

JANKY BÉLA

Összjáték  
Erkölc és racionalitás



JANKY BÉLA

# Összjáték Erkölc és racionalitás

Egy kutatási program regénye

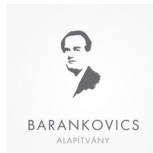
Gondolat Kiadó  
Budapest, 2019

A kötet megjelenését  
a BME GTK Szociológia és Kommunikáció Tanszék



és a

Barankovics István Alapítvány



támogatta.

© Janky Béla, 2019

Minden jog fenntartva. Bármilyen másolás, sokszorosítás,  
illetve adatfeldolgozó rendszerben való tárolás  
a kiadó előzetes írásbeli hozzájárulásához van kötve.

[www.gondolatkiado.hu](http://www.gondolatkiado.hu)  
[facebook.com/gondolat](https://facebook.com/gondolat)

A kiadásért felel Bácskai István  
Szöveggondozó Gál Mihály  
Tördelő: Sörfőző Zsuzsa

ISBN 978-963-693-935-9

# Tartalom

Köszönetnyilvánítás	9
Előszó a magyar kiadáshoz	13
<b>I. ERKÖLCS ÉS RACIONÁLIS DÖNTÉS: A KÉRDÉSEKTŐL A VÁLASZOKIG</b>	17
<b>I.1. A vizsgálódás tárgya</b>	17
I.1.1. Egy hétköznapi rejtély	17
I.1.2. Az erkölcsös viselkedés	19
I.1.3. Az erkölcsi motiváció forrása	23
<b>I.2. A közelítésmód</b>	24
I.2.1. Nagy adat vagy nagy elmélet?	24
I.2.2. Melyik elmélet?	29
<b>I.3. Honnan hová jutunk?</b>	34
I.3.1. A szituációs logikától a pszichológiai modellek felé	35
I.3.2. Vissza egy komplexebb szituációs logikához	41
I.3.3. Összjátékok, eredmények	44
<b>II. 1944 – A RACIONÁLIS POLGÁR ERKÖLCSE</b>	46
<b>II.1. Az önérdék</b>	46
II.1.1. Néhány szó a(z elmélet)történeti kontextusról	46
II.1.2. A racionális döntési modell a közgazdaságtanban	49
II.1.3. Önérdék a strukturalista elemzésben	57

<b>II.2. Mások érdeke</b>	60
II.2.1. Játékelmélet és mások érdekeinek figyelembevétele	60
II.2.2. A Nash-egyensúly	63
II.2.3. Első példázat: egy csodálatos elme csajozási tanácsai	65
II.2.4. Mások érdeke a strukturalista elemzésben: az új gazdaságzociológia	76
<b>II.3. Nyitva maradt kérdés: tudomány-e a döntés logikája?</b>	78
II.3.1. A racionális döntésemélet kettős természete	78
II.3.2. A Fradi vezérszurkoló kettős természete	81
<b>III. 1965 – AZ ÖNZŐ POLGÁR ERKÖLCSE</b>	84
<b>III.1. Önzés és önzetlenség: az erkölcsi dilemma létrehozása</b>	84
III.1.1. A társadalmi háttér	84
III.1.2. A fogolydilemma	86
<b>III.2. Jössz te még az én utcámba!</b>	93
III.2.1. Kétszemélyes játszma	93
III.2.2. Sokszemélyes játszma	103
<b>III.3. Nyitva maradt kérdés: ki jön az én utcámba?</b>	
<b>Az önző szavazó</b>	112
III.3.1. Egy kihívás	112
III.3.2. Sok matek és kevés eredmény	113
<b>IV. 1984 – A TISZTESSÉGES POLGÁR ERKÖLCSE</b>	117
<b>IV.1. Kezdetben vala az szemet szemért paranca</b>	117
IV.1.1. Választás az egyensúlyok közül	117
IV.1.2. Irány a labor!	122
<b>IV.2. Makacs mítoszok és makacs tények</b>	126
IV.2.1. Az egalitarianizmus mítosza	126
IV.2.2. Erős reciprocitás	132
IV.2.3. A méltányos szavazó	140

<b>IV.3. Nyitva maradt kérdés: méltányos döntések, vagy tisztességtelen értelmezések?</b>	145
IV.3.1. Racionális-e a méltányos?	145
IV.3.2. Amikor az élet betör a laborba	149
IV.3.3. Altruista büntetés, végállomás?	154
<b>V. 2001 – AZ ÖSZTÖNÖS POLGÁR ERKÖLCSE</b>	156
<b>V.1. Az erkölcs új pszichológiája</b>	156
V.1.1. A morális alapok elmélete	156
V.1.2. Az undor	159
<b>V.2. Ösztönök és társadalmi attitűdök</b>	164
V.2.1. Kilépés a hétköznapi interakciók világából: ösztönök és politikai ideológia	164
V.2.2. Visszavonulás a védelmi vonalak mögé: undor és etnikai diszkrimináció	171
<b>V.3. Nyitva maradt kérdés: az ösztön a végső mozgató?</b>	177
V.3.1. Most akkor jobbra vagy balra?	177
V.3.2. Az oksági irány megkérdőjelezése	182
<b>VI. 2006 – A NORMAVEZÉRELT POLGÁR ERKÖLCSE</b>	185
<b>VI.1. Vissza a Nash-egyensúlyhoz</b>	185
VI.1.1. Mit tanultunk eddig?	185
VI.1.2. Mit adtak nekünk a matematikusok?	188
VI.1.3. Egyensúly elérése csendes adaptáció révén	191
<b>VI.2. A normák koreográfiája</b>	194
VI.2.1. Koreográfusok és korrelált egyensúlyok	194
VI.2.2. Harmadik példázat: nyulak a delphoi jósdában	196
VI.2.3. A szociológiai koncepciók felé	201
VI.2.4. A koreográfiák koreográfiája	207
VI.2.5. Nem megyek a te utcádba! A normavezérelt szavazó	217
<b>VI.3. Nyitva maradt kérdés: a normakövetés rejtélye</b>	221
VI.3.1. A korrelált egyensúly alkalmazásának problémái	221
VI.3.2. Egyensúly vagy normatív ösztönök?	228

<b>VII. KONKLÚZIÓ: A BEÁGYAZOTT ÖNZŐ POLGÁR ERKÖLCSE</b>	236
<b>VII.1. Egy döntés, több játszma: javaslat egy új     modellezési keretre</b>	236
VII.1.1. A modell alapjai	236
VII.1.2. Erkölcs és racionális döntés: a definíciók	242
<b>VII.2. Erkölcs alapú érdek, érdek alapú erkölcs</b>	246
VII.2.1. Függőség, identitás, erkölcs	246
VII.2.2. Az erkölcs hatalma	250
VII.2.3. A hatalom erkölce?	253
<b>VII.3. Nyitott kérdés (?): a morális indulat</b>	256
VII.3.1. Érdekek és érzelmek	256
VII.3.2. Érzelmek, érdekek és normatív erő	260
<b>VIII. EPILÓGUS: A FIAMRÓL, A NORVÉG ÍRÓSZTÁRRÓL, ÉS A TUDOMÁNYTALAN TUDOMÁNYOK FONTOSSÁGÁRÓL</b>	264
Az epilógus epilógusa	268
<b>Hivatkozások</b>	269
<b>Névmutató</b>	301



## Köszönetnyilvánítás

Mindenekelőtt köszönetet szeretnék mondani kollégámnak, Szakadát Istvánnak, aki a BME tanszékvezetőjeként sabbatical félévet adott nekem e könyv befejezésének idejére. Akik ismernek, azok tudják, hogy feletteseimet nem minden nap dicsérem szembe (a többi kedves olvasó ugorja át a következő mondatot!): István nemcsak kolléga, de harcostárs is a társadalmi cselekvések racionális megértését célzó alap kutatások létjogosultságáért folytatott küzdelemben, amelyben itt a „racionális”, ott a „megértés”, amott pedig az „alapkutatások” legitimitását kell védenie egy fájóan fogyatkozó csapatnak. Köszönöm továbbá Csigó Péternek a helyettesítést legfontosabb kurzusomon a könyv írása alatt, ami igen nagyvonalú vállalás. Nagyon bízom közös projektünk sikerében, és már eddig is sokat tanultam tőle. Továbbá itt, az elején kell végtelen hálámat kifejeznem Ágának, Csillának és Zsombornak azért, hogy egy ideig teljes erőbedobással saját mániámmal tudtam foglalkozni. Remélem, nem bánják meg nagyvonalúságukat!

Külön köszönöm Bognár Bulcsú, Csigó Péter, Hamp Gábor, Kisfalusi Dorottya, Kocsis János Balázs, Lakatos Zoltán, Lengyel György, Szakadát István, Takács Erzsébet, Tardos Róbert és Wessely Anna tanácsait és észrevételeit, akik kézirat első változatához szóltak hozzá szóban és/vagy írásban. A megjegyzések nyomán sok helyen módosítottam a megfogalmazáson, és a kéziratvita különösen sokat segített a cím és az I. fejezet megformálásában. Az elkészült kézirat magyarítását Márton Róza Krisztina végezte el – nyelvi és szakmai hozzáértéssel,

és az olvashatóság eléréséhez szükséges bátorsággal. Az ő munkáját is nagyon köszönöm. Köszönettel tartozom továbbá a Gondolat Kiadó munkatársának, Csizmadia Nórának, aki a második körös olvasószervezési feladatokat végezte el.

Fontosak a könyv megírásához nyújtott közvetlen támogatások. Azonban az elemzés huszonöt év kutatómunka tapasztalatait hasznosítja. A tanulmányom alapjait jelentő módszertani individualista közelítéssel és a racionális döntésemeléttel Szántó Zoltán és Helmich Dezső kurzusán ismerkedtem meg még a kilencvenes évek elején a „Közgáz” Rajk Szakkollégiumban. Nagyon fontosak számomra ezek az alapok. Zoli pedagógiai és tudományszervezési tevékenysége, amellyel ezt az irányzatot népszerűsítette itthon a kilencvenes években, külön kiemelendő. A metaelméleti alapokról és a (társadalom)-tudományhoz való hozzáállásról Bertalan László legendás kurzusain próbáltam minél többet megtanulni. Bertalan tanár urat méltatlan ilyen röviden méltatni – nagyon jelentős kutató és pedagógus volt, a könyvben bemutatott kutatási program hazai művelésének megalapozója. Itt kell megemlékezni szakkollégiumi diáktársamról, a már diákként is jelentős közgazdász pedagógus Eső Péterről is, aki sokkal többre tanított minket, mint a játékelmélet technikai alapjai.

A könyvben összefoglalt elméleti kutatásaimat részben közpolitikai attitűdök megértése motiválta. Nem magam találtam ezt a témát, hanem örökös témavezetőmtől, Lengyel Györgytől kaptam, aki tudományszervezőként akkor (is) azon dolgozott, hogy összehangolja a hazai szociológia elmélettörténet-orientált hagyományát a kurrens nemzetközi tudományos diskurzushoz köthető módszerekkel és témákkal. Kutatóként, és később vezetőként is komolyan befolyásoltak a szociológus feladatára és a magyar szociológia pozicionálására vonatkozó elvei. Lengyel György a kéziratot is mintegy témavezetőként széljegyzetelte végig, és az alábbi elemzés megfogalmazásaira gyakorolt közvetlen hatása is jelentős.

Hálával tartozom azoknak is, akik már részben az én koncepcióimhoz igazodva tették saját jelentős tudásukkal értékesebbé e könyvhöz kötődő elméleti és empirikus munkáimat: Bakó Boglárkának, Bognár Adriennek, Kopasz Mariannak, Rajkó Andreának és Varga

Dánielnek. E könyvben felhasználom mindazt, amit a jóléti preferenciák közgazdasági megközelítéséről Gál Róberttől és Simonovits Andrástól tanultam – hálás vagyok nekik is.

A külön fejezetet kapó morális pszichológiára Tardos Róbert hívta fel a figyelmem. Az ő műveltségéből és gondolataiból közvetetten is profitálok, amikor tanítványaival működöm együtt e könyv számára is fontos kutatásaimban. Közülük Lakatos Zoltán segített a legtöbbet.

Minden szakmai, családi és vezetői támogatás mellett sem indulhattam volna el a pályán, ha Takács Károly nem húz be a szakkollégiumba, és nem ad mércét a tudományos munkában. Már a szakkollégiumban megtapasztaltam, hogyan stimulálja az embert a szakmailag igényes baráti társaság, és e könyv megírásakor a Károly által létrehozott MTA Recens kutatócsoport biztosította e stimulációt. Ilyen „peer-pressure” a *Módszeresen* című beszélgetéssorozat stábjából is ér, Gárdos Judit és Németh Renáta részéről.

Könyvemben ugyanakkor fontos szerepet kap az „egzakt társadalomtudományok” egyes képviselőinek ifjútörök hevületével szembeni kétkedés, és a visszakanyarodás a szociológiai szemléletmódhoz. Ebben szerepe lehetett azoknak a kollégáknak, akik az általam alkalmazottól eltérő szociológiai hagyományokat képviselnek hitelesen, és mindig nyitottak a tudományt éltető párbeszédre. Közülük két kollégám hatása kötődik a könyv gondolatmenetéhez. Wessely Anna már ifjútörök diákkoromban lenyűgözött, amikor egy szakkollégiumi kurzus udvariatlan vendégeként értő alaposággal mutatott rá a racionális döntésemélet Achilles-sarkára. A könyv részben az általa anno felvetett kérdésekre adott válasz. Neményi Mária pedig – aki mindig eljön, meghallgat és kérdez – arra figyelmeztet, hogy a tudomány nem beauty contest, és mi egy közösség vagyunk, amelynek feladata van; és a kutatási módszerekkel és társadalmi funkciókkal kapcsolatos nézeteltéréseink másodlagosak ahhoz képest, hogy *közösségként* valamilyen *feladatunk* van.

A társadalomtudós a társas életből a szakmáról is tanul. Én mindenképpen tanultam. Meg nem mondom, hogy ki az, aki megtanított arra, hogy a jó nem abból születik, ha kérdés nélkül átcibáljuk a vak

nénit a zebrán, hanem abból, ha képesek vagyunk odafigyelni embertársaink szándékaira és a társas élet szabályaira. Ez a tézis a könyv talán legfontosabb tartalmi mondanivalója, huszonöt éves szakmai útkeresésem aktuális konklúziója. Köszönet annak, akit illet.

A végére egy távoli kolléga maradt, aki tapasztaltabb szakkollégistaként berakott a fent említett Bertalan László kurzusára, később pedig beajánlott a pályámon szintén meghatározó Kemény Istvánhoz. Kézdi Gábor, aki az egyik legfontosabb szakmai referenciapontom, később is sokat segített. Eddig ezt nem tudtam meghálálni, máig sem tudom, hogyan tehetném. Legyen ő az, akinek ajánlom ezt a könyvet.

## Előszó a magyar kiadáshoz

Meggyőződésem, hogy a társadalomtudomány művelése (is) nemzetközi vállalkozás, és a különböző témákra fókuszáló kutatási programok „láthatatlan intézetei” nem ismernek országhatárokat. Régóta azt az álláspontot képviselem, hogy eredményeinket meg kell osztani a nemzetközi tudományos közösséggel. Praktikusan ez ma azt jelenti, hogy meg kell írni angolul azt, amire jutottunk, és meg kell próbálni eljuttatni olyan disszeminációs fórumokra, amelyeket a láthatatlan intézetünk tagjai valóban figyelnek. E könyvben is beszámolok olyan korábbi elemzéseimről, melyek sokak által olvasott nemzetközi folyóiratokban jelentek meg, utalok olyan kutatásokra is, amelyek eredményeit mostanában próbálom eljuttatni a szélesebb nemzetközi olvasóközönséghez, továbbá prezentálok olyan téziseket, amelyeket a közeljövőben kívánok angol nyelven is papírra vetni. De ezt a könyvet magyarul írtam, az itthoni kollégáim, a hallgatók és az érdeklődő tágabb közönség számára. Hogy miért?

Elmélettörténeti kutatásokból tudjuk (a témánkban lásd például Fontaine 2006), hogy az általunk kanonizált irodalomként olvasott szakcikk jelentős része nem izolált körülmények között született. A kutatók beszélgetnek közeli kollégáikkal szóban, és írásban is. A cikkek egy része is e beszélgetésekben elhangzó állítások letisztázott kivonata. Jómagam két, egymástól sok tekintetben különböző ethoszú és szerveződésű kutatási területen (romakutatás és hálózatelemzés) is megtapasztaltam az elmúlt két évtizedben, ahogy a téma iránt érdeklődő hazai kutatók lokális szóbeli és írásbeli eszmecseréje és ak-

tív oktatási tevékenysége kiforrja magából azokat az eredményeket, amelyek végül érdemben képesek hozzájárulni az adott téma nemzetközi tudományos diskurzusához is.

Ezt a könyvet, a fentiek figyelembevételével, a nemzetközi diskussziót előkészítő hazai beszélgetések stimulálására, valamint a (tovább)képzési tevékenység támogatására írtam. A konkrét nemzetközi és hazai kapcsolódási pontokról az I. fejezetben lesz szó.

A kötet egyes elmélettörténeti áttekintései segítséget nyújthatnak az adott résztema iránt érdeklődő, de irodalmában még nem jártas kutatótársaknak. Például itthon csak mostanában kaptak lendületet a viselkedési közgazdaságtan paradigmáját követő, kísérleteken nyugvó kutatások. A kapcsolódó módszerekkel és elméletekkel ismerkedő kollégák számára hasznos lehet a társas preferenciákra vonatkozó empirikus kutatások kritikai áttekintése (IV. fejezet). Az értékekutatással foglalkozó szociológus kollégák körében pedig egyre nő az érdeklődés egyes újabb elméletek, kiemelten Jonathan Haidt erkölcspszichológiája iránt. Számukra e terület eredményeinek kritikai elemzése lehet érdekes (V. fejezet). De más kutatási témák iránt érdeklődők is találhatnak a könyvben számukra érdekes információkat.

Az elméletek fejlődését bemutató narratívában több ponton szünetet tartok, hogy egy-egy fontos koncepciót vagy tézist a legegyszerűbb játékelméleti eszközök segítségével tegyek világossá. Ezeket a példákat én is használni szeretném az oktatásban; és ezt javaslom társadalomelméleti kurzusokat vezető kollégáimnak is.

A könyv azonban *nem* korábban megjelent áttekintő esszék és egyetemi jegyzetek gyűjteménye, hanem az első szótól az utolsóig a kötetet átszövő gondolatmenetet szolgáló, frissen készült szöveg. A teljes narratíva végigkövetését négy, jól elkülöníthető célcsoportnak is javaslom. Egyrészt, a társadalomelmélet iránt érdeklődő kollégáimnak szeretném megmutatni, hogy tévesek a racionális döntéselmélettel kapcsolatban közszájon forgó (szociológiai) vélekedések, melyek szerint e paradigma inherensen alkalmatlan az emberi cselekvések társadalmi beágyazottságának vizsgálatára. Továbbá, ajánlom a könyvet azoknak, a diszciplína általánosabb kérdései iránt érdeklődő közgazdászoknak, akik – miközben bíznak a neoklasszikus mikro-

ökonómia széles körű alkalmazási lehetőségeiben – szkeptikusak a játékelmélet „tudományos potenciáljával” kapcsolatban. Nekik többek között meg szeretném mutatni azt, hogy milyen fontos tartalmi állítás rejtőzik az egyértelmű előrejelzések számukra technikai problémának tűnő hiánya mögött. Mellettük ajánlom a könyvet azoknak a kvantitatív empirikus társadalomkutatóknak is, akik abban bíznak, hogy megfelelően nagy adatmennyiség megfelelően fényesre csiszolt statisztikai elemzésével, stabil elméleti háttér nélkül is érdemben gazdagíthatjuk a társadalom működésére vonatkozó ismereteinket. Számukra azt szeretném megmutatni, hogy a társadalmi jelenségekre vonatkozó empirikus eredmények elméleti háttér, és – ami talán még fontosabb –, intencionális magyarázatok nélkül olyan, korlátozott felhasználhatóságú adatsorok, amelyekről valójában azt sem tudjuk, hogy hol húzódnak felhasználásuk korlátai. Végül, ajánlom e tanulmányt minden olyan, közügyek iránt érdeklődő embernek, akit érdekel, hogy baráti társaságoknál nagyobb csoportokban, különösen a modern nagyméretű társadalmakban lehetséges-e elvileg „jó” közösséget építeni: olyat, ahol „jó” emberek egymást nem bántva, hanem inkább segítve küzdenek saját és közös céljaikért. Igyekeztem olyan mankókkal megtámasztani a gondolatmenetet, amelyek a szakmai kérdések részleteiben kevésbé jártas olvasókat is támogatni tudják a történet ívének követésében.

A könyvben több ponton esik szó olyan koncepciókról és tézisekről, melyek részletes elemzéséhez komoly formális matematikai eszköztárat használtak az általam idézett kutatók. Ezek közül a narratíva szempontjából legfontosabb fogalmak és gondolatok ismertetésére vállalkozom, egészen addig a pontig, ameddig a formulák nyelvének alkalmazása elkerülhető. E törekvést részben a megcélzott olvasóközönség „formula-nyelvtudásának” korlátai, részben pedig a verbális nyelv matematikai társadalomtudományok képviselői által néha lebecsült lehetőségei magyarázzák. Egy döntően matematikai kutatási irányzat verbális kifejtése miatt a könyv alcíme lehetne „*Matematikai szociológiaelmélet bölcsészeknek*” is.

E módszertani döntéshez kapcsolódik az előszót záró idézet. 1962-ben David Gale és Lloyd Sahpley a párosítási algoritmusokról írt

cikket a *The American Mathematical Monthly* című folyóiratba (itt közölt matematikai levezetések Sahpley 2012-ben elnyert közgazdasági Nobel-díját alapozták meg). Tanulmányuk összegzéséből az alábbiakat ajánlom az olvasók figyelmébe: „A legtöbb matematikus már legalább egyszer került olyan helyzetbe, ahol azt a nézetet próbálta megcáfolni, hogy a matematikusok a »képletek emberei«, akik »jó sok formulát ismernek«. ...Ilyen helyzetekre [...] ajánljuk az I. tételünk állítását és bizonyítását. Az érvelés nem használ matematikai formulákat, egyszerű angol nyelven íródott. ...Bármilyen elégségesen precíz érvelés matematikai érvelés.”\*

Szentendre, 2019. február 7.

\* A szerző fordítása.



# I. Erkölcs és racionális döntés: a kérdésektől a válaszokig

## I.1. A vizsgálódás tárgya

### I.1.1. EGY HÉTKÖZNAPI REJTÉLY

Egyszer, sok évvel ezelőtt egy, a világ dolgai iránt érdeklődő, de a modern társadalomtudományokban kevésbé járatos barátom megkért, hogy mondjam el, mivel foglalkozom, kutatóként mi érdekel leginkább. Azt feleltem, hogy annak a jelenségnek a magyarázatát keresem, amikor egyes autósok fénykürttel (villogással) jelzik a szembejövőknak, ha azok ideiglenesen felállított sebességmérő radar felé közelednek. Vajon mi motiválja erre az embereket? Sok év elteltével is emlékszem a példára, mert a barátom is emlékszik rá, ahogy arra is, hogy ez számára világos leírása volt az engem érdeklő elméleti problémának.

Most utólag én is azt gondolom, hogy az akkor hirtelen előkapott, konkrét példám jó illusztrációja annak a problémának, amely végigkísérte eddigi pályámat, és amely e könyv központi témája is. A példa tehát a következő: az autós elhalad a szemközti sávot pásztázó rendőrségi radar mellett, majd a radar felé közeledő szembejövőknak a fényszórója villogtatásával jelez. Ez nem kivételes jelenség hazánkban (tehát nem a normális viselkedéstől eltérő „hiba”). De nem is olyan mindenki által ismert, a fejünkbe vert, naponta sokszor követett szabály, mint a piros lámpánál megállás, amelyet látszólag értelmetlen helyzetekben is betartunk (például késő éjjel, teljesen üres, belátha-

tó kereszteződésben, ismerős helyen) – ezek az esetek a megszokással magyarázhatóak. Az öntudatlan utánzás sem jellemzi e viselkedésformát, azaz nem úgy terjed, mint az ástás. A villogás tudatos döntés, amelyhez ki kell zökkennünk a vezetés ritmusából, figyelmünket meg kell osztani. Hiába tűnik kis költségű cselekvésnek, a tapasztalat azt mutatja, hogy jól magyarázható ok nélkül nem csinálunk ilyen vezetésből kikölkentő dolgokat. Ráadásul a közlekedést zavaró villogás büntethető is, bár a büntetés kockázata kétségtelenül csekély.

A biológia, a közgazdaság-tudomány és a politikatudomány már sok évtizede megfogalmazta azokat a faktorokat, amelyek altruista viselkedésre sarkallhatják az élőlényeket. Ebben a helyzetben azonban nehéz fellelni e tényezőket. Tipikus esetben nem ismerjük a szembejövőket, és ők sem ismernek minket, semmi közünk sincs egymáshoz. Gyakorlatilag nulla az esélye, hogy közvetlen viszonzást várhassunk a villogás gesztusáért. Fontos ebből a szempontból, hogy autóban ülünk, nem látjuk rendesen a másikat. Tehát olyan ösztönök sem léphetnek működésbe, amelyeket a kedvezményezettek személyes jelenléte, a szemtől szembeni interakció hív elő. Továbbá a fénykürt használata tudatos döntéssel és pszichológiai költséggel jár, ugyanakkor nincs benne semmi heroikus, szórakoztató vagy izgalmas mozzanat, amely a cselekvő számára hozzáadott értéket jelenthetne. Tehetségünket biztos nem tudjuk vele demonstrálni, még saját magunk számára sem. A cselekvés következményei ugyanakkor kétesek. Az biztos, hogy nincs olyan kézzelfogható haszna, amely a villogó autós-nál realizálódik, és emiatt megérné azt a pár másodperc kikölkentést. Jelentősebb költségei ugyanakkor lehetnek, a villogás esetleges büntetésén túl is: például megzavarhatja valamelyik szembejövőt, vagy a vezető figyelmét vonhatja el a villogás egy hirtelen bekövetkező vészhelyzetről. Az egyetlen haszna egyúttal ellenérv is a villogással szemben: lelassítjuk a gyorsajtásra hajlamos, és ezért veszélyes autósokat. Ugyanis, ha mások gyorsajtása miatt aggódunk, nem akkor kellene villognunk, amikor radar mellett megyünk el, hanem rava-szul akkor, amikor a közlekedésre veszélyesnek látjuk a gyorsajtást, a radar láttán pedig nem kellene jeleznünk, és örülnünk kellene, ha leleplezik a veszélyes gyorsajtókat. Ráadásul a helyzet elemzését és

megértését tovább komplikálja, hogy ez a szokás nem univerzális jelenség a motorizált társadalmakban, csak néhány országban figyelhető meg.

E könyv megírását az ilyen nehezen magyarázható jelenségek motiválták. E jelenségek arra utalnak, hogy létezik bennünk egy olyan *késztetés* (Syi 2014), amely mások javával törődik (akár pozitív akár negatív előjellel). E késztetés megértésének vágya vezérelt számos kutatásomban az elmúlt huszonöt évben. A saját pályáivemnél azonban sokkal fontosabb, hogy egy viszonylag jól körülhatárolható „láthatatlan intézet” kutatói legalább hetvenöt éve nyomoznak szisztematikusan e motiváció után. A nyomozás egymást követő eredményei egymásra épülnek; a kutatók a felmerült ötletek cáfolataiból ugyanúgy igyekeznek tanulni, mint a korábban elfogadott tények pontosításáiból. E hosszú nyomozás eredményeiről szeretnék beszámolni ebben a könyvben.

### I.1.2. AZ ERKÖLCSÖS VISELKEDÉS

A társadalomtudományok területén is vannak olyan kutatási témák, amelyek fogalmilag jól lehatárolhatóak, és tartalmuk is állandó. Ilyen jól lehatárolható jelenség a nők demográfiai magatartása, ezen belül a szülések száma. Emellett vannak olyan jelenségek és intézmények, amelyek kutatói megnevezése állandó, de tartalma történetileg változik. Például ilyen a család. Végül vannak olyan kutatási témák, amelyek egy univerzális jelenség vizsgálatát célozzák, de a kutatási folyamat során időben és térben változik a jelenség pontos lehatárolása. Ilyen vizsgálati téma például a könyvben is központi szerepet játszó racionális döntés.

A könyv központi témája azonban egy olyan jelenség, amelynek a kutatási folyamat során változó tartalmához változó elnevezések is kapcsolódnak. Olyan cselekvéseket vizsgálok e könyvben, amelyek során a cselekvők valamilyen módon figyelembe veszik mások érdekeit illetve elvárásait. Különösen az olyan cselekvések érdekelnek, amelyek során nemcsak egy konkrét személyre figyelünk, ha-

nem egy nagyobb csoport érdekeit és/vagy elvárásait próbáljuk tekintetbe venni. Ilyen típusú cselekvések vizsgálatakor a klasszikus politikai filozófia *morális érzületekről* beszél (Smith 2010 [1759]). A modern társadalomtudomány megpróbálta jobban specifikálni az emberi döntések e típusát. A klasszikus szociológiai irodalom gyakran *normakövető viselkedésre* utal (Durkheim 2003 [1897], Parsons 1951). A biológia, a pszichológia és a közgazdaságtan az önzés ellentétéként, az önzetlenség (*altruizmus*) koncepcióját alkalmazta a múlt század hatvanas éveitől (a fogalmak tömör katalógusát kínálja Crawford és Ostrom 1995; magyarul lásd Bereczkei 2009, Syi 2009). Ez utóbbi diszciplínák képviselői idővel kiszélesítették a nem teljesen önző viselkedés koncepcióját. Ennek eredményeképp ma már inkább *társas preferenciákról* vagy *társas orientációról* beszélnek (Charness és Rabin 2000, Varnum et al. 2010).

