szavazott. Az eredményt mutató táblázatban az *igen* szavazatokat 1, a *nem*-eket 0 jelöli. Mivel a többségi szavazással hozott csoportdöntés minden esetben *igen* (1) volt (6:5 arányban), ezért azt mondhatjuk, hogy az egyéni *igen* (1) szavazatok sikeres voltak, mivel a csoportdöntéssel megegyeztek, az egyéni *nem*-ek pedig sikertelenek.

	A	B	C	D	E	F	G	H	I	J	K	csoportdöntés
1. kérdés	0	1	1	1	1	0	0	1	1	0	0	1
2. kérdés	0	1	0	1	1	0	0	1	0	1	1	1
3. kérdés	0	1	0	0	1	0	0	1	1	1	1	1
4. kérdés	0	1	0	0	0	1	1	0	1	1	1	1
5. kérdés	0	0	0	0	1	1	1	0	1	1	1	1
6. kérdés	0	0	0	0	0	1	1	1	1	1	1	1
7. kérdés	0	0	0	0	0	1	1	1	1	1	1	1
8. kérdés	1	0	0	0	0	0	1	1	1	1	1	1
9. kérdés	1	0	1	0	0	0	0	1	1	1	1	1
10. kérdés	1	0	1	1	0	0	0	1	1	0	1	1
11. kérdés	1	0	1	1	0	0	0	1	0	1	1	1
sikeres szavazat	4	4	4	4	4	4	5	9	9	9	10	

A táblázat legalsó sorából kiolvasható, hogy heten voltak olyanok (A-G), akiknek a szavazata az esetek többségében sikertelen volt. Ez azt jelenti, hogy a többségi szavazás végeredményeként kisebbségbe kerülhetnek azok, akik a többségi vélemények többségével egyetértenének, vagy másként: többen lehetnek olyanok, akiknek a szavazata az esetek többségében sikertelen lesz. Ez az Anscombe-paradoxon lényege.

Jonathan Lamb Gorman válaszcikkében egy jóval egyszerűbb példát javasolt [194], ami áttekinthetőbbé tette az esetet leíró táblázatot is, ezért inkább azt használom a továbbiakban (még pontosabban: Gorman példáját a Carl Wagner által átrendezett formában veszem át [568], mivel az alakilag jobban illeszkedik az itt használt megoldásokhoz). Gorman példájában három alternatíváról (A, B, C) kell szavazni $(igen/nem)$ öt szavazónak, többségi eljárással. A szavazás eredményét mutató táblázatban a ':-)' szimbólum azt jelzi, hogy az adott szavazó adott szavazata megegyezik-e az adott kérdésben hozott többségi döntéssel, vagyis sikeres volt-e a szavazat.

kérdések:	*A*	*B*	*C*
1. szavazó véleménye	igen :-)	igen	nem
2. szavazó véleménye	nem	nem :-)	nem
3. szavazó véleménye	nem	igen	igen :-)
4. szavazó véleménye	igen :-)	nem :-)	igen :-)
5. szavazó véleménye	igen :-)	nem :-)	igen :-)
többségi vélemény	igen	nem	igen

Ebben a táblázatban is a többséggel való egyetértést mutató jel eloszlása zavarbaejtő. Az első, második és harmadik szavazó csak egy-egy esetben került be a többségi véleményt osztók csoportjába, és ők, hárman, ebben a "kisebbségi minőségükben" többségben vannak a maradék kettő "igazi többségi" szavazóval szemben, akik több sikeres szavazatot adtak le, mint sikertelent.

> „When the majority get what they want, they do so at the expense of the minority's frustration, but at least nobody is privileged ... a method which derives the upshot from an equal weight attached to each person's desire must achieve the satisfaction of the desires of the greatest possible number and so be both the best and fairest." [17, 163-164]

Carl Wagner megoldása: 3/4 [568], [569]

Még durvább a helyzet, ha az előző táblázatban bemutatott szavazói vélemények létezését feltételezve és képviseleti demokráciát "szimulálva" olyan jelöltekre kell szavazni, akik a három kérdésben előre kinyilvánítják a preferenciáikat, és ez alapján szavaznak rájuk a szavazók. Tegyük fel, hogy két jelölt van, az alábbi "ígéretekkel".

kérdések:	*A* kérdés	*B* kérdés	*C* kérdés
1. jelölt meghirdetett profilja	igen	nem	igen
2. jelölt meghirdetett profilja	nem	igen	nem

A táblázatból az olvasható ki, hogy az első jelölt a csoporttagok véleményei alapján megállapítható többségi eredmények képviseletét ígéri, míg a második jelölt mindenben pont az ellenkezőjét. Azt várhatnánk, hogy ilyen helyzetben a többségi véleményt képviselő jelöltre érkezik több szavazat, de nem ez a helyzet. A csoportból ugyanis az első három tag két kérdésben inkább egyetért a második jelölt nézeteivel, ezért rá szavaz, és ez már többséget biztosít annak a képviselőjelöltnek, aki minden kérdésben a többségi véleménnyel ellentétes álláspontot hirdet.

Lehetne talán azt gondolni, hogy az osztott kormányzás paradoxona mint a többkérdéses választás paradoxonának egyik altípusa (↬466.o.) és az Anscombe-paradoxon megegyezik egymással, de ez nem így van. Hannu Nurmi mutatott be egy olyan esetet, ami példázza az előbbit, de nem igaz az utóbbira [385].

	1. kérdés	2. kérdés	3. kérdés
1. szavazó véleménye	Y :-)	Y	Y :-)
2. szavazó véleménye	X	X :-)	Y :-)
3. szavazó véleménye	X	Y	Y :-)
4. szavazó véleménye	Y :-)	X :-)	X
5. szavazó véleménye	Y :-)	X :-)	X
többségi vélemény	Y	X	Y

Látható, hogy a többesetes választás paradoxona fennáll, hiszen nyertes kombináció (YXY) egyszer sem fordul elő az egyéni vélemények között, viszont az Anscombe-paradoxon nem teljesül, mivel majdnem mindenki (négyen az öt közül) két kérdésben (tehát többségében) a többségi véleményt vallja.

7.2.18 Osztrogorszkij paradoxon

Noha Mojszej Jakovlevics Osztrogorszkij[37] már az 1900-as évek elején írt arról a paradoxonról, amit később róla neveztek el, a társadalmi választások elméletével foglalkozó kutatók csak a hetvenes években figyeltek fel erre a problémára [385]. A paradoxon lényege a következő. Legyen három kérdéskör, és négy választói csoport (A, B, C, D), amelyek a szavazók 20-20-20-40 százalékát reprezentálják valamint két párt (X és Y). Tegyük fel, hogy a három kérdésben a választói csoportok az alábbi táblázat szerint szavaznak a két párt (programjai) közül valamelyikre.

csoport	1. kérdés	2. kérdés	3. kérdés	támogatott párt
A (20%)	X	X	Y	X
B (20%)	X	Y	X	X
C (20%)	Y	X	X	X
D (40%)	Y	Y	Y	Y
győztes párt (1)	-	-	-	X
győztes párt (2)	Y	Y	Y	Y

Az ellentmondás akkor és abban érhető tetten, amikor dönteni kell arról, hogy melyik párt a győztes az összes kérdés figyelembe vétele mellett. Ha az egyes csoportokra nézzük meg, melyik pártot választanák, akkor az első három csoport esetében az X, a D csoport esetében az Y párt nyerne, és ezek alapján 60:40 arányban az előbbi lenne a végső győztes. Ha a három kérdésben egyenként hoznánk ki részgyőzteseket, akkor mindegyik kérdésben az Y párt jönne ki győztesként, s így természetesen a végső győztes is az lenne. Th. Bezembinder és P. Van Acker a kétféle aggregálási módszer megnevezésére javasolta bevezetni az IP-, illetve PI-egyesítés (IP/PI-amalgamation) kifejezéseket ahhoz igazodóan, hogy előbb a kérdések (Issue), majd a személyek (Person) szerint aggregálunk-e vagy fordítva [50]. Az nem kérdés, hogy a kétféle aggregálási módszert érdemes elkülöníteni és egyértelműen megnevezni, de a Bezembinderék által javasolt rövidítések egyfelől már "foglaltak" a tranzitivitással, indifferenciával kapcsolatos elméletek világában, másfelől a diszkurzív paradoxon jelenségével való párhuzamosság miatt inkább az ott beveze-

[37] Az angolul többféleképpen is hivatkoznak Osztrogorszkij nevére, például így: Moisey Yakovlevich Ostrogorsky/Ostrogorski

tett megkülönböztetést használom itt is (↬474.o.). Az egyes kérdések szerinti összegzés összegzését IBI-módszernek nevezem (az 'issue-by-issue' kifejezésre utalva), míg azt a módszert, amikor előbb az emberek, csoportok, esetek szerint összegeznek, majd az így kialakult részeredményeket vetik össze a második körben, azt nevezem CBC-metódusnak (a 'case-by-case' kifejezés kezőbetűire hivatkozva ezzel).

Hannu Nurmi mutatott egy olyan esetet, ami mind az Osztrogorszkij-paradoxonra, mind a többkérdéses választás paradoxonára (↬466.o.) például szolgálhat [385]. Az osztott kormányzás korábban bemutatott problémáját szemléltetik az alábbi adatok. Két párt, a demokraták (D) és a republikánusok (R) jelöltjeire szavaz 13 szavazó három pozíció betöltése kapcsán (képviselőházi és szenátusi tagság, valamit a kormányzói poszt elnyerése a tét).

szavazók	képviselőház	szenátus	kormányzóság
1-3 szavazó	D	D	D
4 szavazó	D	D	R
5 szavazó	D	R	D
6 szavazó	D	R	R
7-9 szavazó	R	D	R
10-12 szavazó	R	R	D
13 szavazó	R	R	R

Az adatokból kitűnik, hogy a három poszt közül egyet a republikánus párt szerez meg (képviselőház), illetve kettőt a demokrata párt, vagyis RDD kombináció születik. Erre egyénileg senki sem szavazott, tehát fennáll a többesetes választás paradoxona. Viszont meg lehet állapítani az Osztrogorszkij-paradoxont is, hiszen a kétféle aggregálási módszerrel egyszer a demokrata, másszor a republikánus párt jön ki "abszolút" győztesnek. Ennek ellenére a két paradoxon nem azonos egymással, amit önmagában szemléltet az a tény, hogy az Osztrogorszkij-paradoxonra hozott első esetre nem igaz az osztott kormányzás paradoxona, hiszen az egyenkénti összegzéssel kihozott nyertesek (YYY) kombinációja egyben a legtöbb egyéni szavazatot kapta (40%).

Douglas W. Rae és Hans Daudt szerint minden Osztrogorszkij-paradoxon háttérében van egy Condorcet-paradoxon [425], amit közel sem tartottak véletlennek. Szerintük a két jelenségnek közös gyökere van,[38] amire vonatkozóan megfogalmazták azt a sejtésüket, hogy ez talán annak köszönhető, hogy mindegyik jelenség bináris bemeneteket kezel, tehát páros összehasonlításokat követel meg, ezzel szemben a kimeneti oldalon többargumentumú relációkat produkál.

Már az Anscombe-paradoxonra is igaz volt, az Osztrogorszkij-paradoxonra még inkább elmondható, hogy ezek a problémák már nem csak, vagy talán nem is elsősorban a preferenciaaggregálásról szólnak, hanem valami másról. Itt már nem a világállapotok elrendezéseiről, értékeléseiről van szó, hanem inkább a vélemények összesítéséről, amit már az aggregálási problémák egy másik típusaként lehet elemezni, és amit én is külön fejezetben mutatok be.

[38] Pontosabban három jelenség közös gyöketérről beszéltek, mert a vizsgálatukba bevették még a kompozíció problémáját is, de ez utóbbi beemelése a saját gondolatmenetembe fölöslegesnek tűnt.

7.3 Véleményaggregálás

A közösségi döntések alapkérdése évtizedeken kereszül az volt, hogy az egyéni preferenciák aggregálásának milyen jellemzői, esetleg milyen problémái vannak. Sokáig evidencia volt, hogy a közösségi döntéseket az egyéni preferencirendezések alapján kell meghozni. A vélemény egyet jelentett a preferenciarendezések kifejezésével, a véleményaggregáláson pedig a preferenciaaggregálást kellett érteni. Az 1990-es évek azzal hozták el fordulatot e téren, hogy a véleményaggregálás fogalma alá besorolva megjelentek azok a cikkek, amelyek a kijelentések aggregálásának problémáival kezdtek el foglalkozni. Innentől kezdve a **véleményaggregálás** (judgement aggregation) már két dolgot jelenthetett: a hagyományos értelemben vett preferenciaaggregálást és az újonnan "érkezett" kijelentésaggregálást.[39] A következő részben ez utóbbi problémakör legfontosabb kérdéseit tekintem át röviden.

7.3.1 Doktrinális paradoxon

Miközben Wijnand J. Schoenmakers egy 1986-os kétoldalas cikkében az adatbázisok egyesítésének veszélyeire akarta felhívni a figyelmet, a probléma forrását egy olyan képzeletbeli példával szemléltette, ami teljes mértékben alkalmas volt arra, hogy a véleményaggregálás problámájára is rá lehessen mutatni vele [464]. Schoenmakers történetében egy bíró két tanút kérdezett ki különböző időpontokban. Az első tanú azt állította, hogy az A állítás igaz, a második pedig azt, hogy $A \to B$. A két kijelentés alapján a bíró levonta azt a következtetést, hogy B is igaz (és erre hivatkozva halálra ítélte a vádlottat). Később kiderült, hogy mindkét tanú hamisnak tartotta a B állítást, bár más indokok alapján. Az első tanú nem gondolta sem B-t, sem $(A \to B)$-t igaznak, de mivel a bíró ezt nem kérdezte tőle, nem szólt semmit erről. A második tanú pedig úgy vélte, hogy az $A \to B$ állítás igaz ugyan, de tudta azt is, hogy a B nem. Ebből levonhatta magának a következtetést, hogy akkor az A nem lehet igaz, de ezt magától értetődőnek tartotta, ezért ezt nem is mondta a bírónak. A két tanútól átvett állítások alapján a bíró helyes logikai következtetést vont le, ám mégis rossz ítéletet hozott.

Példájával Schoenmakers azt akarta szemléltetni, hogy a különböző forrásokból származó adatok egyesítése könnyen ellentmondásokhoz vezethet. A mérnökök, informatikusok számára azért volt fontos mindez, mert az ilyen paradoxon lehetősége komoly fenyegetést jelenthet az adatbázisok egyesítése során, mivel az egyébként önmagukban konzisztens adatbázisok az egyesítés után (eredményeként) inkonzisztenssé válhatnak. A történet tanulsága azonban messze túlmutat az adatbázisok egyesítésének problémáján. Ugyanez a probléma rejlik például a doktrinális paradoxon mélyében.

A **doktrinális paradoxon** (doctrinal paradox) vagy más néven bírói paradoxon problémáját Kornhauser és Sager fedezte fel [286]. Kornhauserék felismerése az volt, hogy ha egy ügyben nem egyetlen, hanem több bíró dönt valamiért (ahogy Amerikában a fellebbviteli bírósági tárgyalások esetében ez így van), és többségi eljárással összegzik a bírók véleményeit, akkor könnyen paradoxonhoz juthatunk. Nem mindegy ugyanis, hogy a döntéshozatal melyik fázisában összegzik az egyéni véleményeket. Az egyik megoldás szerint előbb közösen tisztázni kell a tényállást. Ehhez a büntethetőséget implikáló egyes

[39] Sajnos újracsak pongyola terminológia alakult ki, és a kijelentések aggregálásának problémáira sokan a véleményaggregálás kérdéseként hivatkoznak.

feltételekről kell közös véleményt kialakítani, és minden tényállási kérdésben többségi szavazást kell alkalmazni. Az egyes feltételekre így megállapított közös vélemény alapján lehet aztán a büntetésre vonatkozó konklúziót levonni. Létezik azonban egy másik módszer is, ami szerint mindenkinek magának kell döntenie a tényállási feltételeről, és a közösségi döntésre csak – az egyéni döntések alapján – a bűnösség kimondásakor, az ítélethozatal során van szükség.

Egy egyszerű példa segítségével szemléltethető, hogy miként juthatunk így ellentétes eredményekhez [286]. Tegyük fel, hogy egy olyan ügyben kell dönteni, amelyben két szempont (A és B) alapján kell megállapítani, hogy bűncselekmény történt-e, amit szankcionálni kell (S) vagy sem ($\neg S$). A két feltétel és a szanció legyen a következő:

A: x és y között érvényes a szerződés
B: x megszegte a szerződést
S: x kártérítésre kötelezett

A döntés logikai szerkezete egyértelmű. Ha A és B feltételek fennállnak, akkor meg kell állapítani a bűncselekmény bekövetkeztét és ki kell szabni a S büntetést. Nézzük meg, mi történik akkor, ha három bírónak kell közösségi döntést hoznia ebben az ügyben.

döntési alternatívák:	A	B	$A \wedge B \rightarrow S$
1. bíró véleménye	igen	nem	nem
2. bíró véleménye	nem	igen	nem
3. bíró véleménye	igen	igen	igen
közösségi vélemény (1)	igen	igen	–
közösségi vélemény (2)	–	–	nem

A paradoxon léte szembeszökő. Ha a premisszák esetében alkalmazzuk a többségi szavazást (1), akkor mindkét feltételre igenlő választ kapunk, ami alapján kötelezően ki kell mondani a bűnösséget és elmarasztaló ítéletet kell hozni (S). Ha az ítélethozatal során támaszkodunk a közösségi döntésre (2), akkor nemleges válaszhoz jutunk a bűnösséget illetően, ami felmentő ítélet eredményez ($\neg S$). Kornhauser és Sager az első esetet **kérdésről kérdésre szavazásnak** (issue-by-issue voting) nevezi. Ekkor minden egyes eldöntendő kérdésről szavazást tartanak, majd az így kialakított közös álláspont alapján vonják le (akkor már a logikai következtetés segítségével) a végleges döntést. A másodikat esetet (ami a konklúzió alapú döntési helyzetnek felel meg) **ügyről ügyre szavazás** (case-by-case voting) címkével illetik, mivel ekkor minden bíró az ügy egészéről mond ítéletet, és aaztán ezeket összegzik többségi szavazással [286]. Azért nevezték el doktrinális paradoxonnak, mert az $A \wedge B \rightarrow S$ formulával reprezentálható jogi doktrinát (legal doctrine) tartották a bírói döntéshozatal alapzatának.

Franz Dietrich is bemutat egy paradoxont, de az ő példájában más a döntésbe vont, egymáshoz kapcsolódó kijelentések viszonya [130]. Képzeljük el azt a helyzetet, amikor egy kormányülésen kell az alábbi ügyben dönteni (amikor feltétezhetjük, hogy a vélemények megoszlását tekintve a kormány egésze három, egyenlő létszámú csoportra osztható). A döntés logikai szerkezete itt más lesz, mint az előző példában, ha úgy viszik be a döntéshozó testület elé az ügyet, hogy az A és a $A \rightarrow B$ kérdéseket jelölik meg premisszaként és a B kérdést konklúzióként. Tegyük fel, hogy az alábbi szavazatokat kapjuk:

A: megengedhető a költségvetési hiány
B: az oktatásra többet kell költeni
$A \to B$: *ha* megengedhető a költségvetési hiány, *akkor* az oktatásra többet kell költeni

döntési alternatívák:	A	$A \to B$	B
1. csoport véleménye	igen	igen	igen
2. csoport véleménye	igen	nem	nem
3. csoport véleménye	nem	igen	nem
közösségi vélemény (1)	igen	igen	–
közösségi vélemény (2)	–	–	nem

A táblázat alapján látszik, hogy ugyanúgy paradoxonhoz jutottunk, mint az előző példában: a kétféle aggregálási technika alkalmazásával eltérő végeredményhez juthatunk. Mindhárom véleménycsoport konzisztens a következtetéseiben, a véleményösszegzés különböző útjai mégis ellentmondáshoz vezetnek. Ez a példa egyébként azt a jelenséget is szemlélteti, hogy a szavazási napirend kialakításával befolyásolhatjuk a végeredményt.

A doktrinális paradoxon lényege az, hogy ha a döntéshozók (politikusok, bírók, szakértők) egy csoportjának logikailag összefüggő állításokról kell közös döntést hoznia, akkor előfordulhat, hogy ha az ügybe vont kijelentésekről (premisszákból) egyenként képeznek közös (többségi elvű) döntéseket, majd ezek alapján vonják le a végkövetkeztetésüket, illetve a premissza-kijelentések alapján levont egyéni konklúziókból közös (többségi elvű) véleményt alkotnak, akkor a kétféle módon képzett közös döntések ellentmondásba kerülhetnek egymással. E rövid jellemzés során azonban már olyan terminusokat használtam, amelyeket a későbbi – a jelenség általánosítását elvégző és a diszkurzív paradoxon fogalmát középpontba helyező – elméletek termeltek ki. Nézzük meg, hogyan épült tovább a "paradoxonok elmélete".

7.3.2 Diszkurzív paradoxon, List-Pettit lehetetlenségi tétele

2002-ben Christian List és Philip Pettit a doktrinális paradoxon "belső logikáját" elemezve, és a problémát általánosítva bevezette a diszkurzív paradoxon fogalmát, majd megfogalmazott egy lehetetlenségi tételt, amire azóta List-Pettit tételként hivatkoznak a szakirodalomban [309]. A vizsgált szakterület is saját nevet "kapott": egyre inkább úgy hivatkoztak erre a vizsgálati irányra, mint a **véleményaggregálás** (judgement aggregation) elemzésével foglalkozó tudományterületre.

Egy 2002-es cikkben List és Pettit Kornhauserék doktrinális paradoxonát értékelve rögtön bemutattak egy másik példát, amely a doktrinális paradoxont "kitermelő" élethelyzethez képest teljesen eltérő világból, a gazdaság területéről származott. Ezzel a kiterjesztéssel azt akarták szemléltetni, hogy az élet minden területén felbukkanhat a doktrinális paradoxonhoz hasonló jelenség. A példában szereplő kisvállalat dolgozóinak, akik egyben tulajdonosai is a cégnek, közösen kell dönteniük arról, lemondjanak-e fizetésük egy részéről annak érdekében, hogy egy biztonsági berendezést felszereljenek a cég telephelyén. Három szempontot mérlegelhet mindenki egyéni döntése során: *(i)* mennyire komoly az elhárítandó veszély, *(ii)* mennyire ígérkezik hatékonynak a berendezés, illetve *(iii)* személy szerint képes-e viselni a lemondás pénzügyi terheit. Ezek alapján – feltételezve, hogy hárman vesznek részt a döntésben – az alábbi táblázat szerint összegezhetjük

az egyéni és közösségi döntéseket (itt is elég az *igen* vagy *nem* döntéseket feltüntetni).

szempontok:	komoly veszély	hatékony berendezés	viselhető anyagilag	lemond-e
1. személy véleménye	igen	nem	igen	nem
2. személy véleménye	nem	igen	igen	nem
3. személy véleménye	igen	igen	nem	nem
közösség véleménye (1)	igen	igen	igen	–
közösség véleménye (2)	–	–	–	nem

A paradoxon egyértelműen kiolvasható a táblázatból. Ha a tagok egyénileg döntenek, és ez alapján aggregálnak közösségi döntést, akkor mindegyik döntéshozó számára létezhet olyan szempont, amely szerint nem támogatja a tehervállalást, és emiatt a három – a beruházást ellenző – egyéni döntésből származtatott közösségi döntés is csak elutasító tartalmú lehet. Miközben minden egyes szempont alapján többségi vélemény volt arról köztük, hogy az adott szempont szerint támogatni kellene a beruházást. A kétféle döntési technikát meg is nevezték, és elkülönítették a **premissza alapú megközelítést** (premise-based approach), illetve **konklúzió alapú megközelítést** (conclusion-based approach) egymástól. A beszédes nevek arra utalnak, hogy az összetett döntési tér egyes elemeiről döntenek-e egyenként, majd részdöntések alapján hoznak az egészre vonatkozó csoportos döntést, vagy minden egyes részkérdésről meghozzák az arra vonatkozó csoportos döntést, és ezeket aggregálják az egészre vonatkozó döntéssé második lépésben. Ez a kétféle döntési lehetőség, illetve a különböző döntési eljárásoknak a végeredményekre gyakorolt hatása ismert volt a doktrinális paradoxon felfedezése idején is, csak akkor még nem nevezték el így ezeket a technikákat.

A véleményaggregálási folyamattal, techikával szemben két szinten lehet követelményeket megfogalmazni. Egyfelől rögzíteni lehet racionalitási feltételeket az egyéni szavazókkal szemben, másfelől elvárásokat lehet megfogalmazni a közösségi aggregálással szemben. Az egyéni racionalitásra vonatkozóan a többféle racionalitáskritériumot is meg kell fogalmaznunk.[40]

(AL.60) $\textsc{complete}(\Phi_i) \stackrel{\text{def}}{=} \forall i \forall k (\phi_k \in \Phi_i \vee \neg\phi_k \in \Phi_i)$

teljes (HUN) – complete (ENG)

(AL.61) $\textsc{consistent}(\Phi_i) \stackrel{\text{def}}{=} \forall i \forall k \neg(\phi_k \in \Phi_i \wedge \neg\phi_k \in \Phi_i)$

konzisztens (HUN) – consistent (ENG)

(AL.62) $\textsc{deductively_closed}(\Phi_i) \stackrel{\text{def}}{=} \forall i \forall k (\Phi_i \vdash \phi_k \rightarrow \phi_k \in \Phi_i)$

deduktíve zárt (HUN) – deductive closed (ENG)

A teljességi feltétel szerint az i. szavazó Φ_i véleményprofilja **teljes**, ha tartalmaz minden propozciót vagy annak negáltját (akár mindkettőt). Ez a feltétel azt biztosítja, hogy mindenki állást foglaljon valahogy az adott véleménytér minden propozíciókettősével kapcsolatban, de ez az elvárás önmagában még nem eredményez racionális véleményprofilt,

[40]Ebben a kontextusban az általánosabb tárgyalásmód érzékeltetése végett az egyéni és csoportos vélemények jelzésére nem a preferenciákra utaló **R**, hanem az általánosabb konnotációval bíró Φ szimbólumokat fogom alkalmazni.

hiszen ez a kritérium "megengedi", hogy valaki egyszerre elfogadon egy állítást és annak tagadását. Ezt "korrigálja" a konzisztenciakritérium, amely azt mondja ki, hogy az i. szavazó Φ_i véleményprofilja akkor **konzisztens**, ha egy propozíció és annak negáltja közül csak az egyiket tartalmazza.[41] Ez annyit jelent, hogy csak az ellentmondó állítások lehetőségét kizáró profilokat tartjuk racionális véleménynek. A deduktív zártság feltétele pedig azt "garantálja", hogy ha valamely ϕ_k propozíció logikailag levezethető a véleményprofilba tartozó propozíciókból, akkor a propozíció is beletartozik a véleményprofilba, azaz a szavazók egyénileg mind képesek levonni a logikailag szükséges következtetéseket.[42] A három fenti tulajdonság alapján már megfogalmazhatjuk mind az egyéni, mind a csoport véleményprofilokra vonatkozó racionalitáskritériumot.

(AL.63) $\text{RATIONAL}(\Phi_i) \stackrel{\text{def}}{=} \text{COMPLETE}(\Phi_i) \wedge \text{CONSISTENT}(\Phi_i) \wedge \text{DEDUCTIVELY_CLOSED}(\Phi_i)$

egyénileg racionális (HUN) – individually rational (ENG)

Akkor beszélhetünk arról, hogy a Φ_i egyéni véleménykészlet (personal set of judgement) racionális, ha arra igaz, hogy teljes, konzisztens és deduktíve zárt. Az aggregált (csoportos, közösségi, kollektív) véleményprofilra vonatkozóan értelmét tekintve ugyanezt a racionalitáskritériumot követelhetjük meg, bár utána ki kell még egészítenünk azt valami másssal is.

(AL.64) $\text{RATIONAL}(\Phi) \stackrel{\text{def}}{=} \text{COMPLETE}(\Phi) \wedge \text{CONSISTENT}(\Phi) \wedge \text{DEDUCTIVELY_CLOSED}(\Phi)$

csoportosan racionális (HUN) – groups rational (ENG)

A Φ csoportos véleménykészlet (collective set of judgement) racionalitásához szükséges feltételként kell megfogalmaznunk, hogy az legyen teljes, konzisztens és deduktíve zárt. Elégséges feltételként meg kell még fogalmaznunk azt az elvárást is, hogy az aggregáló függvénynek is racionálisnak kell lennie. Ehhez az univerzalitás (↬432.o.) és az anonimitás tulajdonságainak (↬441.o.) megkövetelésén túl egy további feltételt ki kell kötnünk, a szisztematikusság, szisztematicitás kritériumát.[43]

(AL.65) $\text{SYSTEMATIC}(\mathbf{F}(\Phi)) \stackrel{\text{def}}{=} \forall \Phi, \Phi' \in \mathbb{D}_F : \forall \phi, \psi \in X : (\forall i : \phi \in \Phi_i \leftrightarrow \psi \in \Phi_i') \rightarrow (\phi \in F(\Phi) \leftrightarrow \psi \in F(\Phi_i'))$

szisztematikus (HUN) – systematic (ENG)

Ez a kritérium az összegző függvényre vonatkozik. Az $\mathbf{F}(\Phi)$ függvényt akkor mondjuk szisztematikusnak, ha van két véleményprofilunk (Φ és Φ'), illetve két állításunk (ϕ és ψ), amelyekre igaz, hogy minden (i.) csoporttag esetén teljesül, hogy az egyik állítás akkor és csak akkor sorolható az egyik egyéni véleményprofilba ($\phi \in \Phi_i$), ha a másik vélemény a másik egyéni véleményprofilba ($\psi \in \Phi_i'$) tartozik, akkor teljesülnie kell annak is, hogy a szóbanforgó két állítás beletartozik a két csoportos véleményprofilba is ($\phi \in \Phi$, illetve $\psi \in \Phi'$). Csáji Balázs Csanád és Rédei Miklós a szisztemetikusság követelményét két tulajdonságra bontja fel.

[41] A szerzők megjegyzik, hogy bármely más logikai kapcsoló is szerepelhetne a konjunkció helyén.

[42] A formulában szereplő $\vdash$ szimbólum a szintaktikai következmény jele.

[43] List és Pettit az 'univerzalitás' terminusa helyett az 'univerzális értelmezési tartomány' (universal domain) kifejezést használja, de ugyanabban az értelemben, ahogy azt korábban már bemutattam.

„A szisztematicitás két tulajdonságot foglal magában:
1. Az, hogy a napirendnek egy bizonyos ϕ állítása igenelve van-e az összegezett véleményben, csak attól függ, hogy hogyan szavaztak a szavazók erre a ϕ napirendi pontra (és így nem fu?gg attól, hogy a napirenden szereplő más állításokkal kapcsolatban ki hogyan szavazott).
2. Az a függvény, amely a ϕ-re leadott N darab szavazat függvényében megadja, hogy ϕ el van-e fogadva az $F(\Phi)$ összesített véleményben, *ugyanaz* a függvény *minden* ϕ napirendi pont esetén." [111, 108]

Ez a feltétel kicsit hasonlít a monotonitás kritériumára (↪432.o.), de nem azonos vele (Csályiék szerint ha az idézetben olvasható második feltételt elhagyjuk, akkor az első tulajdonság gyakorlatilag megegyezik a monotonitással). A fentebb definiált kritériumok segítségével már megfogalmazható a List-Pettit tétel, amely a fent bemutatott racionalitáskritériumok egyidejű kielégíthetetlenségét állítja [309].