A morális érzületek fogalmi és tartalmi lehatárolásai többek között a morál forrása alapján próbálják meg pontosabban lehatárolni ezt a cselekvési kört (lásd a következő alfejezetet). A normakövetés koncepcióját alkalmazó tudósok alapvetően a cselekvőn kívül keresik a morál gyökereit, az altruizmusra és a vele rokon koncepciókra építő kutatók pedig az emberi pszichén belül határozzák meg a morál forrását (Simpson és Willer 2015). Könyvemben a kezdeti, megengedőbb koncepcióhoz (morális érzületek) nyúlok vissza, ugyanis elemzésemben a morál forrásának helye nem alapfeltevés, hanem kérdés.

Bár a vizsgálat tárgyát képező cselekvéseket a téma vizsgálatának egyes pontjain altruistának fogom nevezni, máskor pedig normavezéreltnek, jelezve az adott vizsgálati fázis feltevéseit; de a kötet összességében az *erkölcsös* magatartás koncepcióját (Stets és Carter 2012, Tavory 2011) vizsgálja. E döntésnek jelentősége van, de értelme a gondolatmenet kifejtése során, fokozatosan válik – reményeim szerint –, világossá. A kérdésben az olvasók türelmét kérem; az utolsó fejezetben, a megszerzett ismeretek felhasználásával, összegezni fogom e terminus használatának indokait.

A könyvben tudatosan használom ki a magyar nyelv azon ritka, de nem egyedülálló sajátosságát, hogy azonos, vagy nagyon hasonló jelentéssel használja a morál és erkölcs szavakat a viselkedést befo-

lyásoló tényezők egy bizonyos típusára. A két fogalomnak ugyanis hiába azonos a denotációja, a szavak eltérő konnotációkat hordoznak. A morál kifejezés konnotációi viszonylag szorosan kapcsolódnak a politikai filozófia fent említett hagyományaihoz. Az V. fejezetben jutok el odáig, hogy meghaladom a morál politikai filozófiai értelmezését, és kiszélesítem olyan értelemben, ahogy a hétköznapiokban használják az erkölcsösség fogalmát (vö. Haidt 2008). E meghaladást segítik elő az erkölcs hétköznapi fogalmához kötődő konnotációk.

Úgy érzem, a morál konnotációi a fogalmat a nagy közösségekben értelmezhető közjó eléréséhez kapcsolják – sokkal inkább mint az erkölcs fogalmát. Globális gondolkodást és erős elvi megalapozottságú cselekvést juttat eszünkbe. Ezzel szemben az erkölcs konnotációi jobban kapcsolódnak hétköznapi, az illendőség határaitra vonatkozó megérzésekhez. A politikai ellenállásnak nyelvünkben inkább morálja van, míg az öltözködési szokásoknak inkább erkölcse. Elemzésemben ugyanolyan fontosak a politikai-ideológiai gyökerű erkölcsi döntések, mint az öltözködési szabályok. Az erkölcsi érzületek irodalmának fókuszában azonban a morál konnotációinak irányába torzít (Haidt 2008), és a fókusz szélesítésének szándékát az erkölcs kifejezésének használatával jobban tudom hangsúlyozni.

Gondolatmenetem befogadását remélhetően segítik (vagy legalábbis nem gátolják) az alkalmazott fogalmak hétköznapi használatához kötődő konnotációk. Elemzésemet azonban nem e konnotációkra építem. Az erkölcsi jelenségek vizsgálatának konklúziójaként az erkölcsös viselkedés és -cselekvés pontos, tudományos célú meghatározását fogom adni a VII. fejezetben. Íme a koncepció vázlatos körvonalazása (vö. Tavory 2011):

- a) az erkölcs az itt használt értelmében nem univerzális, az erkölcsös viselkedés és -cselekvés mindig egy konkrét csoporthoz és kontextushoz köthető. Továbbá
- b) minden viselkedés, amely egy szituációra vonatkozó, az adott csoportban elismert erkölcsi szabályokat követ, erkölcsös viselkedésnek tekinthető.
- c) Az erkölcsös viselkedés egyúttal bizonyosan tudatos erkölcsös cselekvés/döntés is, ha a cselekvő egy helyzetben azért viselkedett erkölcsö-

*sen, mert figyelembe vette az említett csoport tagjainak későbbi, más interakciókban várható reakcióit. Azaz e későbbi reakciók nélkül máshogy viselkedett volna a szituációban.*

Másként fogalmazva: egy szituációban az erkölcsös cselekvő az elismert erkölcsi szabályokat követve, de az adott helyzet ösztönzőivel szemben, csoporttársai későbbi reakcióit figyelembe véve dönt. E koncepció az elemzés folyamán lépésről lépésre épül majd fel.

Könyvem nem az erkölcsös emberről szól, hanem az erkölcsös polgárról. Ez fontos distinkció, ugyanis személyközi interakcióink jelentős része közvetlenül kötődik a genetikai állományunk átörökítését segítő tevékenységekhez. E tevékenységek mozgatórugóival a könyv nem foglalkozik. Ezzel szemben kifejezetten érdekelnek azok a helyzetek, amikor a személyközi viszonyokban még nyomokban sem lehet felfedezni a rokoni szálakat, és mindenképpen el kell vetni a társas viselkedés e kötelékekre hivatkozó biológiai magyarázatait. Ugyanakkor a könyv végén, a VII. fejezetben kiderül, hogy az erkölcsös polgári viselkedés motivációiról szerzett ismeretek jól használhatóak a családon belüli viszonyok megértéséhez is.

A közösség polgárának viselkedése nemcsak a civil élet hétköznapi interakcióit fedi le. Különös és fontos típusát alkotja az állampolgári aktivitásnak a politikai tevékenység. Nemcsak a ritka tüntetésen való részvétel, és egyéb mozgalmi aktivitás sorolható ide, hanem a politikai vélemények megformálása is. A közpolitikai attitűdök vizsgálatát nem akarom megkerülni; sőt, kiemelt helyen tárgyalom a könyvben bemutatott nyomozás eredményeinek ide vonatkozó eredményeit, ugyanis a közvélemény az életünk kereteit jelentősen befolyásoló politikai döntéshozatal fontos hivatkozási alapja. Ráadásul a közvélekedés jelentősége a 21. században inkább nő, mint csökken; és a politikai intézményrendszer közvetítő intézményeinek gyengülése felértékeli a döntéstámogató „mérésekben” megnyilvánuló polgári érzületek jelentőségét (például Csigó 2016). Ezek az attitűdök pedig nagyon sok esetben a szavazópolgárok számára ismeretlen emberek sorsa iránti aggodalmat tükröznek (beleértve a negatív érzületeket is).

## I.1.3. AZ ERKÖLCSI MOTIVÁCIÓ FORRÁSA

Az elemzés fókuszában az erkölcsös viselkedés áll, mindazonáltal nem az erkölcs sajátosságainak és típusainak bemutatására töreksem. A vizsgálódás célja az erkölcsös döntések *indítékainak* feltárása. Az indítékok feltárását végigkíséri az *erkölcs forrásának* lokalizálására vonatkozó kérdés. Nevezetesen az, hogy a cselekvőben, vagy rajta kívül álló erőkből kell-e keresni az erkölcsös döntésre ösztönző készleteket?

A fenti kérdésre adott válasz fontos választóvonalat képez az erkölcs kutatói között. A külső erők jelentőségét hangsúlyozó Durkheim és (elsősorban) szociológus követőinek munkássága élesen elkülönül a cselekvők természetét vizsgáló, és a durkheimi nézetekkel vitatkozó Weber, valamint az őt követő szociológusok, közgazdászok és pszichológusok munkájától (Simpson és Willer 2015).

Az erkölcs forrását az emberben kereső elméletek metaelméleti nézőpontból robusztusabbak, mint a kizárólag külső erkölcsi erőket feltételező modellek, ugyanis elvileg elképzelhető, hogy egy csoportban pusztán belső készletekre támaszkodva terjed el valamilyen erkölcsös viselkedési mintázat. Fizikailag azonban csak az egyének léteznek, ezért problémás az erkölcsös magatartásra készítő külső erők egyéni cselekvőkre történő hivatkozás nélküli modellezése (vö. Orthmayr 1997). Mindenképpen hiányos lesz az az elmélet, amely nem mondja meg, hogy mi teszi nyitottá az embereket külső erkölcsi készletek befogadására. Nem véletlen, hogy az erkölcs forrására vonatkozó elméleti viták szolgáltatnak fontos érveket azoknak, akik a legkülönbözőbb társadalomtudományi diszciplínákban a módszertani individualizmus mellett tették le a voksukat.

Ugyanakkor a társadalmi jelenségek empirikus vizsgálatai kézzelfogható bizonyítékot szolgáltatnak az erkölcs erős kontextusfüggőségére. Ezek az empirikus eredmények arra utalnak, hogy az egyes cselekvőn kívül álló környezeti tényezők lényeges szerepet játszanak az erkölcs formálódásában. Ráadásul a feltételezett belső erkölcsi készletek meggyőző empirikus azonosítása sokkal nehezebb, mint az erkölcs sokszínűségére vonatkozó eredmények felmutatása.

A fentiek alapján értelmetlennek tartom azt a kérdést, hogy az erkölcs forrása az emberben magában *vagy* az emberen kívül található-e. A kultúrák sokféleségéről szerzett tudományos ismereteink szerint egyes cselekvőn kívüli tényezők befolyásolják, hogy egy egyén egy adott helyzetben követ-e egy, a helyzetre alkalmazható erkölcsi szabályt. Jogos kérdés marad ugyanakkor, hogy belső késztetések is befolyásolják-e konkrét szabályok követését. Így az erkölcs forrásának lokalizálásával kapcsolatos termékeny kérdés a következőképpen fogalmazható meg: *Léteznek-e olyan univerzális, velünk született belső késztetések, amelyek specifikus erkölcsi szabályok betartására ösztönöznek minket?*

Az ilyen késztetések esetleges feltárása során arra a kérdésre is szükségképpen választ kapunk, hogy milyen jellegű erkölcsi szabályok követésére ösztönöznek ezek a belső késztetések. Ha nem találjuk ilyen belső erkölcsi késztetések nyomát, akkor arra kell választ találnunk, hogy mi vesz rá alapvetően „erkölcstelen”, racionális embereket arra, hogy bizonyos, akár érdekeikkel ellentétes szabályokat kövessenek.

A könyvben található elemzések ezekre a kérdésekre próbálnak választ találni.

## I.2. A közelítésmód

### I.2.1. NAGY ADAT VAGY NAGY ELMÉLET?

#### I.2.1.1. Adatkor

Az alábbi kötet az erkölcsös emberi cselekvésekre vonatkozó egyes alapvető, általános elméletek fejlődését taglalja. Azonban a könyv elméleti érvelése mögött egy metaelméleti mondanivaló is meghúzódik: napjaink „nagy adat” korszakában is szükség van a társas élet általános elméleteinek ismeretére és továbbgondolására (vö. Macy 2015b, Szántó és Syi 2010).

Sok jel utal arra, hogy a 21. században leáldozott a nagy társadalomtudományi elméleteknek, és beköszöntött a nagy adat korszaka



(Macy 2015a). Pontosabban az adatok nagy korszaka – mert távolról sem csak az ún. big data növekvő társadalomtudományi felhasználásáról van szó (Imai 2017). Ez alapvetően pozitív folyamat. A ma már nagyobb bőségben rendelkezésünkre álló adatok segítségével lehetővé válik, hogy számos, korábban erősen vitatott modellt előrejelzéseit mérni tudjunk. A szélsőségekre hajló, a különböző elméleteket gyakran feketén vagy fehéren látó, metaelméleti ellentétekhez gyakran visszakanyarodó vitákat felválthatják a modellek praktikus részleteire koncentráció, közvetlen gyakorlati megoldásokkal kecsegtető párbeszéd. Egyre több lehetőség nyílik arra, hogy teszteléssel próbára tegyünk a legkülönbélebb elméleteket, melyek nyomán az elméletek is finomíthatóak (Macy 2015a). Az adatbőség korszakában egyre ritkábban merül majd fel az a kérdés, hogy egy elmélet az eredeti formájában helytálló-e vagy sem; és sokkal gyakrabban születnek válaszok arról, hogy egy elmélet előrejelzései milyen körülmények között igazak, hol kell megszorító feltevéseket, módosításokat végrehajtani az eredeti modellen. Ezt a folyamatot nevezte Lindenberg (1992) csökkenő absztrakciónak – amely a tudományos ismeretek bővülésének egy ideális útja.

Az adatok nélküli elméleti viták melegágyai a doktrinerségnek, az ideológiai elfogultságnak, és végső soron az áltudományos érvelésnek. A szociálpszichológia válságba jutott (Earp és Trafimow 2015), mivel az adatszűke időszakában elfogadott engedékeny mérési sztenderdek (például kis esetszámú kényelmi minták használata), az elméleti kutatás magasabb presztízsével párosulva önhitté tették a kutatókat, akik nem hisznek az elméleteiket cáfoló adatoknak, és addig „gyűrik” őket, amíg nem igazolják az előzetes elképzeléseiket. A válság abból adódik, hogy az utóbbi években számos, korábban alapvetőnek tekintett, tetszetős és nagyívű elméletekhez szépen simuló eredményről derült ki, hogy megismételhetetlen (Open Science Collaboration 2015).

Az elmúlt évtizedekben a szociálpszichológusok által leírt (Kunda 1999) motivált érvelés mechanizmusa tömegesen volt megfigyelhető a gyenge empirikus lábakon álló pszichológia elméleti intuícióik iránt elkötelezett kutatóinál is – legalábbis erre utalnak az alapo-

sabb vizsgálattal nyert viselkedési adatok. A motivált érvelő számára az információfeldolgozás során fontos, hogy az új eredmények minél jobban illeszkedjenek az elképzeléseihez és elköteleződéséhez. A földrajz, a biológia, a fizika és számos más természettudományos diszciplína egyes nehezen tesztelhető dilemmáiban fellángoló, szélsőséges és végtelen viták (Bosker 2018) arra utalnak, hogy a természettudományos elméletalkotók sem mentesek a motivált érveléstől, azaz a szűkös adatok torzított értelmezésétől. Ugyanakkor könnyű érvelni amellett, hogy a társadalomtudósok gondolkodását – az elvileg minden tudományban jelentkező gondolkodási merevségen túl – könnyen hatalmukba keríthetik a tudományos szempontokon túli, elsősorban ideológiai elfogultságok. Egyes kritikus kutatók néha radikális és megkérdőjelezhető módszerekkel (például áltanulmányok írásával) próbálják meg felhívni a figyelmet erre a jelenségre (Quillette 2018, Sokal és Bricmont 2008 [1997]).

A kutatási dizájn és a statisztikai eljárások fejlesztése, továbbá a válaszadói bázis méretének, diverzitásának és transzparenciájának növekedése ma már megbízhatóbb teszteléseket tesz lehetővé. Hasonló pozitív következményekkel járnak a szociológia és a közgazdaságtan számára a sztenderdizált, bárki számára elérhető, nemzetközi kérdőíves felmérések, valamint egyéb kormányzati adatbázisok is. És persze hasznos a 2010-es években már hatalmas mennyiségben keletkező, ún. digitális lábnyomok elemzése is (Golder és Macy 2014). Ezért kell üdvözölnie az adatbőség korszakát az elméleti érdeklődésű társadalomtudósoknak is.

A társadalomtudományok számos képviselője lelkesen alkalmazza az új lehetőségek nyomán terjedő új kutatási módszereket. Néhányan pedig kifejezetten felhívják a figyelmet azokra a fetelezett veszélyekre, amelyek az emberi motivációkra és interakciókra vonatkozó általános elméletek létrejöttében rejlenek (Watts 2014, vö. Anderson 2008). Szerintük a mindenféle torzításokat tartalmazó intuitív pszichológiai érvelés helyett csak az adatokból kinyerhető összefüggésekben kellene hinnünk.

A mérési lehetőségek bővülésével párhuzamosan került előtérbe a statisztikai tudás a társadalomtudományokon belül (is). A szociál-

pszichológia fent idézett válságát a statisztikusok „aknamunkája” készítette elő (Simmons et al. 2011). És ez a válság csak követője volt a biostatistikusok által az orvosi tudományokon belül elindított drasztikus változásoknak (Ioannidis 2005). De nem csak a tudományos válságok és forradalmak indításában van szerepe néhány lázadó lelkű statisztikusnak. A 21. században a nemzetközi tudományos közösségek által megkövetelt statisztikai ismeretek szintje is jelentősen megnőtt a különböző társadalomtudományi diszciplínákban.

#### I.2.1.2. Elmélet az adatkorban

A trendek tehát egyértelműek: a társadalomkutatáson belül nő a kvantitatív adatok feldolgozásának súlya és jelentősége. Meggyőződésem ugyanakkor, hogy az adatbőség és a statisztikai tesztelés fejlődése csak addig segítik a világ szisztematikus megismerését, amíg a világot leíró, általánosítható tudományos modellek továbbfejlesztését is szolgálják. Megalapozott általános elméletek nélkül az empirikus eredmények olyan hétköznapi megfigyelésekké válnak, amelyek nem nyújtanak segítséget a konkrétan megfigyelt (már megtörtént) eseményeken és mintákon túli, azokhoz hasonló helyzetekre való felkészülésben (Imai 2017). Persze elméletek mindig kéznél lesznek. De az elméletileg felkészületlen empirikus kutatók ki vannak téve az ad hoc hipotézisek és megalapozatlan divatok befolyásának.

Ennek megfelelően meggyőződésem, hogy az emberi cselekvésekkel (tudatos döntésekkel) bármilyen kapcsolatban lévő jelenségeket vizsgáló empirikus kutatásoknak szükségük van az emberi cselekvések megalapozott, széles körben használható modelljeire (Szántó és Syi 2010). Ez akkor is így van, ha egy adott helyzetben az egyéni cselekvések makroszintű következményei nem közvetlenül és/vagy nem intuitíven vezethetőek le az egyéni szándékokból (vö. Watts 2014). A bio- és pszichológiai statisztikusok által elindított, részben válsághoz vezető kritikai mozgalomnak csak az egyik kiváltója volt a dédelgetett elméletéhez adatokat fabrikáló kutató. A statisztikusok

által kritizált másik jelenség az eredményekhez hipotéziseket utólag, ad hoc gyártó, a valós elméletek nélküli „adatbányász” tevékenység, amelynek során a publikálási kényszer alatt dolgozó kutatók mindenáron a szignifikáns összefüggéseket keresik (Banks et al. 2016). Ennek leglátványosabb példáit az epidemiológiában találták meg a kritikus statisztikusok, de a probléma érinti az összes társadalomtudományt is. Az ennek kiküszöbölésére tett erőfeszítések befolyásolták a komolyabb közgazdasági, pszichológiai, politikatudományi és szociológiai folyóiratok statisztikai módszerekre vonatkozó elvárásait, ahogy az Amerikai Statisztikai Társaság teszteléssel kapcsolatos közelmúltbeli állásfoglalását is (Wasserstein és Lazar 2016).

Ebben a könyvben bemutatom, mekkora tétje lehet annak, hogy milyen általános elméleti kiindulópontból értelmezzük adatainkat. Az elméletek meghatározzák, hogy milyen tényezőket vonunk be a vizsgálatunkba, és hogy ezeket hogyan és hova helyezzük a statisztikai modellekben. Az empirikus kutatások elméletileg megalapozott elemzésében az eredményeknek túl kell mutatniuk az adatbázisban talált empirikus trendeken. Meg kell mutatniuk azt, hogy hol lehetnek a feltárt empirikus összefüggések alkalmazási korlátai, és útmutatót kell adniuk a későbbi empirikus kutatásokhoz is. Az adatelemzéshez ad hoc módon elméleteket kereső statisztikacentrikus magatartás éppoly tudománytalan, mint az adatokat megerősítő vagy teljesen ignoráló teoretizálás (Imai 2017).

További érvek helyett inkább szeretném felhívni a statisztika önálló erejében rendíthetetlenül bízó kollégák figyelmét arra, hogy ezekben az években megkezdődött a statisztikai következtetések módszereinek forradalma a társadalomtudományokban is (Pearl 2014, 2018, Pearl et al. 2016). Az új szemléletben a statisztikai modellezés nélkülözhetetlen részévé válik a valódi elméleti gondolkodás. Remélem, hogy ez a könyv közvetett módon hozzájárul a statisztikai gondolkodás említett elméleti fordulatának hazai elterjedéséhez is.

## I.2.2. MELYIK ELMÉLET?

### I.2.2.1. Közgazdaságtani eszköztár, szociológiai elmélet

Az erkölcs forrásával kapcsolatos vita szorosan kötődik ahhoz a szociológiát, és általában a társadalomtudományokat megosztó metodológiai ellentétéhez, amely a módszertani individualizmus és módszertani kollektívizmus között feszül (Orthmayr 1997). Az erkölcs forrását kívül kereső klasszikus elméleteket egyúttal a módszertani kollektívizmus is jellemezte (lásd Durkheim 2003 [1897]); míg az erkölcsöt az emberi természetben kereső elméletek a módszertani individualizmus talaján álltak (lásd Weber 1987 [1922]).

A 20. században a társadalomtudományi diszciplínák többségében a módszertani individualista közelítésmód volt a jellemző, és néhol ez vált kizárólagos metodológiai paradigmává. Ehhez kapcsolódóan az erkölccsel foglalkozó kutatásokban az erkölcsöt belül kereső elméletek népszerűbbek voltak, mint a kizárólag külső erők hatását feltételező modellek.

E könyv szerzőjeként egyetértek azokkal, akik szerint a módszertani individualista közelítésmód termékenyebb tudományos metodológia, mint a kollektívista. E közelítésmódon belül külön kiemelendő a modern közgazdaságtudomány domináns embermodellje és metodológiai közelítésmódja, a racionális döntésemélet. A racionális döntésemélet különösen termékeny általános embermodell, melynek népszerűsége sok évtizeddel elterjedése után sem csökken, és semmi sem utal arra, hogy belátható időn belül más általános elmélet váltaná fel a közgazdaságtudományon belül. Az elmélet széles körben alkalmazzák a politikatudományban, és egyes elemei megjelennek további tárdiszciplínákban is. A racionális döntésemélet viszonyítási pontként szolgál több népszerű cselekvésemélet számára is, a pszichológiától a szociológián át a filozófiáig.

A racionális döntéseméletet gyakran vádolják azzal a kritikusok, hogy a valós folyamatoktól elrugaskodott, absztrakt konstrukció. Ugyanakkor a kvantitatív társadalomtudományi mérések elméleti megalapozásában is kiemelt szerepe van a racionális megközelítés-

nek. A II. fejezetben amellett érvelek, hogy ez részben a vizsgált jelenségek megfelelő kiválasztásának köszönhető. Ugyanakkor a szelektív vizsgálati fókusz figyelembe véve is tagadhatatlan az elmélet kiemelt jelentősége.

Ebben a kötetben nem érvelek sem elvi, sem empirikus alapon amellett, hogy a racionális döntéselmélet az emberi cselekvések és társadalmi jelenségek jelenleg ismert legtermékenyebb modellje, és nem kérem az olvasót, hogy ezt érvek nélkül fogadja el. Adottnak tekintem ugyanakkor, hogy ez az elmúlt nyolcvan év egyik legjelentősebb, töretlen népszerűségű cselekvési modellje, és legitim dolog átgondolni bármely társadalomtudományos kérdést ennek az elméletnek a szemszögéből. Kérem az olvasót, hogy ezt fogadja el. Ugyanis a továbbiakban a racionális döntéselmélet feltevéseiből kiindulva vizsgálom meg az erkölcs forrására vonatkozó, előző alfejezetben ismerttetett kérdéseket. Az utóbbi évtizedekben felhalmozódott, és közel tizenöt év terjedelemben összefoglalt kutatási anyag gazdagsága remélhetően igazolja a vizsgálat – néhány olvasó számára talán szűknek tűnő – elméleti fókuszát.

A gondolatmenet koherenciája és az észszerű terjedelmi korlátok – a talán túl szűknek tűnő elméleti fókusz mellett is –, szükségessé teszik a kutatási programon belüli témák szűrését. Például rengeteg olyan eredmény született az elmúlt évtizedekben, amely az erkölcs vizsgálatok a racionális döntéselméletből indult ki, de evolúciós modelleket alkalmazott. A különböző evolúciós modellek hívei közötti viták ugyanakkor az általam felépített nyomozati narratíva mellékágát alkotják, azaz kevésbé fontosak a konklúzió és az odavezető út szempontjából. Ezért csak akkor és annyit esik szó róluk, amikor és amilyen mértékben csatlakoztak az itt rekonstruált nyomozati munka főáramához. Így ezek a kutatások a láthatatlan intézeten belüli súlyukhoz képest alul lesznek reprezentálva. Egy másik témában is jelentősen szelektáltam az eredmények között. A könyv III. fejezetében kevesebb, majd a VI. és VII. fejezetében több szó esik majd a normákról, amely fontos téma számos kutató számára, mind a racionális döntéselmélet paradigmáján belül, mind azon kívül. Ám e kutatók jelentős része nagyon különböző kérdéseket tesz fel a normákkal

kapcsolatban, mint az általam vizsgált láthatatlan intézet nyomozói. Ezért lesz bizonyos tekintetben felületes a normák tárgyalása.

Mindazonáltal, az V. fejezetben feltűnik a racionális döntésmélet paradigmáján kívül mozgó „vendégnyomozók” egy csoportja, akik amellet, hogy a racionális döntésmélet kutatói által nyitva hagyott kérdésekre reagálnak, fontosak a láthatatlan intézetben belüli munka irányának megértése szempontjából is.

A közgazdaságtan alapjait jelentő individualista és racionalista paradigmára irányuló fókusz nem jelenti azt, hogy adottnak tekintem az erkölcs belső eredetére vonatkozó tézist. Azt gondolom, hogy az erkölcs külső forrásaira vonatkozó durkheimi sejtést számos tudományos megfigyelés erősítette a 20. században, és termékenynek bizonyultak a későbbi, az identitásfüggő szerepelvárásokat (Stets és Burke 2000, Stryker 1968), valamint a valóság társadalmi konstruálását (Berger és Luckmann 1966) tételező szociológiai elméletek is.

Ennek megfelelően az erkölcs forrására vonatkozó kérdést nyitottnak tekintem; a kérdés megválaszolásának célja irányítja a gondolatmenetet. Úgy vélem, hogy „belső erkölcsi forrás” tézis és a (racionális) individualista közelítésmód közötti kapcsolat csak egyfajta heurisztikus vonzódás. A vonatkozó elméletek és empirikus tesztek kritikus áttekintésével igyekszem megszabadítani a téma vizsgálatát e heurisztikai torzítástól.

Azok a szakmabeli olvasók, akik kritikusak a racionális döntésmélettel szemben, ugyanakkor szimpatizálnak az erkölcs külső forrását feltételező klasszikus és kortárs elméletekkel, joggal szkeptikusak azzal kapcsolatban, hogy sikerülhet-e, a fent vázolt keretek között, elfogulatlanul vizsgálni az erkölcs forrására vonatkozó kérdést. E szkeptikus olvasók bizalmát és türelmét kérem, és mindenekelőtt annak belátását, hogy különösen erős érveket szolgáltat a külső forrás tézis mellett egy olyan elmélet, amely a racionális döntésmélet eszköztárát és eredményeit alkalmazva mutatja be, hogy az erkölcsös cselekvés belső erkölcsi készlettel is létrejöhet. A kötetben bemutatott nyomozás megteremti egy ilyen erős konklúzió lehetőségét.

Ez a kötet nem vállalkozik arra, hogy az erkölcs külső forrására vonatkozó racionális döntésméleti modelleket összekösse a velük pár-

huzamosan fejlődő szociológiai elméletekkel. Ez a feladat egy újabb önálló monográfia feladata lehetne. Ugyanakkor a VII. fejezetben bemutatom, hogy a racionális döntéseméleti modelleken nyugvó gondolatmenet következtetései egybevágóan az erkölcs új szociológiájának legfontosabb téziseivel (Abend 2011, Bargheer és Wilson 2018, Hitlin és Vaisey 2010, 2013, Simpson és Willer 2015, Tavory 2011).

Mielőtt bemutatnám, hogy meddig jutott az erkölcs forrására vonatkozó kérdés racionális döntéseméletből kiinduló, immár nyolcvan éve tartó vizsgálata, tisztázni kell, hogy mit jelent és mit nem jelent a racionális döntésemélet. Emellett az is tisztázásra szorul, hogy ezt miért kell tisztázni.

#### I.2.2.2. Racionális döntési eljárás vs. racionális döntésemélet

A racionális döntés vázlatos definíciója egyszerű: olyan cselekvés, amely a cselekvő céljainak eléréséhez az arra legmegfelelőbb eszközöket választja (Elster 1995 [1989]). Azonban ez a meghatározás sokféleképpen operacionalizálható. Több dimenzióban, többféleképpen választható el a racionális a nem racionálistól. Létezik olyan megközelítés, amely egy vadászok által üldözött róka sikeres menekülési cseleit képes racionális döntésként modellezni, ugyanakkor nehezen tudná hasonlóan értelmezni a *Sorstalanság* azon zsidó szereplőinek döntését, akik az óbudai téglagyár területén, hosszas mérlegelés után úgy döntenek, hogy meg kell bízni a munkatáborokba induló vonatokra embereket toborzó német tisztokban (Kertész 1975). A racionális döntések más tudományos megközelítései épp fordított konklúzióra jutnak: a mérlegelő foglyok racionálisak, az ösztönös róka nem. A megközelítések sokféleségének ismertetése szétfeszítené a tartalmi kereteket, ezért csak egyetlen, a könyv gondolatmenetének megértése szempontjából kulcsfontosságú distinkcióra koncentrálok.