(AL.66) $\exists i(\neg(\text{RATIONAL}(\Phi) \wedge \text{RATIONAL}(\Phi_i) \wedge \text{UNIVERSAL}(\mathbf{F}) \wedge \text{ANONYM}(\mathbf{F}) \wedge \text{SYSTEMATIC}(\mathbf{F})))$

List-Pettit tétel (HUN) – List-Pettit theorem (ENG)

A List-Pettit tétel szerint nem létezik olyan $\mathbf{F}$ véleményaggregáló függvény, amely univerzális, anonim és szisztematikus, valamint minden racionális egyéni véleményprofilt racionális csoportos véleményprofillá összegez. Ez egy újabb lehetetlenségi tétel, csak most a racionalitásra vonatkozó intuitív elvárásaink a véleményaggregálás területére vonatkoztatva nem teljesül(het)nek.[44] A tétel nagyon erőset állít, és úgy tűnhet, mintha a társadalmi gyakorlat nem is igazolná vissza az igazát. List és Pettit válasza erre az, hogy a ténylegesen létező, ebben az értelemben működőképes véleményösszegzési eljárások során mindig megsértünk valamilyen racionalitási elvet a fent definiáltak közül. Ha így van, akkor a következő fontos kérdés az lehet, hogy milyen módokon lehet enyhíteni a fenti elvárásokon, hogy még "vállalható" racionalitású és a megvalósíthatósági, lehetségességi elvárásoknak is megfelelő technikákat, eljárásokat kapjunk.[45] Mark Pauly és Martin van Hess a List-Pettit tétel feltételeit az anonimitás elengedésével enyhítette, és következő fontos tételt tudta bebizonyítani [407].

(AL.67) $\forall i(\text{RATIONAL}(\Phi) \wedge \text{RATIONAL}(\Phi_i) \wedge \text{UNIVERSAL}(\mathbf{F}) \wedge \text{SYSTEMATIC}(\mathbf{F})) \rightarrow \text{DIC}(\mathbf{F})$

Pauly-van Hees tétel (HUN) – Pauly-van Hees theorem (ENG)

A Pauly-van Hees tétel azt mondja ki, hogy egy véleményösszegző függvény, amely kielégíti az univerzális értelmezési tartomány és a szisztematicitás feltételeit, akkor és csak akkor összegez racionális véleményprofilokat racionális csoportos véleményprofilba, ha a függvény diktatórikus.

[44] List és Pettit egy tanulmányukban megvizsgálták azt, hogy milyen viszony van a List-Petit tétel és az Arrow-tétel között [310]. Az az állításuk nem meglepő annyira, hogy az Arrow-tételt elméletileg fel lehet fogni a List-Pettit-tétel speciális eseteként, de e kijelentésükkel együtt rögtön felhívták a figyelmet arra is, hogy a két tétel közti megfeleltetés, egy általánosabb keretrendszerbe való ágyazásuk, és az egymáshoz való viszonyuk matematikai kifejezése sok megoldandó kérdést vet fel, ami nem tűnik túl egyszerűnek. Az esetleges végleges válaszokhoz további elemzésekre van szükség.

[45] Csáji Balázs Csanád és Rédei Miklós másfajta eseteket is bemutat, és további problémákat elemez a diszkurzív paradoxon általánosabb témakörén belül [111]. Több minden más mellett vizsgálják azt a fontos kérdést is, hogy milyen módon lehet gyengíteni a List-Pettit tétel feltételein, mi történik akkor, ha elengedünk egyet-egyet a tétel érvényességéhez megkövetelt kritériumok közül.

Bibliography

[1] *2013. évi V. törvény a Polgári Törvénykönyvről*. 2013, Novissima Kiadó.

[2] C. Adam – A. Herzig – D. Longin: A logical formalization of the occ theory of emotions. *Synthese, Knowledge, Rationality & Action*, 168. évf. (2009. May) 2. sz., 201–248. p.

[3] Alfred Adler: *Emberismeret. Gyakorlati individuálpszichológia*. 1933, Magánkiadás.

[4] Aki eladta a Rottenbiller utcát – Legendás szélhámosok 5., 2012. július. URL `http://player.hu/biznisz/aki+eladta+a+rottenbiller+utcat+legendas+szelhamosok+5/#.UZEcjJWgSlA`.

[5] Altrichter Ferenc: A tudat két aspektusa: intencionalitás és qualia. In Vizi E. Szilveszter – Altrichter Ferenc – Nyíri Kristóf – Pléh Csaba (szerk.): *Agy és tudat*. 2002, BIP Kiadó, 135–143. p.

[6] Ambrus-Lakatos Loránd – Meszerics Tamás: Az ultimátumjáték elemzéséhez. *Közgazdasági Szemle*, L. évf. (2003. június), 505–518. p.

[7] Alan Ross Anderson: The logic of norms. *Logique et Analyse*, 2. évf. (1958), 84–91. p.

[8] Alan Ross Anderson: A reduction of deontic logic to alethic modal logic. *Mind, New Series*, 67. évf. (1958) 265. sz., 100–103. p.

[9] Alan Ross Anderson: The formal analysis of normative systems. In N. Rescher (szerk.): *The Logic of Decision and Action*. 1967, University of Pittsburgh Press, 147–213. p.

[10] Alan Ross Anderson: Some nasty problems in the formal logic of ethics. *Noûs*, 1. évf. (1967) 4. sz., 345–360. p.

[11] Alan Ross Anderson – Omar Khayyam Moore: The formal analysis of normative concepts. *American Sociological Review*, 22. évf. (1957) 1. sz., 9–17. p.

[12] Alan Ross Anderson – Omar Khayyam Moore: A normatív fogalmak formális elemzése. In Irving M. Copi – James A. Gould (szerk.): *Kortárs-tanulmányok a logikaelmélet kérdéseiről*. 1985, Gondolat Kiadó, 561–583. p.

[13] Hajnal Andréka – István Németi – Ildikó Sain: Algebraic logic. Jelentés, 1998, Mathematical Institute, Hungarian Academy of Sciences. URL `http://www.math+inst.hu/pub/algebraic+logic/handbook.pdf`.

[14] Hajnal Andréka – M. D. Ryan – P.-Y. Schobbens: Operators and laws for combining preference relations. *Journal of Logic and Computation*, 12. évf. (2002) 1. sz., 13–53. p.

[15] James Andreoni: Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal*, 100. évf. (1990) 401. sz., 464–477. p.

[16] G.E.M. Anscombe: *Intention*. 1963, Harvard University Press.

[17] G.E.M. Anscombe: On frustration of the majority by fulfilment of the majority's will. *Analysis*, 36. évf. (1976) 4. sz., 161–168. p.

[18] Stephen Ansolabehere – James M. Snyder Jr.: Valence politics and equilibrium in spatial election models. *Public Choice*, 103. évf. (2000) 3/4. sz., 327–336. p.

[19] Lennart Åqvist: Good samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs*, 1. évf. (1967) 4. sz., 361–379. p.

[20] Lennart Åqvist: Some results on dyadic deontic logic and the logic of preference. *Synthese*, 66. évf. (1986) 1. sz., 95–110. p.

[21] Dan Ariely: *Kiszámíthatóan irracionális. A racionálisnak vélt döntéseinket alakító rejtett erőkről*. 2011, Gabo Kiadó.

[22] David M. Armstrong: The nature of mind. In John Perry – Michael Bratman (szerk.): *Introduction to Philosophy: Classical and Contemporary Readings*. 4th. kiad. 2007, Oxford University Press, 295–302. p.

[23] Kenneth J. Arrow: A difficulty in the concept of social welfare. *Journal of Political Economy*, 58. évf. (1950) 4. sz., 328–346. p.

[24] Kenneth J. Arrow: Rational choice functions and orderings. *Economica*, 26. évf. (1959), 121–127. p.

[25] Kenneth J. Arrow: *Social Choice and Individual Values*. 1963, John Wiley & Sons.

[26] Robert Audi (szerk.): *The Cambridge Dictionary of Philosophy*. 1999, Cambridge University Press.

[27] David Austen-Smith – Jeffrey Banks: Cycling of simple rules in the spatial model. *Social Choice and Welfare*, 16. évf. (1999) 4. sz., 663–672. p.

[28] John L. Austin: *A tetten ért szavak*. Budapest, 1990, Akadémiai Kiadó.

[29] Robert Axelrod: On six advances in cooperation theory. *Analyse & Kritik*, 22. évf. (2000) 1. sz., 130–151. p.

[30] Robert Axelrod: *The Evolution of Cooperation*. 2006 (1984), Basic Book.

[31] Balázs Zoltán: A szabad akarat hierarchikus elmélete. *Magyar Filozófiai Szemle*, 43. évf. (1999) 1-3. sz., 190–222. p.

[32] Miguel A. Ballester – Guillaume Haeringer: A characterization of the single-peaked domain. *Social Choice and Welfare*, 36. évf. (2011) 2. sz., 305–322. p.

[33] Salvador Barberà: Pivotal voters. a new proof of arrow's theorem. *Economic Letters*, 6. évf. (1980), 13–16. p.

[34] Salvador Barberà: Strategy-proofness and Pivotal Voters: A direct proof of the Gibbard-Satterthwaite theorem. *International Economic Review*, 24. évf. (1983), 413–417. p.

[35] Salvador Barberà – Dolors Berga – Bernardo Moreno: Domains, ranges and strategy-proofness: the case of single-dipped preferences. *Social Choice and Welfare*, 39. évf. (2012), 335–352. p.

[36] Salvador Barberà – Bernardo Moreno: Top monotonicity: a common root for single peakedness, single crossing and the median voter result. *Games and Economic Behavior*, 73. évf. (2011) 2. sz., 345–359. p.

[37] J. Bartholdi – J. Orlin: Single Transferable Vote resists strategic voting. *Social Choice and Welfare*, 8. évf. (1991), 341– 354. p.

[38] J. J. Bartholdi III – C. A. Tovey – M. A. Trick: The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6. évf. (1989) 3. sz., 227–241. p.

[39] Mehdi Dastani Bas R. Steunebrink – John-Jules Ch. Meyer: A formal model of emotion triggers: an approach for bdi agents. *Synthese*, 185. évf. (2012), 83–129. p.

[40] Gary S. Becker: Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature*, 14. évf. (1976. Sep) 3. sz., 817–326. p.

[41] Nuel Belnap – Michael Perloff: Seeing to it that: A canonical form for agentives. *Theoria*, 54. évf. (1988) 3. sz., 175–199. p.

[42] Nuel Belnap – Michael Perloff: Seeing to it that: A canonical form for agentives. In Jr. Henry E. Kyburg – Ronald P. Loui – Greg N. Carlson (szerk.): *Knowledge Representation and Defeasible Reasoning*. 1990, Kluwer Academic Publishers, 167–190. p.

[43] Nuel Belnap – Michael Perloff – Ming Xu: *Facing the Future. Agents and Choices in Our Indeterminist World*. 2001, Oxford University Press.

[44] Bélyácz Iván: Kockázat vagy bizonytalanság? elmélettörténeti töredék a régi dilemmáról. *Közgazdasági Szemle*, LVII. évf. (2010), 652–665. p.

[45] Bélyácz Iván: Kockázat, bizonytalanság, valószínűség. *Hitelintézeti Szemle*, 10. évf. (2011) 4. sz., 289–313. p.

[46] Bereczkei Tamás: *Az erény természete*. Budapest, 2009, Typotex.

[47] Isaiah Berlin: *Négy esszé a szabadságról*. 1990, Európa Könyvkiadó.

[48] Helen Bernhard – Urs Fischbacher – Ernst Fehr: Parochial altruism in humans. *Nature*, 442. évf. (2006) 7105. sz., 912–915. p.

[49] Bertalan László: Előadások Max Weberről: Max Weber cselekvéselmélete és cselekvéstipológiája. *Szociológiai Szemle*, 11. évf. (2001) 3. sz., 113–136. p. URL `http://www.szociologia.hu/dynamic/0103bertalan.htm`.

[50] Th.G.G. Bezembinder – P. Van Acker: The ostrogorski paradox and its relation to nontransitive choice. *Journal of Mathematical Sociology*, 11. évf. (1985) 2. sz., 131–158. p.

[51] Jean-Yves Béziau: New light on the square of oppositions and its nameless corner. *Logical Investigations*, 10. évf. (2003), 218–233. p.

[52] Jean-Yves Béziau: The power of the hexagon. *Logica Universalis*, 6. évf. (2012), 1–43. p.

[53] Cristina Bicchieri: *The Grammar of Society. The Nature and Dynamics of Social Norms*. 2006, Cambridge University Press.

[54] Meghyn Bienvenu – Jérôme Lang – Nic Wilson: From preference logics to preference languages, and back. In *Proceedings of the Twelth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)* (konferenciaanyag). 2010, 214–224. p.

[55] Ken Binmore: The evolution of fairness norms. *Rationality and Society*, 10. évf. (1998), 275–301. p.

[56] Ken Binmore: *Game Theory. A Very Short Introduction*. 2007, Oxford University Press.

[57] Biró Péter: Stabil párosítási modellek és ezeken alapuló központi párosító programok. *Szigma*, XXXVII. évf. (2006) 3-4. sz., 153–175. p.

[58] Biró Péter – Csóka Péter – Kóczy Á. László – Radványi Anna Ráhel – Sziklai Balázs: Közgazdasági nobel-emlékdíj2012: Alvin e. roth és lloyd s. shapley. *Magyar Tudomány*, 174. évf. (2013) 2. sz., 190–199. p.

[59] Thomas Bittner – Maureen Donnelly – Barry Smith: Individuals, universals, collections: On the foundational relations of ontology. In Achille C. Varzi – Laure Vieu (szerk.): *Formal Ontology in Information Systems. Proceedings of the Third International Conference (FOIS 2004)*. Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2004, IOS Press, 37–48. p.

[60] Duncan Black: On the rationale of group decision-making. *Journal of Political Economy*, 56. évf. (1948), 23–34. p.

[61] Patrick Blackburn – Maarten de Rijke – Yde Venema.: *Modal Logic*. 2001, Cambridge University Press.

[62] Julian H. Blau: Liberal values and independence. *The Review of Economic Studies*, 42. évf. (1975) 3. sz., 395–401. p.

[63] Guido Boella – Gabriella Pigozzi – Leendert van der Torre: Normative systems in computer science. ten guidelines for normative multiagent systems. In Guido Boella – P. Noriega – Gabriella Pigozzi – H. Verhagen (szerk.): *Normative Multi-Agent Systems* (konferenciaanyag). 2009.

[64] Walter Bossert – Kotaro Suzumura: Social norms and rationality of choice. ISSN 0709-9231. Jelentés, 2007. 7, Département de sciences économiques Université de Montréal.

[65] Walter Bossert – Kotaro Suzumura: *Consistency, Choice, and Rationality*. 2010, Harvard University Press.

[66] Pierre Bourdieu: *A társadalmi egyenlőtlenségek újratermelődése*. 1978, Gondolat.

[67] Pierre Bourdieu: *A gyakorlat elméletének vázlata. Három kabil etnológiai tanulmány*. 2009, Napvilág Kiadó.

[68] Pierre Bourdieu – Loïc J.D. Wacquant: *An Invitation to Reflexive Sociology*. 2009, Polity Press.

[69] Craig Boutilier: Toward a logic for qualitative decision theory. In Jon Doyle – Erik Sandewall – Pietro Torasso (szerk.): *Principles of knowledge representation and reasoning*. 1994, Morgan Kaufmann Publishers, 75–86. p.

[70] Robert Boyd: Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game. *Journal of Theoretical Biology*, 136. évf. (1989), 47–56. p.

[71] Robert Boyd – Jeffrey P. Lorberbaum: No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature*, 327. évf. (1987. May), 58–59. p.

[72] Ronald J. Brachman – Hector J. Levesque: *Knowledge Representation and Reasoning*. 2004, Morgan Kaufmann Publishers.

[73] S.J. Brams – D.M. Kilgour – W.S. Zwicker: The paradox of multiple elections. *Social Choice and Welfare*, 15. évf. (1998) 2. sz., 211–236. p.

[74] Steven J. Brams: *Theory of moves*. 1994, Cambridge University Press.

[75] Steven J. Brams – William V. Gehrlein – Fred S. Roberts (szerk.): *The Mathematics of Preference, Choice and Order. Essays in Honor of Peter C. Fishburn*. 2009, Springer.

[76] Michael Bratman: *Intention, Plans, and Practical Reason*. Cambridge, MA, 1987, Harvard University Press.

[77] Robert Bredereck – Jiehua Chen – Gerhard J. Woeginger: Are there any nicely structured preference profiles nearby? In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI '13)* (konferenciaanyag). 2013, AAAI Press.

[78] Robert Bredereck – Jiehua Chen – Gerhard J. Woeginger: A characterization of the single-crossing domain. *Social Choice and Welfare*, 41. évf. (2013), 989–998. p.

[79] Franz Brentano: A pszichikai és fizikai jelenségek közötti különbségrõl. In Csaba Pléh (szerk.): *Pszichológiatörténeti szöveggyûjtemény*. I. köt. 1983, Tankönyvkiadó, 157–163. p.

[80] Franz Brentano: *Az erkölcsi ismeret eredete*. 1994, Kossuth Könyvkiadó.

[81] John Broome: Reasoning with preferences? In Serena Olsaretti (szerk.): *Preferences and Well-Being*. 2006, Cambridge University Press, 183–208. p.

[82] Jerome Bruner: *Valóságos elmék, lehetséges világok*. 2005, Új Mandátum Könyvkiadó.

[83] R. Bshary – A.S. Grutter: Image scoring and cooperation in a cleaner fish mutualism. *Nature*, 441. évf. (2006) 22. sz., 975–978. p.

[84] James M. Buchanan – Gordon Tullock: *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, Michigan, 1962:, University of Michigan Press.

[85] James N. Buchanan: A törvény és a láthatatlan kéz. In *Piac, állam, alkotmányosság*. 1992, KJK, 39–51. p.

[86] Robert A. Bull – Krister Segerberg: Basic modal logic. In Dov M. Gabbay – Franz Guenthner (szerk.): *Handbook of Philosophical Logic*. 3. köt. 2nd edition. kiad. 2001, Kluwer Academic Publishers, 1–82. p.

[87] Eugenio Bulygin: On norms of competence. *Law and Philosophy*, 11. évf. (1992) 3. sz., 201–216. p.

[88] B.J. Bushman – C.A. Anderson: Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review*, 108. évf. (2001) 1. sz., 273–279. p.

[89] Cristiano Castelfranchi – Fabio Paglieri: The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. *Synthese,*, 155. évf. (2007) 2. sz., 237–263. p.

[90] Brian F. Chellas: Imperatives. *Theoria*, 37. évf. (1971) 2. sz., 114–129. p.

[91] Brian Farrell Chellas: *The logical form of imperatives*. 1968, Dept. of Philosophy, Stanford University.

[92] Brian Farrell Chellas: *Modal Logic: An Introduction*. 1980, Cambridge University Press. ISBN 0-521-22476-4.

[93] Herman Chernoff: Rational selection of decision functions. *Econometrica*, 22. évf. (1954) 4. sz., 422–443. p.

[94] Roderick M. Chisholm: Contrary-to-duty imperatives and deontic logic. *Analysis*, 24. évf. (1963) 2. sz., 33–36. p.

[95] Roderick M. Chisholm: Supererogation and offence: A conceptual scheme for ethics. *Ratio*, 5. évf. (1963), 1–14. p.

[96] Roderick M. Chisholm: The ethics of requirement. *American Philosophical Quarterly*, 1. évf. (1964) 2. sz., 147–153. p.

[97] Roderick M. Chisholm – Ernest Sosa: Intrinsic preferability and the problem of supererogation. *Synthese*, 16. évf. (1966) 3/4. sz., 321–331. p.

[98] Roderick M. Chisholm – Ernest Sosa: On the logic of "intrinsically better". *American Philosophical Quarterly*, 1966., 244–249. p.

[99] Robert B. Cialdini – R.R. Reno – C.A. Kallgren: A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58. évf. (1990) 6. sz., 1015–1026. p.

[100] Tim Clutton-Brock: Cooperation between non-kin in animal societies. *Nature*, 462. évf. (2009) 5. sz., 51–57. p.

[101] Philip R. Cohen – Hector J. Levesque: Intention is choice with commitment. *Artificial Intelligence*, 42. évf. (1990) 213-261. sz.

[102] Tom Coleman – Vanessa Teague: On the complexity of manipulating elections. In *Proceeding CATS '07 Proceedings of the thirteenth Australasian symposium on Theory of computing* (konferenciaanyag), 65. köt. 2007, Australian Computer Society, 25–33. p.

[103] Andrew M. Colman – Jonathan A. Stirk: Singleton bias and lexicographic preferences among equally valued alternatives. *Journal of Economic Behavior & Organization*, 40. évf. (1999), 337–351. p.

[104] Vincent Conitzer – Tuomas Sandholm – Jérôme Lang: When are elections with few candidates hard to manipulate. *Journal of the Association for Computing Machinery*, 54. évf. (2007) 3. sz.

[105] Clyde H. Coombs: *A Theory of Data*. 1964, Mathesis Press.

[106] David Copp: Reasons and needs. In R.G. Frey – Christopher W. Morris (szerk.): *Value, Welfare, and Morality*. Cambridge, 1993, Cambridge University Press, 112–137. p.

[107] Arthur Corbin: Legal analysis and terminology. *Yale Law Journal*, 1919., 163–173. p.

[108] Arthur Corbin: Jural relations and their classification. *Yale Law Journal*, 1921., –238. p.

[109] Gary W. Cox – Kenneth A. Shepsle: Majority cycling and agenda manipulation: Richard mckelvey's contributions and legacy. In John H. Aldrich – James E. Alt – Arthur Lupia (szerk.): *Positive Changes in Political Science*. 2007, The University of Michigan Press, 19–40. p.

[110] Charles B. Cross: The modal logic of discrepancy. *Journal of Philosophical Logic*, 26. évf. (1997) 2. sz., 143–168. p.

[111] Csáji Balázs Csanád – Rédei Miklós: A racionális demokratikus véleményösszegzés korlátairól. *Magyar Filozófiai Szemle*, 55. évf. (2011) 2. sz., 97–121. p.

[112] Cserne Péter: Gary stanley becker, a preferenciák és a közgazdaságtan imperializmusa. *Szociológiai Szemle*, 10. évf. (2000) 1. sz., 113–127. p.
URL `http://www.szociologia.hu/dynamic/0001cserne.htm`.

[113] Csontos László: Szempontok a privatizáció és a hierarchikus gazdasági szervezetek belső szerkezetének változásai közti összefüggések tanulmányozásához. *Szociológiai Szemle*, 4. évf. (1994) 4. sz., 83–95. p.
URL `http://www.mtapti.hu/mszt/19944/csontos.htm`.

[114] Csontos László: A politika tanulmányozása és a közgazdaságtan. *Közgazdasági Szemle*, XLIV. évf. (1997), 557–568. p.

[115] Csontos László – Király Júlia – László Géza: Az ezredvégi nagy borzongás. *Közgazdasági Szemle*, XLIV. évf. (1997) július-augusztus. sz., 569–596. p.

[116] Antonio Damasio: *Descartes tévedése. Érzelem, értelem és az emberi agy*. 1996, Aduprint.

[117] Justin D'Arms – Daniel Jacobson: The significance of recalcitrant emotion (or, anti-quasijudgmentalism). In Anthony Hatzimoysis (szerk.): *Philosophy and the Emotions*. 2003, Cambridge University Press, 127–145. p.

[118] Donald Davidson: The logical form of action sentences. In Nicholas Rescher (szerk.): *The Logic of Decision and Action*. 1966, University of Pittsburgh, 81–95. p.

[119] Donald Davidson: The logical form of action sentences. In *Essays on Actions and Events*. 1967, Clarendon Press, 105–121. p.

[120] Donald Davidson: Cselekvések, indokok és okok. *Thél@geme*, 2. évf. (1998) 4. sz., 7–21. p.

[121] Donald Davidson: Actions, reasons, and causes. In *Essays on Actions and Events*. 2002, Clarendon Press, 3–19. p.

[122] Donald Davidson: Psychology as philosophy. In *Essays on Actions and Events*. 2002 (1974), Clarendon Press, 229–244. p.

[123] Otto A. Davis – Morris H. DeGroot – Melvin J. Hinich: Social preference orderings and majority rule. *Econometrica*, 40. évf. (1972) 1. sz., 147–157. p.

[124] Robyn M. Dawes: Social dilemmas. *Annual Review of Psychology*, 31. évf. (1980), 169–93. p.

[125] Richard Dawkins: *Az önző gén*. 2005, Kossuth Könyvkiadó.

[126] Howard L. Dazeley – Wolfgang L. Gombocz: Interpreting Anselm as Logician. *Synthese*, 40. évf. (1979) 1. sz., 71–96. p.

[127] Dick de Jongh – Fenrong Liu: Preference, priorities and belief. In T. Grüne-Yanoff – S. O. Hansson (szerk.): *Preference Change: Approaches from Philosophy, Economics and Psychology*. 2009, Springer, 85–108. p.

[128] Ronald de Sousa: *The Rationality of Emotion*. 1987, MIT Press.

[129] Andreas Diekmann: Volunteer's dilemma. *Journal of Conflict Resolution*, 29. évf. (1985) 4. sz., 605–610. p.

[130] Franz Dietrich: Judgment aggregation: (im)possibility theorems. *Journal of Economic Theory*, 126. évf. (2006), 286–298. p.

[131] Franz Dietrich – Christian List: Where do preferences come from? *International Journal of Game Theory*, 2012.

[132] Dombi J. – Vincze N.: A lexikografikus döntési eljárás kapcsolata a hagyományos döntési eljárásokkal. In *XXV. Magyar Operációkutatási Konferencia Előadáskivonatok, Debrecen, Hungary* (konferenciaanyag), 32. köt. 2001.

[133] Jon Doyle – Michael Wellman: Representing preferences as ceteris paribus comparatives. In Steve Hanks – Stuart Russell – Michael Wellman (szerk.): *Working Notes of the AAAI Spring Symposium on Decision-Theoretic Planning* (konferenciaanyag). 1994, AAAI Press.

[134] Pierre Dragicevic: Játékelméleti jelölések.
URL `http://www.lri.fr/~dragice/gameicons/`.

[135] Lester E. Dubins – David A. Freedman: Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88. évf. (1981) 7. sz., 485–494. p.

[136] Charles Duhigg: *A szokás hatalma*. 2012, Casparus Kiadó.

[137] Michael Dummett – Robin Farquharson: Stability in voting. *Econometrica*, 29. évf. (1961) 1. sz., 33–43. p.

[138] Émile Durkheim: Az erkölcsi tény meghatározása. In *A társadalmi tények magyarázatához*. Budapest, 1978, KJK, 161–195. p.

[139] Émile Durkheim: A szociológia módszertani szabályai. In *A társadalmi tények magyarázatához*. 1978, KJK, 21–160. p.

[140] Émile Durkheim: *Az öngyilkosság. Szociológiai tanulmány*. 1982, KJK.

[141] Émile Durkheim: *A vallási élet elemi formái*. 2002, L'Harmattan Kiadó.

[142] Paul Ekman: Basic emotions. In T. Dalgleish – M. Power (szerk.): *Handbook of Cognition and Emotion*. 3 fejezet. 1999, John Wiley and Sons, 45–60. p.

[143] Norbert Elias: *A civilizáció folyamata. Szociogenetikus és pszichogenetikus vizsgálódások (1937-1939)*. 1987, Gondolat Kiadó.

[144] Jon Elster: *Ulysses and the Sirens*. 1979, Cambridge University Press.

[145] Jon Elster: *A társadalom fogaskerei*. 1997, Osiris.

[146] Jon Elster: *Alchemies of the Mind. Rationality and the Emotions*. 1999, Cambridge University Press.

[147] Seymour Epstein: Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, 49. évf. (1994), 709–724. p.

[148] Amitai Etzioni: *The Moral Dimension. Toward a New Economics*. 1988, Free Press.

[149] E. E. Evans-Pritchard: Science and sentiment an exposition and criticism of the writings of pareto. *Revue européenne des sciences sociales, Herméneutique paretienne*, 16. évf. (1978) 43. sz., 51–73. p.

[150] Ronald Fagin – Joseph Y. Halpern – Yoram Moses – Moshe Y. Vardi (szerk.): *Reasoning about Knowledge*. 1995, MIT Press.

[151] Piotr Faliszewski – Edith Hemaspaandra – Lane A. Hemaspaandra: How hard is bribery in elections? *Journal of Artificial Intelligence Research*, 35. évf. (2009), 485–532. p.

[152] Piotr Faliszewski – Ariel D. Procaccia: AI's war on manipulation: Are we winning? *AI Magazine*, 31. évf. (2010) 4. sz., 53–64. p.

[153] Armin Falk – Ernst Fehr – Urs Fischbacher: Driving forces behind informal sanctions. *Econometrica*, 73. évf. (2005) 6. sz., 2017–2030. p.

[154] Robin Farquharson: *Theory of Voting*. 1969, Yale University Press.

[155] Ernst Fehr – Urs Fischbacher: Third-party punishment and social norms. *Evolution and Human Behavior*, 25. évf. (2004), 63–87. p.

[156] Ernst Fehr – Urs Fischbacher – Simon Gächter: Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, 13. évf. (2002), 1–25. p.

[157] Ernst Fehr – Simon Gächter: Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14. évf. (Summer 2000) 3. sz., 159–181. p.

[158] Ernst Fehr – Karla Hoff – Mayuresh Kshetramade: Spite and development. *American Economic Review*, 98. évf. (2008) 2. sz., 494–499. p.

[159] Joel Feinberg: Supererogation and rules. *Ethics*, 71. évf. (1961) 4. sz., 276–288. p.

[160] Joel Feinberg: *Társadalomfilozófia*. 1999, Osiris Kiadó.

[161] Todd E. Feinberg: *Altered Egos. How the Brain Creates the Self*. 2001, Oxford University Press.

[162] Ferenczi Miklós: *Matematikai logika*. 2002, Műszaki Könyvkiadó.

[163] FIFA. *A labdarúgás játékszabályai 2009/2010*. 2010. URL `http://jatekvezetok.blsz.hu/szabalyzatok/szk2009magyar.pdf`.

[164] Peter C. Fishburn: On the foundations of game theory: The case of non-archimedean utilities. *International Journal of Game Theory*, 1. évf. (1971) 1. sz., 65–71. p.

[165] Peter C. Fishburn: A study of lexicographic expected utility. *Management Science*, 17. évf. (1971. July) 11. sz., 672–678. p.

[166] Peter C. Fishburn: Lexicographic orders, utilities and decision rules: A survey. *Management Science*, 20. évf. (1974. July) 11. sz., 1442–1471. p.

[167] Peter C. Fishburn: Axioms for lexicographic preferences. *The Review of Economic Studies*, 42. évf. (1975. July) 3. sz., 415–419. p.

[168] Peter C. Fishburn – Steven J. Brams: Paradoxes of preferential voting. *Mathematics Magazine*, 56. évf. (1983) 4. sz., 207–214. p.

[169] Peter C. Fishburn – Irving H. LaValle: Subjective expected lexicographic utility: Axioms and assessment. *Annals of Operations Research*, 80. évf. (1998), 183–206. p.

[170] Frederic B. Fitch: A revision of hohfeld's theory of legal concepts. *Logique et Analyse*, 1967., 269–276. p.

[171] Dagfinn Follesdal: The status of rationality assumptions in interpretation and in the explanation of action. In Michael Martin – Lee C. McIntyre (szerk.): *Readings in the Philosophy of Social Science*. 1994, MIT Press, 299–310. p.

[172] Forrai Gábor: Az intencionalitás (bevezető tanulmány). In Ambrus Gergely – Demeter Tamás – Forrai Gábor – Tőzsér János (szerk.): *Elmefilozófia*. 2008, L'Harmattan, 137–158. p.

[173] James William Forrester: Gentle murder, or the adverbial samaritan. *The Journal of Philosophy*, 81. évf. (1984) 4. sz., 193–197. p.

[174] Robert Frank – Thomas Gilovich – Dennis Regan: Does studying economics inhibit cooperation? *Journal of Economic Perspectives*, 7. évf. (1993) 2. sz., 159–171. p.

[175] Harry G. Frankfurt: Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68. évf. (1971) 1. sz., 5–20. p.

[176] Marcus R. Frean: The prisoner's dilemma without synchrony. *Proc. R. Soc. B*, 257. évf. (1994), 75–79. p.

[177] James W. Friedman: A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38. évf. (1971. Jan) 1. sz., 1–12. p.

[178] Nico H. Frijda: *The Emotions*. 1986, Cambridge University Press.

[179] Drew Fudenberg – Eric Maskin: Evolution and cooperation in noisy repeated games. *The American Economic Review*, 80. évf. (1990) 2. sz., 274–279. p.

[180] David Gale – Lloyd S. Shapley: College admissions and the stability of marriage. *American Mathematical Monthly*, 69. évf. (1962) 1. sz., 9–15. p.