A distinkció megértéséhez mindeneke előtt látni kell a racionális döntési modellek egy közös pontját: a fent említett „megfelelő eszköz” mindig beazonosítható a logika univerzális szabályait követő



levezetés révén. E levezetés néha pofonegyszerű, más helyzetekben csak szakemberek szűk csoportja számára követhető, komplex következtetések láncolata. De minden esetben igaz, hogy a levezetéshez sosem elég a beleérző megértés, miszerint azért észszerű a döntés, mert azt érzem, hogy én is így cselekdnék, de nem tudom megmagyarázni, hogy miért.

A „megfelelő” döntés logikus levezethetőségének követelményéhez azonban két interpretáció is tartozik. E két interpretációt gyakran keverik a racionális döntésmélet kritikusaival.

Egyrészt a racionális döntés bizonyos modelljei a fenti említett logikus levezetést a létező emberi döntési eljárások egyik alapvető típusának tekintik. Ebben a megközelítésben a racionális döntési modell egy, a döntések meghozatalának módjára vonatkozó tipológiát építő cselekvésmélet része. Ilyen cselekvésméletet épített Weber (1987 [1922]), és ilyen típusú cselekvésméletekkel dolgoznak mindmáig a szociológusok (például Strand és Lizardo 2015, 2017). A pszichológiai döntésméletek jelentős része is a döntési eljárásokból indul ki, és az eljárások tipológiájába illeszti a racionális döntést (Kahneman 2011). Olyan elméletet nem ismerek, amely a döntések meghozatalának kizárólagos vagy domináns módjának tekintené a mérlegelő, kalkulatív döntéshozatalt.

Másrészt léteznek a döntéseknek egy olyan racionális elmélete, amely a fent említett logikus kalkulációt használja az egyének várható viselkedésének előrejelzésére. Ez az elmélet nem foglalkozik azzal, hogy miképpen születnek meg azok a döntések, amelyek a feltételezett célokból logikusan levezethető viselkedéshez vezetnek (Friedman 1953). E racionális elmélet egyes képviselői ugyan kiegészítő feltevéseket fogalmaznak meg arra vonatkozóan, hogy milyen mechanizmusok terelhetik a racionális eredmény irányába a viselkedést (Alchian 1950, Binmore 1998), ám e kiegészítő feltevések nem tartoznak az elmélet alapvető építőelemei közé. Ilyen, a racionalitás irányába terelő mechanizmus lehet a kevésbé racionális szereplők kiszelektálódása, vagy utánzás-, illetve próba-szerencse alapú adaptáció révén az irracionális viselkedési minták elhagyása. A könyv végén szót ejtek azokról a pszichológusokról is, akik amellet érvelnek,

hogyan tudatos kalkuláció nélkül is képes a bejövő információk racionális feldolgozására, ráadásul ezek eredményét épp a tudatos kalkulációt elnyomó érzelmek közvetítik (Lazarus 1991).

A döntések racionális elmélete a logikusan kalkulálható végeredményt nem a döntések egy típusához köti, hanem azt felételezi, hogy az emberi viselkedés általában jól megjósolható a döntések célokból és lehetőségekből történő logikus levezetésével. Ezt a viselkedési elméletet *racionális döntésselméletnek* nevezzük. A fent említett szociológiai és pszichológiai cselekvésselméletek nem sorolhatóak a racionális döntésselmélet kategóriájába. A racionális döntés azokban az elméletekben a döntési eljárások egy típusa.

Ez a könyv döntően az erkölcs racionális döntésselméleti kutatásai-val foglalkozik. Nem foglalkozik a döntési eljárásokra vonatkozó cselekvésselméletekkel, amelyekben a racionális döntés a döntési eljárások egyik típusa. Az itt tárgyalt distinkció megértése és észben tartása különösen fontos azon olvasók számára, akik ismerik és méltányolják a racionális döntésselmélet közkeletű szociológiai (és egyéb) kritikáit. E kritikák ugyanis gyakran keverik a racionális döntési eljárás típusának modelljét a döntések racionális elméletével, azt feltételezve, hogy a racionális döntésselmélet a racionális döntési eljárás dominanciáját posztulálja.\*

### I.3. Honnan hová jutunk?

A kötetet egybefűző retorikai logika a termékeny, a tudományos állítások pontosításához vezető feleselés analógiájára épül. Tehát ebben a könyvben *nem* a kanonizált klasszikus eredmények statikus bemutatására vállalkozom (vö. Syi 2008), és *nem* a tudomány jelenlegi állásának (a legújabb tézisek és eredmények) keresztmetszeti bemutatására törekszem. Továbbá a kötet *nem* is örök hitviták elmélet-történetének a bemutatása (vö. Fontaine 2006, 2012). Ez utóbbit

\* A racionális döntésselmélet lehatárolása még e distinkció megértése és elfogadása után sem egyértelmű. Erről a II. fejezet végén lesz szó.

azért kell hangsúlyozni, mert több, öröknek tűnő metaelméleti ellentét árnyéka vetül rá a könyvben bemutatott felfedezési folyamatra: egyrészt a módszertani individualizmus és kollektivizmus közötti ellentét (Orthmayr 1997), másrészt a pszichologizáló megközelítések képviselői és a szituációs logika alkalmazói közötti vita (Popper 1989 [1957]), harmadrészt a gondolkodási modulok szerepét hangsúlyozó magatartástudományi közelítés és a konstrukcionizmus között feszülő ellentmondás (Schwandt 1994). Ezek metaelméleti szinten örök vitának tűnhetnek, de az itt bemutatott tudományos nyomozás során egy fejlődéstörténet bontakozik ki, melynek jelenkori állása állásfoglalásra késztet majd engem is az említett metaelméleti kérdésekben.

Az elmélet fejlődésének időbeli tagolódását jelzem a fejezetek címeiben, amelyek a téma megjelölése előtt egy-egy, a fejezetben bemutatott kutatási távlatot megnyitó, emblematisz tanulmány publikálási évét jelzik. A történet a második világháború végén kezdődik, és a jelenkorban ér véget. Egyes fejezetek elején néhány elnagyolt megjegyzéssel utalok arra a sejtésemre, miszerint az új irányok keresése talán nem volt teljesen független az aktuális társadalmi változásoktól. Ennek felvetése – mégoly kidolgozatlan formában is –, azért fontos, mert a konklúzióban az erkölcsi elvek erős kontextusfüggősége mellett foglalom állást; és az elvekről való tudományos gondolkozás esetleges kontextusfüggősége az erkölcs társadalmi beágyazottságának különös, de erős és különlegességében szép bizonyítéka lehetne.

### I.3.1. A SZITUÁCIÓS LOGIKÁTÓL A PSZICHOLÓGIAI MODELLEK FELÉ

#### I.3.1.1. A külső erkölcsi ösztönzők játékelméleti modelljei

Az alábbi áttekintést követő II. fejezet körülhatárolja azt a „láthatlan intézetet”, amelynek az erkölcsös polgárral kapcsolatos kutatási eredményeit összegzem. Ez az „intézet” az európai politikai filozófia több évszázados hagyományaira támaszkodva, a racionális döntésemélet paradigmájának keretein belül alakult ki a második világháborút kö-

vetően, és elsősorban – de nem kizárólagosan – közgazdaság-tudományi tanszékekről verbuválta a tagjait. A fejezet a fenti lehatárolás mellett három tézis kifejtésére vállalkozik. Egyrészt arra mutat rá, hogy a neoklasszikus közgazdaságtan és általában a racionális döntéselmélet külső kritikussai hibásan vádolják a paradigmát azzal, hogy az önző ember modelljét alkották meg. A homo oeconomicus alapmodellje ugyanis hallgat az ember céljairól. Ez nagyon fontos, explicit alapvetése e paradigmának. A vágyakra vonatkozó feltevések az egyes részterületekre fókuszáló, középszintű elméletekben jelennek meg. Másrészt arra is a felhívom figyelmet, hogy a modern közgazdaságtan képviselői jellemzően elmosás és technikai jellegűnek minősítik racionális döntéselmélet két nagy modelljét, a neoklasszikus mikroökonómia és a játékelmélet emberképe közötti, az erkölcsi viselkedés szempontjából alapvető különbséget. A játékelmélet ágense ugyanis – szemben a neoklasszikus mikroökonómia homo oeconomicusával –, jellemétől függetlenül rákényszerül arra, hogy átgondolja mások övétől eltérő vágyait, és arra törekszik, hogy ne keresztesse mások céljait. Ezt persze nem jóindulatból, hanem saját jól felfogott érdekében teszi, abban a tudatban, hogy a többiek is figyelnek az ő céljaira. Végül a fejezetből az is kiderül, hogy az emberi viselkedés a szociológiában népszerű – a racionális döntéselmélettel szemben álló, leegyszerűsítve itt strukturalistának nevezett – középszintű elméletei talán még inkább támaszkodnak az anyagi érdekek primátusának kiegészítő feltevésére, mint az alkalmazott közgazdaságtani modellek. Ugyanakkor a kérdést elmosó közgazdaságtan ignoranciájával szemben, az atomizált aktorokat feltételező (közgazdaságtani és szociológiai) modellek szociológiai kritikussai régóta hangsúlyozzák, hogy az erkölcsös cselekvés forrását az egyén társadalmi környezetéből/-ről szerzett információkban, és az onnan érkező nyomásban kell keresni.

Ha túllépünk a biológiai altruizmus szűk érvényességi körén, akkor már nem a jóindulatunk, hanem az egymás eltérő céljaira figyelő ún. stratégiai döntések teszik lehetővé az erkölcsös magatartást; a játékelmélet pedig az ilyen stratégiai döntések formális modellezését. Ez a könyv egyik fő tézise, ezért ennek sok teret szentelek: egy egysze-

rú, a tudományos ismeretterjesztésben is sokszor alkalmazott játszma példáján keresztül nemcsak azt mutatom be részletesen, hogy a morálban nem a jóindulat, hanem az odafigyelés a fontos, hanem azt is, hogy a neoklasszikus közgazdaságtan legtöbb képviselője nem érzékeny a stratégiai interakciókban rejlő inherens morális tartalomra, és a racionális döntéssel szembeni sztereotípiáknak megfelelően inkább az önzőség következményeit keresi mindenhol.

A II. fejezetben tehát arról lesz szó, hogy már a 20. század közepén adott volt egy olyan analitikai eszköztár, amely képes volt egzakt módon kezelni az erkölcs „társadalmi térletén” alapuló leírását. Továbbá a társadalom tényleges működésével, annak teljes sokszínűségével foglalkozó szociológusok körében is sokan voltak, akik az egyén jól körülírható társadalmi környezetében, és az egyén e környezet iránti érzékenységében keresték az erkölcs forrását. A III. fejezet először azt mutatja meg, hogy a megfelelő eszköztár léte, valamint az atomizált modelltől gyorsan elforduló szociológusok figyelmeztetése ellenére a racionális individualizmus paradigmája az egyén saját késztetéseiben kezdte keresni az erkölcsösség és erkölcstelenség forrását. A nyomozás első lépéseként definiálták azt a speciális játszmát, ahol a társadalmi környezetére érzékeny polgár pont úgy viselkedik, mint a mások szándékaira vak, neoklasszikus homo oeconomicus. Ez a játszma, az ún. fogolydilemma, ahol az egyéni érdekkövetés világosan ütközik a csoportszinten értelmezhető közjóval, és ahol a racionális cselekvő nem foglalkozik azzal, hogy mit csinálnak a többiek. A fogolydilemma vonzza azt az értelmezést, hogy csak érdekeinken túlmutató motivációk sarkallhatnak a köz javát szolgáló cselekedetekre. Ez azért fontos, mert könnyű felismerni számos kisebb és nagyobb jelentőségű emberi játszmában a fogolydilemma szerkezetét; ráadásul az önző és az erkölcsös viselkedés fogolydilemma alapján történő szétválasztása illeszkedik a klasszikus politikai- és erkölcsfilozófia, valamint a klasszikus közgazdaságtan duális emberképéhez.

Ugyanakkor a játékelmélet korai alkalmazói már a hatvanas években felismerték a situációs logikai elemzés perspektíváit a fogolydilemma speciális szerkezetében. Egyrészt megmutatták, hogy a neoklasszikus közgazdaságtan biológiából átvett, egyszerre önző és

altruista cselekvője már viszonylag kis csoportban is úgy viselkedik, mint a teljesen önző homo oeconomicus. A kihívás ezután az volt, hogy miképpen lehet átalakítani a vak önzőségre biztató fogolydilemmát úgy, hogy a tisztán önérdékkövető emberek a köz javát szolgálják. A kihívás jelentősége az volt, hogy amennyiben e szélsőséges esetben is sikerül az átalakítás, akkor az nagyon erős érv a racionális döntéshozó modellje mellett. A megoldás egyik fele a játszma folytatásában keresendő: idővel a dilemma foglyai, fenyegetések elvileg végtelen láncolatával rávehetik egymást a közös érdek szolgálatára. „Ha te a közös legelőnkön csak a magad dolgával törödsz, akkor (és csak akkor) a jövőben én is azt teszem – és így mindenki rosszul jár.” Ameddig van esély újabb és újabb játszmákra, ezt a fenyegetést érdemes komolyan venni. A megoldás másik fele a felelősség meghatározása: akármilyen sokan szembesülünk a fogolydilemmával, minden tékozló fiúnak (és lánynak) minden időpillanatban legyen egyetlen, és ezért egyértelműen felelős, hitelesen fenyegető „gyámja”. E megoldásokkal az önzés és az altruizmus ellentétét definiáló fogolydilemmát a szituációs logikai mentén elemző tudósok viszonylag rövid idő alatt azonosították az önérdékkövető emberek együttműködésének néhány fontos társadalomszerkezeti feltételét. A játékelméleti modellekben definiált mechanizmus jól leírta azt, ahogyan sok társadalomkutató a normák fenntartását elképzei.

#### I.3.1.2. Belső erkölcsi késztetések beillesztése: kifelé a racionális döntéselméletből

Az altruizmus jelentőségével szembeni játékelméleti érvek nem szegték a kedvét azoknak, akik belső okokban keresték az emberi együttműködés gyökereit. Ők úgy vélték, hogy a Jóisten nem bízhatta pusztán a próféták munkájára és az emberek belátására az egyénileg és kollektíve is racionális stratégiák megtalálását és megvalósítását. Azt feltételezték, hogy az emberiség hosszú őstörténete során szelekciós előnyt élveztek azok, akik az együttműködési dilemmákkal szembe-sülve, hosszas morfondírozgatás helyett reflexszerűen alkalmazták a

fenti fenyegető stratégiákat. A IV. fejezet áttekinti, hogy a feltételes együttműködésre született ember modelljét kidolgozó kutatók miképpen alakították át a nyolcvanas évektől az ismételt fogolydilemmában együttműködést eredményező döntési elveket ún. társas preferenciákká. A feltételezett preferenciák azt célozzák, hogy az emberi játszmákban senki se járjon se túl jól, se túl rosszul. Eszerint az ember se nem önző, se nem altruista: csak tisztességesen/méltányosan akar viselkedni.

Mivel a tisztességes ember elméletének felvetése idején az önző ember modellje uralta a racionális döntéelmélet legtöbb alkalmazási területét, nagy volt a bizonyítási teher az új megközelítés hívein. Így ők a tisztességes magatartás ösztönét preferáló evolúciós erők modelljeinek kidolgozása mellett gyorsan a pszichológia kísérleti megközelítése felé fordultak. Úgy látták, hogy a szituációs logikának már nincs több mondanivalója az emberi együttműködésről, és innentől a közgazdaságtan is pszichológiai elméletként fejlődhet tovább, egyre jobban megismerve a velünk született preferenciák konkrét paramétereit.

A társas preferenciák után nyomozó kísérleti pszichológiai megközelítés a 21. században nagy, máig tartó népszerűsége tette szert a közgazdaságtanon belül, és lassan terjedni kezdett a politikatudományban is. A kísérletek eredménye ugyanakkor felemás volt. Komoly erőfeszítések árán sem sikerült azonosítani, hogy pontosan mire vágyik a méltányos ember. Néhány perdöntő kísérlet azt mutatta, hogy leginkább semmi olyasmire, ami méltányosak nevezhető. Ellenben a laboratóriumok embere fél a büntetéstől. És jogosan. Ugyanis a laboratóriumok embere gyakran érdek nélkül bünteti az aszociális magatartást. Ez elég stabil eredmény, azzal együtt, hogy nem sikerült azonosítani olyan univerzális célokat, amelyek érdekében büntetnek.

Ekkor – már a 21. század elején –, két szögesen ellentétes irányban haladt tovább a nyomozás. Pszichológusok egy csoportja úgy gondolta, hogy a racionális döntéelmélet még erősen pszichologizált formájában sem tudott elszakadni a klasszikus politikai filozófiai alapjaitól, és nem válhatott az erkölcsi motivációk általános pszichológiájává. Az V. fejezet láthatatlan intézetünk pszichológus „vendég-

kutatóinak” munkásságát mutatja be. Az erkölcs új pszichológiájának képviselői úgy vélik, hogy saját értékeik, valamint egy torzított társadalomfilozófiai hagyomány rabjai azok, akik számára az erkölcs kizárólag torták felosztásáról, és a szenvedők felemeléséről szól. Arra mutatnak rá, hogy a fogolydilemmához többé-kevésbé hasonló erkölcsi játszmák nemcsak gazdagságról és szegénységről, gazdasági erőfeszítésekről és pénzre váltható hasznokról szólhatnak, hanem akár arról is, hogy mennyi energiát szánjanak a tisztálkodásra. A koszos ember, a „sztenderd” potyautasokhoz hasonlóan, árthat a környezetnek. De ami az igazi váltás: nemcsak a tudatos potyautazó lehet fertőző, hanem az ártatlan beteg is, aki semmilyen fenyegetés hatására sem tud úgy dönteni, hogy holnaptól nem fertőz. Ahogy az erkölcspszichológia, úgy a fejezet is kiemelten foglalkozik az ilyen „ártatlan potyautasokkal”, és a más, üldözés helyett inkább elkerülendő emberekkel szemben ébredő undor érzésével.

Az undor tárgyalásával az erkölcs új pszichológiája az erkölcs és altruizmus közötti utolsó kötelekeket is elvágja. Talán még fontosabb a racionális döntésemélet szempontjából, hogy az undor kezelhetetlen az anyagi javak elosztásának fogalmaival. A pénznek nincs szaga. Azaz a tisztátalansággal szembeni undor egy olyan társas preferencia, amely tesztelhetetlen azokkal az eszközökkel, amelyekkel az önző és a tisztességes polgár modelljeit tesztelték a közgazdászok. A racionális döntésemélet következetes pszichologizálása végül hasznavehetlenné teszi a játékelméleti alapú kísérletezést. Az undort és a többi, az erkölcspszichológia által azonosított erkölcsi ösztönt más, hagyományos pszichológiai módszerekkel próbálják mérni.

Az új erkölcspszichológia hasonló ütemben és mértékben vált népszerűvé az egyetemek falain belül és kívül, mint a tisztességes ember modellje. Ahogy a méltányos polgár modelljét az ortodox játékelmélet hívei kritizálták, úgy éri egyre több komoly kritika az erkölcspszichológiát a pszichológia egyéb irányzatainak képviselői részéről. Az irányzat által azonosított erkölcsi ösztönök nem tűnnek élesen elkülöníthető viselkedési moduloknak, és egyre több olyan kísérleti eredmény születik, amely megkérdőjelezi, hogy ezek az erkölcsi intuíciók döntéseink végső mozgatói lennének.



### I.3.2. VISSZA EGY KOMPLEXEBB SZITUÁCIÓS LOGIKÁHOZ

#### I.3.2.1. Társadalmi normák önző emberek társadalmában

A méltányos polgár modelljére vonatkozó ellentmondásos eredmények nemcsak a pszichológusokat bátorították továbblépésre, de a racionális döntésemélet ortodox híveit is. A VI. fejezet ismerteti azt a kutatási irányt, amely a kísérleti laborokból az önző polgárok által fenntartott normák játékelméleti modelljeihez tér vissza. Ez, a fenténél kevésbé népszerű irány arra törekszik, hogy teljessé tegye a normák fenntartásának szituációs logikai modelljét, azaz pontosabban határozza meg azokat a körülményeket, amelyek mellett az emberek rávehetőek az együttműködő magatartásra. Az önző polgár tézisének alapuló korábbi erkölcsmodellek nem adtak útmutatót arra vonatkozóan, hogy miképpen juthatunk el a normákhoz, és mi lesz azok tartalma. Ez adott teret azoknak a pszichológiai spekulációknak, melyeket a IV. és V. fejezet ismertet: ha a szituációk logikája nem ad pontos útmutatót az erkölcsös viselkedéshez, akkor annak konkrét formái csak belülről fakadhatnak.

A normakövetés egyes játékelméleti modelljei az apró és csendes korrekciók által kitaposott egyensúlyi ösvényekre mutatnak rá. Többen vannak, akik ezt elégtelennek tartják, és az emberek közötti kommunikáció szerepét mutatják be a normák létrehozása és fenntartása terén. Elméleti szinten bizonyítják, hogy a kommunikáció megfelelő struktúrája elvezetheti a racionális cselekvőket az erkölcsös viselkedéshez. Ezek az elméletek nem a pszichológiai, hanem a szociológiai szemlélet irányába tolják a játékelméleti modellezést. Kulcsfogalom má válik a közös tudat, ahol nemcsak a társadalmi szabályok ismerete fontos, hanem az is, hogy mit gondolok arról, hogy mások mit gondolnak a szabályokról – és a többiek gondolatairól. A megfelelő közös kommunikációs térben elhangzó normatív kijelentések olyan közös kincsek e modellekben, amelyek nem értelmezhetőek egyéni gondolatokként – ugyanis az egyének fejében is mások gondolataiként nyerik el értelmüket. Önmagában annak nincs jelentősége az én viselkedésem szempontjából, hogy én egyetértek-e egy szabállyal.

Az erkölcs új pszichológiájának egyik lényegi következtetése volt, hogy többféle erkölcs létezik. Azonban azt is állította, hogy megszámlálhatóan sok erkölcsi elv van – feltételezett evolúciós folyamatok határozták meg az erkölcsi elvek körét. A normák játékelméleti modelljei ebben a tekintetben jóval tovább mennek. Mivel az embereket nem korlátozzák belső erkölcsi elvek, a racionális döntéshozók közötti játszmák laza keretei között elvileg bármi erkölcsössé válhat. Erkölcsös döntés az, amiről egy csoportban mindenki azt gondolja, hogy erkölcsös, és mindenki reálisan várja, hogy a többiek meghozzák e döntést – azaz a döntés meghozatala nem túl költséges a cselekvők számára (Bicchieri 2006, 2016). Az erkölcsi pluralizmus közös jellemzője a két, egymással élesen szemben álló metodológiai alapokra épülő, 21. századi erkölcselméletnek (a pszichológiainak, illetve a játékelméletinek).

A normakövetés játékelméleti modelljeinek ambíciója, hogy teljessé tegyék az erkölcsi viselkedés szituációs logikai elemzését. A teljességhez azonban szükséges az is, hogy elszakadjanak az önző polgárok együttműködését leíró fogolydilemma modellektől. Továbbá fel kell tenni a kérdést, hogy mi történik azokban a helyzetekben, ahol szabad szemmel erkölcsi viselkedésmintákat látunk, ám a játszmán belül nehéz felfedezni a racionális normafenntartáshoz szükséges, végtelen ismétlődési láncolatot. Az ilyen szituációk létezését elismerő kurrens elméletek azt feltételezik, hogy valamiféle erkölcsi motiváció mégiscsak működik az emberekben. Azonban ez nem az emberi játszmák konkrét kimenetelére irányuló preferencia, hanem a közös tudatban létrejövő szabályok iránti lojalitás (Gintis 2009, 2016). Tehát a normakövetés végső biztosítéka – a komprehenzív szituációs logikai elemzés ellenére –, az ember veleszületett normakövetési vágya. Kétségtelen, hogy a méltányos polgár modelljével kapcsolatban végzett számos empirikus teszt eddig még nem cáfolta az inherens normakövetési hajlandóságra vonatkozó elméleti állításokat. Azonban nem is ez az állítás állt eddig az empirikus vizsgálatok célkeresztjében.

### I.3.2.2. Egy szociológiai játékelmélet: önérdék és racionális erkölcs beágyazott játszmákban

A nyomozás eddigi történetét lezáró VII. fejezetben fejtem ki a saját téziseimet. A fejezet célja, hogy úgy egészítse ki a normakövetés korábbi fejezetben bemutatott játékelméleti modelljét, hogy a szituációs logikai elemzés minden szempontból teljes legyen, és a környezet pontos leírásának köszönhetően szükségtelenné váljon belső, erkölcsi motivációk feltételezése. A teljes leírás révén visszatérhetünk az önző polgár modelljéhez, ahol az erkölcs belső forrása nem az erkölcsi motiváció, hanem az odafigyelés a hosszú távú perspektívákra, mások érdekeire, és a közösségekben kialakuló szabályokra.

A játékelméleti normaelméletek általam javasolt kiegészítése az emberi játszmák egy olyan modelljén alapul, amelyben a polgárok egy-egy erkölcsi jelentőségű döntésükkel két játszmában vesznek részt egyszerre. Az egyik játszma az a konkrét helyzet, amelyre a közösségekben kialakított erkölcsi szabályok vonatkoznak. A másik játszma egy metajátszma, amely az aktuális szituáció mellett az adott közösségekben az egyénre váró összes jövőbeni játszmát tartalmazza. Az aktuális játszma lehet bármilyen struktúrájú szituáció, a legkülönbélebb érdekeltségekkel és döntési lehetőségekkel. A metajátszmában azonban csak a közösséghez tartozást jelző „erkölcsös”, és az azt elutasító „erkölcstelen” döntés között lehet választani. Az, hogy mi az erkölcsös, azt az aktuális konkrét játszma szabályai mondják meg. Egy metanorma határozza meg, hogy miképpen kell elbánni azokkal, akik nem tartják be az egyes konkrét játszmák erkölcsi szabályait. Ebben a kettősjátszma-modellben kezelhető a társas helyzetek sokfélesége, miközben meg lehet tartani a foglydilemma szituációs logikai elemzése során feltárt, az interakciók ismétlődésén nyugvó normafenntartó fenyegetési láncot.

A könyv konklúziója az, hogy akár belső erkölcsi motivációk nélkül is ki lehet alakítani és fenn lehet tartani erkölcsi rendet. Azonban, ha nincsenek belső erkölcsi motivációk, az erkölcsi szabályok elveinek kevés dolog szab határt. A vonatkozó szakirodalomban fontos kérdés, hogy mégis milyen elvek formálódhatnak erkölcsi sza-

bályokká, ha azok nem születnek velünk. Az erkölcs forrását kereső nyomozati munkát bemutató tanulmányom azonban ezen a ponton véget ér, és későbbi tanulmányokra hagyja a kérdést.

### I.3.3. ÖSSZJÁTÉKOK, EREDMÉNYEK

E könyv elegáns és tömör címe akár az is lehetne, hogy „Az erkölcs forrása”. E cím utalhatna a legtömörebben a monográfiában vizsgált kérdésre: honnan jön az erkölcsi készítés? A választott címmel („Összjáték”) inkább az elemzés során született válaszokat hangsúlyozom, nem a vizsgálandó kérdést.\* Tanulmányom konklúziója három szinten is összjátékokat azonosít.

Mindenekelőtt a vizsgálódás konklúziójaként leírt kettős játszma modelljében világossá válik, hogy miközben az erkölcsös és önérdékkövető viselkedés fogalmilag értelmesen szétválasztható, erkölcs és racionalitás kéz a kézben jár. Szélesebb kontextusban értelmezve a rövid távú önérdékkövetéstől megkülönböztetett erkölcsös viselkedés is racionális. Erkölcs nélkül vagyunk racionálisak, amikor egy élethelyzetben a saját érdekeinket követjük. Ez az élethelyzet értelmezhető hosszú távú játszmaként vagy játszmasorozatként is; továbbá az önérdékkövetés magában foglalhatja a szituáció társutasainak megértését, és a feljüket – saját érdekünkben –, tett gesztusokat is. Ezzel szemben erkölcsösen vagyunk racionálisak, amikor globálisan gondolkozunk, és – a saját érdekünkben –, egy csoport tagjai akarunk maradni. A csoporttagság sok-sok helyzetben realizálódó előnyeit akkor élvezhetjük, ha az egyes konkrét élethelyzetekben betartjuk az adott helyzetre vonatkozó, a csoport tagjai által érvényesnek gondolt erkölcsi szabályokat.

Továbbá erkölcs és racionalitás összjátékának megértéséhez fel kell adnunk a társadalmi jelenségek individualista és kollektivista megközelítésének éles szétválasztását. Az emberi játszmák, illetve a játszmákban érintett csoportok individualista és kollektivista szem-

\* Tardos Róbert javasolta e koncepció címbe építését.

lélete együtt magyarázza meg a racionális erkölcsöt. A betartandó erkölcsi szabályok nem csupán az egyes egyének fejében léteznek. Ha csak ott léteznek, akkor nem bírnak valódi erkölcsi erővel. A normák játékelméleti vizsgálata megmutatja, hogy ezek a szabályok a csoport-tagok közötti kommunikációs térben jönnek létre és maradnak fenn. Az egyének fejében létező vélekedések nem magukra a szabályokra vonatkoznak, hanem a többiek vélekedésére arról, hogy mások hogyan vélekednek a szabályok elfogadottságára vonatkozó vélekedésekről. Az individualista megközelítés segít annak megértésében, hogy a mechanikusan nem dezaggregálható kollektív tudat miképpen jön létre egyéni gondolatok interakciójaként, és hogyan tartja benn azt a közös térben az egyéni racionális reakciók sokasága.

Végül, az egyéni és kollektív gondolatok interakciójának megértésével világossá válik, hogy az erkölcsös cselekvés kevésbé tekinthető ideák üldözésének, sokkal inkább közösségek egymásra utalt tagjainak egymáshoz illeszkedéséből jön létre. Az illeszkedés a közös kommunikációs térben fenntartott közös mítoszokhoz való igazodáson keresztül történik. Az erkölcsi szabályok a tagok egymáshoz illeszkedéséből létrejövő összjáték szabályai.