[181] David Gale – Marilda Sotomayor: Ms. machiavelli and the stable matching problem. *American Mathematical Monthly*, 92. évf. (1985) 4. sz., 261–268. p.

[182] Francis Galton: Vox populi. *Nature*, 75. évf. (1907), 450–451. p.

[183] A. Gardner – S.A. West: Spite. *Current Biology*, 16. évf. (2006) 17. sz., R662–R664. p.

[184] Nicholas Georgescu-Roegen: The pure theory of consumer's behavior. *Quarterly Journal of Economics*, 50. évf. (1936) 3. sz., 545–593. p.

[185] Nicholas Georgescu-Roegen: Choice, expectations, and measurability. *Quarterly Journal of Economics*, 68. évf. (1954) 2. sz., 503–534. p.

[186] Allan Gibbard: Social choice and the arrow conditions (1968-69). mimeograph, 1969. URL `http://www+personal.umich.edu/~gibbard/Gibbard1968+Social+Choice+Arrow+Conditions.pdf`.

[187] Allan Gibbard: Manipulation of voting schemes: A general result. *Econometrica*, 41. évf. (1973), 587–601. p.

[188] Herbert Gintis: Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206. évf. (2000) 2. sz., 169–179. p.

[189] Herbert Gintis: The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology*, 220. évf. (2003), 407–418. p.

[190] Herbert Gintis – Samuel Bowles – Robert Boyd – Ernst Fehr: Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24. évf. (2003), 153–172. p.

[191] P. Girard: Modal logic for lexicographic preference aggregation. In J.V Benthem – A Gupta – E. Pacuit (szerk.): *Games, Norms and Reasons*. 2011, Springer Verlag, 97–117. p.

[192] Patrick Girard: *Modal logic for belief and preference change*. Illc dissertation series ds-2008-04 (Institute for Logic, Language and Computation Universiteit van Amsterdam). 2008.

[193] Alan H. Goldman: *Reasons from Within. Desires and Values*. 2009, Oxford University Press.

[194] Jonathan Lamb Gorman: A problem in the justification of democracy. *Analysis*, 39. évf. (1978) 1. sz., 46–50. p.

[195] Alvin W. Gouldner: The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25. évf. (1960), 161–78. p.

[196] Umberto Grandi – Ulle Endriss: First-order logic formalisation of arrow's theorem. In *Logic, Rationality, and Interaction*. 2009, Springer, 133–146. p.

[197] Umberto Grandi – Ulle Endriss: First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 42. évf. (2013), 595–618. p.

[198] Donald W. Griesinger – James W. Livingston: Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, 18. évf. (1973) 3. sz., 173–188. p.

[199] Gulyás Attila: A méltányosságelmélet alapjai. modellek és nézõpontok. *Közgazdasági Szemle*, LIV. évf. (2007. február), 167–183. p.

[200] Gulyás Attila: Diktátor a barátom? a barátság és a méltányosság kapcsolata. *Közgazdasági Szemle*, LVIII. évf. (2011), 430–444. p.

[201] Werner Güth – Rolf Schmittberger – Bernd Schwarze: An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3. évf. (1982) 4. sz., 367–388. p.

[202] Jürgen Habermas: *A kommunikatív cselekvés elmélete*. Budapest, 2011, Gondolat Kiadó.

[203] Ian Hacking: *Rewriting the Soul. Multiple Personality and the Sciences of Memory*. 1995, Princeton University Press.

[204] Ian Hacking: Kinds of people: Moving targets. In *Proceedings of the British Academy* (konferenciaanyag), 151. köt. 2007, 285–318. p.

[205] Sören Halldén: *On the Logic of 'Better'*. Lund, 1957, Cwk Gleerup.

[206] W. D. Hamilton: Selfish and spiteful behaviour in an evolutionary model. *Nature*, 288. évf. (1970) 5277. sz., 1218–1220. p.

[207] William Donald Hamilton: The genetical evolution of social behavior I. *Journal of Theoretical Biology*, 7. évf. (1964), 1–16. p.

[208] William Donald Hamilton: The genetical evolution of social behavior II. *Journal of Theoretical Biology*, 7. évf. (1964), 17–52. p.

[209] Hámori Balázs: *Érzelemgazdaságtan – A közgazdasági elemzés kiterjesztése*. 2003, Kossuth Kiadó.

[210] Hámori Balázs: Kísérletek és kilátások. Daniel Kahneman. *Közgazdasági Szemle*, L. évf. (2003), 779–799. p.

[211] J. Hansen – G. Pigozzi – L. van der Torre: Ten philosophical problems in deontic logic. In G. Boella – L. van der Torre – H. Verhagen (szerk.): *Normative Multiagent Systems* (konferenciaanyag), 07122. köt. 2007.

[212] Jörg Hansen: *Imperatives and Deontic Logic. On the Semantic Foundations of Deontic Logic*. PhD értekezés (Der Fakultat für Sozialwissenschaften und Philosophie der Universitat Leipzig). 2008.

[213] Bengt Hansson: Fundamental axioms for preference relations. *Synthese*, 18. évf. (1968), 423–442. p.

[214] Bengt Hansson: An Analysis of Some Deontic Logics. *Noûs*, 3. évf. (1969) 4. sz., 373–398. p.

[215] Bengt Hansson: Deontic logic and different levels of generality. *Theoria*, 36. évf. (1970. December) 3. sz., 241–248. p.

[216] Sven Ove Hansson: A new semantical approach to the logic of preference. *Erkenntnis (1975-)*, 31. évf. (1989) 1. sz., 1–42. p.

[217] Sven Ove Hansson: The revenger's paradox. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61. évf. (1991) 3. sz., 301–305. p.

[218] Sven Ove Hansson: Changes in preference. *Theory and Decision*, 38. évf. (1995) 1. sz.

[219] Sven Ove Hansson: Preference logic. In DM. Gabbay – F. Guenthner (szerk.): *Handbook of Philosophical Logic*. 4. köt. 2001, Kluwer Academic Publishers, 319–393. p.

[220] Garret Hardin: Discriminating Altruism. *Zygon*, 17. évf. (1982) 2. sz., 163–186. p.

[221] Garrett Hardin: The tragedy of the commons. *Science*, 162. évf. (1968), 1243–1248. p.

[222] Russel Hardin: Collective action as an agreeable n-prisoners' dilemma. *Behavioral Science*, 16. évf. (1971), 472–481. p.

[223] Gilbert Harman: *Explaining Value and Other Essays in Moral Philosophy*. 2000, Clarendon Press.

[224] John C. Harsanyi: Rational choice models of political behavior vs. funcionalist and conformist theories. *World Politics*, 21. évf. (1969), 513–538. p.

[225] John C. Harsanyi: Advances in understanding rational behavior. In *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht, 1976, Reidel, 89–117. p.

[226] H.L.A. Hart: Bentham on legal rights. In A.W.B. Simpson (szerk.): *Oxford Essays in Jurisprudence, Second Series*. Oxford, 1973, Clarendon Press, 171–201. p.

[227] H.L.A. Hart: Definition and theory in jurisprudence. In *Essays in Jurisprudence and Philosophy*. 1983, Clarendon Press, 21–48. p.

[228] H.L.A. Hart: *A jog fogalma*. Budapest, 1995, Osiris.

[229] H.L.A. Hart: Felelősség. In Zsolt Krokovay (szerk.): *Felelősség*. 2006, L'Harmattan, 51–70. p.

[230] Hegyi beszéd, 2013. URL `http://szentiras.hu/SZIT/Mt5`.

[231] Héjjas István: *A keleti vallások filozófiája és világképe*. 2007, ANNO Kiad.

[232] Carl G. Hempel: Aspects of scientific explanation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, 1965, Free Press, 466–467. p.

[233] Joseph Henrich – Robert Boyd – Samuel Bowles – Colin Camerer – Ernst Fehr – Herbert Gintis (szerk.): *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. 2004, Oxford University Press.

[234] F.A. Hindriks: Constitutive rules, language, and ontology. *Erkenntnis*, 71. évf. (2009), 253–275. p.

[235] Albert O. Hirschman: *Az érdekek és a szenvedélyek. Politikai érvek a kapitalizmus mellett annak győzelme előtt*. 1998, Jószöveg Műhely Kiadó.

[236] Albert O. Hirschman: A mértékletesség ellen: három könnyű módszer a gazdaságtan néhány kategóriájának bonyolítására. In *Versengő nézetek a piaci társadalomról*. 2000, Jószöveg Műhely Kiadó, 153–171. p.

[237] Jack Hirshleifer: The expanding domain of economics. *The American Economic Review*, 75. évf. (1985) 6. sz., 53–68. p. Centennial Essays and 1985 Survey of Members.

[238] Jack Hirshleifer: On the emotions as guarantors of threats and promises. In J. Dupre (szerk.): *The Latest on the Best*. Cambridge, Mass., 1987, MIT Press, 307–326. p.

[239] Thomas Hobbes: *Leviatán*. Budapest, 1999, Kossuth Kiadó.

[240] Jerry R. Hobbs: Ontological promiscuity. In *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics* (konferenciaanyag), 23. köt. Chicago, Illinois, 1985, Association for Computational Linguistics, 61–69. p. URL `www.isi.edu/~hobbs/op+acl85.pdf`.

[241] Wesley Newcomb Hohfeld: Alapvető jogi fogalmak a bírói érvelésben. In Miklós Szabó – Csaba Varga (szerk.): *Jog és nyelv*. 2000, kiadó nélkül, 59–96. p.

[242] Robin Horton: African traditional thought and western science. In Bryan R. Wilson (szerk.): *Rationality*. 1971, Harper Torchbooks, 131–171. p.

[243] John F. Horty: *Agency and Deontoic Logic*. 2001, Oxford University Press.

[244] Harold Hotelling: Stability in competition. *The Economic Journal*, 39. évf. (1929), 41–57. p.

[245] G. E. Hughes – M. J. Cresswell: *A New Introduction to Modal Logic*. 1996, Routledge.

[246] Huoranszki Ferenc: Kinyilvánított preferenciák és irracionalitás. *Szociológiai Szemle*, 3. évf. (1993) 3-4. sz., 61–70. p.

[247] Huoranszki Ferenc: Döntéselmélet és erkölcsi normák. In Gál Róbert Iván – Szántó Zoltán (szerk.): *Cselekvéselmélet és társadalomkutatás. In memoriam Csontos László*. Budapest, 2003, Közgazdasági Szemle Alapítvány, 107–133. p.

[248] Husserl Edmund: Logikai vizsgálódások. v. vizsgálódás. In Varga Péter András – Zuh Deodáth (szerk.): *Husserl és a Logikai vizsgálódások. Ismeretfilozófia és fenomenológiai filozófia*. 2009, L'Harmattan, 21–46. p.

[249] Husserl Edmund: Logikai vizsgálódások. vi. vizsgálódás. In Varga Péter András – Zuh Deodáth (szerk.): *Husserl és a Logikai vizsgálódások. Ismeretfilozófia és fenomenológiai filozófia*. 2009, L'Harmattan, 47–92. p.

[250] Edmund Husserl: Teleológia a filozófiatörténetben. In Péter András Varga – Deodáth Zuh (szerk.): *Az új Husserl*. 2011, L'Harmattan, 171–198. p.

[251] Ken-ichi Inada: A note on the simple majority decision rule. *Econometrica*, 32. évf. (1964) 4. sz., 525–531. p.

[252] Ken-ichi Inada: The simple majority decision rule. *Econometrica*, 37. évf. (1969) 3. sz., 490–506. p.

[253] Index.hu: Szándékos öngól a koreai ligában. URL `http://sportgeza.hu/futball/2013/07/05/koreai_ongol/`.

[254] Index.hu: Különc orosz zsenié az egymillió dolláros matematikai díj, 2010. URL `http://index.hu/tudomany/2010/03/19/kulonc_orosz_zseninek_iteltek_az_egymillio_dollaros_matematikai_dijat/`.

[255] Jakab András: A norma szerkezetének vizsgálata. *JESZ*, 2. évf. (2001) 4. sz. URL `http://jesz.ajk.elte.hu/jakab8.html`.

[256] Susan James: Szenvedélyek, vágyak és a cselekvés magyarázata. *Magyar Filozófiai Szemle*, 39. évf. (1995) 5-6. sz., 743–757. p. 'Desires, Passions and the Explanation of Action' in S. Gaukroger (ed.), The Soft Underbelly of Reason (Routledge, 1998).

[257] Susan James: *Passion and Action. The Emotions in Seventeenth-Century Philosophy*. 1997, Clarendon Press.

[258] Janky Béla: *Szolidaritás és jóléti preferenciák*. Philosophiae Doctores sorozat. Budapest, 2005, Akadémiai Kiadó.

[259] Richard Jeffrey: *The Logic of Decision*. Chicago, 1983, University of Chicago.

[260] R. E. Jennings: Can there be a natural deontic logic? *Synthese*, 65. évf. (1985) 2. sz., 257–273. p.

[261] Ray E. Jennings: Preference and choice as logical correlates. *Mind*, 76. évf. (1967) 304. sz., 556–567. p.

[262] Keith Jensen: Punishment and spite, the dark side of cooperation. *Phil. Trans. R. Soc. B*, 365. évf. (2010), 2635–2650. p.

[263] Jan C. Joerden: Deontological square, hexagon, and decagon: A deontic framework for supererogation. *Logica Universalis*, 2012. February., 1–16. p.

[264] David B. Johnson: *A közösségi döntések elmélete*. 1999, Osiris.

[265] R. A. Johnstone – R. Bshary: Evolution of spite through indirect reciprocity. *Proc. R. Soc. Lond. B*, 271. évf. (2004), 1917–1922. p.

[266] Thomas Joiner: *Why People Die by Suicide*. 2005, Harvard University Press.

[267] Andrew J.I. Jones – Marek J. Sergot: A formal characterisation of institutionalised power. *Journal of the IGPL*, 4. évf. (1996) 3. sz., 429–445. p.

[268] Jörgen Jörgensen: Imperatives and logic. *Erkenntnis*, 7. évf. (1937/193) 8. sz., 288–296. p.

[269] Daniel Kahneman – Amos Tversky: Choices, values, and frames. *American Psychologist*, 39. évf. (1984) 4. sz., 341–350. p.

[270] Kahneman Daniel – Tversky Amos: Kilátáselmélet: a kockázatos helyzetekben hozott döntések elemei. In Csontos László (szerk.): *A racionális döntések elmélete*. 1998, Osiris – Láthatatlan Kollégium, 82–112. p.

[271] Walter J. Kamba: Legal Theory and Hohfeld's Analysis of a Legal Right. *Judicial Review*, 19. évf. (1974), 249–262. p.

[272] Michihiro Kandori: Social Norms and Community Enforcement. *The Review of Economic Studies*, 59. évf. (1992. January) 1. sz., 63–80. p.
URL `http://www.jstor.org/pss/2297925`.

[273] Stig Kanger: New foundations for ethical theory. In R. Hilpinen (szerk.): *Deontic Logic: Introductory and Systematic Readings*. 1971, Reidel, 36–58. p.

[274] Stig Kanger: Law and logic. *Theoria*, 38. évf. (1972), 105–132. p.

[275] Stig Kanger – Helle Kanger: Rights and parliamentarism. *Theoria*, 32. évf. (1966), 85–115. p.

[276] Hans Kelsen: *Tiszta jogtan*. Budapest, 2001, Rejtjel Kiadó.

[277] Robert O. Keohane: *After Hegemony. Cooperation and Discord in the World Political Economy*. 1984, Princeton University Press.

[278] Elias L. Khalil: What is altruism? *Journal of Economic Psychology*, 25. évf. (2004), 97–123. p.

[279] Bettina Klaus: Population-monotonicity and separability for economies with single-dipped preferences and the assignment of an indivisible object. *Economic Theory*, 17. évf. (2001), 675–692. p.

[280] Bettina Klaus – Hans Peters – Ton Storcken: Strategy-proof division of a private good when preferences are single-dipped. *Economics Letters*, 55. évf. (1997), 339–346. p.

[281] Stephen Cole Kleene: *Mathematical Logic*. New York, 2002, Dover.

[282] William Kneale – Martha Kneale: *A logika fejlődése*. Budapest, 1987, Gondolat Kiadó.

[283] Albert Kocourek: The hohfeld system of fundamental legal concepts. *Illinois Law Review*, 15. évf. (1920) 1. sz., 24–39. p.

[284] Kóczy László Á.: Központi felvételi rendszerek. taktikázás és stabilitás. *Közgazdasági Szemle*, 56. évf. (2009) 5. sz., 422–442. p.

[285] Kóczy László Á.: A magyarországi felvételi rendszerek sajátosságai. *Közgazdasági Szemle*, 57. évf. (2010) 2. sz., 142–164. p.

[286] L. A. Kornhauser – L. G. Sager: The one and the many: Adjudication in collegial courts. *California Law Review*, 81. évf. (1993), 1–59. p.

[287] Körösényi András: Lehangolótudomány, vagy a politikaművészete? Vezetés, manipuláció és demokrácia. *Politikatudományi Szemle*, XVIII. évf. (2009) 3. sz., 33–56. p.

[288] Erzsébet Kovács: A kockázat mint látens fogalom. *Hitelintézeti Szemle*, 10. évf. (2011) 4. sz., 349–359. p.

[289] Kovács Máté: Kinyilvánított preferencia és racionalitás. *Közgazdasági Szemle*, LVI. évf. (2009) 6. sz., 546–562. p.

[290] David Kraines – Vivian Kraines: Pavlov and the Prisoner's Dilemma. *Theory and Decision*, 26. évf. (1989), 47–79. p.

[291] David Kraines – Vivian Kraines: Learning to cooperate with pavlov an adaptive strategy for the iterated prisoner's dilemma with noise. *Theory and Decision*, 35. évf. (September, 1993) 2. sz., 107–150. p.

[292] Matthew H. Kramer: Rights without trimmings. In M. Kramer – N. Simmonds – H. Steiner (szerk.): *A Debate over Rights. Philosophical Enquiries*. 2000, Oxford University Press, 7–111. p.

[293] Matthew H. Kramer – N.E. Simmonds – Hillel Steiner: *A Debate over Rights. Philosophical Enquiries*. 2000, Oxford University Press.

[294] Dennis L. Krebs: Altruism: An examination of the concept and a review of the literature. *Psychological Bulletin*, 73. évf. (1970) 4. sz., 258–302. p.

[295] Christen Krogh – Henning Herrestad: Getting personal. some notes on the relationship between personal and impersonal obligation. In Mark A. Brown – Jose Carmo (szerk.): *Deontic Logic, Agency and Normative Systems. International Workshop on Deontic Logic in Computer Science*. 1996, Springer, 134–153. p.

[296] Christen Krogh – Henning Herrestad: Hohfeld in cyberspace and other applications of normative reasoning in agent technology. *Artificial Intelligence and Law March*, 7. évf. (1999) 1. sz., 81–96. p.

[297] David Lane: *Hierarchy in Natural and Social Sciences*. 4. köt. Hierarchy, Complexity, Society fejezet. 2006, Springer, 81–119. p.

[298] Irving H. Lavalle – Peter C. Fishburn: Lexicographic state-dependent subjective expected utility. *Journal of Risk and Uncertainty*, 4. évf. (1991) 3. sz., 251–269. p.

[299] Irving H. Lavalle – Peter C. Fishburn: State-independent subjective expected lexicographic utility. *Journal of Risk and Uncertainty*, 5. évf. (1992) 3. sz., 217–240. p.

[300] Marc Lavoie: Post Keynesian consumer theory: Potential synergies with consumer research and economic psychology. *Journal of Economic Psychology*, 25. évf. (2004), 639–649. p.

[301] E. J. Lemmon: Moral dilemmas. *The Philosophical Review*, 71. évf. (1962) 2. sz., 139–158. p.

[302] Edward Lemmon: Deontic logic and the logic of imperatives. *Logique et Analyse*, 8. évf. (1965), 39–71. p.

[303] David K. Levine: Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1. évf. (1998), 593–622. p.

[304] Michael E. Levine – Charles R. Plott: Agenda influence and its implications. *Virginia Law Review*, 63. évf. (1977) 4. sz., 561–604. p.

[305] David K. Lewis: *Convention. A Philosophical Study*. Cambridge (MA), 1969, Harvard University Press.

[306] Amir N. Licht: Games commissions play: 2x2 games of international securities regulation. *The Yale Law Journal*, 24. évf. (1999) 61. sz., 61–125. p.

[307] Wim B. G. Liebrand: The effect of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology*, 14. évf. (1984), 239–264. p.

[308] Lars Lindahl: *Position and Change*. 1977, Springer.

[309] Christian List – Philip Petit: Aggregating sets of judgments : an impossibility result. *Economics and philosophy*, 18. évf. (2002) 1. sz., 89–110. p.

[310] Christian List – Philip Pettit: Aggregating sets of judgement: Two impossibility results compared. *Synthese*, 140. évf. (2004), 207–235. p.

[311] Fenrong Liu: *Reasoning about Preference Dynamics*. 2011, Springer.

[312] Hugo Lövheim: A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses*, 78. évf. (2012) 2. sz., 341–348. p.

[313] R. Duncan Luce: Semiorders and a theory of utility discrimination. *Econometrica*, 24. évf. (1956) 2. sz., 178–191. p.

[314] Steven Lukes: Some problems about rationality. In Bryan R. Wilson (szerk.): *Rationality*. 1970s, Blackwell, 194–213. p.

[315] David Lyons: The correlativity of rights and duties. *Noûs*, 4. évf. (1970) 1. sz., 45–55. p.

[316] Roger D. Maddux: The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica*, 50. évf. (1990), 421–455. p.

[317] Magyarkúti Gyula: *A racionalitás fogalmának axiomatikus megközelítése*. PhD értekezés (BKÁE Közgazdaságtani PhD-program). 2001.

[318] D. Makinson: On the formal representation of rights relations. *Journal of Philosophical Logic*, 15. évf. (1986), 403–425. p.

[319] David Makinson: Stenius' approach to disjunctive permission. *Theoria*, 50. évf. (1984) 2-3. sz., 138–147. p.

[320] David Makinson: On a fundamental problem of deontic logic. In Paul McNamara – Henry Prakken (szerk.): *Norms, Logics and Information Systems. New Studies in Deontic Logic and Computer Science*. 1999, IOS Press, 29–53. p.

[321] Mala József: Arrow-típusú lehetetlenségi tételek. *Szigma*, XXXVIII. évf. (2007) 3-4. sz., 77–110. p.

[322] Málik József Zoltán: *Analitikus társadalomelmélet*. Budapest, 2006, szerzői kiadás.

[323] Stephen A. Marglin: Mit tesznek a főnökök? a hierarchia eredete és funkciói a tőkés termelésben. In *Munka, hatalom, gazdaság (munkaszociológiai tanulmányok)*. 1990, BKE TEK, 1–47. p.

[324] Frank W. Marlowe – J. Colette Berbesque – Clark Barrett – Alexander Bolyanatz – Michael Gurven – David Tracer: The 'spiteful' origins of human cooperation. *Proc. R. Soc. B*, 278. évf. (2011), 2159–2164. p.

[325] Marcel Mauss: A test technikái. In *Szociológia és antropológia*. Budapest, 2004, Osiris Kiadó, 425–446. p.

[326] Kenneth O. May: A set of independent necessary and sufficient conditions for majority decision. *Econometrica*, 20. évf. (1952) 4. sz., 680–684. p.

[327] John Maynard Smith – G. R. Price: The logic of animal conflict. *Nature*, 246. évf. (1973) 5427. sz., 15–18. p.

[328] Ernst Mayr: Cause and effect in biology. *Science, New Series*, 134. évf. (1961) 3489. sz., 1501–1506. p.

[329] Ernst Mayr: Teleological and teleonomic, a new analysis. *Boston Studies in the Philosophy of Science*, 14. évf. (1974), 91–117. p.

[330] Ernst Mayr: The idea of teleology. *Journal of the History of Ideas*, 53. évf. (1992) 1. sz., 117–135. p.

[331] Charles G. McClintock: Social motivation – a set of propositions. *Behavioral Science*, 17. évf. (1972) 5. sz., 438–454. p.

[332] Joseph E. McElliskem: Affective and predatory violence: a bimodal classification system of human aggression and violence. *Aggression and Violent Behavior*, 10. évf. (2004), 1–30. p.

[333] R.D. McKelvey: Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12. évf. (1976), 472–482. p.

[334] Richard D. McKelvey: General conditions for global intransitivities in formal voting models. *Econometrica*, 47. évf. (1979) 5. sz., 1086–1112. p.

[335] P. McNamara: Making room for going beyond the call. *Mind*, 105. évf. (1996) 419. sz.

[336] Paul McNamara: Doing well enough: Toward a logic for common-sense morality. *Studia Logica: An International Journal for Symbolic Logic*, 57. évf. (1996) 1. sz., 167–192. p.

[337] George Herbert Mead: *A pszichikum, az én és a társadalom*. 1973, Gondolat.

[338] Medvegyev Péter: Néhány megjegyzés a kockázat, bizonytalanság, valószínűség kérdéséhez. *Hitelintézeti Szemle*, 10. évf. (2011) 4. sz., 314–324. p.

[339] Mérő László: *Mindenki másként egyforma*. Budapest, 1996, Tertium.

[340] Mérő László: *Az érzelmek logikája*. 2010, Tericum Kiadó.

[341] David M. Messick – Charles G. McClintock: Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, 4. évf. (1968), 1–2. p.

[342] Mészáros József: *Játékelmélet*. Budapest, 2005, Gondolat Kiadó.

[343] Mészáros József – Szakadát István: *Választási eljárások – választási rendszerek*. A BME Szociológia Tanszék kiadványai sorozat, 2. köt. 1993, BME Szociológia Tanszék. URL `http://mycite.omikk.bme.hu/doc/88355.pdf`.

[344] Mészáros József – Szakadát István: *Magyarország Politikai Atlasza*. Budapest, 1995, Konrad Adenauer Stiftung.

[345] Mészáros József – Szakadát István: *Parlamenti képviselői helyek megoszlásának becslése közvélemény-kutatási adatok alapján*. Tárki társadalompolitikai tanulmányok sorozat, 6. köt. Budapest, 1998, TÁRKI.

[346] Mészáros József – Szakadát István: Választási eredmények területi megoszlása. In Kolosi Tamás – Tóth István György – Vukovich Györgyi (szerk.): *Társadalmi Riport*. Budapest, 1998, TÁRKI, 412–433. p.

[347] Mészáros József – Szakadát István: *Magyarország Politikai Atlasza 1998*. Budapest, 1999, Osiris Kiadó.

[348] Mészáros József – Szakadát István: Territorial distribution of election results. In Kolosi Tamás – Tóth István György – Vukovich Györgyi (szerk.): *Social Report 1998*. Budapest, 1999, TÁRKI, 409–430. p.

[349] Mészáros József – Szakadát István: A választási rendszer elemei. In Dezső Márta – Kukorelli István (szerk.): *Választástudományi tanulmányok*. 1999, OVB, 217–242. p.

[350] Mészáros József – Szakadát István: *Magyarország Politikai Atlasza 2004*. 2005, Gondolat Kiadó.

[351] Meszerics Tamás: Stratégiai viselkedés és bizottsági döntés. *Közgazdasági Szemle*, XLIV. évf. (1997) 7-8. sz., 687–697. p.

[352] Meszerics Tamás: A manipuláció és a mögöttes normák. *Beszélő*, 8. évf. (2003) 11. sz., 18–25. p.

[353] John-Jules Ch. Meyer: A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29. évf. (1988) 1. sz., 109–136. p.

[354] John-Jules Ch. Meyer: Reasoning about emotional agents. *International Journal of Intelligent Systems*, 21. évf. (2006) 6. sz., 601–619. p.

[355] John-Jules Ch. Meyer – W. van der Hoek – B. van Linder: A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113. évf. (1999), 1–40. p.

[356] E. Midlarsky – E. Kahana: Theories and concepts of altruism and helping. In E. Midlarsky – E. Kahana (szerk.): *Alruism in Later Life*. Library of Social Research sorozat, 8. köt. 1994, Sage, 11–40. p.

[357] Ruth Garrett Millikan: *Language, Thought and Other Biological Categories*. 1984, MIT Press.

[358] Ruth Garrett Millikan: Thoughts without laws; cognitive science with content. *The Philosophical Review*, 95. évf. (1986) 1. sz., 47–80. p.

[359] Ruth Garrett Millikan: Bioszemantika. In Gergely Ambrus – Tamás Demeter – Gábor Forrai – János Tőzsér (szerk.): *Elmefilozófia*. 2008, L'Harmattan, 252–270. p.

[360] Gina Mireault – John Sparrow – Merlin Poutre – Brittany Perdue – Laura Macke: Infant humor perception from 3- to 6-months and attachment at one year. *Infant Behavior and Development*, 35. évf. (2102), 797–802. p.

[361] Philippe Mongin: The doctrinal paradox, the discursive dilemma, and logical aggregation theory. *Theory and Decision*, 73. évf. (2012) 3. sz., 315–355. p.

[362] Kristen Renwick Monroe: A Fat Lady in a Corset: Altruism and Social Theory. *American Journal of Political Science*, 38. évf. (1994) 4. sz., 861–893. p.

[363] Alessio Moretti: *The Geometry of Logical Opposition*. PhD értekezés (University of Neuchâtel). 2009.

[364] Alessio Moretti: The geometry of standard deontic logic. *Logica Universalis*, 3. évf. (2009), 19–57. p.

[365] Alessio Moretti: The critics of paraconsistency and of many-valuedness and the geometry of oppositions. *Logic and Logical Philosophy*, 19. évf. (2010), 63–94. p.

[366] Alessio Moretti: Why the logical hexagon? *Logica Universalis*, 6. évf. (2012), 69–107. p.

[367] Pierre Le Morvan: Intentionality: Transparent, translucent, and opaque. *Journal of Philosophical Research*, 30. évf. (2005), 283–302. p.

[368] Pierre Le Morvan: Selfishness, altruism, and our future selves. *Australasian Journal of Philosophy*, 87. évf. (2009) 3. sz., 409–424. p.

[369] Hervé Moulin: On strategy-proofness and single peakedness. *Public Choice*, 35. évf. (1980), 437–455. p.

[370] Roger B. Myerson: Axiomatic derivation of scoring rules without the ordering assumption. *Social Choice and Welfare*, 12. évf. (1995), 59–74. p.

[371] Jan Narveson: *The Libertarian Idea*. 2001, Broadview Press.

[372] John Nash: Non-cooperative games. *The Annals of Mathematics, Second Series*, 54. évf. (1951. Sep) 2. sz., 286–295. p.

[373] John F. Nash: The bargaining problem. *Econometrica*, 18. évf. (1950) 2. sz., 155–162. p.

[374] John F. Nash: Equilibrium points in n-person games. *Proc. Natl. Acad. Sci.*, 36. évf. (1950), 48–49. p.

[375] Klaus Nehring: Self-control through second-order preferences, 2006.

[376] Némedi Dénes: *Émile Durkheim*. Budapest, 1996, Áron Kiadó.

[377] Yew-kwang Ng: Sub-semiorder: A model of multidimensional choice with preference intransitivity. *Journal of Mathematical Psychology*, 16. évf. (1977) 1. sz., 51–59. p.

[378] M. Nowak – K. Sigmund: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364. évf. (1993), 56–58. p.

[379] M. A. Nowak – K. Sigmund: The alternating prisoner's dilemma. *Journal of Theoretical Biology*, 168. évf. (1994), 219–226. p.

[380] M. A. Nowak – K. Sigmund: Evolution of indirect reciprocity by image scoring. *Nature*, 393. évf. (1998), 573–577. p.

[381] Martin A. Nowak: Evolutionary dynamics of cooperation. In *Proceedings of the International Congress of Mathematicians* (konferenciaanyag). 2006, 1523–1540. p.

[382] Martin A. Nowak – Karl Sigmund: Tit for tat in heterogeneous populations. *Nature*, 355. évf. (1992. January), 250–253. p.

[383] Martin A. Nowak – Corina E. Tarnita – Edward O. Wilson: The evolution of eusociality. *Nature*, 466. évf. (2010) 7310. sz., 1057–1062. p.

[384] Lauri Nummenmaa – Enrico Glerean – Riitta Hari – Jari K. Hietanen: Bodily maps of emotions. *PNAS*, 110. évf. (2013).