Az összjáték ez utóbbi koncepciója rávilágít a tanulmány egyik alapvető, eddig nem hangsúlyozott tézisére. Nevezetesen: erkölcsi jellegű belső indíték hiányában is körülhatárolható az erkölcsös viselkedésre készítő belső képesség. Ez a képesség valamiféle „társadalmi térlátás”, amely segít felismerni a velünk interakcióban lévő cselekvők céljait és felénk irányuló elvárásait. Az alábbi elemzés fontos állítása, hogy ez az odafigyelési képesség adja az erkölcsös viselkedés belső forrását. A betartandó erkölcsi szabályok természetét ellenben a csoporttagok érdekeit közvetítő diskurzus alakítja ki.

Kezdődjék hát a nyomozati cselekmények bemutatása!

## II. 1944\* – A racionális polgár erkölcsé

### II.1. Az önérdék

#### II.1.1. NÉHÁNY SZÓ A(Z ELMÉLET)TÖRTÉNETI KONTEXTUSRÓL

Európában a korai kapitalista társadalmak lakói (különösen – de nem kizárólag –, az a növekvő csoport, amelyet akkor és/vagy később polgárságként azonosítottak), a túlélés érdekében hozott mindennapi döntéseik során az elemekkel való küzdelem, és saját kisebb közösségükben történő lavírozás mellett egyre többet foglalkoztak a többé-kevésbé idegen emberekkel folytatott, hosszabb-rövidebb együttműködésekkel. Ez utóbbi tranzakciókat egyre nagyobb arányban jellemezte, hogy az együttműködések pontos terminusai nem voltak szilárd szabályok által kőbe vésve. E helyzetekben a megtanult rutinokba, a szokáserkölcsbe és más szabályokba kapaszkodás helyett egyre gyakrabban kellett a polgároknak explicit módon feltenniük maguknak a kérdést, hogy mi is a saját érdekük egy-egy helyzetben. Továbbá az egyének egyre inkább képesek voltak pénzben kifejezni a saját egyéni érdekeiket ezekben az tranzakciókban – és rá is kényszerültek erre.

Az időszakban kibontakozó új világi politikai filozófia képviselőit egyszerre rémítették meg és készítették csodálatra azok a kihívások, amelyeket sok, egymásnak idegen és eltérő célokat követő ember

\* Von Neumann, J., & Morgenstern, O. (1944): *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

együtműködése, és azok lehetséges megoldásai jelentettek (Hobbes 1970 [1651], Smith 1992 [1776]). Gyorsan körvonalazódott az elképzelés, hogy az akkoriban megerősödő államok, és az általuk fenntartott formális intézmények nélkülözhetlenek az idegenek közötti tranzakciók olajozásában. Azt is hozzátették, hogy a formális intézmények mellett saját belső erkölcsi korlátaink szabhatnak határt a tranzakciókban a nyers érdekérvényesítésnek, fenntartva a békés együtműködést (Smith 2010 [1759]).

Később, a 19. század végén, a szociológia intézményesülése annak a tézisnek a kifejtéséhez köthető, hogy az élet sokkal gazdagabb az állami szabályozás által lefedett világnál, és a lefedetlen területet nem a belső jószágunk, hanem társadalmi kötelékeink szabályozzák (Durkheim 2001 [1893]). Az egyik ilyen kötelék a (gazdasági) tranzakciók kényszereiből adódik, a másik köteléket a közösségek/társadalmak tagjainak fejében lévő közös képzetek ereje alkotja. A diszciplínát megalapozó durkheimi hitvallás alapján a szociológia feladata e cselekvést befolyásoló struktúrák keresése, figyelése és megértése.

A másik társalapító, Max Weber (akit a közgazdászok, jogászok és politikatudósok is saját klasszikusuknak tekintenek) osztotta Durkheim érdeklődését az állami intézményeken túli érdekszabályozó mechanizmusok iránt. Ugyanakkor e mechanizmusok kutatása kapcsán a klasszikus politikai filozófia és az emelkedő közgazdaságtan individualista nézőpontját védte a durkheimi állásponttal szemben – rendkívül színvonalasan. Arra hívta fel a figyelmet, hogy a közös képzeteink is csak a fejünkben léteznek, és ilyenformán egyéni jellemzőként érdemes vizsgálni őket (Weber 1987 [1922]). Ennek következtében a társas lét szabályozó mechanizmusait, köztük az erkölcsöt is, az emberek fejében kell keresni.

Durkheim (ahogy Weber is) nagyon érzékenyen reagált a (nagy)-városi modern élet új, speciális kihívásaira. Azonban módszertani kollektivistá közeli módja nem találkozott a korszellemmel (diszciplínárisan is a modernitás által nem érintett társadalmakat kutató kulturális antropológiához kötődött inkább a közgazdász-jogász Weberrel szemben). A fent vázolt bizonytalan környezetben a szellemi életet, például a kortárs irodalom és képzőművészet alkotóit és

közönségét is az egyéniség titkának feltárása izgatta, és utána még sokáig ez határozta meg a kulturális átalakulások trendjeit. Ma már kimondhatjuk hogy a 20. században a (tárgan értelmezett) társadalomtudomány individualista megközelítései is jóval több követőt szereztek, mint az összefoglalóan kollektivistának nevezett irányzatok. Az erkölcs gyökereinek kutatásában is nagyobb szerepet kapott az individualista irányzat.

Néhány évvel Durkheim, majd Weber halála után, 1926. december 7-én, egy huszonhárom éves, frissen doktorált magyar matematikus, Neumann János előadást tartott kollégáinak Göttingenben *Zur theorie der gesellschaftsspiele (A társasjátékok elméletéről)* címmel, melyet két évvel később jelentetett meg írásban (Neumann 1928). Az előadásában arra tett javaslatot, hogy miképpen lehet egyszerűen modellezni, és ezáltal elméletben megoldani az emberi játszmák végtelen regresszusnak tűnő problémáját, miszerint én annak függvényében lépnék, hogy te mit lépsz, de te annak függvényében lépsz, hogy én mit lépek. Tehát nekem úgy kell lépnem, hogy reagálok arra, ahogy te reagálsz arra, ahogy én reagálok a te lépésedre – amelyik maga is reagálás (és így tovább a végtelenségig). Azt mutatta meg – egy viszonylag speciális esetre vonatkozóan –, hogy létezik megoldás a végtelenségű regresszus feloldására, azaz lehet olyan lépéseket javasolni nekem és neked, amelyek tényleg kölcsönösen reagálnak a másik fél reagálásokra reagáló lépésére anélkül, hogy végig kellene gondolatban haladni a reagálások végtelen sorozatán.

Szerény ismereteim szerint Neumannt a legcsekélyebb mértékben sem érdekelte a társadalmak vizsgálatának individualista és kollektivisták megközelítései között feszülő ellentét, amely egymással szemben álló táborokba terelte Durkheimet és Webert, majd utánuk rengeteg szociológust és más társadalomtudóst. Ettől függetlenül ma sok társadalomtudós véleménye az, hogy Neumann megoldása, mellyel lefektette a játékelmélet alapjait, olyan eszköztárat adott a társadalomkutatók kezébe (és az eszköztár használatától függetlenül: olyan szemléletmódra irányította rá a figyelmet), amely sok-sok egyéb alkalmazási lehetősége mellett képes feloldani az erkölcs és a társadalmi rend forrásával kapcsolatos vitát is.



De ne szaladjunk előre! Ebben a kötetben végig szeretném vezetni az olvasókat annak a kalandos nyomozásnak a fontos lépésein, amelynek során a weberi, módszertani individualistának (Orthmayr 1997) is nevezett társadalomkutatási irányzatok képviselői az erkölcs forrását keresték az emberben az elmúlt háromnegyed évszázadban; és amelynek jelenlegi fázisában egyesek még mindig a lélekben próbálnak még mélyebbre ásni, míg mások immár inkább az emberek közötti térben kutakodnak, egyre közelebb kerülve a Durkheim által százhusz éve bemutatott állásponthoz – amelynek tagadása a bennünk rejlő erkölcsi processzor keresésének egyik fő mozgatórugója volt („valahol itt házon belül kell lennie, kint nem lehet!”). Ám ez a könyv nem egy hiábavaló nyomozásról szól. Azt szeretném, ha a lépésről lépésre haladó áttekintésem világossá tenné, hogy csak a bennünk rejlő erkölcsi motiváció keresése során magunkról szerzett ismereteink teszik érthetővé az emberek közötti térben kialakuló közerkölcs száz éve még rejtélyesnek tűnő erejét.

Az olvasók látni fogják, hogy az emberben rejlő (állampolgári) erkölcsi érzék utáni nyomozás általam jelentősnek tekintett lépései nagyrészt a szociológiai diszciplínán kívül történtek. Ugyanakkor a durkheimi tézisekhez közelítő erkölcskutatók ma már a szociológusoknál kopogtatnak a hosszú kutatási program eredményeiből lesűrhető tanulságokkal (Gintis 2016). Kérdés, hogy e diszciplína képviselői be tudják-e fogadni az erkölccsel foglalkozó elméleti és empirikus játékelméleti kutatások eredményeit, és hajlandóak-e használni e kutatások elemzési eszköztárát, hogy hozzátegyék a legjobb szándékok ellenére is hiányzó szociológiai szemléletmódot és tudást az erkölcs játékelméleti alapú, interdiszciplináris kutatási programjához?

## II.1.2. A RACIONÁLIS DÖNTÉSI MODELL A KÖZGAZDASÁGTANBAN

A mai tankönyvekből tudjuk, hogy a piaci szereplők viselkedését elemző mikroökonómia alapjait a 20. század közepén rakták le. A második világháború után indult hódító útjára az axiomatikus alapokon nyugvó neoklasszikus közgazdaságtan, amely téziseit a ma-

tematika nyelvén fejezi ki. Az emberi döntések logikájának matematikai nyelvre fordításában oroszlánrészt vállalt Neumann János magyar matematikus, és Oscar Morgenstern osztrák közgazdász a játékelméletet megalapozó, 1944-ben megjelent művükkel (Neumann és Morgenstern 1944). Alaptételeiket nem csupán az akkoriban még a közgazdaságtan perifériáján elhelyezkedő játékelmélet alkalmazta, hanem azok minden mikroökonómiai modell építőköveivé váltak. Ugyanakkor a szakmára, és annak matematikai orientációjára ebben az időszakban közvetlenebb és nagyobb hatást gyakorolt Paul A. Samuelson *Foundations of Economic Analysis* című műve (1947), amely a matematikai közgazdaságtan egyfajta manifesztója volt, és a következő évtizedek közgazdasági tankönyveinek mintájává vált. Ezután nagyon gyorsan a diszciplína perifériájára szorultak a kizárólag verbális érvelést alkalmazó elméletek. A közgazdaságtan gyorsan és visszavonhatatlanul csatlakozott a matematika nyelvét értő és beszélő tudományok csoportjához.

Az alapfeltevések és a modellezési logika rendkívül sikeresnek bizonyultak. A közgazdaságtan gyorsan fejlődött; egyre több gazdasági és társadalmi folyamatra kínált vonzó magyarázatokat. Ezt a gyors fejlődést és a „tudomány univerzális nyelvének” használatát ismerte el a Nobel Alapítvány, amikor az 1968-ban befogadta a svéd jegybank által alapított közgazdasági Nobel-éremdíjat. Ma a legkiválóbbnak minősített közgazdászokat évente együtt ünneplik a természettudományok legkiválóbb tudósaival. A Nobel-éremdíj nemcsak elismerést, de valószínűleg újabb lökést is adott a közgazdaságtudománynak, és tovább növelte az axiomatikus, matematikai nyelven megfogalmazott elméletek súlyát a diszciplínán belül.

Az elemzésem szempontjából külön említésre méltó, hogy az ember viselkedésének új típusú közgazdasági modelljeit néhány kutató gyorsan elkezdte használni politikai kérdések analizálására is. A jóval később Nobel-díjjal jutalmazott Thomas Schelling és Harsányi János már az ötvenes és hatvanas évek fordulóján foglalkozott a korábban csak politikatudósokat érdeklő dilemmák racionális döntéseméleti elemzésével (Harsányi 1962, Schelling, 1960). E dilemmák elsősorban tárgyalásokhoz, alkukhoz és konfliktusokhoz kapcsolódtak. Még

érdekesebb Anthony Downs (1990 [1957]) korai elemzése a választói viselkedés természetéről. Downs a pártok ideológiai versengését, a pártprogramok összeállításának stratégiáját vezette vissza a közgazdasági modelleknek megfelelő, racionálisan mérlegelő választók preferenciáira. A racionális fogyasztó és vállalkozó modellje mellett tehát nagyon gyorsan megjelent a racionális választópolgár és politikus matematikai modellje is.

A modern racionális döntések elméletében az egyéni döntések matematikai megfogalmazásából eltűntek az ember kettős természetére (egyszerre önző és morális lény) vonatkozó feltevések, amelyek még a 20. század elején is jelen voltak a gazdasági cselekvésről szóló elméletekben (Pareto 1917, Weber 1987 [1922]). Azonban a kettős természetre vonatkozó feltevéseket nem az egyféle, önző motivációra vonatkozó feltevések váltották fel. Az ekkor kidolgozott általános modellek nem szigorították a motivációs feltevéseket. Mi sem állt távolabb tőlük. A neoklasszikus közgazdaságtan (ahogy idővel elnevezték a modernizált diszciplínát) alapmodelljei *nem* tartalmaznak motivációs feltevéseket. Ahogy az I. fejezetben hangsúlyoztam: a racionális döntésemélet (és az arra épülő neoklasszikus közgazdaságtan) nem a racionálisan meghozott döntések modellje, hanem az emberi döntések általános, racionalista elmélete.

A már említett Samuelson által kidolgozott, és a fogyasztói viselkedés közgazdasági elemzésének alaptézisévé vált kinyilvánított preferencia elmélete (Samuelson 1948) bármilyen cél követését megengedi. Az általános elmélet szintjén a neoklasszikus közgazdaságtan és a játékelmélet sem foglalkozik az emberek motivációinak és vágyainak természetével. Buchanan (1992 [1969]) megkülönbözteti a „döntések logikáját” az erős motivációs feltevésekkel rendelkező magatartási modellektől. Egyértelmű, hogy a racionális döntés múlt század közepén lefektetett általános elmélete a döntések logikáját írja le. Empirikus feltevései a döntések koherenciájára vonatkoznak, és a modell korai kritikái is a döntések inkoherenciáját próbálják bemutatni (például Ellsberg 1961).

Már Max Weber (1987 [1922]) megkülönböztette egymástól a célok és az eszközök racionalitását. Weber terminológiáját alkalmaz-

va azt mondhatnánk, hogy a neoklasszikus közgazdaságtan általános cselekvésemélete semmilyen megszorítást nem tartalmaz az egyéni célokra vonatkozóan. A racionális döntésemélet e matematikai modellje a célokból, lehetőségekből és korlátokból logikusan következő döntéseket határozza meg. A célokról a konkrét vizsgált jelenségre vonatkozóan kell feltevéseket megfogalmazni, hogy az optimális döntés ezekből levezethető legyen. A downsi modellt tekintve is egyértelmű, hogy a racionális választó általa leírt modellje *nem* az önző választó modellje. Downs racionális választói nyugodtan mehetnek szembe saját anyagi érdekeikkel. A lényeg, hogy saját céljaikkal összhangban választanak pártot maguknak. Az egyéni céljaik és az (anyagi) érdekeik viszonyára vonatkozóan a modell nem tartalmaz feltevéseket.

Az új közelítésmód korai támogatói közül néhány jelentős kutató élesen tiltakozott az ellen is, hogy a racionális döntés modelljét az emberi döntési folyamatok módjára vonatkozó empirikus modellnek tekintsék. Milton Friedman (1953) klasszikus tanulmányában érvelt a fény felé törekvő növény racionalitása mellett. A racionális döntés modellje – mondja Friedman –, a célokból levezethető végeredményt mutatja meg, és nemcsak a célokról hallgat, de a célok eléréséhez használt eszközökről is. Ha definiáltuk egy növény célját (fény felé kíván növekedni), és ismerjük viselkedésének korlátait (például fizikai akadályok a fény felé vezető úton), akkor előre tudjuk jelezni növekedésének irányát. A döntési folyamat racionalitására vonatkozó empirikus feltevések feleslegességének másik nagy hatású hirdetője volt Armen Alchian, aki az evolúciós közgazdaságtan egyik megalapozója volt (Alchian 1950). Alchian érvelése szerint a nem racionális viselkedésformák kiszelektálódnak, és aggregált szinten racionális reakciókat fogunk tapasztalni akkor is, ha egyéni szinten nem is jellemző a racionális döntéshozatal.

A sok kritikus által a neoklasszikus közgazdaságtannak tulajdonított, szigorú motivációs feltevések a szűkebb területekre alkalmazott, konkrét magyarázó modellekben jelentek meg. Leginkább kiterjesztően talán épp az imént említett Alchian alkalmazta az anyagi önértékét követő döntéshozó modelljét. Ő a legkülönbözőbb ver-

sengő piacok szereplőivel kapcsolatban vetette fel általánosságban, hogy populációs szinten, a piaci szelekció hatására, a profitérdekeltségből levezethető racionális döntéseknek megfelelő, (átlagos) viselkedéssel reagálnak a környezeti korlátokra és lehetőségekre. De mint említettem, amikor például a politikai viselkedésre alkalmazták a neoklasszikus közgazdaságtant, nem feltételezték, hogy kizárólag az anyagi önérdék motiválja az emberek döntéseit (Downs 1990 [1957]). Továbbá amikor Gary Becker, a „közgazdaságtudományi imperializmus” legjelentősebb képviselője (az 1992. évi közgazdasági Nobel-díj kitüntetettje) a családi életet kezdte elemezni a neoklasszikus közgazdaságtan eszköztárával (Becker 1974, 1981), értelemszerűen hozzáigazította motivációs feltevéseit a kutatott területen jellemző cselekvési mintázatokhoz. Az ő nevéhez fűződik az altruizmus beemelése a közgazdasági modellezésbe (Becker 1974; de lásd még Phelps 1975). Sőt, már az ötvenes években is felvetette a diszkriminációról írt disszertációjában, hogy egyes gazdasági szereplők nemcsak a profitjukat kívánják maximalizálni, hanem boldogságot okoz nekik, ha ki tudnak zárni bizonyos társadalmi csoportokat a környezetükből (Becker 1957).

Tagadhatatlan, hogy a neoklasszikus közgazdaságtan kritikusai sokszor tulajdonítottak leszűkítő motivációs feltevéseket a racionális döntéshozatalnak a múlt század második felét végigkísérő metaelméleti vitákban (például Green és Shapiro 1996, England 1989). Nem véletlen, hogy a hetvenes években, az első, e tárgyban született, tisztázó szándékú metodológiai elemzések után Gary Becker a később szintén Nobel-díjjal jutalmazott Georg Stiglerrel próbálta újra megfogalmazni az Amerikai Közgazdasági Társaság lapjában, hogy a neoklasszikus közgazdaságtan által alkalmazott racionális döntéshozatali modell semmit nem mond az emberek céljairól (Becker és Stigler 1977: „De gustibus non est disputandum”).

A kritikák mindazonáltal nem légből kapottak voltak. A folyóiratokban megjelenő racionális döntéshozatali elemzések döntő többsége feltételezte, hogy az egyének az általuk vizsgált konkrét társadalmi helyzetben anyagi önérdelmeiket követik (Fontaine 2006). Ez annyiban nem meglepő, hogy a racionális döntéshozatali elméletet elsősorban a

döntően piaci folyamatokkal foglalkozó közgazdaságtan alkalmazta. Annyiban viszont magyarázatra szorul az önzőségre korlátozó motívációs feltevések dominanciája, hogy ahogy már Becker feltételezte az ötvenes években (Becker 1957), és ahogy Fehr és követői hangoztatták a kilencvenes évektől, a gazdasági élet tranzakcióit is átszövik a direkt profitérdekeken túlmutató megfontolások.

Max Weber (1987 [1922]) már a század elején rámutatott, hogy az anyagi érdekeket követő gazdasági cselekvések a legjobban előrejelezhető viselkedésformák, ezek engedik magukat a legkönnyebben tudományos vizsgálat tárgyává tenni. A gazdasági élet döntési helyzetekben az anyagi érdekeket befolyásoló tényezőket lehet a legkönnyebben és a legeggyértelműbben beazonosítani és mérni. Az árak, bérek és erőforrásköltségek a gazdasági szereplők gondolkodásának megismerése nélkül is rendelkezésre állhatnak. E tényezők kapcsolódása az önérdekhez univerzális, az előrejelzések készítéséhez nem szükségesek a vizsgált tranzakció kulturális közegére és a vizsgált szereplők lelki beállítottságára vonatkozó információk. Versengő piacokon a szereplők kognitív képességei sem érdekesek: a piaci szelekció racionálisan „dönt” a korlátokhoz és lehetőségekhez történő alkalmazkodásról.

A neoklasszikus közgazdaságtan művelőinek közössége kritizálható volt amiatt, mert elfeledkeztek az önérdeken túlmutató motívációkról, és a piaci viselkedés elemzésekor szinte mindig mindenki az anyagi önérdek szerepére koncentrált. Azonban az önérdek kutatásokban tapasztalható dominanciája analitikusan nem vezethető le a neoklasszikus közgazdaságtan által alkalmazott racionális döntéselméleti alapmodellből. Továbbá nem mindenki feledkezett el más célokról, és nem minden területen zárták ki az anyagi önérdeken túlmutató motívációkat. A 20. század második felében jelentősek voltak, és nagy hatást gyakoroltak azok a részben itt is megemlített racionális döntéselméleti kutatások is, amelyekben az emberek anyagi önérdekén túlmutató célok is nagy szerepet kaptak. Mivel a társas preferenciák kutatási programja az erkölcs és a jóindulat fogalmából indult ki (erkölcs alatt a kutatói közösség által fontosnak tartott morális célokat értve), kevésbé volt feltűnő számukra, hogy például

a diszkrimináció iránti vágy következményeinek vizsgálata jelentős szerepet kapott éppen azon a kutatási területen, a munkaerőpiac és a munkaügyi kapcsolatok vizsgálatában, ahol Fehr és munkatársai később megalapozták a társas preferenciák kutatási programját (lásd például Arrow 1973, 1998, Cain 1986). Korai munkájában már Becker (1957) is elkülönítette a profit maximalizálását szolgáló statisztikai diszkriminációt, a profitérdekekkel szemben álló, diszkrimináció iránti vágytól. Ráadásul sok ember amorálisnak tartja a diszkriminációt, és csak bűnös ösztönként vagy döntési hibaként szeretné kezelni, ezért számukra – sok kulturális kontextusban és sok élethelyzetben –, a diszkrimináció ugyanúgy erkölcsi elvekből vezethető le, mint a javak egyenlőségének eszméje (vö. Haidt 2012). Gondoljunk a „hazai termelőket” vagy „határon túli magyarokat” preferáló egyéni és kollektív döntésekre!

A fentiekén túlmenően, az anyagi önérderekre koncentráló elemzések sem feltételezik, hogy csak az önző motivációk befolyásolják a vizsgált gazdasági szereplők viselkedését. Csupán azzal a feltevésével élnek, hogy az anyagi önérderekre koncentrálva a viselkedések változékonyságának viszonylag jelentős hányadát meg tudják magyarázni. Ne felejtjük el, hogy a neoklasszikus közgazdaságtan virágzása a második világháború után indult. Racionális kutatókat feltételezve (vagy legalább Alchian megengedő, evolúciós értelmezésében, egy kompetitív tudományos piacot feltételezve) érthető, ha az új eszköztár alkalmazásának első néhány évtizedében a legkönnyebben mérhető, és a viselkedésminták változásának viszonylag nagy hányadát magyarázó motivációs tényezőre koncentrálnak a kutatások. Amikor a kizárólag az önérderekre fókuszáló újabb gazdasági modellek addicionális magyarázóereje (közgazdasági terminológiával határhaszna) fokozatosan lecsökken, egyre többek szerint érdemes az újabb motivációs elemek bevonásával kísérletezni, melyek kezdeti alkalmazásai relatíve nagy határhasznot hozhatnak. A társas preferenciák kutatási programjának növekvő népszerűsége felfogható egy ilyen szerves fejlődési folyamat eredményének is. A történet ilyen keretezése szemben áll a Fehr és kollégái által kínált, az ősökhoz visszatérő, hanyatlást megállító lázadási narratívával.

Az általam leírt fejlődési folyamat összhangban van a Lindenberg (1992) tudományfejlődési modelljében megfogalmazott csökkenő absztrakció elvével. Ennek értelmében egy tudományos paradigma a kezdetben általános gondolkodási és elemzési kereteket meghatározó, ún. kevésbé tartalmas modellek kidolgozásától halad az egyre több, empirikusan cáfolható feltevést és predikciót magukban foglaló, tartalmasabb modellek fejlesztéséig. A korábbi eredmények nem egészében cáfolódnak meg az új elméletek által, hanem előrejelzéseik hatóköre szűkül, vagy éppen nem történik más, csak pontosabbá válnak az előrejelzések az új szempontok bevonásával, a valóság komplexitásához közelebb álló modellek alkalmazásával.

Állításom tehát az, hogy a neoklasszikus közgazdaságtannak nem volt olyan általános érvényességet célzó állítása az emberi motivációról, amelyet a morális megfontolások nyomát kereső kutatások eredményei cáfolni tudtak volna. Kétségtelen ugyanakkor, hogy a közgazdasági kutatások többsége a neoklasszikus közgazdaságtan fejlődésének első öt évtizedében kevés figyelmet szentelt az anyagi önérdéken túlmutató preferenciáknak, a morális megfontolások hatásának.

A neoklasszikus közgazdaságtan racionális döntéelméleti modelljének ugyanakkor valóban volt egy olyan sajátossága, amely individualizálta az elemzések szemléletmódját, és jelentősen szűkítette a döntést befolyásoló különböző motivációk értelmezését. Ez a sajátosság érdemtelenül kevés figyelmet kap az emberi motivációk modellezéséről szóló vitákban. A kortárs közgazdaságtan már rég túl lépett a neoklasszikus közgazdasági döntési modelljének e korlátján, de a gondolkodási kereteket még befolyásolja, és jelentős hatása volt a társas preferenciák kutatási programján dolgozók gondolkodására is (vö. Binmore 2010). Mielőtt e korlát bemutatására rátérek, bemutatom azt, hogy az egyéni anyagi érdekekre fókuszáló elemzések túlsúlya nem csak a közgazdaságtani kutatások területét jellemezte a háború utáni időszakban. Emiatt is indokolatlan az önző ember modellje miatt támadni a racionális döntéelméletet.



## II.1.3. ÖNÉRDEK A STRUKTURALISTA ELEMZÉSSEN

A társas preferenciák utóbbi időben népszerűvé váló kutatási programjának neoklasszikus közgazdaságtannal szembeni kritikája több olyan elemet is tartalmaz, amelyet a társtudományok képviselői, különösen egyes szociológusok már jóval korábban felvetettek (England 1989). Röviden érdemes kitérni arra is, hogy a (politikai) szociológia hogyan tekintett az önzőség és erkölcsösség kérdésére.

Ahogy a közgazdaságtan, úgy a szociológia esetében sem térek ki a 20. század első felében lezáruló, klasszikus korszak képviselőinek nézeteire. A szociológiai elmélettörténet ismerői számára világos, hogy a klasszikus közgazdaságtan alapjaival foglalkozó tudósokhoz hasonlóan a szociológia megalapozói – pontosabban akiket ma annak tekintünk –, és egyes képviselői is hangsúlyozták az ember kettős természetét. A korábban idézett weberi cselekvéseméletben a célracionális és az értékracionális döntés distinkciója mutatja ezt (Weber 1987 [1922]).

Ugyanakkor úgy látom, hogy a neoklasszikus közgazdaságtannal szemben megfogalmazott szociológiai kritikák, az általuk sugalmazott helyzetek, valamint a Weber iránti, már-már vallásos tisztelet ellenére is sok hasonlóság van abban, ahogy a két diszciplína emberképe alakult a második világháborút követő évtizedekben.

Mindenekelőtt látni kell, hogy a világháborút követő két-három évtized a szociológia számára is a presztízs, a kutató közösség és a hallgatói létszám jelentős emelkedését hozta (Szelényi 2015). A szociológia felépítése sokkal kevésbé centralizált, mint a közgazdaságtané, és a sokközpontú kutatási térben kevésbé centrális az egyéni cselekvések motivációs gyökereinek problémája, mint a közgazdaságtanban. Az alábbiakban elsősorban a politikai szociológiára és a rokon területekre vonatkozó tendenciákat mutatom be.

A (politikai) szociológiában preferenciák helyett értékekre hivatkoztak. Míg a kinyilvánított preferenciák közgazdász kutatói elsősorban a (gazdasági) viselkedést figyelték meg, a szociológusok gyakran kérdezték meg az embereket a viselkedésüket (a feltételezésük szerint) motiváló értékeikről is. Az értékek struktúrájának kutatása

amúgy később, a nemzetközi tudományos kapcsolatok erősödésével és a számítógépes adatfeldolgozás fejlődésével kapott új lendületet a hetvenes évek végétől – ezúttal már politikatudósok (Inglehart 1981) és pszichológusok (Schwartz és Bilsky 1987) főszereplésével. De a közéleti attitűdök értékalapú motivációk feltételezésén nyugvó vizsgálata jóval megelőzte e fejleményeket (Berelson et al. 1954).