[385] Hannu Nurmi: Compound majority paradoxes and proportional representation. *European Journal of Political Economy*, 13. évf. (1997), 443–4454. p.

[386] Keith Oatley – Jennifer M. Jenkins: *Érzelmeink*. 2001, Osiris.

[387] Óészaki mitológiai és hősi énekek: 2. a fenséges beszéde, 42–43. versszak, 2003. January. URL `http://mek.oszk.hu/00300/00377/`.

[388] Hisashi Ohtsuki – Yoh Iwasa: How should we define goodness? – reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231. évf. (2004. November) 1. sz., 107–120. p.

[389] Hisashi Ohtsuki – Yoh Iwasa: The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239. évf. (2006. April) 4. sz., 435–444. p.

[390] Hisashi Ohtsuki – Yoh Iwasa: Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *Journal of Theoretical Biology*, 244. évf. (2007) 3. sz., 518 – 531. p. ISSN 0022-5193. URL http://www.sciencedirect.com/science/article/pii/S0022519306003900.

[391] Andrew Oldenquist: The possibility of selfishness. *American Philosophical Quarterly*, 17. évf. (1980) 1. sz., 25–33. p.

[392] Pamela Oliver: Rewards and Punishments as Selective Incentives. An Apex Game. *The Journal of Conflict Resolution*, 28. évf. (1980. May) 1. sz., 123–148. p.

[393] Pamela Oliver: Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Ivestigations. *The American Journal of Sociology*, 85. évf. (1980. May) 6. sz., 1356–1375. p.

[394] Mancur Olson: *A kollektív cselekvés logikája*. Budapest, 1997, Osiris.

[395] Peter C. Ordeshook: *Game theory and political theory. An introduction*. 1986, Cambridge University Press.

[396] Orthmayr Imre: A sztoikus szenvtelenség. *Századvég*, 4. évf. (1999) 13. sz. URL http://www.c3.hu/scripta/szazadveg/13/orthmayr.htm.

[397] Orthmayr Imre: Erkölcsi normák. *Világosság*, 44. évf. (2003) 5-6. sz., 149–161. p.

[398] Orthmayr Imre: A társadalmi normák döntéselméleti és evolúciós magyarázata. *Szociológiai Szemle*, 14. évf. (2004) 3. sz., 3–22. p.

[399] Orthmayr Imre: Konfliktusok. *Világosság*, 48. évf. (2007) 4. sz., 71–89. p.

[400] A. Ortony – G. L. Clore – A. Collins: *The cognitive structure of emotions*. 1988, Cambridge University Press.

[401] Jorge M. Pacheco – Francisco C. Santos – Fabio A.C.C. Chalub: Stern-Judging: a Simple, Successful Norm which Promotes Cooperation under Indirect Reciprocity. *PLoS Computational Biology*, 2. évf. (2006. december) 12. sz., 1634–1638. p.

[402] Allan Paivio: *Mental representations: A dual-coding approach*. 1986, Oxford University Press.

[403] Karthik Panchanathan – Rob Boyd: Human cooperation: Second-order free-riding problem solved? (reply). *Nature*, 437. évf. (2005. 09) 7058. sz., E8–E9. p. URL http://dx.doi.org/10.1038/nature04202.

[404] Karthik Panchanathan – Robert Boyd: Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432. évf. (2004. November 25.), 499–502. p.

[405] Talcott Parsons – Edward A. Shils: *Toward a General Theory of Action. Theoretical Foundations for the Social Sciences*. New Brunswick (USA) and London (UK), 2007 (1951), Transactions Publisher.

[406] Terence Parsons: The traditional square of opposition. in stanford encyclopedia of philosophy. URL http://plato.stanford.edu/entries/square/.

[407] Marc Pauly – Martin van Hees: Logical constraints on judgment aggregation. *Journal of Philosophical Logic*, 35. évf. (2006) 6. sz., 569–585. p.

[408] Sam Perlo-Freeman: The topology of conflict and co-operation, 2006. May.

[409] Michael Perloff: Stit and the language of agency. *Synthese*, 86. évf. (1991) 3. sz., 379–408. p.

[410] Thomas D. Perry: A paradigm of philosophy: Hohfeld on legal rights. *American Philosophical Quarterly*, 14. évf. (1977) 1. sz., 41–50. p.

[411] Pete Krisztián: A társadalmi konvenciók természetéről. *Infotárs*, 2008., 73–99. p.

[412] R. W. Picard: Affective computing. No. 321. Jelentés, 1995, M.I.T Media Laboratory Perceptual Computing Section Technical Report.

[413] Jane Allyn Piliavin – Hong-Wen Charng: Altruism: A review of recent theory and research. *Annual Review of Sociology*, 16. évf. (1990), 27–65. p.

[414] John Platt: Social traps. *American Psychologist*, 28. évf. (1973) 8. sz., 641–651. p.

[415] Ifjabb Plinius: *Levelek*. Budapest, 1981, Európa Könyvkiadó.

[416] Charles R. Plott – Michael E. Levine: A model of agenda influence on committee decisions. *American Economic Review*, 68. évf. (1978) 1. sz., 146–70. p.

[417] Robert Plutchik: *Title: Emotion: A Psychoevolutionary Synthesis*. 1980, Harper & Row.

[418] Pokol Béla: *Modern francia szociológiaelméletek*. Miskolc, 1996, Bíbor Kiadó. URL http://mek.oszk.hu/02000/02027/02027.pdf.

[419] Polányi Mihály: *Személyes tudás I-II.* 2004, Atlantisz Kiadó.

[420] Ingmar Pörn: *The Logic of Power*. 1970, Blackwells.

[421] William Poundstone: *Prisoner's Dilemma*. 1992, Anchor.

[422] A. N. Prior: The paradoxes of derived obligation. *Mind, New Series*, 63. évf. (1954) 249. sz., 64–65. p.

[423] A. N. Prior: Escapism: The logical basis of ethics. In A.I. Melden (szerk.): *Essays In Moral Philosophy*. Seattle, 1958, University of Washington Press, 135–146. p.

[424] Warren S. Quinn: The puzzle of the self-torturer. *Philosophical Studies*, 59. évf. (1990) 1. sz., 79–90. p.

[425] Douglas W. Rae – Hans Daudt: The ostrogorski paradox: a peculiarity of compound majority decision. *European Journal of Political Research*, 4. évf. (1976) 4. sz., 391–399. p.

[426] N. J. Raihani – K. McAuliffe: Human punishment is motivated by inequity aversion, not a desire for reciprocity, 2012.

[427] J. M. Ramírez – J. M. Andreu: Aggression, and some related psychological constructs (anger, hostility, and impulsivity). some comments from a research project. *Neurosci Biobehav Rev*, 30. évf. (2006) 3. sz., 276–291. p.

[428] Ranschburg Jenő: *Félelem, harag, agresszió*. 1979, Tankönyvkiadó.

[429] Joseph Ransdell: Constitutive rules and speech-act analysis. *The Journal of Philosophy*, 68. évf. (1971) 13. sz., 385–400. p.

[430] Anatol Rapoport: Three modes of conflict. *Management Science*, 7. évf. (1961) 3. sz., 210–218. p.

[431] Anatol Rapoport: Optimal policies for the prisoner's dilemma. *Psychological Review*, 74. évf. (1967) 2. sz., 136–148. p.

[432] Anatol Rapoport: A kizsákmányoló, a vezér, a hős és a mártír: a 2 x 2-es játszma négy archetípusa. In László Csontos (szerk.): *A racionális döntések elmélete*. Budapest, 1998, Osiris Kiadó – Láthatatlan Kollégium, 129–134. p.

[433] John Rawls: Kétfajta szabályfogalom. In Mária Lónyai (szerk.): *Tények és értékek. A modern angolszász etika irodalmából*. Budapest, 1981, Gondolat, 364–416. p.

[434] Joseph Raz: *Practical Reason and Norms*. 1975/2002, Oxford University Press.

[435] Joseph Raz: *The Morality of Freedom*. 1989, Clarendon Press.

[436] Hyne W. Reese: Teleology and teleonomy in behavior analysis. *The Behavior Analyst*, 17. évf. (1994) 1. sz., 75–91. p.

[437] Réz Anna: Makacs érzelmek és morális felelősség. *Magyar Filozófiai Szemle*, 55. évf. (2011) 1. sz., 23–33. p.

[438] Réz Anna (szerk.): *Vétkek és választások - A felelősség elméletei*. 2013, Gondolat Kiadó.

[439] Marcel K. Richter: Revealed preference theory. *Econometrica*, 34. évf. (1966) 3. sz., 635–645. p.

[440] Matt Ridley: *The Origins of Virtue*. 1997, Softback Preview.

[441] William H. Riker: *The Art of Political Manipulation*. 1986, Yale University Press.

[442] Kevin W.S. Roberts: Voting over income tax schedules. *Journal of Public Economics*, 8. évf. (1977) 3. sz., 329–340. p.

[443] David Robinson – David Goforth: *The Topology of the 2x2 Games*. 2005, Routledge.

[444] David Robinson – David Goforth – Matt Cargill: Toward a topological treatment of the non-strictly ordered 2x2 games, 2007.

[445] Alejandro Rosas: Evolutionary game theory meets social science: Is there a unifying rule for human cooperation? *Journal of Theoretical Biology*, 264. évf. (2010), 450–456. p.

[446] Alf Ross: Imperatives and logic. *Philosophx of Science*, 11. évf. (1944) 1. sz., 30–46. p.

[447] David Ross: *The Right and the Good*. Oxford, 2002 (1930), Clarendon Press.

[448] A. E. Roth: Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory, Special Issue in Honor of David Gale on his 85th birthday*, 36. évf. (2008) 3-4. sz., 537–569. p.

[449] Alvin E. Roth: The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92. évf. (1984) 6. sz., 991–1016. p.

[450] Paul Rothstein: Representative voter theorems. *Public Choice*, 72. évf. (1991), 193–212. p.

[451] M. Roubens – Ph. Vincke: *Preference modelling*. Lecture Notes in Economics and Mathematical Systems sorozat, 250. köt. Berlin, 1985, Springer.

[452] N. Rouland: Az emberi jogok antropológiai alapjai. In István H. Szilágyi (szerk.): *Jog és antropológia*. 2005, Szent István Társulat, 275–294. p.

[453] Stuart J. Russel – Devika Subramanian: Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, 2. évf. (1995), 575–609. p.

[454] Ruzsa Imre: *Bevezetés a modern logikába*. 2000, Osiris Kiadó.

[455] Gilbert Ryle: *A szellem fogalma*. Budapest, 1974, Gondolat.

[456] Donald G. Saari: *Decisions and Elections: Explaining the Unexpected*. Cambridge, 2001, Cambridge University Press.

[457] Marshall D. Sahlins: Törzsek. In *Vadászok, törzsek, parasztok*. Budapest, 1973, Kossuth Könyvkiadó, 135–315. p.

[458] David H. Sanford: Contraries and subcontraries. *Noûs*, 2. évf. (1968) 1. sz., 95–96. p.

[459] Alejandro Saporiti: Strategy-proofness and single-crossing. *Theoretical Economics*, 4. évf. (2009), 127–163. p.

[460] Mark Allen Satterthwaite: Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10. évf. (1975), 187–216. p.

[461] Frederick Schauer: A szabályok és a jog uralma. In *A Hart utáni jogelmélet alapproblémái*. 2004, Bíbor Kiadó, 407–457. p.

[462] Th.C. Schelling: *The Strategy of Conflict*. Cambridge (MA), 1980, Harvard University Press.

[463] Walter Schneider – Richard M. Shiffrin: Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, 84. évf. (1977), 1–66. p.

[464] Wijnand J. Schoenmakers: A problem in knowledge acquisition. *SIGART Newsletter*, 95. évf. (1986), 56–57. p.

[465] Arthur Schopenhauer: *Az akarat szabadságáról*. 1913, Franklin Társulat.

[466] George Schwab: Enemy vagy foe: a modern politika konfliktusa. In Márton Szabó (szerk.): *Az ellenség neve*. 1998, Jószöveg Műhely, 39–55. p.

[467] Thomas Schwartz: The paradox of representation. *The Journal of Politics*, 57. évf. (1995) 2. sz., 309–323. p.

[468] John R. Searle: *The Construction of Social Reality*. 1995, The Free Press.

[469] John R. Searle: *Elme, nyelv, társadalom. A való világ filozófiája*. Budapest, 2000, Vince Kiadó.

[470] John R. Searle: *Beszédaktusok*. 2009, Alkalmazott Kommunikációtudományi Intézet – Gondolat Kiadó.

[471] Krister Segerberg: Validity and satisfaction in imperative logic. *Notre Dame Journal of Formal Logic*, 31. évf. (1990), 203–221. p.

[472] Krister Segerberg: Getting started: Beginnings in the logic of action. *Studia Logica: An International Journal for Symbolic Logic*, 51. évf. (1992) 3/4. sz., 347–378. p.

[473] A.K. Sen: Internal consistency of choice. *Econometrica*, 61. évf. (1993) 3. sz., 495–521. p.

[474] Amartya K. Sen: A possibility theorem on majority decisions. *Econometrica*, 34. évf. (1966) 2. sz., 491–499. p.

[475] Amartya K. Sen: Isolation, assurance and the social rate of discount. *The Quarterly Journal of Economics*, 81. évf. (1967) 1. sz., 112–124. p.

[476] Amartya K. Sen: Quasi-transitivity, rational choice and collective decisions. *The Review of Economic Studies*, 36. évf. (1969) 3. sz., 381–393. p.

[477] Amartya K. Sen: *Collective Choice and Social Welfare*. 1970, Holden-Day.

[478] Amartya K. Sen: The impossibility of a paretian liberal. *Journal of Political Economy*, 78. évf. (1970) 1. sz., 152–157. p.

[479] Amartya K. Sen: Choice functions and revealed preference. *Review of Economic Studies*, 38. évf. (1971) 3. sz., 307–317. p.

[480] Amartya K. Sen: Is a paretian liberal really impossible: A reply. *Public Choice*, 21. évf. (1975), 111–113. p.

[481] Amartya K. Sen: Liberty, unanimity and rights. *Economica*, 43. évf. (1976) 171. sz., 217–245. p.

[482] Amartya K. Sen: Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6. évf. (1977) 4. sz., 317–344. p.

[483] Amartya K. Sen: Maximization and the act of choice. *Econometrica*, 65. évf. (1997) 4. sz., 745–779. p.

[484] Amartya K. Sen – Prasanta K. Pattanaik: Necessary and sufficient conditions for rational choice under majority decision. *Journal of Economic Theory*, 1. évf. (1969), 178–202. p.

[485] Amartya Kumar Sen: Racionális bolondok. a közgazdaságtan magatartásbeli megalapozásának kritikája. *Café Bábel*, 17. évf. (2008) 55. sz., 93–104. p.

[486] Marek J. Sergot: A computational theory of normative positions. *ACM Transactions on Computational Logic*, 2. évf. (2001) 4. sz., 581–622. p.

[487] Richard M. Shiffrin – Walter Schneider: Controlled and automatic human information processing: II. perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84. évf. (1977), 127–190. p.

[488] Herbert Simon: A behavioral model of rational choice. *Quarterly Journal of Economics*, 69. évf. (1955), 99–118. p.

[489] Herbert A. Simon: Bonyolult rendszerek szerkezete. In *Korlátozott racionalitás*. 1982, KJK, 87–120. p.

[490] Herbert A. Simon: A racionalitás mint folyamat és gondolkodási produktum. In *Korlátozott racionalitás*. Budapest, 1982, KJK, 59–86. p.

[491] Herbert A. Simon: Rationality in psychology and economics. *The Journal of Business*, 59. évf. (1986) 4. sz., S209–S224. p.

[492] Simonovits András: Bevezetés a játékelméletbe: vázlat, 2007.

[493] Peter Simons: *Parts: A Study in Ontology*. 2000 (1987), Oxford University Press.

[494] Brian Skyrms: *The Stag Hunt and the Evolution of Social Structure*. 2004, Cambridge University Press.

[495] Brian Skyrms: Evolution and the social contract, 2007.
URL `http://tannerlectures.utah.edu/_documents/a+to+z/s/Skyrms_07.pdf`.