A racionális döntésemélet, és vele a neoklasszikus közgazdaságtan szociológiai kritikái nagy vehemenciával támadták az általuk megrajzolt, önző homo oeconomicus modelljét, noha a kortárs szociológia sem feltétlenül a weberi racionális-értékalapú distinkcióban gondolkodott. Amikor a szociológia az egyének sokszínű értékeiből indult ki, pontosan annyira ignorálta az önzőség-erkölcsösség dilemmáját, mint az akkori közgazdaságtan. A szociológia kutatási paradigmája meglehetősen közel állt Samuelson kinyilvánított preferencia elvéhez: az ember motivációiról a viselkedése ad információt, azt definiáljuk emberi értékeknek, és a neoklasszikus közgazdaságtanhoz hasonlóan az emberi viselkedés alapmodelljében nincs szerepe annak, hogy önző vagy önzetlen értékeket követnek-e az egyének. A modellezési és mérési eszköztár tette nagyon különbözővé a két diszciplínát. A közgazdászok, szemben a szociológusokkal, a döntések axiomatikus matematikai modellezésében hittek. Továbbá a közgazdászok nagyon sokáig nem akarták elhinni azt, amit az emberek mondtak, és csak arra voltak hajlandóak figyelni, amit csináltak. A szociológusok ellenben nagyon nagy figyelmet szenteltek az emberek beszámolóinak arról, hogy mit szeretnének. A szociológiai mérés középpontjába a kérdőív és az interjú került. Ezek a jelentős különbségek elfedték azt a hasonlóságot, hogy legsikeresebb expanziós időszakában egyik diszciplína sem szentelt kiemelt figyelmet az önzőség-önzatlenség kérdésének; modellezési paradigmáik semlegesek voltak ebben a tekintetben.

További hasonlóság a két diszciplína emberképének második világháborút követő fejlődésében, hogy a konkrét, erősen materialista szemléletű, alkalmazott kutatási programokban az önérdék empirikus definiálása nélkülözhetetlen volt. A szakszociológiákban a mobilitáskutatásokról a szegénységkutatásokig – a pénzügyi,

munka- stb. gazdaságtan kutatásaihoz hasonlóan –, az anyagi javak primátusát feltételezték. A kutatók az emberek helyzetét aszerint értékelték, hogy anyagilag hova jutottak (Aron 1950, Blau 1977, Ferge 1966).

A politikai szociológiában az osztályszavazás modellje volt a leginkább meghatározó közelítésmód (Alford 1962, 1963). E modell szerint az emberek jellemzően osztályérdekeik szerint alakították ki a közéleti kérdésekkel kapcsolatos véleményeiket. A szavazás kérdésében különösen feltűnő a diszciplínák közötti metaelméleti viták háttérében húzódó hasonlóság az egyéni döntéseket befolyásoló tényezőkre vonatkozó feltételezésekben. A közvélemény-kutatások által tesztelt statisztikai modellekben a racionális döntéseméleti hipotézisek nagyjából kongruensek voltak az osztályszavazás szociológiai modelljének előrejelzéseivel. Az átlagosnál kisebb keresetű, bérből és fizetésből élő emberek osztályérdekei e modellekben megegyeztek ezeknek az embereknek az egyéni anyagi érdekeivel. Ugyanakkor mind a neoklasszikus közgazdaságtanon nyugvó racionális döntéseméleti politikatudomány, mind pedig a szociológia elviekben megengedte, hogy a szavazók fejében más szempontok felülírják az anyagi önérdeket. Praktikusan feltételezték, hogy az anyagi érdekek fontosak, de elvileg más irányú preferenciák illetve értékek mégis eltéríthették tőlük az embereket (Berelson et al. 1954).

A két diszciplína közötti szemléletbeli különbség ott mutatkozik meg, hogy a közgazdászok által megengedett nem anyagi preferenciák belülről fakadnak. Pontosabban: nem akarnak azzal foglalkozni, hogy honnan származnak. Ezzel szemben a szociológia határozott álláspontja, hogy az értékek kívülről érkeznek, és a szociológusokat érdekli is az értékátadási folyamat. Ez valóban jelentős különbség, amely fontos szerephez jut majd az erkölcsös viselkedés motivációi utáni kutatásom konklúziójában.

A belső motivációk és a külső elvárások közötti különbség nem kapott fontos szerepet a politikai viselkedés elemzésében. De bizonyos más területeken határozottan megmutatkozott. Említettem már Becker (1981) családgazdaságtanát. Becker kiindulópontja – figyellem! –, a csoportérdek volt, hasonlóan az osztályszavazás marxi ih-

letésű modelljéhez. De a család tekintetében ez nehezen volt összeegyeztethető az egyéni érdekekkel

A közgazdaságtan és szociológia szemléletbeli különbségeinek illusztrálása szempontjából fontos az a Becker nevéhez fűződő motivációs elem, amely az akkori tudományos fejlemények alapján belső, velünk született adottságnak tekinthető (Becker 1974). Az altruizmus matematizált, evolúciós biológiai modelljeit ugyanis nem sokkal azelőtt dolgozták ki (Hamilton 1964). Bár azt is látni kell, hogy Becker a tartós együttélésen alapuló „önző” altruizmust ugyanúgy elképzelhetőnek tartotta (erről nemsokára szó lesz). Ugyanekkor a családi élet szociológiai vizsgálatára a legnagyobb hosszú távú hatással Sheldon Stryker (1968) családi (és egyéb) szerepelvárásokkal kapcsolatos elméletei voltak. Stryker modelljében nem genetikai kódjaink utasítanak az (anyagi) érdekeinkkel ellentétes családi munkamegosztás elfogadására, hanem a társadalom különböző szerepelvárásainak felelünk meg. A társadalmi nyomásnak engedünk. Ez a különbség segít megérteni azt, hogy miért tudta befogadni a közgazdaságtan a társas preferenciák kutatási programját – és hogy a szociológiai szemléletmód miért segít e program, illetve a vele rokon pszichológiai programok kritikai értékelésében.

## II.2. Mások érdeke

### II.2.1. JÁTÉKELMÉLET ÉS MÁSONK ÉRDEKEINEK FIGYELEMBEVÉTELE

Ahogy említettem, a racionális döntések matematikai elemzésének alapjait Neumann János és Oscar Morgenstern (1944) fektették le a játékelméletről szóló művükben. Kezdetben azonban a matematika nyelvén gondolkodó modern közgazdaságtan fejlődésére másoknak nagyobb befolyása volt. Például Paul Samuelsonnak (1948), aki a termodinamikai rendszerek analógiájára építve mutatta be a piacok működését. Neumann és Morgenstern (1944) a világmindenséget magában foglaló preferenciák matematizálhatóságát, és a racionális döntések matematikai kezeléséhez szükséges néhány alapvető axió-

mát szolgáltatták a közgazdászoknak, maga a játékelméleti gondolkodás azonban a közgazdaság-tudomány perifériáján maradt egészen a nyolcvanas évekig.

A játékelmélet a stratégiai gondolkodás elmélete. A neoklasszikus mikroökonómia modelljeiben ki-ki a saját céljait üldözi a környezeti tényezők adta lehetőségek és korlátok között. Ezzel szemben a játékelméleti modellekben az egyén figyelembe veszi, hogy a lehetőségeit korlátozó vagy kiterjesztő környezeti tényezőknek részei más gondolkodó lények is, akik hozzá hasonlóan a saját céljaikért küzdenek, és figyelembe veszik, hogy másoknak is saját céljaik vannak.

A társas szituációk játékelméleti modellezése nyilvánvalóan komplexebb, mint a piacok és a piacokon kívüli helyzetek hagyományos mikroökonómiai leírása. A második világháborút követő évtizedekben, a csökkenő absztrakció elvének (Lindenberg 1992) megfelelően, a neoklasszikus közgazdaságtan előbb használta ki azokat a lehetőségeket, amelyeket az egyszerűbb, sokkal könnyebben kezelhető, ún. parametrikus döntési modellek nyújtottak, és csak később terjedtek el szélesebb körben a játékelméleti alkalmazások – miközben számos területen továbbra is az egyszerűbb, hagyományos parametrikus döntési modelleket alkalmazták. A múlt század utolsó két évtizedében a játékelmélet alkalmazásának széles körű terjedéséhez nem kapcsolódott hasonló metaelméleti vita, mint az anyagi önértéken túlmutató preferenciák bevonását célzó törekvésekhez. Az utóbbi dilemmával kapcsolatban a társtudományok képviselői folyamatosan támadták a közgazdaságtan egészét (Camic 1979). A motivációs feltevéseket tagadó metaelméleti tanulmányok (Becker és Stigler 1977) részben e külső támadásokra reagáltak. A játékelmélet klasszikus mikroökonómiával szembenálló szemléletmódját nem támogatta meg hasonló, a diszciplínán kívülről szított, hosszú kritikai diskurzus. A játékelmélet „csendes forradalmának” egyik következménye azonban, hogy az önértékkövetésre vonatkozó feltevésekről szóló tudományos vitában elsikkadt a játékelmélet és a hagyományos mikroökonómia társas orientációja közötti különbség.

A neoklasszikus közgazdaságtanban korábban domináns, és ma is elterjedt hagyományos mikroökonómiai modellek döntéshozói akkor

is individualisták, amikor másokat akarnak megsegíteni. Ez a modell nem veszi figyelembe, hogy milyen vágyai vannak annak, akit a döntéshozó nem célzott meg külön a jó- vagy rosszindulatával. Sőt, a hagyományos mikroökonómiai döntéshozó számára a saját vágyai által érintett emberek akarata is zavaró tényező lehet; már az is megzavarja, ha a döntéshozó szerettei viszonszeretik őt, és ezért nem feltétlenül szeretnék, ha feláldozná magát értük. A játékelmélet alaptézise ellenben, hogy figyeljünk mások vágyaira. Az odafigyelés háttérében állhat teljesen önző motiváció is, de akkor is odafigyelés.

A neoklasszikus közgazdaságtan megengedi, hogy az egyént önzetlen vágyak vezéreljék, de csak individualista módon lehet önzetlen. Önzetlensége nemcsak szelektív, de egyoldalú is. Képes kezdeni valamit mások materiális fogyasztásával, de mások vágyaival már kevésbé; másoknak a feléje mint döntéshozó egyén felé irányuló elvárásaival pedig még kevésbé. A játékelméleti modellek döntéshozója lehet végtelenségig materialista és önző, azonban még saját önző materialista céljainak elérése során is átgondolja, hogy mi a jó másoknak, és kerül azokat a helyzeteket, amelyeket a vele interakcióban lévő emberek nem tartanának elfogadhatónak.

A fentiek végtelenül triviálisnak tűnhetnek mindazok számára, akik akár egy bevezető kurzus erejéig foglalkoztak már játékelmélettel. Mégis, ismereteim szerint a racionális döntéselmélet metaelméleti diskurzusából hiányzik annak elemzése, hogy a játékelmélet és a neoklasszikus közgazdaságtan hagyományos mikroökonómiája mennyire másképp modellezi az önzőséget és a morált (vö. Binmore 2010). Ebben a könyvben hangsúlyos szerepet fog kapni a hagyományos mikroökonómiától eltérő játékelméleti szemléletmód szerepe az erkölcsös cselekvések megértésében. Ezért fontos, hogy mélységében is megértsük a két szemlélet közötti különbséget, és azt, hogy a közgazdaságtani diskurzusban nem tárgyalják e fontos különbségeket.

### II.2.2. A NASH-EGYENSÚLY

Egy népszerű film egyik ikonikus jelenetének felhasználásával illusztrálom majd a neoklasszikus közgazdaságtan individualista közelítésmódja és a játékelmélet társas szemlélete közötti különbséget, és a (játékelmélethez értő) közgazdászok ezzel kapcsolatos értetlenségét. A játékelmélet Neumann és Morgenstern utáni legfontosabb elméleti megalapozója John Nash amerikai matematikus, aki az ötvenes években általánosította Neumann egyensúly-konceptióját (Nash 1950, 1951). Az 1994. évi közgazdasági Nobel-díj egyik kitüntetettjének életéről nagy sikerű (négy Oscar-díjat nyert) film készült 2001-ben, *Egy csodálatos elme* címmel, Russel Crowe főszereplésével. A film egyik ikonikus jelenetében a tudós legnagyobb hatású koncepcióját, a Nash-egyensúlyt próbálják bemutatni. A kikényszeríthető szerződések nélküli emberi játszmákat leíró játékelméleti modellek mindmáig e koncepció segítségével választják el a játszmák lehetséges kimeneteleinek halmazát azoktól a helyzetektől, amelyeket a racionális szereplők, szándékaik összjátéka nyomán, várhatóan elkerülnek. Az ikonikus filmjelenet körül kialakult szakmai vita, valamint ennek kapcsán a neoklasszikus közgazdaságtani és a játékelméleti szemlélet különbségének megértéséhez ismerni kell ezt az egyensúlyfogalmat. Nem bonyolult, ellenben fontos szerepet fog játszani a könyvben.

Nash-egyensúly minden olyan, több ember cselekvése által befolyásolt helyzet (játszma, játék), amely úgy jött létre, hogy mindegyik, a helyzetet befolyásolni képes szereplőnek a kialakult helyzethez (nevezzük kimenetelnek) vezető döntése a szereplő számára optimális döntés volt. Egy adott kimenetelnek megfelelő optimális döntést pedig minden szereplő úgy határoz meg, hogy elsőként elképzeli, hogy mi lenne, ha a többiek mind az adott kimenetelhez vezető döntéseket hoznák meg, és e feltevés mentén mérlegeli, hogy ebben az esetben neki is érdemes lenne-e az ide vezető döntést választania (létrehozva az adott kimenetelt), vagy inkább mást kellene csinálnia. Ha tehát egy adott kimenetellel kapcsolatban minden érintett szereplő, a saját gondolatkísérlete alapján, úgy látja, hogy neki is érdemes a kimene-

telhez vezető döntést hoznia (pontosan fogalmazva: nem kifejezett érdeke, hogy más tegyen), akkor az a kimenetel Nash-egyensúly.

A játékelmélet nagy kérdése, hogy a valóságban hogyan jutnak egyensúlyi helyzetekbe az emberek (Binmore 2006). Csak akkor van a véletlen próbálkozásnál nagyobb esélye annak, hogy egymástól függő szereplők elsőre, hibák nélkül egyensúlyi kimenetelhez vezető döntéseket hozzanak, ha az egyes szereplők a lehetséges kimenetekhez kapcsolódóan nemcsak saját gondolatkísérletüket játsszák le, hanem tudnak a többiek gondolatkísérletéről is, és nemcsak azokat a döntési opciókat zárják ki, amelyek saját szempontjukból rosszak, de azokat is, amelyek ugyan számukra kedves kimenetekhez vezethetnek, de mások ép ésszel nem hoznák meg az ide vezető döntéseket (tehát mások miatt nem jönne létre a vágyott kimenetel). Például a sebességmérővel felszerelt rendőrök felé közelítő, száguldozni szerető autós számára tetszetős kimenetel lenne, ha ő száguld, miközben a rendőrök kikapcsolják a készüléküket és félrenéznek. Viszont ha belegondol a rendőrök preferenciáiba, és belátja, hogy a rendőrök számára ez nem lenne optimális lejátszás, akkor kerülni fogja a számára kedves kimenetelhez szükséges saját döntést (száguldás).

A Nash-egyensúly tehát – szemben a hagyományos mikroökonómia egyensúlyával –, nem vezethető le pusztán az egyének optimális döntéseiből. A játékelmélet emberi játszmáiban ugyanis nem minden optimális reakció vezet egyensúlyhoz. Csak a többiek optimális döntéseire adott optimális válaszok vezetnek egyensúlyi állapothoz. Ha egy adott kimenetel csak úgy jöhet létre, hogy akár csak egyetlen szereplő is szuboptimális döntést hoz, akkor e kimenetel nem lehet racionális döntések eredője – a játékelméleti gondolkodás az ilyen helyzetet nem tekinti egyensúlyinak.

Fontos, hogy lássuk, hogy a neoklasszikus közgazdaságtan döntéshozójáról feltehetjük, hogy már megvalósult körülményekre reagál (és e feltevést finomíthatjuk kedvünk szerint úgy, hogy a jövőbeni körülmények most várható tulajdonságait veszi figyelembe). A Nash-féle döntéshozó mindenképpen még meg nem valósult, csak elképzelhető helyzetek sorában végez képzeletbeli optimalizációt. És míg a neoklasszikus közgazdaságtan döntéshozója csak „a saját feje



után megy”, a Nash-féle döntéshozó kénytelen mások fejével is gondolkozni, és a többiek optimalizációs feladatát is elvégezni ahhoz, hogy a sajátját elvégezhesse.

A mindjárt kifejtésre kerülő filmes példa, a hagyományos neoklasszikus és a játékelméleti szemlélet közötti különbség, valamint a további elméleti elemzések megértése érdekében azt is tisztázni kell, hogy a Nash-egyensúly lokális optimalizáció eredménye. Ha egy helyzet úgy jött létre, hogy a szereplők egymás érdekeinek figyelembevételével optimalizáltak, és végül mindenki az optimális döntést hozta, az nem jelenti azt is, hogy ez a helyzet feltétlenül a lehető világok legjobbjika az összes szereplő számára. Még azt sem, hogy nincs a lehetséges döntéseknek olyan kombinációja, amellyel mindenki jobban járna, mint azzal, ami konkrétan megvalósult. És még csak azt sem jelenti, hogy az egyensúlyok szűkebb halmazán belül ne lehetne mindenki számára jobb helyzetet találni. A Nash-egyensúly csak egy olyan helyzetet jelent, amelyben mindenki feltette, hogy a többiek e helyzet irányába mozdulnának, és ezért azt gondolta, hogyha így lenne, akkor ő is azt csinálná, ami végül ehhez a helyzethez vezetne. De a többiek döntésére vonatkozó más feltevésekhez más reakciók tartoznak, és az emberi játszmákban tipikus, hogy több, kölcsönösen optimális reakció kombináció létezik, és ezek között lehetnek olyanok is, amelyek akár mindenki számára jobbak, mint néhány további, kölcsönösen optimális reakciókombináció. De lássuk végre a konkrét példát!

### II.2.3. ELSŐ PÉLDÁZAT: EGY CSODÁLATOS ELME CSAJÓZÁSI TANÁCSAI

#### II.2.3.1. Egyensúlyvesztés: John Nash és Russel Crowe a kocsmában

A szakmai körökben ikonikussá vált (kitalált) filmjelentében Nash/Crowe éppen néhány évfolyamtársával sörözik egy bárban (pontosabban Nash/Crowe inkább „szakmázik”, amíg a többiek szórakoznak), amikor belép néhány attraktív fiatal lány, köztük egy szőke és több barna hajú, akik rögtön felkeltik a fiúk figyelmét (és az érdek-

lődés kölcsönös).<sup>\*</sup> Gyorsan kiderül, hogy a szőke lány izgatja leginkább a csajozni vágyó fiúkat, és egyikük a közgazdaságtan alapítójára, Adam Smith-re hivatkozva arról viccelődik, hogy társadalmilag is az az optimális, ha egymást taposva versengenek a szőkéért. Nash/Crowe azonban ekkor hirtelen megvilágosodik, és a konkrét helyzet kapcsán rájön, hogy az emberi játszmákban gyakran szuboptimális lehet, ha a többiek szándékainak figyelembevétel nélkül üldözzük saját céljainkat. Egyrészt Nash/Crowe a szőkével együtt érkező barna lányok fiúzási szempontjaira hívja fel a figyelmet (a fiúk csajozási szándékainak szempontjából): bizonyosan el fogják utasítani azt a fiút, aki csak azután próbálkozik náluk, hogy kudarcot vallott a szőkénél. E korlátozó feltétel mellett hívja fel a fiúk figyelmét arra, hogy mivel a többiek is a szőkét nézték ki, a tülekedésben nagy eséllyel egymást fogják akadályozni abban, hogy becsajozzanak. Azaz, ha Smith-t követve csak a saját fejük után mennek, és mindenki rányomul a szőkére, a végén jó eséllyel senkinek sem jut lány aznap este. Nash/Crowe azt mondja a többieknek, hogy Adam Smith tévedett, mert valójában nem az vezet a legjobb eredményre, ha mindenki a saját érdekeit követi („ez így nem teljes” – mondja), hanem ha mindenki azt nézi, hogy mi jó neki és a többieknek.<sup>\*\*</sup> Azt tanácsolja tehát a barátainak, hogy hagyják a szőkét, mindenki válasszon egy barna lányt, mert csak így juthatnak nőhöz aznap este.

A forgatókönyvíró, Akiva Goldsman saját bevallása szerint komoly kihívásnak érezte, hogy gyenge matematikai érzékkel kell egy komoly matematikai koncepció lényegét elmagyaráznia a nagyközönségnek (Script 2012). Elmondása alapján a jelenetet nem szakértők bevonásával írta, de nyilván olvasott már a játékelméletről.

A jelenetet és a benne szereplő „csajozási dilemmát” számos ismeretterjesztő blog tárgyalta a bemutatást követő években; és még a legutóbbi időben is vannak, akik foglalkoznak a 2001-es film e

<sup>\*</sup> [https://www.youtube.com/watch?v=2d\\_dtTZQyUM&t=63s](https://www.youtube.com/watch?v=2d_dtTZQyUM&t=63s) (Letöltve: 2018. szeptember 10.)

<sup>\*\*</sup> E ponton korrigáltam a film félreérthető fogalmazását, ahol „a csoport” érdekéről van szó.

jelenetével (Hodgson 2017, Liu 2016, Quora é. n., Rey 2008, Rogers 2018). A szakmai szempontokat vizsgáló kommentátorok között egyetértés van abban a tekintetben, hogy a filmben Nasht alakító Russell Crowe szájába olyan javaslatot adott Goldsman, amely egyensúlyon kívüli magatartásra buzdítja a többieket. A kritikusok érvelését (és még inkább magát a Nash-egyensúlyt) akkor lehet megérteni, ha a játékelméleti tankönyvek által használt ábrázolásmód segítségével mutatjuk be a játszma egy egyszerűsített változatát. Ebben két fiú és két lány szerepel csak. Felteszem (ahogy a jelenetben is feltette Nash), hogy a barnák közül már mindenkinek más-más tetszik (és viszont), így eltekintek a barnákért való versengéstől és az elutasítás lehetőségétől, és csak a felmerült szőke vagy barna dilemmára koncentrálok. A jelenetet feldolgozó ismeretterjesztő tartalmak külső, fixált feltételnek tekintik a lányok preferenciáit, és a fiúk egymás közötti játszmájára koncentrálnak (ahogy a jelenetben Nash is tette).

1.a ábra. A Nash és barátaiféle „csajozási játszma” szövegesen

		Te	
		szőkére	másik barnára
Én	szőkére	egyikünknek se lesz partnere, mindkettőnknek a legrosszabb	én a szőkével, te barnával, nekem nagyon jó, neked kicsit kevésbé jó
	egyik barnára	én barnával, te a szőkével, nekem kicsit jó, neked nagyon jó	lesz partnere, mindkettőnknek, neked is és nekem is jó valamennyire

1.b ábra. A Nash és barátaiféle „csajozási játszma”  
– boldogságértékekkel kifejezve

		Te	
		szőkére	másik barnára
Én	szőkére	Én: 0, Te: 0	Én: 2, Te: 1
	egyik barnára	Én: 1, Te: 2	Én: 1, Te: 1

1.c ábra. A Nash és barátai-féle „csajozási játszma” boldogságértékekkel – az egyensúly felé ható „erőkkel”

		Te	
		szókére	másik barnára
Én	szókére	Én: 0 ↓, Te: 0 →	Én: 2, Te: 1
	egyik barnára	Én: 1, Te: 2	Én: 1 ↑, ← Te: 1

A fenti három ábra szövegesen és számokkal is bemutatja ezt az egyszerűsített, fiúk közötti játszmát. Mindkét fiú vagy a szókére, vagy a neki tetsző barna lányra nyomul (aki más, mint a másik fiúnak tetsző lány). Versengés esetén (ha mindkét fiú az egyetlen szókére hajt), egymást akadályozzák, és nem jönnek össze a fiatalok, de egyedüli kezdeményezés esetén igen. A szóke mindkét fiúnak még jobban tetszik, mint a neki tetsző barna. Nash/Crowe javaslata az, hogy mindketten a nekik tetsző barnához menjenek oda.

A kommentátorok helyesen mutatnak rá, hogy egy ilyen szituációban Nash/Crowe filmbeli javaslata nem Nash-egyensúly. Ugyanis, amint azt gondolom, hogy te egy barnához fogsz odamenni, nekem nem optimális barnát választanom; ehelyett a szókéhez kell odamennem. Persze ha tudnám, hogy a szókéhez mész, akkor semmiképpen sem mennék a szókéhez. A fenti dilemmában a barna–barna döntési kombináció nem egyensúlyi kimenetel. Ehelyett a barna–szóke és a szóke–barna a két egyensúlyi helyzet. Tehát Nash/Crowe filmbeli állításával szemben mindannyian partnerhez juthatnak úgy is, hogy egyikük a szókével jön össze, a többiek pedig egy-egy barnával. Sőt, mutatnak rá a fent említett szakmai kommentátorok, csak az ilyen kombinációk tekinthetők Nash-egyensúlynak. Néhányan továbbmennek, és figyelmeztetnek arra, hogy a forgatókönyvíró épp a racionális döntésméletben alapvető önértékkövető gondolkodást hagyta figyelmen kívül. A jelenetben javasolt, egyensúlyon kívüli állapotban legalább egyvalaki (de a Nash/Crowe/Goldstein által javasolt megoldásban mindenki) szuboptimális döntést hoz, a többiek javaslattal egybevágó döntését (barna) feltételezve.

Goldstein tehát a fenti, plauzibilis modell alapján a Nash-egyensúlyt félremagyarázó szavakat adott Nash/Crowe szájába. Azonban úgy vélem, hogy a forgatókönyvíró a hibásnak tűnő technikai leírás ellenére jól ragadta meg a (neo)klasszikus közgazdaságtan és a játékelmélet szemléletmódja közötti különbséget. A Nash-egyensúly verbális leírásakor, és azt a neoklasszikus közgazdaságtan által vizsgált parametrikus döntéssel összevetve hangsúlyoztam, hogy a neoklasszikus homo oeconomicusszal szemben a játékok résztvevői tekintettel vannak mások vágyaira is. A híres filmjelenetben Nash/Crowe erről próbálja meggyőzni társait: ha jót akarunk magunknak, gondoljuk át azt is, hogy mit akar a másik (első körben a lányok, utána a többi fiú), mi neki a jó. Ez nem azt jelenti, hogy a másik javát akarjuk, csak azt, hogy nem tudjuk elérni a saját céljainkat, ha nem vesszük figyelembe, hogy a saját céljaikat követő mások hogyan keresztezhetik a mi céljainkat, és ha nem okoskodjuk ki, hogy miképp kerüljük el egymás céljainak keresztezését.

Könyvem egyik legfontosabb, a végkövetkeztetéseket is megalapozó állítása, hogy a játékelméleti modellek szereplői, szemben a neoklasszikus közgazdaságtan parametrikus döntéshozóival, figyelembe veszik mások vágyait is, mielőtt meghozzák a döntéseiket. Ezt különösen fontosnak tartom az erkölcsös viselkedésformák motivációinak megértése szempontjából. Úgy látom továbbá, hogy a társas preferenciákkal, és általában az erkölcsös viselkedéssel foglalkozó (közgazdaságtani és egyéb társadalomtudományi) kutatási programok nem szentelnek elég figyelmet a játékelmélet által modellezett ún. stratégiai döntések e megkülönböztető jellegzetességének.

Mindazonáltal érthető a stratégiai (azaz nem parametrikus) döntések erkölcsösség szempontjából alapvető jellegzetességeinek ignorálása. A megértéshez látni kell egy fontos distinkciót. Mások vágyainak figyelembevétele ugyanis (legalább) két alapvető mechanizmuson keresztül történhet. Az egyik lehetőség, hogy az egyén azt szeretné, hogy másoknak is jó legyen. Azaz a saját vágyainak része, hogy mások jól (vagy épp rosszul) érezzék magukat. Az utóbbi két évtizedben az ilyen típusú vágyak gyűjtőneve lett a közgazdaságtanban a társas preferencia. Ahogy már említettem, társas preferenciákkal

igen régóta foglalkozik a neoklasszikus közgazdaságtan, bár kétségtelen, hogy a legutóbbi időkig kis súlyt képviseltek a diszciplínán belül az ilyen irányú kutatások. A másik lehetőség, hogy bár csak a saját javunkat akarjuk, döntéseinkben figyelembe vesszük a céljainkat keresztetni képes más szereplők céljait is, azt elkerülendő, hogy döntéseink kombinációja számunkra (túlságosan) káros következményekkel járjon.

A társas preferenciák ma népszerű kutatási programja csak arra a lehetőségre koncentrál, amikor mások jólléte iránti vágyaink sarkallnak erkölcsös viselkedésre. Azt is említettem, hogy nem ebben a programban kezdték ezen preferenciák modellezését és a következményeik kutatását: egyes neoklasszikus közgazdászok már évtizedek dolgoztak a mások jólléte iránt aggódó embereket feltételező modellekkel (Becker 1974 stb.).

### II.2.3.2. Keverjük a kártyákat: Crowe első megmentése

Térjünk vissza a filmben bemutatott szituációhoz! Joggal adódik a kérdés, hogy amennyiben a bölcész forgatókönyvíró jól érzett rá, hogy a játékelmélet modellezési logikája magában hordozza a társas orientációt, akkor miért nem „működik” a szituációt modellező játék? Miért nem Nash-egyensúly a stratégiáknak az a kombinációja, amelyet a filmbeli Nash/Crowe javasolt az egymást keresztelő célok figyelembevételével?

A válaszomhoz látni kell, hogy még egy olyan egyszerű, kevés szereplőt és döntési lehetőséget magában foglaló élethelyzet is, mint az idézett filmben szereplő „csajozási dilemma”, jóval komplexebb annál, mint amit a fent bemutatott egyszerű játékmóddal kezel. Ebből az is következik, hogy a komplexitás redukciójának (a modellezésnek) több (sok) útja lehetséges. A filmjelenet játékelméletben jártas kritikusai egy plauzibilis egyszerűsítési utat választottak, de nem ez az egyetlen út (vö. Rubinstein 1991). Lehetséges, hogy előbb érezték úgy, hogy rossz a Nash-egyensúly interpretációja, és ehhez illesztettek egy plauzibilis és egyszerű modellt, de azt nem keresték, hogy mi-

lyen modell írná le Nash/Crowe filmbeli érvelését a Nash-egyensúly paradigmáján belül maradva.

Ha jóindulatúak vagyunk a forgatókönyvíróval, akkor elfogadhatjuk, hogy a filmbéli Nash/Crowe valóban Nash-egyensúlyi stratégiákat javasolt barátai számára. Ezzel egyben az is kimutatható, hogy a közgazdaságtan sok képviselője hogyan értelmezi félre a játékelméleti modellezés értelmét.

Induljunk ki először a szakmai kommentárok által javasolt szerkezetből (1.a, 1.b és 1.c ábra)! Ennek a játéknak alapesetben két egyensúlyi stratégia kombinációja van, és egyik sem az, amelyet Nash/Crowe javasol a filmben. Ha viszont egy játszmában több egyensúly van, akkor a bizonytalanság óhatatlanul szerepet kap a döntésekben. Ugyanis a játékosok nem kapnak egyértelmű „utasítást” arra vonatkozóan, hogy mit kell tenniük. A számukra optimális döntés a másik döntésétől függ. A bizonytalanságot ebben a helyzetben csak a játék szekvenciálissá alakításával lehet kizárni, azaz úgy, hogy valamelyik fél képes legyen előbb lépni. És az előbb lépő fél lesz előnyben, hiszen neki a szőke lányhoz kell közelednie. Tehát ha komolyan gondoljuk a kritikusok modelljét az 1.a, 1.b és 1.c ábrán, és nem akarunk bizonytalansággal számolni, akkor az egyik (a legeszesebb) fiúnak már Nash/Crowe magyarázata közben el kellett volna indulnia (rohannia) a szőke lány felé. A filmből úgy tűnik azonban, hogy nem akartak ilyen megelőzőes játékot játszani, és betartották, hogy egyszerre indulnak el, miután megtárgyalták a kérdést. Ha viszont szimultán döntést hoznak, akkor a bizonytalanság elkerülhetetlen.

Ha bizonytalanság van, akkor a racionális szereplőknek valószínűségeket kell becsülniük, és az átlagosan várható boldogság szintek alapján hozzadják meg döntésüket. A bizonytalanság mellett várható kifizetések kalkulálásához viszont nem csupán a bizonytalanságot kell számszerűsíteni (valószínűség-értékekkel), hanem a boldogságértéket is. Tehát nem elég azt meghatározni, hogy egyik kimenetel jobb vagy rosszabb, mint a másik, hanem azt is számszerűsíteni kell, hogy mennyivel. Csak így lehet várható értéket számítani. Ebben az esetben tehát feltevéseket kell tennünk arra vonatkozóan is, hogy mit mennyire szeretnének a fiúk.

Nash/Crowe monológiával nem ellenkezik egy olyan feltevés, hogy a fiúk, tekintettel a lányok jelenlétére, nagyon szívesen becsajoznának. Legalábbis Nash/Crowe tanácsa arra az esetre vonatkozik, hogy miként kerülhető el a legrosszabb: mindenki hoppon marad. Az egyszerűség kedvéért egy ponton változtassuk meg a játszmát a szokásos interpretációkban leíró mátrixot: magas negatív számmal jelöljük annak értékét, ha valaki hoppon marad (a 2. ábrában ezt  $-100$ -as értékkel jelezzük). Ez a kimenetel – a filmjelenet kritikusaival konszenzusban tett egyszerűsítő előfeltevéseink szerint – akkor következik be, ha mindkét fiú a szóke lányra hajt rá, ugyanis ő ekkor elutasítja őket, és nyomulásuk hiteltelenné teszi őket a nekik tetsző barna lányoknál is. Hogyan keresünk megoldást egy ilyen bizonytalan esetben?

Nashnek volt megoldása a bizonytalanságra: az ún. kevert stratégiák halmazán értelmezett egyensúly. Ha a felek egymás választását valószínűségekkkel írják le, akkor stratégiai döntés esetén – amikor a másik fél tudatosan, a mi dilemmáinkat figyelembe véve dönt –, ezt úgy kell tekinteni, hogy a másik tudatosan „keveri a kártyákat”. A kártyakeverések kombinációja akkor egyensúlyi, ha fenntartják a többiek dilemmáját a helyes döntéssel kapcsolatban, és ezzel a többiek számára is lehetővé teszik, hogy kártyakeveréssel bizonytalanítsák el a szereplőket.

Tehát a bizonytalanságból is lehet stratégiai játszmát csinálni. Nem törvényszerű, de ha mindenki átlátja a helyzetben rejlő bizonytalanságot (mint az általunk vizsgált játszmában), akkor plauzibilis feltevés, hogy minden fél a bizonytalansággal is játszmázik.

A játszmánkban az egyik fiúnak olyan módon érdemes meghatároznia, hogy mekkora eséllyel választja a szóke, illetve a neki tetsző barna lányt, hogy a másik fiú ne legyen biztos abban, hogy mit tegyen, és ő is „keverje a kártyákat”. Ha ego túl nagy eséllyel választja eleve a barna lányt, akkor a másikat arra ösztönzi, hogy mindenképpen a szóket válassza, és akkor már egónak nem is érdemes kevernie, hanem eleve el kell fogadnia, hogy lemond a szókérről. Ha viszont túl nagy eséllyel tenne a szókére, az eleve hiteltelen, hiszen irracionálisan vállal túl nagy kockázatot. Tehát az optimális kártyakeverés valamennyire tükrözi a másik racionális várakozásait is (amit racionálisan érdemes betartani).



Ha egy játszma sokat ismétlődik (mondjuk kő-papír-olló játékot játszunk sorozatban), akkor evidens módon tudjuk befolyásolni a másik bizonytalanságérzetét a „keveréssel”. Egy egyszeri játszmában leginkább az előzetes kommunikáció adhat lehetőséget arra, hogy befolyásoljuk mások viselkedésünk esélyeire vonatkozó becsléseit.

A konkrét helyzetben az egyik fiú a várakozások szerint olyan valószínűséggel hajt a szőkére, hogy a másik fiú számára egyenlővé váljon a szőkére-, illetve a barnára hajtás várható haszna. Ekkor ez a másik fiú is racionálisan keverheti a kártyákat (hiszen számára mind-egy, hogy miképp dönt) –, mégpedig olyan arányban, hogy az előbb említett „egyik” fiú számára is egyenlővé váljon a szőkére- illetve a barnára hajtás várható haszna. Ha  $p$  annak a valószínűsége, hogy az egyik fiú a szőkére hajt rá, akkor a másik fiú számára a 2. ábra alapján akkor válik egyenlővé a két opció várható haszna, ha  $-100p + 2(1-p) = 1p + 1(1-p)$ . (Az egyenlet baloldala a szőkére hajtás, a jobboldala pedig a barnára hajtás várható hasznát fejezi ki a szőkére hajtás valószínűségének függvényében).

Ebben az esetben a (kölcsonös keverést eredményező), racionálisan választott  $p$  (szőkére hajtás esélye) kisebb mint 1%. Azaz ebben a modellben már jól mondta, Nash/Crowe: ha mindenki lányhoz akar jutni (még hozzá nagyon akar – ezt fejezi ki a  $-100$ -as boldogtalanság érték a hoppon maradás esetén), akkor nem érdemes kockáztatni, és mindenkinek a neki leginkább tetsző barna lányhoz kell menni udvarolni (ez 99% esélyt jelent). Ha a „mindenképpen lányhoz akarunk jutni” kifejezést úgy fordítjuk le, hogy  $-100$  helyett végtelenül nagy negatív számot helyettesítünk be a karambol boldogságértékéhez, akkor kb. a 100% is meglesz.

2. ábra. A Nash és barátai-féle „csajozási játszma”  
– boldogságértékekkel kifejezve

		Te	
		szőkére	másik barnára
Én	szőkére	Én: $-100$ , Te: $-100$	Én: $2$ , Te: $1$
	egyik barnára	Én: $1$ , Te: $2$	Én: $1$ , Te: $1$

## II.2.3.3. Viselkedés vs. stratégia: Crowe második megmentése

De nem ez az egyetlen interpretációs lehetőség. Most tekintsünk el a bizonytalanságtól, de vegyük figyelembe, hogy a fiúk nem rohanták le azonnal a szőke lányt, amint Nash/Crowe felvázolta nekik a helyzetet. Ehelyett tárgyaltak egymással, ígéreteket, javaslatokat tettek, illetve fenyegették egymást. Éppen egy ilyen fenyegetésre reagált Nash/Crowe a saját koncepciójával, és a vita során a következő javaslattal állt elő: mindenki válassza a neki tetsző barna lányt. Ha a többiek megfogadják a tanácsát, akkor azt lehet úgy értelmezni, ahogy a jelenet szakember kritikusai is tették, miszerint mindenki feltétel nélkül követi a javaslatot. De lehet úgy is értelmezni, hogy ez a javaslat a közös térben hangzott el, mindenkinek szól, azaz a teljesítése is a többiek teljesítéséhez kötött. Tehát Nash/Crowe javaslatának elfogadása azt jelenti, hogy részt veszek abban a közös stratégiában, hogy mindenki a neki tetsző barna lányhoz megy oda, és senki sem a szőkéhez. Ez nem azonos az 1-2. ábrákban bemutatott „szőkére” stratégiával. Ez egy ott nem szereplő stratégia (vö. Binmore és Shaked 2010, Rubinstein 1991), amelyet hozzá kell adni a mátrixhoz. Ebben az esetben a játszma modellje a 3.a. és 3.b. ábra szerint alakul.

3.a ábra. A Nash és barátai-féle „csajozási játszma”  
– boldogságértékekkel kifejezve

		Te		
		szőkére	másik barnára	Nash-javaslat
Én	szőkére	Én: 0, Te: 0	Én: 2, Te: 1	Én: 0, Te: 0
	egyik barnára	Én: 1, Te: 2	Én: 1, Te: 1	Én: 1, Te: 1
	Nash-javaslat	Én: 0, Te: 0	Én: 1, Te: 1	Én: 1, Te: 1

Nash javaslat: a nekem tetsző barnára hajtok rá, ha a másik is a neki tetsző barnára hajt rá, de a szőkére, ha a másik is a szőkére.

3.b ábra. A Nash és barátaí-féle „csajozási játszma” – boldogságértékekkel kifejezve, az egyensúly felé ható „erőkkel”

		Te		
		szőkére	másik barnára	Nash-javaslat
Én	szőkére	Én: 0 ↓, Te: 0 →	Én: 2, Te: 1	Én: 0 ↓, ← Te: 0
	egyik barnára	Én: 1, Te: 2	Én: 1 ↑, ← Te: 1	Én: 1, ← Te: 1
	Nash-javaslat	Én: 0 ↑, Te: 0 →	Én: 1 ↑, Te: 1	Én: 1, Te: 1

A 3.b ábrán az is látható, hogy a Nash/Crowe által javasolt stratégia-kombinációból nem mutatnak kifele nyílak, az ugyanúgy egyensúlyi helyzet, mint azok, amikor az egyik fiú a szőkére hajt, a másik pedig valamelyik barna lányra.

Nem véletlenül szenteltem ilyen nagy teret ennek az egyszerű illusztrációnak. A könyv konklúziója szempontjából kulcsfontosságú az az állítás, hogy az önérdüket követő emberek interakciójának elemzése is megfelelő lehet az erkölcsös, csoportérdekeket figyelembe vevő viselkedés vizsgálatára. Ráadásul ehhez nem kell feltétlenül külön definiálni az önző és önzetlen magatartást. A filmbeli példában Nash/Crowe végig a barátaí józan belátására hagyatkozva próbálta meggyőzni őket, hogy a csoportszinten értelmezhető optimumot keressék a szemellenzős önzőség helyett. Az sem véletlen, hogy két megoldást is bemutattam arra, hogy miképpen lehet a kooperatív megoldást modellezni. Mindkét megoldási típus elő fog kerülni a könyv későbbi fejezeteiben, általános érvényű dilemmák megoldásaként. E megoldások illusztrálását persze számos más példával is megtehettem volna, de az sem véletlen, hogy egy olyan szituációt választottam, ahol a témához hozzászóló közgazdászok nem akarják látni a kooperatív megoldások racionalitását. Ugyanis nemcsak a dilemmát, és a lehetséges megoldásokat akartam illusztrálni, hanem a közgazdaságtan főáramának vakságát is a játékelméletben rejlő lehetőségekre, amelyekkel az erkölcsös viselkedés az erkölcsi motivációkra vonatkozó feltevések nélkül kezelhető.

#### II.2.4. MÁSOK ÉRDEKE A STRUKTURALISTA ELEMZÉSBN: AZ ÚJ GAZDASÁGSZOCIOLÓGIA

Ennek a fejezetnek az az alapvető mondanivalója, hogy az emberi interakciók vizsgálata megoldást jelenthet az erkölcsi viselkedés megértéséhez (is). Ide tartozik a korábbi fejezetben kifejtett kritika a neoklasszikus mikroökonómiával szemben, miszerint a csak magával törődő homo oeconomicus modellje kevésbé alkalmas a közérkölc szerinti cselekvés megértésére. Ugyanakkor amellet is érveltem, hogy a második világháború utáni survey-szociológia sok tekintetben felzárkózott a közgazdaságtanhoz az individualista szemlélet terén.

Ha ebben a formában foglaljuk össze az emberi játszmák vizsgálatának jelentőségét és relatív előnyeit, akkor világossá válik, hogy érvelésem egy nagyon régi szociológiai gondolatmenet reprodukciója. A kapcsolatháló elemzés módszerét alkalmazó, ún. új gazdaságszociológia a hetvenes évektől hirdette a személyközi interakciók jelentőségét a társadalmi normák megértésében (Granovetter 1973).

Később, már a nyolcvanas években Granovetter (1985) explicit módon fejtette ki fenntartásait mind a neoklasszikus közgazdaságtan, mind a survey-szociológia atomizált emberképével szemben. A kritizált atomizált emberképet szembeállította az új gazdaságszociológia által preferált, beágyazott cselekvés modelljével. Világossá tette azt is, hogy a neoklasszikus közgazdaságtan belsőnek tekinti az önző preferenciákat, míg a survey-szociológia szerint rejtélyes külső erők formálják értékeinket. De a két modell közös jellemzője, hogy a cselekvő nem veszi figyelembe azt a közvetlen emberi környezetet, amelyben döntését meghozza. Ezzel szemben az új gazdaságszociológia az emberi döntések társadalmi beágyazottságából indul ki, és a beágyazottság nem megfoghatatlan, makroszintű jelenségek befolyását jelenti, hanem a döntések közvetlen társadalmi környezetét jelentő emberek befolyását.

Az új gazdaságszociológia úttörő munkájában Granovetter többek között a munkakeresés meghatározó tényezőivel is foglalkozott (Granovetter 1973). A másik példája azonban a kollektív érdekérvényesítéshez, és ezen keresztül az erkölcsös viselkedéshez kapcsolódott. Ez

a kutatási program megmutatta, hogy a csoportérdeket követő erkölcsös viselkedés létrejöttének, illetve elmaradásának magyarázatához nem feltétlenül szükséges az egyéni erkölcsösségre vonatkozó feltevéseket tenni. Granovetter példájában – egy, a lakónegyed eredeti formájában való fennmaradásáért küzdő közösségben –, nem feltétlenül a lakónegyed egymás számára ismeretlen lakói iránti elkötelezettség magyarázza a sikeres kollektív cselekvést. Sokkal inkább számítanak az egyének rokonsági-szomszédsági kapcsolataiból felépülő hálózat jellegzetességei.

Az utóbbi pont megmutatja a hangsúlybeli különbséget a játékelmélet, valamint az új gazdaságzociológia beágyazott cselekvés koncepcióján nyugvó hálózatelemzés között. A játékelmélet alkalmazói a lehetőségei ellenére nehezen lépnek túl a neoklasszikus közgazdaságtan individualista emberképén. Az új gazdaságzociológia a szociológia klasszikus makroszintű orientációjának bűvkörében gondolkodik, ahogy a továbbfejlesztése nyomán világhódító útjára indult szociológiai hálózatelemzés is. A két kutatási terület csak a 21. században találkozott, és ma is nagyon kevés az interakció közöttük.

Összefoglalva – és vállalva a pedagógiai célú leegyszerűsítést: a játékelmélet modellje lehetővé teszi, hogy az emberi döntések társas és erkölcsi beágyazottságát modellezzük, ugyanakkor a játékelmélet kutatóit ez, nagyon kevés kivételtől eltekintve, nem érdekli. Az új gazdaságzociológia által elindított hálózatelemzési paradigma ellenben explicit módon állítja, hogy a döntések társadalmi beágyazottságának vizsgálatával meg tudja oldani az erkölcsi viselkedés modellezésének problémáját, ám ennek a tudományterületnek a fejlődése a strukturális kérdések irányába fordult: a kapcsolathálókat szerkezeti sajátosságainak viselkedésre gyakorolt hatásait vizsgálják. Azonban nincs egységes egyéni döntési modelljük, amelyre stabilan építkezni lehetne (Szántó és Syi 2010). Ezzel együtt az új gazdaságzociológia képviselői tárják fel az erkölcsös viselkedés legtöbb olyan mozgatórugóját, amelyekre a társadalmi normák játékelméleti elemzése (VI. fejezet) valamint e kötet konklúziójának tekinthető, a beágyazott önző polgár játékelméleti modellje (VII. fejezet) is felhívja a figyelmet (lásd például Granovetter 2017).

Ennek a fejezetnek az lehet az összegzése, hogy a játékelmélet modellezési eszköztára különösebb speciális motivációs előfeltevések alkalmazása nélkül is megfelelő az erkölcsös, csoportorientált viselkedés tanulmányozására. De az erkölcsös viselkedés társadalomtudományi vizsgálata nem ebből az állításból indult ki. Sőt, az erkölcsös viselkedés játékelméleti vizsgálata sem. Csak évekkel később jutott el ide a kutatók egy része. A következő fejezetekben azt a nagy kerülőutat mutatom be, amelyből rengeteget tanultunk, és amelynek tanulságai megerősítik az erkölcs játékelméleti elemzésének téziseit.

### **II.3. Nyitva maradt kérdés: tudomány-e a döntés logikája?**

#### **II.3.1. A RACIONÁLIS DÖNTÉSELMÉLET KETTŐS TERMÉSZETE**

A racionális döntésemélet kritikusai gyakran keverik a következő fejezetben bemutatásra kerülő önző polgár modelljét a jelen fejezet tárgyát képező racionális polgár modelljével. A fejezet elején már hosszabban írtam erről a félreértésről; és később többször is megemlítem majd. A modellek összekeverése részben terminológiai problémákra vezethető vissza, hiszen a racionális döntésemélet széles körben alkalmazott koncepciója mindkét modellt magában foglalja (Csontos 1996). A konfúzió ugyanakkor a két modell szükségszerű összefonódását is tükrözi. Az irányzat legtöbb, általam ismert kritikus nem vesz tudomást a terminológiai zavarról. Az irányzat általam ismert követőinek egy jelentős része pedig nem tárgyalja a terminológia által összekötött két modell tartalmi összefonódását.

A fejezet korábbi részében ismertetett racionális polgár modellje alapvetően egy metodológiai paradigma (Csontos 1996). Buchanan (1992 [1969]) fent már idézett művében a döntések logikájáról beszél (szembeállítva a magatartás-tudománnyal). Lindenberg (1992) már többször is idézett tanulmánya egyszerűen kevésbé tartalmas modellként hivatkozik arra az elméletre, amelyre mi a racionális polgár modelljeként tekintünk. Tekintve, hogy e monográfia elméleti, és

nem metaelméleti tanulmány, itt nem tárgyalom ezeket az elemzéseket (a témához lásd még Janky 2001, 2005, Syi 2008, 2014, Szántó 1993, 1999). Az alábbiakban Csontos (1996) követve nem egyszerűen üres modellként, hanem metodológiai paradigmaként hivatkozom a racionális polgár koncepciójára.

Szeretném hangsúlyozni, hogy a tárgyalt vizsgálati téma vonatkozásában a racionális polgár itt bemutatott alapkonceptiója nem valódi tudományos modell, azaz nem lehet vele popperi értelemben tudományos (azaz tesztelhető) hipotéziseket felállítani. Ahhoz cáfolható tézisek kellenének (a kockázatos döntésekre vonatkozóan tartalmaz ilyen tesztelhető állításokat az alapmodell, de ezek kívül esnek kutatási programunk fókuszán). Az önző (vagy talán még találóbban: a materialista) polgár modellje ellenben valódi tudományos modell. Erre a tudományos modellre épülnek a közgazdaságtan legtöbb szakterületének elméletei. Erről a fejezet elején már volt szó.

Az erkölcsös polgár tudományos elmélete a fentiek alapján nem alapozódhat pusztán a racionális polgár modelljére. Épp emiatt problematikus, ha az erkölcs kapcsán az önző polgár modelljét érő kritikákra válaszul a racionális polgár koncepciója mögé vonulnak vissza a racionális döntéelmélet követői (vö. Binmore 1998, 2010). Ugyanakkor attól, hogy „üres” ez a modell, még nem értelmetlenség, nem üres tautológia (vö. Green és Shapiro 1996). Röviden elmagyarázom, hogy miért nem. A magyarázat révén az is érthetővé válik, hogy miért kapcsolódik törvényszerűen össze a racionális és az önző polgár modellje (miközben a kettő analitikusan világosan elválasztható!).

Csontoshoz (1996) hasonlóan úgy tekintem, hogy a racionális polgár modellje egy metodológiai paradigma. Másként fogalmazva, egy kutatási eljárásokra vonatkozó útmutató. A sokak által „üres” modellnek nevezett koncepció adja az útmutatásokat. Ez az útmutató két szigorú korlátot alkalmaz (ezért lehet megtévesztő az „üres modell”, vagy akár a „döntési logika” kifejezés). Az egyik a szituációs logika alkalmazásának követelménye (Popper 1989 [1957]), a másik pedig az intencionális magyarázatok bevonásának követelménye (Von Wright 1987 [1971], Weber 1987 [1922]).

A szituációs logika követelménye azt jelenti, hogy ne a résztvevők lelki alkatának változatosságából és változékonyságából vezessük le társadalmi jelenségek változatosságát és változékonyságát, hanem inkább a résztvevők környezetének mérhető sajátosságaiban keressük a változatosság és változékonyság okait. Az intencionális magyarázatok követelménye azt fogalmazza meg, hogy ne tekintsük az embert mechanikus reagálógépnek. Tételezzük fel, hogy az embereknek céljaik vannak, amelyeket el *akarnak* élni. A környezeti tényezőkre a saját céljaik tükrében reagálnak. Ez különös jelentőséget kap akkor, ha a releváns környezeti tényezők között más emberek viselkedése is szerepel. Ugyanis az intencionális magyarázat ekkor magában foglalhatja azt, hogy a modellezett döntéshozó maga is intencionális magyarázatokat alkalmaz az emberi környezet hatásainak elemzése során. A fejezet elején megénekelt Neumann János (1928) a játékelmélet megalapozásával ennek az intencionális magyarázatba beágyazott, intencionális elemzésnek egy algebráját alkotta meg. Másfelől közelítve: a játékelmélet egyszerre képes a társadalmi beágyazottság (Granovetter 1985) és az intencionalitás kezelésére.

Ha elengedjük a szituációs logika követelményét, akkor az emberek személyiségének változatosságából magyarázhatjuk a viselkedési változatosságokat, és hangulatuk változékonyságából a viselkedésük megváltozását. A személyiség egyes sajátosságai, és a pillanatnyi érzelmi állapotok egyaránt mérhetőek, és az erkölcsös polgár kutatásában ez egyre népszerűbb irányzat. Erről lesz szó az V. fejezetben. Ha a pszichológiai tényezők alapján előrejelzéseket tudunk adni, akkor az erre épülő elméletek tudományosak. Ám mindenképpen eltávolodnak a racionális polgár alapmodelljétől.

Ha elengedjük az intencionális magyarázatok követelményét, akkor a környezeti tényezőkből közvetlenül is következtethetünk viselkedési következményekre. Ha mérhető környezeti sajátosságokból következtetünk mérhető viselkedési következményekre, akkor kétséget kizáróan tudományos munkát végzünk. Azonban az ilyen munka sem tartozik a racionális polgár paradigmájához.



### II.3.2. A FRADI VEZÉRSZURKOLÓ KETTŐS TERMÉSZETE

Ebben a könyvben nem érvelek általában a racionális polgár modellje mellett. Ahogy említettem, ez egy elméleti, és nem metaelméleti munka. Az olvasó a könyv kézhezvételekor már helyesen tippelhet meg, hogy termékenynek tartom ezt a megközelítést (magyarul is kiváló metaelméleti tanulmányok készültek: Csontos 1996, Syi és Szántó 2010, Szántó 1999). Ehelyett azt szeretném bemutatni, hogy a szituációs logika és az intencionális magyarázat követelményei egymással ellentétes irányba vezetik a kutatót. Ezért – ismétlem magam –, ez az „üres” és „tautologikus” modell valójában nagyon szigorú eljárási szabályokat határoz meg.

A szituációs logika akkor tud jól működni, ha univerzális célokat feltételezünk, azaz nagyjából mindenki ugyanazt akarja. Az emberek természete közötti különbségek nem takarják el a környezeti tényezők különbségeinek hatásait. Az intencionális magyarázat teljességéhez pedig éppenhogy az emberek egyedi szempontjainak figyelembevétele tartozik hozzá. Az talán világos, hogy általánosabb érvényű elméletekhez jutunk, ha e szigorú szabályokat be tudjuk tartani, akkor, mint ha nem. Egyrészt van egy univerzális lista az emberek céljairól. Ez adhat egy univerzális listát a releváns környezeti tényezőkről, amelyek változatosságát és változásait megfigyelve bármilyen kontextusban előrejelzéssel tudunk szolgálni.

A környezet-viselkedés kapcsolatok megfigyelése univerzális cselekvési modell nélkül nem nyújt biztonságot abban a tekintetben, hogy máshol és máskor is ugyanezt a kapcsolatot fogjuk megfigyelni. A pusztán személyes szándékokra és értelmezésekre építő viselkedés-magyarázatokból pedig még nagyon hasonló kontextusban is nehéz aggregált trendeket előrejelezni – egyének aktuális lelkiállapotát még saját maguk jövőbeni helyzetére sem vetíthetjük ki tudományos igénnyel, nemhogy másokéra.

A fenti, egymásnak ellentmondó követelmények miatt kötődik törvényszerűen a racionális polgár modellje (mint kutatási útmutató) az önző/materialista polgár (tudományos) modelljéhez. És ebből következik, hogy a racionális polgár modelljét mindenekelőtt a gaz-

dasági viselkedés magyarázatára kezdték alkalmazni – meglehetősen sikerrel. Még Teréz anyja is, ha kimegy a piacra, akkor árusként pénzt akar keresni, vevőként pedig a legolcsóbban akarja a legjobbat beszerezni. Az embereket a piacon az egyéni pszichológiai jellemzőik vizsgálata nélkül is megértjük, és jól be tudjuk azonosítani a mindenkinek számára fontos környezeti tényezőket (a piacon csereberélt javak árai, és a résztvevők költségvetési korlátai). Itt világosan látszik a megértés haszna: az nem öncélú „lelkizés”, hanem éppen a releváns környezeti tényezők beazonosítását szolgálja.

Vegyük Kalapot (Zöllei János), a hetvenes-nyolcvanas évek legendás fradista vezérszurkolóját. Kalap a vezérszurkolás mellett fruttit és szotyit árult, meccseken kívül pedig emléktárgyakat is. Ha *értjük*, hogy emléktárgyárusként Kalap szándéka a pénzkereset, akkor nem igazán releváns, hogy Fradi-drukker-e, ezért azt várjuk, hogy piaci áron ad el mindenkinek mindenféle emléktárgyat. Ezt én tízévesen, a badacsonyi kikötő mellett értettem meg, amikor minden bátorságomat összeszedve a Fradi ősellenségének, a Dózsának a zászlaját kértem tőle (pénzért), és kifejezetten kedvesen szolgált ki, láthatóan örülve annak, hogy a zászlót eladhatta nekem. Kalap lelátói viselkedésének megértéséhez és előrejelzéséhez vezető út egyáltalán nem olyan egyértelmű mint Kalap, a piaci árus modellezése. Nem csoda, hogy az erkölcsös polgár racionális döntéseméleti alapú kutatása – szemben a szokványos piaci viselkedés vizsgálatával –, sok évtized után is alapkérdéseken rágódik.

Ebben semmi újdonság nincs. Weber (1987 [1922]) ezt egészen világosan leírta. Ezzel előre jelezte a közgazdaságtan tartós házasságát a racionális döntésemélettel, és egyúttal jó pár társtudomány fenntartásait is a számukra néha lehetetlen követelményeket szabó metodológiai paradigmával szemben.

A fentiekből következik, hogy a könyvben tárgyalt kutatási programnak csak a kiindulópontja lehet a racionális polgár modellje. Innen mindenképpen tovább kell lépni tudományos elméletek felé. Ebben a könyvben azt az irányzatot mutatom be, amely egy idő után azért engedett a szituációs logika követelményeiből, hogy az emberek céljait egyre pontosabban írhasse le. Ez a kutatási irány az erköl-

csökkel foglalkozó szociológusok többsége számára kevésbé ismert. A szociológusok ugyanis egyrészt a saját, a piaci viselkedésen túlterjeszkedő érdeklődési területeik tudományos vizsgálatához ragaszkodtak. Azaz mérték a mérhető empirikus összefüggéseket akkor is, ha hipotéziseiket nem tudták felépíteni egy megalapozott cselekvési modellre (elhagyták a megértés követelményét). Másrészt a megértő társadalomtudományt preferáló irányzatok (például különböző kulturális tanulmányok) kifejezetten a lokális kontextusok értelmezését keresték az univerzális célok helyett (engedtek a popperi tudomány követelményeiből). Ezekből a perspektívákból érthető a szociológia társtudományok között is kiemelkedő ellenállása a racionális döntésemeléttel szemben. Ugyanakkor ebben a könyvben amellet érvelek, hogy az erkölcs racionális döntéseméleti alapú vizsgálatának egyes új fejleményei kifejezetten igénylik mindkét szociológiai hagyományban felhalmozódott tudást. A továbblépést e tudások nélkül nehézkesnek látom.