[496] Paul Slovic – Melissa L. Finucane – Ellen Peters – Donald G. MacGregor: Rational Actors or Rational Fools? Implications of the Affect Heuristic for Behavioral Economics. *Journal of Socio-Economics*, 31. évf. (2002), 329–342. p.

[497] Hans Smessaert: On the 3d visualisation of logical relations. *Logica Universalis*, 3. évf. (2009), 303–332. p.

[498] Adam Smith: *A nemzetek gazdagsága. E gazdagság természetének és okainak vizsgálata*. 1959, Akadémiai Kiadó.

[499] John H. Smith: Aggregation of preferences with variable electorate. *Econometrica*, 41. évf. (1973) 6. sz., 1027–1041. p.

[500] John Maynard Smith: *Evolution and the Theory of Games*. 1982, Cambridge University Press.

[501] Michael Smith – David Lewis – Mark Johnston: Dispositional theories of value. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 63. évf. (1989), 89 –111+113–137+139–174. p.

[502] Glenn H. Snyder – Paul Diesing: *Conflict Among Nations*. Princeton, NJ, 1977, Princeton University Press.

[503] Elliott Sober – David Sloan Wilson: *Unto Others: The Evolution and Psychology of Unselfish Behavior*. 1998, Harvard University Press.

[504] Solt Kornél: *Jogi logika I-II.* 1996, MTA Állam- és Jogtudományi Intézete and Seneca.

[505] Somlai Péter: Triádok, közvetítők, koalíciók. konfliktusok és egyensúlyok kettőnél több tagú szociális rendszerekben. *Szociológiai Szemle*, 2. évf. (1992) 4. sz., 3–14. p.

[506] Chun Siong Soon – Marcel Brass – Hans-Jochen Heinze – John-Dylan Haynes: Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11. évf. (2008), 543–545. p.

[507] Wolfgang Spohn: An analysis of hansson's dyadic deontic logic. *Journal of Philosophical Logic*, 4. évf. (1975. May) 2. sz., 237–252. p.

[508] J. A. Srejder: *Egyenlőség, hasonlóság, rendezés*. Budapest, 1975, Gondolat.

[509] Dennis W. Stampe: The authority of desire. *Philosophical Review*, 96. évf. (1987) 3. sz., 335–381. p.

[510] Steiger Kornél: Megjegyzések Orthmayr Imre "a sztoikus szenvtelenség" c. tanulmányához. *Századvég*, 1999.
URL `http://www.c3.hu/scripta/szazadveg/13/steiger.htm`.

[511] Robert Sugden: Spontaneous order. *The Journal of Economic Perspectives*, 3. évf. (1989) 4. sz., 85–97. p.

[512] Robert Sugden: *The Economics of Rights, Co-operation and Welfare*. 2004 (1986), Palgrave MacMillan.

[513] Cass R. Sunstein: A pozitív jogok ellen. *Fundamentum*, 2. évf. (1998) 3. sz., 11–14. p.

[514] Kotaro Suzumura: *Rational choice, collective decisions, and social welfare*. 1983, Cambridge University Press.

[515] Syi: *Egyben az egész. Egytől egyig*. 2007, Typotex.

[516] Syi: Akarat-erő-logika. In Katalin S. Nagy – Annamária Orbán (szerk.): *Értékek és normák interdiszciplináris megközelítésben*. Budapest, 2008, Gondolat Kiadó, 50–68. p.

[517] Syi: *Cselekvéselmélet dióhéjban*. 2008, Typotex.

[518] Syi: Önzetlenségelméletek. *BUKSZ*, 21. évf. (2009) 4. sz., 226–233. p.

[519] Syi: Max weber cselekvéselméleti ontológiája. *Szociológiai Szemle*, 21. évf. (2011) 2. sz., 87–110. p.

[520] Syi: Feléd fordítom, újra. In Bajnok Andrea – Korpics Márta – Milován Andrea – Pólya Tamás – Szabó Levente (szerk.): *A kommunikatív állapot: Diszciplináris rekonstrukciók*. 2012, Typotex Kiadó, 122–134. p.

[521] Syi: A racionális aracionális. *BUKSZ*, 24. évf. (2012) 2. sz., 118–125. p.

[522] Syi: A reciprocitás rejtett mintázatai. *Szociológiai Szemle*, 23. évf. (2013) 3. sz., 30–50. p.

[523] Syi: Hidden patterns of reciprocity. *Journal of Theoretical Biology*, 2014.

[524] Szakadát István: *A kommunista uralmi rend és a nómenklatúraelv*. PhD értekezés (MTA szociológiai tudományok kandidátusa). 1999.

[525] Szakadát István: Hierarchia. In Helmich Dezső – Szántó Zoltán (szerk.): *Metodológia, társadalom, gazdaság. Im memoriam Bertalan László*. 2004, Közgazdasági Szemle Alapítvány, 139–166. p.

[526] Szakadát István: Reláció, szintaktika, szemantika. *Tudományos és Műszaki Tájékoztatás*, 51. évf. (2004) 12. sz., 531–540. p.

[527] Szakadát István: Réteges struktúra, alaprelációk. In Csendes Lilla – Alexin Zoltán (szerk.): *III. Szám´ógépes nyelvészeti konferencia* (konferenciaanyag). Szeged, 2005, SZTE.

[528] Szakadát István: A politikai cselekvések szabályszerűségei. In S.Nagy Katalin (szerk.): *Szociológia*. 2006, Typotex Kiadó, 99–114. p.

[529] Szakadát István: Politikai gazdaságtan. In S.Nagy Katalin (szerk.): *Szociológia közgazdászoknak*. 2007, Typotex Kiadó, 213–231. p.

[530] Szántó Zoltán: A racionális döntések elméletén nyugvó társadalomtudomány. In Csontos László (szerk.): *A racionális döntések elmélete*. 1998, Osiris – Láthatatlan Kollégium, 7–24. p.

[531] Száz János: Valószínűség, esély, relatív súlyok. opciók és reálopciók. *Hitelintézeti Szemle*, 10. évf. (2011) 4. sz., 336–348. p.

[532] Szécsi Gábor: Okság és intencionalitás. *Magyar Filozófiai Szemle*, 39. évf. (1995) 3-4. sz., 529–561. p.

[533] Szent-Györgyi Albert: *Az őrült majom. Írta egy biológus*. 1989, Magvető Kiadó.

[534] Szily László: Az apostoli magyarok leváltották schmittet.
URL `http://index.hu/belfold/2012/03/01/schmitt_palt_levaltottak_lazar_es_tarsai_haditorvenyszek_elott/`.

[535] Szolcsányi Tibor: *A feljogosító következtetések logikája*. PhD értekezés (Pécsi Tudományegyetem Bölcsészettudományi Kar Filozófia Intézet). 2009.

[536] Takács Károly: Hálózati kísérletek. *Közgazdasági Szemle*, LVII. évf. (2010), 958–979. p.

[537] N. Takahashi – R. Mashima: The emergence of indirect reciprocity: Is the standing strategy the answer?, 2003. Aug.
URL `http://www.allacademic.com/meta/p107391_index.html`. Paper presented at the annual meeting of the American Sociological Association, Atlanta Hilton Hotel, Atlanta, GA Online <.PDF> Retrieved 2006-10-05.

[538] Alfred Tarski: On the calculus of relations. *Journal of Symbolic Logic*, 6. évf. (1941), 73–89. p.

[539] Alan D. Taylor: *Social Choice and the Mathematics of Manipulation*. Cambridge, 2005, Cambridge University Press.

[540] Charles Taylor: *Sources of the Self. The Making of the Modern Identity*. 1989, Harvard University Press.

[541] Temesi József: *A döntéselmélet alapjai*. 2002, Aula Kiadó.

[542] Warren D. TenHouten: *A General Theory of Emotions and Social Life*. 2007, Routledge.

[543] James E. Tomberlin: Good samaritans and castañeda's system of deontic logic. In James E. Tomberlin (szerk.): *Héctor-Neri Castañeda*. 6. köt. 1986, Springer, 255–272. p.

[544] Tőzsér János: Az intencionális állapotok szerkezete. In Varga Péter András – Zuh Deodáth (szerk.): *Husserl és a Logikai vizsgálódások. Ismeretfilozófia és fenomenológiai filozófia*. 2009, L'Harmattan, 165–177. p.

[545] Judy Trevena – Jeff Miller: Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition*, 19. évf. (2010) 1. sz., 447–456. p.

[546] Robert L. Trivers: A reciprok altruizmus evolúciója. In Zsuzsanna Kulcsár (szerk.): *Morális fejlődés, empátia és altruizmus*. 1999, ELTE Eötvös Kiadó, 245–272. p.

[547] Gordon Tullock: Problems of majority voting. *Journal of Political Economy*, 67. évf. (1959) 6. sz., 571–579. p.

[548] Amos Tversky: Intransitivity of preferences. *Psychological Review*, 76. évf. (1969) 1. sz., 31–48. p.

[549] Ullman Tamás: *A láthatatlan forma. Sematizmus és intencionalitás*. 2010, L'Harmattan.

[550] Edna Ullmann-Margalit: *The Emergence of Norms*. 1977, Clarendon Press.

[551] James O. Urmson: Saints and heroes. In A.J. Melden (szerk.): *Essays in Moral Philosophy*. Seattle, 1958, University of Washington Press, 198–216. p.

[552] James O. Urmson: Szentek és hősök. *Café Bábel*, 13. évf. (1994) 3. sz., 53–62. p.

[553] Vámos Tibor: Kooperatív rendszerek, új fejődési távlatok. In *Hazánk és a műszaki haladás*. Gyorsuló idõ sorozat. 1984, Magvetõ, 126–167. p.

[554] Johan van Benthem – Patrick Girard – Olivier Roy: Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, 38. évf. (2009) 1. sz., 83–125. p.

[555] Johan van Benthem – Fenrong Liu: Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17. évf. (2007) 2. sz., 157–182. p.

[556] Paul A. M. van Lange: The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77. évf. (1999) 2. sz., 337–349. p.

[557] Paul A. M. van Lange – D. De Cremer – E. Van Dijk – M. Van Vugt: Self-interest and beyond: Basic principles of social interaction. In A. W. Kruglanski – E. T. Higgings (szerk.): *Social Psychology: Handbook of Basic Principles*. 2nd edition. kiad. 2007, Guilford, 540–561. p.

[558] Paul A. M. van Lange – W. Otten – E. M. N. De Bruin – J. A. Joireman: Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73. évf. (1997) 4. sz., 733–746. p.

[559] A.C. Varzi: Parts, wholes, and part-whole relations: The prospects of mereotopology. *Data and Knowledge Engineering*, 20. évf. (1996), 177–198. p.

[560] Bas Verplanken – Wendy Wood: Interventions to break and create consumer habits. *Journal of Public Policy and Marketing*, 25. évf. (2006), 90–103. p.

[561] William Vickrey: Utility, strategy, and social decision rules. *The Quarterly Journal of Economics*, 74. évf. (1960) 4. sz., 507–535. p.

[562] G. H. von Wright: A note on deontic logic and derived obligation. *Mind, New Series*, 65. évf. (1956) 260. sz., 507–509. p.

[563] Georg Henrik von Wright: *An Essay in Deontic Logic*. 1951, North-Holland Publishing Company.

[564] Georg Henrik von Wright: *Norm and Action*. London, 1963, Routledge and Kegan Paul.

[565] Georg Henrik von Wright: *The Logic of Preference*. Edinburgh, 1963, Edinburgh University Press.

[566] Georg Henrik von Wright: The logic of preference reconsidered. *Theory and Decision*, 3. évf. (1972) 3. sz., 140–169. p.

[567] Georg Henrik von Wright: Deontikus logika. In Irving M. Copi – James A. Gould (szerk.): *Kortárs-tanulmányok a logikaelmélet kérdéseiről*. Budapest, 1985, Gondolat, 531–554. p.

[568] Carl Wagner: Anscombe's paradox and the rule of three-fourths. *Theory and Decision*, 15. évf. (1983) 3. sz., 303–308. p.

[569] Carl Wagner: Avoiding Anscombe's paradox. *Theory and Decision*, 16. évf. (1984) 3. sz., 233–238. p.

[570] Douglas N. Walton: Logical form and agency. *Philosophical Studies*, 29. évf. (1976), 75–89. p.

[571] Douglas N. Walton: St. Anselm and the logical syntax of agency. *Franciscan Studies*, 36. évf. (1976), 298–312. p.

[572] Douglas N. Walton: On the logical form of some commomplace action expression. *Grazer Philosophische Studien*, 10. évf. (1980), 141–148. p.

[573] Peng-Hsiang Wang: Anderson's reduction and kelsen's normativism. In Arend Soeteman (szerk.): *Pluralism and Law* (konferenciaanyag). 2004, 104–110. p.

[574] Max Weber: *Economy and Society*. 1978, University of California Press.

[575] Max Weber: *Gazdaság és társadalom I. A megértő szociológia alapvonalai*. 1987, KJK.

[576] Daniel M. Wegner: *A tudatos akarat illúziója*. 2009, Kossuth Kiadó.

[577] Daniel M. Wegner – John A. Bargh: Control and automaticity in social life. In D. Gilbert – S. T. Fiske – G. Lindzey (szerk.): *Handbook of social psychology*. 1. köt. 4th. kiad. 1998, McGraw-Hill, 446–496. p.

[578] Leif Wenar: Rights, 2005 (2011).
URL `http://plato.stanford.edu/entries/rights/`.

[579] S.A. West – A. Gardner: Altruism, spite and greenbeards. *Science*, 327. évf. (2010), 1341–1344. p.

[580] Stuart A. West – Claire El Mouden – Andy Gardner: Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32. évf. (2011), 231–262. p.

[581] A.N. Whitehead: *An Introduction to Mathematics*. 1911, Henry Holt and Company.

[582] G.S. Wilkinson: Reciprocal food-sharing in the vampire bat. *Nature*, 308. évf. (1984), 181–184. p.

[583] Timothy Williamson: Converse relations. *The Philosophical Review*, 94. évf. (1985) 2. sz., 249–262. p.

[584] Edward O. Wilson – David Sloan Wilson: Rethinking the theoretical foundation of sociobiology. *Q Rev Biol*, 82. évf. (2007) 4. sz., 327–348. p.

[585] Edward.O. Wilson: *Sociobiology: the new synthesis*. 1975, Harvard University Press.

[586] Peter Winch: Understanding a primitive society. In Bryan R. Wilson (szerk.): *Rationality*. 1970, Blackwell, 78–111. p.

[587] Jonathan Wright: Altruism as a signal: Zahavi's alternative to kin selection and reciprocity. *Journal of Avian Biology*, 30. évf. (1999) 1. sz., 108–115. p.

[588] Larry Wright: Functions. *The Philosophical Review*, 82. évf. (1973) 2. sz., 139–168. p.

[589] H. P. Young: Social choice scoring functions. *SIAM Journal of Applied Mathematics*, 28. évf. (1975) 4. sz., 824–838. p.

[590] H. Peyton Young: An axiomatization of borda's rule. *Journal of Economic Theory*, 9. évf. (1974), 43–52. p.

[591] Amotz Zahavi: Mate selection: Selection for a handicap. *Journal of Theoretical Biology*, 53. évf. (1975) 1. sz., 205–214. p.

[592] Amotz Zahavi: Altruism as a handicap: The limitations of kin selection and reciprocity. *Journal of Avian Biology*, 26. évf. (1995) 1. sz., 1–3. p.

[593] Robert B. Zajonc: Érzelem és gondolkodás (a preferenciák nem igényelnek következtetéseket). In György Hunyady (szerk.): *Szociálpszichológiai tanulmányok*. Budapest, 1991, Tankönyvkiadó, 7–55. p.

[594] Robert B. Zajonc: A puszta kitettség attitűdre gyakorolt hatásai. In *Érzelmek a társas kapcsolatokban és megismerésben*. 2003, Osiris Kiadó, 45–98. p.

[595] Berislav Žarnić: Dynamic models in imperative logic (imperatives in action: Changing minds and norms). In A. Brożek – J. Jadacki – B. Žarnić (szerk.): *Theory of Imperatives from Different Points of View* (konferenciaanyag). 2011, 60–120. p.

[596] Zsolnai László: *Boldogság és gazdaság: A buddhista közgazdaságtan eszméi*. 2010, Typotex.

[597] Gary Zukav: *The Dancing Wu Li Masters. An Overview of the New Physics*. 1984, Bantam Books.