Szándékom, hogy a racionális döntésemeléttel szembeni, ismerethiánytól fűtött, zsigeri elutasítással nem törődve, e kutatási program eddigi eredményeinek és jelenlegi kihívásainak megértésén keresztül mutassam be a szociológiai tudás jelentőségét az erkölcs vizsgálatában. Meggyőződésem, hogy a szociológiai tudás nem a racionális individualizmus helyébe lépve, hanem arra építve szolgálhatja társadalmunk jobb megértését és jobbá tételét.

Lássuk, hogy honnan jutott el a racionális individualizmus a szociológiai koncepciókig!

## III. 1965\* – Az önző polgár erkölcsse

### III.1. Önzés és önzetlenség: az erkölcsi dilemma létrehozása

#### III.1.1. A TÁRSADALMI HÁTTÉR

Az előzőekben viszonylag részletesen kifejtettem, hogy sem a neoklasszikus közgazdaságtan alapmodelljében, sem a játékelméleti modellekben nincs ab ovo jelen az önző és önzetlen, vagy erkölcsös motiváció közötti különbségtétel. Ezt a dichotómiát azonban létre lehet hozni. Utaltam a fentiekben a neoklasszikus közgazdaságtan kiemelkedő képviselőjének, Gary Beckernek (1974, 1981) a családgazdaságtanára, amelynek keretei között megfogalmazta az önzőség és altruizmus dilemmáját.

Abban az időszakban – a hetvenes években –, Becker a közgazdaságtanon belül sem volt egyedül az önzetlenség modellezésének dilemmáival. Az altruizmusmodell közvetlen ihletői biológiai modellek voltak. Ám a tudományos változásokat abban az időszakban bizonyos társadalmi változások is segíthették.

Korábban kifejtettem, hogy a neoklasszikus közgazdaságtanon nyugvó politikatudomány és a survey-szociológia egymástól eltérő

\* Olson, M. (1997 [1965]): *A kolektív cselekvés logikája. Közjavak és csoportelmélet*. Budapest: Osiris [*The Logic of Collective Action: Public Goods and the Theory of Groups*]. Cambridge, MA: Harvard University Press.

elméleti gyökerű modelljei jól illeszkedtek az osztályszavazás gyakorlatához. Az Amerikában a harmincas és hatvanas évek között virágkorát élő ipari társadalom a hetvenes évek elejétől nagy sebességgel bomlott szét. A hagyományos nagyipar külföldre települése és gépesítése sok mindent megváltoztatott (Wilson 2009). A világ nagy államai közül elsőként az USA alakult át ipariból posztindusztriális társadalommá.

Egy példa az átalakulásokra: az Egyesült Államokban a második világháborút követően a középosztály széles rétegeire terjedt ki az egykeresős családmodell, ahol még a képzett nők is háztartásbeliként töltötték aktív életük nagy részét. Ehhez tökéletesen illeszkedett Becker családi szintű optimalizációt célzó munkamegosztás modellje. Azonban a hetvenes évek elejétől gyorsan változott a tipikus középosztálybeli családok belső munkamegosztása, és egyre több nő kezdett el dolgozni a gyermekvállalás mellett is. Ez értelemszerűen új helyzetet teremtett a családokon belül. A „mindenki egyért, egy mindenkiért” idilli elvével hirtelen nehezen lett összeegyeztethető sok családi probléma. Az érdekellentétek nyilvánvalóvá váltak, és a nők alkupozíciójának javulásával a válások száma is gyors emelkedésnek indult. Nyilvánvalóvá vált, hogy a családon belül is élesen szétválhatnak az egyéni érdekek, és az önzőség vs. önfeláldozás dilemmája sokak számára kézzelfoghatóvá vált.

Az osztályszerkezet és a szavazói magatartás változása is hasonló dilemmák felszínre kerülését hozta. Az osztályszerkezet bomlásával kezdett – egyre feltűnőbbben, és egyre több dilemmát okozva – szétválni a szavazók egyéni anyagi érdeke a társadalmi környezetük elvárásaitól (a társadalmi kontextus ittenivel rokon, de mégis különböző interpretációjához lásd Fontaine 2006). Ronald Inglehart politikatudós már a hetvenes évek elején megfogalmazta a posztmaterialista értékek terjedésére vonatkozó hipotézisét (Inglehart 1971). Szerinte a gazdag demokráciában élő választók képzetesebb és anyagi biztonságban élő része egyre kevésbé a közvetlen anyagi érdekeire támaszkodva fogalmazza meg a közéleti kérdésekre vonatkozó véleményét. A posztmaterialista szavazók számára az anyagiakon túlmutató társadalmi értékek jelentősége nagyobb, mint a korábbi osztályszavazók

számára. Ez volt az első elmélete annak a hosszú távon végül tényleg érvényre jutó tendenciának, amelynek során az anyagi érdekek súlya csökkent a közéleti attitűdök kialakítása és a pártválasztás során (Franklin 1985). A folyamat során a baloldali szavazók státusza fokozatosan emelkedett, míg jobboldaliaké csökkent, és ennek megfelelően az újraelosztási kérdések jelentősége is mérséklődött a pártpreferenciák kialakításában (Manza és Brooks 1999).

Ebben a folyamatban egyes újraelosztási kérdések is kifejezetten szolidaritási kérdéssé váltak. A posztindusztriális átalakulásban megjelent az ún. új szegénység, amely az ipari társadalom dolgozó szegényeivel és rövid távú munkanélkülijeivel szemben a munkaerőpiacról tartósan kiszorult embereket érintette (Wilson 2009). A felénk irányuló segítség megítélését sokkal nehezebb beleerőltetni a hagyományos, önérdek-alapú osztályszavazói modellekbe, különös tekintettel arra, hogy az empirikus eredmények szerint az új szegények nem a hozzájuk anyagilag közelebb álló, szegényebb dolgozó osztályok segítségére számíthattak, hanem a magasabban képzett rétegek demonstratív szolidaritására (Takács 2018, vö. Baldwin 1990).

Az együttműködés és szolidaritás elméleti modelljei tehát egy olyan társadalmi környezetben születtek meg, ahol a társadalomban mikro- és makroszinten is feltűnőbbé vált a különbség az önzőség és önzetlenség között.

### III.1.2. A FOGOLYDILEMMA

A játékelmélet által modellezett stratégiai gondolkodás lényege, hogy saját céljaink érvényesítésekor, éppen azok érdekében, figyelembe vesszük társaink céljait, és igyekszünk elkerülni, hogy tetteinkkel keresztezzük egymás útjait. Ebben a modellben nem triviális az individualista, illetve a másokkal törődő cselekvő szétválasztása.

Azokban vannak olyan játszmák, ahol a kétféle magatartás ütközése feltűnővé válik: például ha egy játszmában a részt vevő játéko-

sok optimális döntése független a többiek döntésétől, miközben ezzel az optimális döntésükkel megkárosítják a többieket. Játékelméleti terminusokkal: a játékosoknak domináns stratégiájuk van, ugyanakkor a domináns stratégiák alapján létrejövő egyensúlyi helyzet Pareto-szuboptimális. A domináns stratégia azt jelenti, hogy az optimális választás független mások döntésétől. Tehát az adott domináns stratégiát kell választanom függetlenül attól, hogy a többiek mit csinálnak. Nem kell morfondíroznom a többiek lehetséges döntéseiről. A Pareto-szuboptimális helyzet pedig azt jelenti, hogy létezik olyan kimenetel, amely minden résztvevő játékos számára jobb a jelenlegi helyzetnél. Tehát dönthetnének úgy is, hogy összességében jobban járjanak, mint most.

Egy ilyen játszmában a szereplők tudják, hogy döntéseikkel keresztezik egymás céljait, mégis az a racionális számukra, ha nem törődnek ezzel, és az egyéni érdekeiket követve egy kollektíve rossz kimenetelhez járulnak hozzá – előre látva, hogy ez fog bekövetkezni.

Már az ötvenes években megjelent az ilyen típusú játszmák legegyszerűbb formája, az ún. fogolydilemma (Flood 1958, Flood és Dresher 1952, Luce és Raiffa 1957, Tucker 1950). A közismert példában két, egymástól elkülönített bűnelkövetőnek kell döntenie a hallgatás (együttműködés a másik rabbal), és a vallomástétel (betyárbecsület megtörése) között.

4.a ábra. A fogolydilemma típusú játszma

		Te	
		hallgatsz	vallasz
Én	hallgatok	mindkettőnket csak tiltott fegyverviselésért ítélnék el, nem kell börtönbe vonulni	mehetek börtönbe kettőnk bűnéért, téged vádalku keretében elengednek
	vallok	engem vádalku keretében elengednek, te mehetsz börtönbe kettőnk bűnéért	mindketten börtönbe megyünk, de megoszlik a felelősség és a büntetés

4.b ábra. A fogolydilemma típusú játszma – boldogságértékekkel kifejezve

		Te	
		hallgatsz	vallasz
Én	hallgatok	Én: 3, Te: 3	Én: 1, Te: 4
	vallok	Én: 4, Te: 1	Én: 2, Te: 2

4.c ábra. A fogolydilemma típusú játszma – az egyensúly felé ható „erőkkel”

		Te	
		hallgatsz	vallasz
Én	hallgatok	Én: 3 ↓, Te: 3 →	Én: 1 ↓, Te: 4
	vallok	Én: 4, Te: 1 →	Én: 2, Te: 2

Ebben a szituációban mindkét játékos tudja, hogy a vallomástétellel egymás céljait keresztezik, és mindketten jobban járnának az együttes hallgatással, mint az együttes vallomástétellel. De az egyéni racionalitás azt diktálja, hogy az egyén ne foglalkozzon a másik (vagy a közösség) céljaival, mert nem jutna vele előre. Bármit is csinál a másik, az egyénnek a vallomástétel az optimális válasza. Ha a másik hallgat, akkor kifejezetten jó „potyautasnak” lenni, és a másik fél betyárbecsületet követő magatartását kihasználva vallani, és így megúszni a börtönt. Ha a másik vall, akkor így lehet menteni a menthetőt, és csökkenteni a börtönbüntetést.

Ez az a speciális helyzet, amelyben létrejön az ellentmondás a közösségre való figyelés és a saját érdekeinkre való koncentráció között. Ebben a helyzetben – szemben az általam módosított Nash/Crowe-féle játékkal, és rengeteg más típusú emberi játszmával – csak a közösségi érdek ignorálása a racionális döntés, míg a közösségi érdek figyelembevétele teljességgel irracionális.

Az utóbbi fél évszázadban ez a helyzet lett a legfontosabb illusztrációja az önzés és az önzetlenség dilemmájának. Vegyük észre, hogy a



Nash/Crowe-féle „csajozási játszmában” mások céljainak figyelmen kívül hagyása (például gondolkodás nélkül ráhajtani a szőke lányra) nem racionális önzőség, hanem irracionális mohóság; míg a többiek céljainak figyelembevétele racionális mérsékletesség. Tehát a megfelelő szituáció kiválasztása szükséges volt ahhoz, hogy a racionális önzőség és az irracionális önzetlenség dilemmáját fel lehessen vetni.

A dilemma jelentőségét több kutató is azonnal felismerte (például Kaplan 1960, Lave 1960). A potyautazás problémájának hódító útja a közgazdaságtanban és a politikatudományban azonban 1965-ben Mancur Olson Harvard Egyetemen írt disszertációjával indult (Olson 1997 [1965]). Olson az érdekcsoportok megszervezésének nehézségeiről írta munkáját, és azokat a peremfeltételeket kereste, amelyek egy-egy csoport megszervezéséhez szükségesek. A kollektív cselekvés logikájának nevezte el a potyautas magatartás irányába ható racionális motivációt.

Olson hagyományos mikroökonómiai eszközökkel tárta fel a racionális potyautazás és a közérdek konfliktusát, és nem használta a fogolydilemma játékot, de szemlélete teljesen játékelméleti volt. Talán szerepe volt ebben témavezetőjének, Thomas Schellingnek, aki a játékelmélet politikatudományi alkalmazásának úttörője volt, és 2005-ben Nobel-díjat kapott a munkásságáért. Mindenképpen szimbolikus, hogy hagyományos mikroökonómiai modellben írták le azt az emberi játszmát, amelynek segítségével az egyéni motivációk irányából lehet megközelíteni az egyéni és csoportérdek dilemmáját. Technikailag az tette lehetővé ennek a játszmának a hagyományos, neoklasszikus mikroökonómiai modellel történő elemzését, hogy a játékosoknak domináns stratégiáik vannak, tehát nem szükséges használniuk a játékelmélet alapját jelentő stratégiai gondolkodást az optimális döntés meghozatala során (nem kell azon morfondíroznik, hogy mit reagálna a másik az ő különböző döntéseikre).

A kollektív cselekvés logikájának mikroökonómiai elemzését hamarosan felváltotta a fogolydilemma alapú játékelméleti elemzés (Hardin 1968). A következő évtizedekben ez az egyszerű játszma fontos szerepet játszott a káros hatású, önző racionalitás, valamint a csoportot szolgáló önzetlen, de irracionális viselkedés dichotómiá-

jának értelmezésében. Újra hangsúlyozom: ez a szétválasztás nem következik általában az emberi játszmák természetéből, ez ennek a speciális játszmának a sajátossága.

Ez a játszma hivatott cáfolni Adam Smith tételét a közjóról, amely a saját boldogulásukat kereső gazdasági szereplők viselkedése nyomán jön létre a piacokon. „Nem a pék jóindulatától lesz kenyérünk” – állította Smith (1992 [1776]). Erre az volt a közjóról a fogolydilemma paradigmájában gondolkodó társadalomtudósok válasza, hogy de, a pék jóindulatára igenis szükség van; a nyers, önző racionalitás időnként hátráltatja a közösséget. Megjegyzem, a Nash-film forgatókönyvírójának visszaemlékezéseit olvasva lehetséges, hogy valójában a fogolydilemma járt a fejében (és azt oldotta meg helytelenül), amikor Nash/Crowe szájába adta Adam Smith fenti tételének cáfolatát (Script 2012).

Olson disszertációjában elsősorban a kollektivista, csoport/osztályérdek alapú társadalomtudományi elemzéseket kritizálta. Ez akkoriban uralkodó közelítésmód volt, amelyet például a politikai viselkedés empirikus vizsgálatai nem cáfoltak. Olson munkásságának, és az annak nyomán kibontakozó fogolydilemma paradigmának elsődleges tanítása az volt, hogy a moralitás mindenképpen több, mint stratégiai gondolkodás. Ezzel a közösségi erkölccsel foglalkozó játékelméleti kutatásokat visszahúzta a neoklasszikus mikroökonómia és a klasszikus politikai filozófia szemléletéhez, ahol az önző önérdék-követésen túlmutató motivációk játszanak kulcsszerepet a morális viselkedésben.

De a kollektív cselekvési dilemma játékelméleti megfogalmazása ezzel együtt túlmutat a neoklasszikus közgazdaságtan mikroökonómiai modelljeinek következtetésein. A fogolydilemma típusú szituációk elemzéséből kiinduló elméleti vizsgálatok ugyanis gyorsan megmutatták, hogy a hagyományos mikroökonómia sugallatával szemben az önzetlenség sem feltétlenül oldja meg a potyautazás dilemmáját – ezt a hagyományos mikroökonómia eszköztárával csak jóval később mutatta meg Lindbeck és Weibull (1988) (lásd erről még Fontaine 2012). Hiába vezetjük be a fogolydilemmába a jóindulatot: a jóindulatú, de továbbra is racionális játékosok esetén könnyen újabb

potyautazási dilemmákhoz juthatunk. Az egyik legismertebb, jóindulatú racionális játékosokat magában foglaló helyzet a hiányzó hős problémája (5. ábra). Ebben a helyzetben mindenki törődik a közös érdekekkel, ugyanakkor annyiban racionális, hogy ha mások már eleget tettek a köz érdekében, akkor (és csak akkor) a potyautazást választják. Itt létrejön egy ahhoz hasonló bizonytalanság, amely az egyik interpretációban a Nash/Crowe-féle „csajozási játszmat” is jellemezte. Ennek a bizonytalanságnak a következménye pedig az lehet, hogy a jóindulatú, de racionálisan takarékos szereplők egymásra várnak, és végül nem cselekszenek kellő számban a köz érdekében (a számításoktól most eltekintek – ez a játszma is közkedvelt példája játékelméleti tankönyveknek).

5. ábra. Hiányzó hős típusú játszma – az egyensúly felé ható „erőkkel”

		Te	
		önfeláldozó vagy	potyautazol
Én	önfeláldozó vagyok	Én: 2 ↓, Te: 2 →	Én: 2, Te: 3
	potyautazom	Én: 3, Te: 2	Én: 1 ↑, ← Te: 1

Elmélettörténeti áttekintésünkben már utaltunk a 21. század első évtizedében népszerűvé váló, a társas preferenciákat vizsgáló kutatási programra, melyről a következő fejezetben lesz szó. Ennek a kutatási programnak az ember kettős természetére vonatkozó alapfeltevése a 18. századig nyúlik vissza – erről volt szó korábban. Ugyanakkor a dichotómia egyik közvetlen elméleti megalapozója a fogolydilemma paradigma volt. A másik megalapozó elmélet a már bemutatott Becker-féle családgazdaságtan önzőség és altruizmus kettőségén nyugvó motivációs modellje. De a társas preferenciák programja több közvetlen következtetést vett át a fogolydilemma paradigmakutatásaiból, mint Beckertől és a neoklasszikus közgazdaságtan többi altruizmus-kutatójától. Ezeket az eredményeket ebben és a következő fejezetben fogjuk megismerni.

A kollektív cselekvés problémájának Olson-féle közgazdasági elemzése cáfolja az önédekkövetés és a közérdek összehangolódá-

sáról szóló, Adam Smithnek tulajdonított téziseket, és így látszólag nyit a szociológia szemléletmódjának irányába. Valójában azonban Olson ugyanúgy a közgazdasági imperializmus képviselője volt, mint később Becker, akinek munkásságához kapcsolódóan e fogalom elterjedt (a közgazdasági imperialisták saját diszciplínájuk szemléletmódját kívánták más diszciplínák hagyományos kutatási területein alkalmazni). Olson egyébként nyíltan kritizálta és korlátoztnak tekintette a munkások együttműködését vizsgáló szociológusokat (Olson 1997 [1965]). Az elmélete pont arra kívánt rámutatni, hogy miképp lehet megérteni a szakszervezet fenntartásában kollektíve érdekelt munkásság tartózkodását a szakszervezet aktív támogatásától, és kritizálta azokat, akik irracionálisnak tekintették a munkások potyautas magatartását.

Mindazonáltal azt is látni kell, hogy Olson modellje nyitotta meg az utat a fogolydilemmához hasonló társadalmi csapdahelyzetek kezelési mechanizmusainak elemzéséhez. Bevezette ugyanis a társadalmi szelektív ösztönzők fogalmát. Ezek olyan, a kollektív cselekvésben érdekelt közösség tagjai által a cselekvőnek kínált, nem anyagi jellegű költségek és jutalmak, amelyek a potyautazás helyett a kollektív cselekvésben való részvételre ösztönzik a racionális döntéshozót – aki a szelektív ösztönzők nélkül egy egyszerű fogolydilemmával szembesülne, és a domináns stratégiát követné. Olson ezzel elismerte, hogy a társadalmi környezet elvárásai fontos szerepet játszhatnak az emberi együttműködés létrejöttében, az erkölcsös döntések meghozatalában. Ez szemben áll a neoklasszikus közgazdaságtan belső preferenciákon alapuló közelítésmódjával.

Olson – a társadalmi szelektív ösztönzők szerepének elismerése mellett –, jobban érdekelték az intézményi megoldások segítségével létrejövő, együttműködést támogató szelektív ösztönzők, beleértve a szavazással meghatározott kötelességek intézményi segítséggel történő kikényszerítését is. A csoportérdek és a szavazáson alapuló politikai rendszerek kapcsolatáról később még szó lesz.

Ezen a ponton egyelőre az a fő kérdés, hogy az önző emberek miként szabadulnak ki a fogolydilemma jellegű társadalmi csapdából, kívülről érkező varázslatos szelektív ösztönzők és intézményi kény-

szerek nélkül? Nemsokára kiderül, hogy miért tekinthetőek varázslatosnak a szelektív ösztönzők a szigorú, racionális döntéseméleti elemzés nézőpontjából. Elég csak abba belegondolni, hogy a szelektív ösztönzők nyújtása is költség valakinek, aki ezzel nem a saját egyéni hasznát növeli, hanem a közösségnek tesz jót – a potyautazás kérdése tehát elkerülhetlenül felmerül a társadalmi szelektív ösztönzők nyújtásának vizsgálatakor is. Egy következetes, racionális döntéseméleti elemző számára az intézményi megoldások kapcsán is felmerül a kérdés, hogy ki és miért tartaná be az intézményesen kötelezővé tett, költséges szabályokat? Nyilván a büntetéstől fél, de valakiknek fogantatosítani kell a potyautasokkal szembeni büntetést, akikkel kapcsolatban ugyanúgy feltehető a fenti kérdés. A válasz racionális döntéseméleti (játékelméleti) keretekben megadható, de a megértéséhez végig kell menni azon az úton, amelyen a fogolydilemma problémájának kutatói is végigmentek a múlt század hatvanas és kilencvenes éve között.

### III.2. Jössz te még az én utcámba!

#### III.2.1. KÉTSZEMÉLYES JÁTSZMA

##### III.2.1.1. A jövő árnyéka

A fogolydilemma más perspektívába kerül, ha a két szereplő közötti interakció sokszor megismétlődik egymás után. Ismétlődő interakciókban a szereplők a másik természetére vonatkozó tapasztalattal, és a saját maguk természetére vonatkozó reputációval érkeznek a döntési helyzetbe. Ezen túlmenően a döntés során kalkulálni tudnak azzal is, hogy nem utoljára találkoznak egymással. Politológusok és szociológusok ezt később úgy fogalmazták meg, hogy az ismétlődő fogolydilemmában a múlt és a jövő árnyéka vetül a döntésekre (Axelrod 1984, Raub és Weesie 1990).

A fogolydilemmában a domináns stratégián nyugvó egyensúly, és a kölcsönösen előnyös kimenetel feszültsége definiálhatóvá tette az

önző és az erkölcsös magatartás dichotómiáját. Ugyanakkor a racionális döntéelmélet azonnal választ is adott a dichotómiából következő dilemmára: a racionális ember csakis a potyautazást választhatja. A múlt és a jövő árnyéka ugyanakkor az önző racionális ember számára is valódi dilemmává teszi a döntést.

Három tényezőt érdemes kiemelni. 1. Idővel nő a különbség az egymás céljait keresztező potyautasok és a kölcsönösen jóindulatú szereplők boldogsága között (Lave 1962). A potyautasok vesztesége összeadódik, a kölcsönösen jóindulatú játékosok előnye hosszú távon egyre látványosabb; és ezt már a játszma legelején előre láthatják a szereplők. 2. Bizonyos megszorítások mellett észszerű lehet a becsületes ember reputációját felépíteni. 3. Lehetőség van arra, hogy elkerüljük a becsületes balek szerepét, mivel a későbbi interakciók során reagálni tudunk arra, ha a másik szereplő nem díjazza becsületességünket, és továbbra is potyautazik („Jössz te még az én utcámba!”).

Néhány kutató rögtön a dilemma bevezetését követő években elkezdett foglalkozni az ismétlődés lehetőségével (Lave 1962, 1965, Morehous 1966, Rapoport és Dale 1966). E korai tanulmányokban diákok körében végzett kísérletekben vizsgálták a fogolydilemma ismétlődésének hatását az együttműködési hajlandóságra. A kísérletek eredményei azt mutatták, hogy az iterált fogolydilemmában a várakozásoknak megfelelően növekszik a résztvevők együttműködési hajlandósága.

Az ismétlés nem a valóságtól elrugaszkodott elméleti konstrukció. Hétköznapi interakcióink jelentős részét tudjuk ismétlődő játszmákkal modellezni. A számunkra fontos társas tevékenységeink jellemzően tartalmaznak ismétlődő struktúrákat. A fogolydilemmával közelíthető társadalmicsapda-helyzetekben is gyakori az ismétlődés. Sokszor annyira gyakori, és oly sokszor ismétlődik, hogy nem is érezzük, hogy a másik iránt jóindulatú döntéseinkkel egy társadalmi csapdából mászunk ki nap mint nap.

III.2.1.2. Második példázat: *carpe diem* a napfényben?

## A nyaralási dilemma

Az ismétlődés vonzó magyarázat arra, hogy önző és racionális emberek miként kerülnek ki a fogolydilemma által létrehozott csapdából. Azonban az imént bemutatott egyszerű formájában ez a magyarázat elégtelen. Lássuk az alábbi példát!

Egy turistaparadicsomban az egyik turnusban egy kétlakásos apartmanházban egymás mellé kerül két pihenni vágyó család, két különböző országból. A tengerparti ház kisebb hátránya, hogy nincs külön tárolóhelyiség a felfújott gumicsónakoknak és egyéb strandkellékeknek, továbbá az apartmankomplexum központi szemeteskukája is messze van. A lusta és önző lakók az apartman közös előszobájában tárolják a strandkellékeket és a szemeteszacskókat. Ha mindkét család így tesz, akkor hamarosan már a ki- és bejutás is nehézkes lesz. Ha a szemetet sem viszik el több napig, még bűdös is van. Kulturált előszobájuk a nyaralás alatt akkor lesz, ha mindkét család időt szán minden reggel és este a kellékek felfújására, majd elrakására (és rendszeresen elslattyognak a központi kukához).

Tegyük fel, hogy a szóban forgó „Apartman I”-be csak individualista családokat szállásolnak el. Közöttük vannak olyanok, akik akkor is megtesznek mindent a tiszta és rendezett előszobáért, ha nincs másik család az apartmanban, mert szeretik a rendet, undorodnak a szemét bűzétől, és nem lusták. De vegyük most azt az esetet, amikor két országból egy-egy lusta család kerül egymás mellé egy hétre! Vagy akár két hétre. Az világos, hogy ha csak két napra kerülnének egymás mellé, egyik sem törődne a másikkal okozott kellemetlenséggel; és első nap este mindent széthagynának.

Mi történik akkor, ha egy-két hétig laknak egymás mellett? A két családot foglyul ejtő fogolydilemma ez esetben ismétlődő játszma; minden nap végén döntést kell hozniuk. Lesz-e rend és tisztaság a fentiek értelmében az előszobában? A játszma játékelméleti modellje az ismétlődéssel kapcsolatban megfogalmazott korai intuíciókkal és kísérleti eredményekkel szemben azt implikálja, hogy amennyiben hűek maradnak önző elveikhez, és nem engednek lustaságukból

sem, akkor bizony rendetlenség és bűz lesz az előszobában a nyaralás alatt.

Ennek megértéséhez a szekvenciális játzmákra értelmezhető visszafelé göngyölítés eljárását a legcélszerűbb alkalmazni (például Selden 1978). Mi történik az utolsó este, amikor a jövő árnyéka már nem vetül rá a lemenő nap vöröses fényében úszó előszobára? Egy igazán önző és lusta családot bizony semmi sem ösztönöz arra este, hogy az egész napos pihenéstől fáradtan még a rendrakással is foglalkozzon (főleg, ha másnap délelőtt, indulás előtt még csobbannának egyet). Másnap már senki sem jön az utcájukba, nem büntetheti vissza őket többet a szomszédjuk. Tehát az utolsó napot kihúzzhatjuk az együttműködés sorozatából. Mindezek tudatában mi történik az utolsó előtti este? Számíthat-e „hálára” a rákövetkező, utolsó napon a szomszéd (szintén önző) családtól az önző család? Inkább mindenképpen „büntetésre” számíthat: bármit tesz most, a másik mindenképpen szét fogja hagyni a cuccait az utolsó este. Ennek következtében – ösztönzés híján –, a lusta és individualista családok az utolsó előtti este is széthagynak mindent az előszobában. De ha az utolsó előtti este is mindenképpen rendetlenség és bűz lesz az előszobában, akkor mi ösztönözné e családokat arra, hogy két nappal a nyaralás vége előtt elpakoljanak maguk után?

A fenti gondolatmenetet folytatva, a visszafelé göngyölítés logikájával eljutunk az első naphoz. A játékelmélet szigorú logikáját alkalmazva a konklúziónk az, hogy hiába ismétlődik meg többször a fogolydilemma, a játékosok az első játzmától az utolsóig a lusta stratégiát választják. Lássuk be, hogy a visszafelé göngyölítés logikája a tetszőleges alkalommal megismételt fogolydilemmára alkalmazható. Tehát két ilyen család lakhat egymás mellett akár egész nyáron, már a legelső napon a lusta stratégiát kellene választaniuk. De tekinthetnének akár két, egy kollégiumban vagy magánszálláson hasonló elrendezésben egymás mellé kerülő, (és a rend kedvéért két különböző országból származó, és különböző szakon tanuló, egymástól eltérő érdeklődésű) diák játzmáját is, akik akár több szemesztert is egymás mellett töltenek. A játékelméleti logika eredménye ugyanaz.



Könnyen belátható, hogyha az ismétlődő fogolydilemmákból létrejövő játszmat egyetlen játéknak tekintjük, akkor a fent bemutatott visszafelé göngyölítés módszerével e játék egyetlen Nash-egyensúlyát találjuk meg. Tehát, ha a fent megismert mátrixformában felírjuk a játékot, és keressük az optimális reakciókra adott optimális reakciókat, akkor ugyanezt a megoldást kapjuk: soha senki nem kooperál.

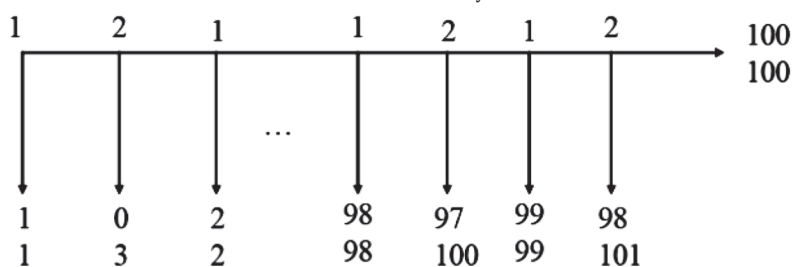
Azt is látnunk kell, hogy a fogolydilemma és a belőle származtatott játszmák ebből a szempontból is speciális eseteknek tekinthetők. A szekvenciális formában felírható játékokra általában nem igaz, hogy a visszafelé göngyölítéssel kapott megoldások lefedik az összes Nash-egyensúlyt. De az kétségtelen, hogy a játékelmélet főárama „kiemelt” egyensúlyként kezeli a visszafelé göngyölítés eredményét, amelyet a valódi, észszerűen várható kimenetelnek tekint a Nash-egyensúlyok időnként meglehetősen nagy halmazában (Harsányi és Selten 1988). Végül, e könyv gondolatmenetének szempontjából az is fontos, hogy vannak olyan, a kutatási programunkban fontos szerepet betöltő kutatók, akik nem tartják legitimnek az ilyen, és ehhez hasonló, a Nash-egyensúlyok halmazát szűkítő megoldásokat (Binmore 2010). Erre a vitára még többször visszatérünk.

Azonban az vitán felül áll, hogy bármilyen nagy, de véges számú alkalommal ismétlődő fogolydilemmában a folyamatosan dezertáló (nem együttműködő) magatartás a racionális. A játékelmélet legnagyobb gondolkodóit is zavarta, hogy ebben a helyzetben a játékelméleti logika konklúziója eltér a hétköznapi intuíciótól és a korai kísérleti eredményektől. Ne feledjük: általános esetben sokan épp a visszafelé göngyölítést tekintik a többi Nash-egyensúlynál jobban védhető megoldásnak a hétköznapi észszerűség talaján. A Nash-egyensúly finomításaiért később Nobel-díjjal kitüntetett Reinhard Selten (1978) kifejezetten paradoxonnak nevezte ezt az ellentmondásos helyzetet.

A játékelméleti szakirodalomban a százlábú játékkal illusztrálták a visszafelé göngyölítés problémáját a hosszan elhúzódó, de domináns stratégiákon nyugvó, Pareto-szuboptimális egyensúllyal rendelkező játszmákban (Aumann 1995, 1998, Binmore 1997, Rosenthal 1981; 6. ábra). A játéknak 198 „lába” van, és az olvasó az ábrában a jobb-

oldali utolsó döntési ponttól indulva végigkövetheti, hogy a játékelméleti logikát követő szereplők hogyan szűrnak ki magukkal.

6. ábra. A százlábú játék



A játékban 1-es és 2-es játékos játszik egymással. Az elágazásokban döntési pontok vannak, ahol 1-es vagy 2-es játékos dönt. A nyilak az egyes lehetséges kimenetekre mutatnak. Az 1-es játékos boldogságszintje felül, a 2-es játékosé pedig alul szerepel.

Selten a játékosok racionalitására vonatkozó feltevésekkel próbálta megoldani a paradoxont, de hosszabb távon nem az ő megoldása mutatkozott termékenynek. A megoldás kulcsa a legtöbb játékelméleti kutató szerint az információs feltevésekben rejlik. David Kreps és munkatársai (1982) a véges sokszor lejátszott fogolydilemmára vonatkozóan megmutatták, hogy elég egy kisebb bizonytalanság a másik játékos racionalitásával kapcsolatban, és a kölcsönösen együttműködő stratégiakombináció is Nash-egyensúlyt eredményez. A bizonytalanság tehát nem zárja ki a kölcsönösen előnytelen kimenetelt, de megnyitja az utat a kölcsönösen segítő viselkedés irányába is. Nem arról van szó, hogy a modellező feltételezi, hogy az egyik játékos valamekkora eséllyel irracionális jóindulatra hajlamos. Két racionális játékos van, akik közül az egyik bizonytalankodik kicsit. Ráadásul az együttműködéshez vezető bizonytalanság létrejöhet úgy is, hogy mindkét játékos bizonyos abban, hogy a másik is racionális, csak épp az egyikük bizonytalan abban a tekintetben, hogy a másik mit gondol őrá (és a másik tud-e erről a bizonytalanságról). És a bizonytalanság adott esetben jelentkezhet még magasabb áttételen is („nem vagyok biztos benne, hogy ő biztos benne, hogy én biztos vagyok benne, hogy ő racionális” stb.).

A fogolydilemma ismétlődésével foglalkozó korai tanulmányok (Lave 1965) is utaltak rá, hogy az ismétlés lehetőséget ad arra, hogy a játékosok a korai fordulóknban jelzéseket adjanak saját jellemükről, ezzel befolyásolva a másik racionális kalkulációját a rákövetkező fordulóknban. Ken Binmore (1997) szerint ezért inherens az ismétlésben a bizonytalanság: racionális lehet megingatni a másik bizonyosságát a mi tökéletes racionalitásunkban (vagy bizonyosságunkban az ő racionalitásáról stb.), mindkettőnket előnyhöz juttatva ezzel. Ezért tartja értelmetlennek azt a levezetést, miszerint ha mindkét fél tudja, hogy mindkét fél tudja stb. a végtelenségig, hogy mindketten racionálisak, akkor az ilyen ismétlődő társadalmi csapdáknban nem indul el az együttműködés (Aumann 1995, 1998). A visszafelé göngyölítés logikáján nyugvó egyensúlyok speciális erejét hangsúlyozzók a hiteltelen fenyegetések tényleges hiteltelenségére mutatnak rá. Ha tudjuk, hogy mindenkinek világos, hogy mindenkinek világos (stb.), hogy egy bizonyos lépés nem éri meg valakinek, akkor úgy kell kalkulálni, hogy az a lépés nem fog megtörténni, és az a lépés az adott feltételek között nem is fog megtörténni a racionalitás világában. A valóságban azért nehéz megfigyelni a visszafelé göngyölítés alkalmazását az ismételt társadalmi csapdahelyzetekben, mert ilyen szintű teljes kölcsönös bizonyosság a másik racionalitásával kapcsolatban nem létezik. A játékelmélet alapkérdéseivel foglalkozók körében talán ez tekinthető az elfogadottabb álláspontnak. A másik oldal ellenben az irracionális ígéretének észszerűségére mutat rá: nem zárhatjuk ki a racionális döntési opciók közül azt, hogy elbizonytalanítom a másikat a saját racionalitásommal kapcsolatban. Az ismétlődő interakciók kezdetén lehetséges a teljes bizonyosság a szereplők racionalitásával kapcsolatban; ezt a bizonyosságot racionális módon le lehet rombolni.

A racionalitás kérdésében elfoglalt kétféle álláspont ellentéte később még szerepet kap a társadalmi együttműködés gyökereinek kutatásában. Az önző egyének közötti együttműködés kutatásának ebben a fázisában azonban a két oldal közötti hasonlóság a fontos: az ismételt fogolydilemmában az önző racionalitással kapcsolatos kételemek vezethetnek – önbeteljesítő várakozásokon keresztül –, kölcsönösen előnyös együttműködéshez.

A bizonytalanság ugyanakkor vonatkozhat arra is, hogy mire vágyunk az önző cselekvők. A modellezés szempontjából ekvivalens lehet a preferenciákkal szembeálló irracionalitás, és a fogolydilemmáétól eltérő egyéni ízlés. Például a nyaralós dilemmában az egyik lusta család feltételezheti azt a másikról, hogy talán ők is lusták, de eléggé szeretik a rendet: a tiszta előszobát látva inkább elpakolnak, és csak akkor hagyják az egészet, ha látják, hogy tőlük függetlenül is rendetlenség van, és ezért fáradásuk ellenére is káosz és bűz lesz az előszobában. Ebben az esetben az egyéjszakás nyaralás esetén még mindenképpen rendetlen család egy hosszabb turnus elején megfontolhatja, hogy a hosszú távon elég kellemetlen rendetlenség és bűz elkerülése érdekében megpróbál rendet tartani, aztán csak a végefelé kanászodik el.

Egy közelmúltban publikált metaelemzés segít rávilágítani a visszafelé göngyölítés vita mindkét álláspontjának igazságára. Embrey és szerzőtársai (2017) öt korábbi, egymástól független kutatás adatait egyesítve és újraelemelve vizsgálták, hogy miképpen reagálnak a kísérleti alanyok az itt tárgyalt, ismételt fogolydilemmákban. Azt találták, hogy a szigorú visszafelé göngyölítési hipotézissel szemben a kísérletekben részt vevő játékosok reagálnak az együttműködés hosszára. Ugyanakkor, ahogy nő a játékosok tapasztalata, úgy hozzák előrébb (messzebb a játék végétől) kilépésüket az ismétlődő együttműködésből.

Van azonban az ismétlődő emberi interakciókban egy olyan, gyakran jellemző bizonytalanság is, amely alapvetően változtatja meg az emberi játszmák karakterét. Életünk sok fontos és kevésbé fontos társas interakciója tekinthető olyan ismétlődő játszmának, amelynek nem ismerjük a végét. Amikor a lusta család hazaért a nyaralásról, többek között a szomszédaival, rokonaival, munkatársaival és a helyi boltosokkal kell megtalálnia a közös hangot apró-cseprő ügyekben, és komoly dilemmákban is. És akkor még nem beszéltünk a család tagjainak egymás közötti játszmáiról. Ezekről az ismételt dilemmákból álló játszmákról tipikus esetben nem tudjuk, hogy mikor érnek véget.

## III.2.1.3. Végtelen perspektívák egy véges világban

A játékelmélet korai és fontos eredménye az ún. „népi tétel”, amely alapján a fogolydilemmára vonatkozóan is megfogalmazhatták, hogy végtelen ismétlés esetén a folyamatos dezertálás mellett a kölcsönösen előnyös együttműködés stratégiái is Nash-egyensúlyt eredményezhetnek (Fudenberg és Maskin 1986, Rubinstein 1979). Végtelen ismétlés esetén ugyanis nincsen olyan utolsó, elutazás előtti játszma, ahonnan a lusta és rendetlen stratégia parancsát vissza lehet vezetni a legelső estéig. Mindig van holnap, és ha bármikor letérek az együttműködés rövid távon költséges útjáról, hosszú ideig (örökké) tartó büntetést kaphatok.

Az örök kárhozat fenyegetése rövid földi életünk körülményei között azonban hiteltelen (csak Istennel érdemes erre alapozott racionálisan együttműködő játszmát játszani, lásd Brams 2018, Johnson és Krüger 2004, Pascal 1670). Azonban azok az iterált véges játszmák, ahol minden egyes iteráció után azonos a következő alkalom valószínűsége, modellezhetőek végtelenszer ismétlődő interakcióknak (Binmore 1998, Fudenberg és Maskin 1986) – habár még ez is viszonylag merész, életidegen feltevésnek tekinthető. De gondoljunk bele: például a hosszú távú, és egyúttal intenzív emberi kapcsolataink esetében nagyon (érzékeltlenül) lassan változik annak a valószínűsége, hogy másnap nem folytatódnak-e mindennapi játszmáink. Egy naponta a közeli pékhez járó nagyon idős és beteg ember talán érzékeltően nagyobb eséllyel nem megy másnap kenyérért, mint a nemrég a környéken lakást vásárló, stabil állással rendelkező, és megállapodott, a pék friss áruját szintén kedvelő fiatalember. De egyik napról, hétről vagy hónapról a másokra ezek az esélyek nem változnak (pontosabban kis eséllyel változnak, nagy eséllyel érzékeltetlen mértékben csökkennek csak). A pék tehát nyugodtan szemlélheti úgy, hogy végtelen iterációjú játszmákban szerepel a helyi vevőivel, amelyekben rosszkódásának (például rásózza az előző napról megmaradt száraz pékárut a benne megbízó vevőkre) következménye örök kárhozat lehet (ha például soha többet nem jön a becsapott helyi vevő).

A bizonytalan ideig tartó ismételt fogolydilemma játszma már a hetvenes években kulcsszerepet játszott annak illusztrálásában, hogy racionális és önző emberek hogyan tudnak együttműködni egymással külső beavatkozás nélkül. Nem csupán közgazdászok (Kurz 1978), de a játékelmélet iránt érdeklődő politikatudósok is (Axelrod 1984, Taylor 1976) intenzíven foglalkoztak a racionális emberi együttműködés modellezésével. Utóbbiak műveikben azt sugallták, hogy az ismételt fogolydilemmával megtalálták az emberi együttműködés titkát, amely nem az emberi jóindulatban rejlik hanem a hosszan tartó társas interakciók együttműködésre készítő struktúrájában. Erről a kutatási irányról a következő fejezetben lesz szó.

Ugyanakkor problematikus a végtelenszer lejátszott fogolydilemma olyan interpretációja, amelyből egyenesen következik a tartós kölcsönös együttműködés. Egyrészt, ahogy említettem, a tartós tartózkodás az együttműködéstől ugyanúgy egyensúly marad, mint a tartós együttműködés – bár kétségtelenül kevesebbet hoz a konyhára. Komolyabb probléma, hogy a végtelen játékban végtelen sok egyensúlyi stratégiakombináció azonosítható: ezek közül számtalan alig tartalmaz együttműködő döntéseket. Ráadásul a tartós együttműködés is sokféle stratégiakombináció eredménye lehet: a potyautasokat sokféleképpen lehet sikeresen fenyegetni, és különböző fenyegető stratégiák vezethetnek ugyanolyan együttműködő viselkedéshez. A tiszta modellek ideális, teljes információs környezetében talán kisebb a jelentősége annak, hogy milyen fenyegetésekkel tartják fenn a résztvevők a tartós együttműködést. Azonban a való világ zavaró tényezőinek jelenlétében jelentőséget nyerhet, hogy milyen stratégia alapján reagálunk például dilemmapartnerünk hibáira. Például, ha az egyébként kiváló pék egy nap *bepróbálkozik* azzal, hogy a száraz maradékot sózza ránk; vagy ha az egyébként jó fej szomszédunk előzetes figyelmeztetés nélkül tart pihenésünket zavaró partit; vagy ha az egyébként megbízható munkatársunk nem készül el határidőre az aktuális közös projektünk rá eső részével, minket is befeketítve a főnök előtt. Mit tegyünk ebben az esetben? Mennyi hiba után hagyjuk abba a kooperációt, és mennyi ideig büntessünk dezertálással? A játékelmélet nem ad útmutatást erre. Ha hiszünk a visszafelé gön-

gyölítés erejében, akkor a teljes információ és tökéletes racionalitás világában az örök haragtartó stratégia tudja elrettenteni a másik felet a bepróbálkozástól (Binmore 1998). De ha a legkisebb hiba, bizonytalanság is előfordulhat, akkor már radikálisan változhat a helyzet.

A végtelenszer ismételt fogolydilemma korai elemzői azt feltételezték, hogy a játékosok önzőek, de korlátozott a racionális kalkulációs képességük. Nevezetesen, igen egyszerűen látják a lehetséges opciókat, és a végtelen választékból csak néhány egyszerű stratégia közül választanak (Kurz 1978, Taylor 1976). Robert Axelrod, a Michigani Egyetem politikatudósa más utat választott a reálisan várható stratégiák szűkebb halmazának meghatározására: a hetvenes évek végén programozói versenyt hirdetett amerikai akadémiai körökben az ismételt fogolydilemmában alkalmazható stratégiák megalkotására (Axelrod 1984). A verseny győztese Anatol Rapoport kanadai pszichológus lett, aki régóta foglalkozott az alapvető emberi játszmák játékelméleti modellezésével, például az ismételt fogolydilemmával is (Rapoport és Dale 1966), és egy pár soros, Basic nyelven írt programot adott le. Axelrod az automaták versenyével valójában túllépett a racionális döntéshozatal vizsgálatán. Eredményei és közelítésmódja megnyitotta az utat a fogolydilemma evolúciós elemzése, továbbá az önérdékkövetésen túlmutató emberi motivációk intenzívebb vizsgálata felé. A következő fejezetben a nem kis részben az Axelrod által mozgásba lendített új elméleti, majd később empirikus kutatási irányról lesz szó. Kiderül majd az is, hogy mit tartalmazott Rapoport pár soros programja. Ám az ismételt fogolydilemma evolúciós vizsgálatának intenzívvé válásával még nem zárult le a racionális együttműködés lehetőségeinek kutatása.

### III.2.2. SOKSZEMÉLYES JÁTSZMA

#### III.2.2.1. Ki bünteti a vétkezőt?

A hetvenes évek végén a fent nevezett politikatudósok publikációi nyomán könnyen lehetett azt a következtetést levonni, hogy közel

jutottunk az emberek közötti jóindulatú viselkedés megértéséhez. Azonban Pamela Oliver, a Wisconsin Egyetem szociológusprofesszora, a kollektív cselekvés formális modelljeinek kutatója felhívta a figyelmet egy komoly hiányosságra (Oliver 1980): a sokszereplős interakciókban nem triviális, hogy miként működik a racionális büntetés kétszereplős fogolydilemmában megfigyelhető mechanizmusa. Másodrendű potyautas problémának nevezte el azt a kétfősnél nagyobb csoportokban megjelenő dilemmát, amely arról szól, hogy ki büntesse meg a csoport egésze ellen vétkező önző embert. A másodrendű potyautas probléma nem fogolydilemma, inkább a hiányzó hős típusú játékkal írható le (5. ábra). Ha tudom, hogy más büntetni fog, akkor én már nem fáradok azzal, de ha tudom, hogy más nem teszi meg, akkor magamra vállalom. Tehát hiába jutott el a fogolydilemma a végtelen ismétléses modellekhez, az csak egyes erős társas (családi, baráti és néhány kollegiális) köteleink működését magyarázza meg, de keveset mond az idegenekkel szembeni jóindulat gyökereiről, és a nagyobb közösségeket összetartó erőről.

A sokszereplős helyzetek problémájáról a játékelmélet közgazdász kutatói is tudtak, és a nyolcvanas évek végére előálltak a megoldással. Rasmusen és Hirshleifer (1989) sokszereplős fogolydilemma játékukba bevezették az ókori görög demokráciákból ismert osztrakiszmosz intézményét: a potyautasokat szavazással ki lehet zárni a közösségből. Modelljükben egy ember kizárása csökkenti az együttműködésből elérhető haszon nagyságát. Azonban maga a kizárási aktus nem költséges – és bárki kizárhat bárkit. Megmutatják, hogy ezekkel a feltételekkel még véges sokszor lejátszott, ismétlődő sokszereplős fogolydilemmában is létrejöhet az együttműködés (az utolsó előtti fordulóig) egy olyan stratégiakombináció következtében, ahol az első fordulótól kezdve mindenki kooperál, továbbá a második fordulótól kezdve mindenki kész arra, hogy kizárja azt, aki nem működött együtt az előző fordulóban, majd a következő körtől kezdve mindenki kész arra is, hogy kizárással büntesse azokat is, akik az előző körben nem büntették meg azt, aki két körrel ezelőtt dezertált – és így tovább. Rasmusen és Hirshleifer (1989) megoldása egyrészt közelít a közösségek valós normafenntartó mechanizmusainak modellezé-



séhez, másrészt az ingyenes osztrakizmosz feltételezésével sikeresen megkerüli a másodrendű potyautas problémát. A büntetési láncolat bevezetésével (büntetem a nem büntetőt is) ugyanakkor mégis közelebb jutottunk a racionális együttműködést fenntartó mechanizmus leírásához.

#### III.2.2.2. A racionális büntetési mechanizmus

A fogolydilemma két stratégiájának (együttműködés vs. potyautazás) keretei között maradván végül a Princeton Egyetemen tanító japán közgazdász, Kandori Michihiro mutatta meg a racionális közösségi normafenntartás egy életképes mechanizmusát (Kandori 1992). Modelljének kulcsa az ún. matching game (párosítós játszma) bevezetése volt. A szociológia ekkoriban megerősödő hálózatelemzési paradigmájának követőihez (Granovetter 1985) hasonlóan azt feltételezte, hogy a sokszereplős együttműködési dilemmákban is kétszereplős interakciók során érintkeznek egymással a csoport tagjai, és a társadalmi normák érvényesítése ezekben az interakciókban történik (lásd még Coleman 1990). Megengedte ugyanakkor, hogy a kétszereplős interakciók alkalmiak legyenek. Mindig a csoport más tagjával találkozunk, és e találkozások során kell döntenünk arról, hogy betartjuk-e a csoport számunkra költséges viselkedési normáit, és jóindulatúan viselkedünk a másikkal, vagy önző módon dezertálunk.

A Kandori (1992) által alkalmazott párosítós játszma játékelméleti modellje úgy néz ki, hogy a végtelenszer ismétlődő játék minden fordulójában a szereplőket véletlenszerűen sorsolják egymáshoz, és a kisorsolt párok egy egyszeri, kétszemélyes fogolydilemma játékot játszanak. Az együttműködést fenntartó egyensúlyi stratégiakombináció a Rasmusen és Hirshleifer (1989) által is alkalmazott büntetési láncot tartalmazza, és a párosítással oldja meg a másodrendű potyautas problémát. Aki potyautasként viselkedett a párjával szemben egy adott fordulóban, azt a következő fordulóban kötelessége megbüntetnie az új párjának (a büntetés az, hogy az új párja potyautas stratégiát

választ a korábban vétkezővel szemben). Ha az új párja nem bünteti meg a vétkezőt, akkor őt is büntetnie kell annak, akit a rákövetkező fordulóban számára kisorsoltak párnak. És ez így megy tovább a végtelenségig (a potyautast nem büntetöt meg nem büntetöt is büntetni kell stb.). Könnyű belátni, hogy az ilyen, áttételes büntetést is magában foglaló stratégiák kombinációja Nash-egyensúly. Ha mindenki más ezt alkalmazza, akkor nekem sem éri meg eltérni ettől (primer dezertálással, vagy a büntető dezertálás elmulasztásával), ugyanis a szóban forgó stratégia szerint azonnal büntetést kapok a következő fordulóban nekem kisorsolt páromtól. Ráadásul ez a stratégiakombináció a visszafelé göngyölítéssel is elérhető aljáték tökéletes egyensúly is egyben, amely sokak által más egyensúlyoknál hitelesebbnek, „erősebbnek” tekintett megoldás. Az együttműködő egyensúly ereje abból ered, hogy a velem szemben kilátásba helyezett büntetés hiteles fenyegetés, ugyanis tudom, hogy a büntetés elmulasztóját is bünteti majd egy harmadik ember, aki szintén hitelesen fenyeget, hiszen ő is érzi egy negyedik, hitelesen fenyegető ember hideg leheletét a tarkóján – és így tovább, a végtelenségig.

Kandori teljes informáltságot feltételező modellje jól írja le a belátható nagyságú közösségekben működő kiközösítés mechanizmusát, és megmutatja azt az erőt, amely önző és racionális embereket a normák betartására ösztönöz. A vétkezők egy erős normákat fenntartó közösségben páriákká válnak, akivel senki sem szívesen mutatkozik, de aki mégis megteszi, az maga is könnyen páriává válik. Ha mindenki hisz abban, hogy a közösségben mindenki hisz abban (stb.), hogy működik a kiközösítési mechanizmus, akkor az ettől való félelem távol tartja az embereket a normasértéstől, és azoktól, akik mégis megsértették a közösségi normákat. Nem köszönünk nekik, vagy legalábbis nem csevegünk velük, nem segítünk nekik, ha épp ránk lenne szükségük. Ez az informális kiközösítés, amelyben a közösségi játszma vélelenszerű „párosításai” megfelelnek a csoport tagjai közötti, hétköznapi interakcióknak.

Kandori (1992) cikkének, illetve a modelljét továbbfejlesztő tanulmányoknak (például Ellison 1994) az volt a fő törekvése, hogy az egymásról kevesebb ismerettel rendelkező emberek nagyobb kö-

zösségeire vonatkozóan is megmutassák, hogy működhet ez a racionális együttműködést fenntartó mechanizmus. El kell ismernünk a kooperáció korlátozott információk melletti peremfeltételeinek sikeres meghatározását, ugyanakkor azt is be kell látnunk, hogy az együttműködés robusztus, és intuitívan könnyen belátható mechanizmusa csak az átlátható, kisebb közösségekre alkalmazható hitelesen.

A modell nagyobb közösségekre történő alkalmazásának problémáját a fenti közgazdász kutatók az ismeretlen egyének korábbi cselekedeteire vonatkozó információk korlátosságában látták. Közgazdászként az érdekelte őket, hogy nagyobb méretű piacokon is kikényszeríthető-e az állam kényszerítő intézményei nélkül a tisztességes piaci magatartás. A meglehetősen komplex modellek optimizmusával szemben Avner Greif (1994) sokat idézett történeti összehasonlító elemzése a szkeptikusok táborát erősítette. Greif a 12. századi genovai és a marokkói zsidó tengeri távolsági kereskedő-közösségeket hasonlította össze. Arra volt kíváncsi, hogy miképpen birkóztak meg a hajótulajdonosok az értékes áruk, a nagy távolságok és a hosszú hajóutak miatt jelentkező megbízó-ügynök problémával. Azt állítja, hogy a muszlim környezetben bezárkózó, marokkói kereskedők kis közössége intenzív levelezéssel létrehozta a Kandori-féle alapmodellben szereplő normafenntartó mechanizmust. Ha valaki becsapott egy kereskedőt, azt azonnal kipletykálták, és később egyetlen kereskedő sem üzletelt vele többet. A nagyobb, nyitottabb genovai kereskedőközösség viszont képtelen volt a megfelelő információs hálózatot, és az arra épülő informális büntetési mechanizmust fenntartani, ami idővel a kényszerítő (város)állami intézmények kiépüléséhez vezetett.

Ugyanakkor a nagyobb csoportokban létező bizonytalanságnak van egy olyan dimenziója, amelyet az idézett modellek nem érintettek. Nevezetesen az, hogy a csoport tagjai mit gondolnak arról, hogy a csoport többi tagja mit gondol a kiközösítési mechanizmus létezéséről. Emlékezzünk rá: végtelenített játszmaiban a lehetséges egyensúlystratégia-kombinációk száma is könnyen végtelen lesz. A vívódások nélküli döntéshez tudnunk kell, hogy mások is

azon az ösvényen haladnának, mint mi, és ők is tudják a többiekről mindezt, és így tovább. Kisebb közösségekben könnyen kialakulhat ezzel kapcsolatban egy közös vélekedés, nagyobb csoportban azonban ez már kérdéses lehet. A VI. fejezetben még visszatérek erre a problémára.

### III.2.2.3. Normák az önző polgárok világában

A társadalmi normák racionális döntésméleti elemzése felé az ismételt fogolydilemma analízisen nyugvó, Kandori-féle modell, és a hozzá hasonló elemzések egyengették az utat. Ez fontos volt, mivel a neoklasszikus közgazdaságtan racionális döntésméleti paradigmája nyíltan szembehelyezkedett a normáknak kiemelt befolyásoló szerepet tulajdonító parsonsi társadalommodellel; miközben a közjavak problémája iránt már régóta érdeklődő közgazdászok állandóan beleütköztek a társadalmi normák kérdésébe. Ezt a kollektív cselekvés úttörő elemzőjének, Mancur Olsonnak (1965) is el kellett ismernie, amikor létrehozta a társadalmi szelektív ösztönzők kategóriáját. Elinor Ostrom pedig empirikusan elemezte, hogy miképpen gondoskodnak a közösségek saját normafenntartó eszközeikkel a fontos közjavaikról (Ostrom 1990).

A játékelméleti gondolkodás tehát nyitott a normák felé, és ez találkozott néhány, a normákkal foglalkozó filozófus és politikatudós azon törekvéseivel, hogy – a parsonsi hagyományokkal szemben –, egyéni preferenciák és társas interakciók terminusaiban beszéljenek a normákról. Ullmann-Margalit (1977) és Taylor (1976) korai elemzéseit követően Cristina Bicchieri foglalkozott az interaktív episztemiológia társadalmi normákat érintő vonatkozásaival (Bicchieri 1990, 1997, 2006, 2016). Ez a kutatási hagyomány a fogolydilemma paradigmát követte a társadalmi normák szempontjából releváns szituációk leírásában. Azt feltételezték – összhangban az ismételt fogolydilemmák néhány egyszerűsített értelmezésével –, hogy a fogolydilemma helyzetek normatív jelentőségűvé válásakor az eredeti dilemma átértelmeződik a közösségben, és mások rosszállásától való

félelmükben az emberek úgy tekintenek az adott döntési helyzetre, mint egy ún. biztosítási játszámára (Taylor 1976, Bicchieri 2006), ahol megéri a normát betartani, ha mások is azt teszik, de a norma megsértése a racionális döntés, ha mások sem tartják be. Ez a modell rezonál Axelrod (1984) végtelen sokszor ismételt fogolydilemmákat feltételező, egyszerűsített értelmezésére.

A játékelmélet közgazdász képviselőit az együttműködés és a normafenntartás racionális mechanizmusának megfejtése után már nem érdekelték különösebben a társadalmi normák. Ha azonban normagyánús jelenségekbe ütköztek, bátrabban hivatkozhattak rájuk azután, hogy a racionális döntéssel képviselői számára értelmezhető fogalomként tarthatták számon. Ezek a hivatkozások az utóbbi években, a kutatási program regényének a végén megszaporodnak, és ezért még lesz szó róluk. Ízelítőként Uri Gneezy és Aldo Rustichini (2000) nagy visszhangot kiváltó kutatását említem. Ők egy izraeli óvodában vizsgálták, hogy miképp lehet rávenni a dolgozó szülőket arra, hogy délután időben elvigyék gyermekeiket, és ne kényszerítsék kifizetetlen túlórára az óvónőket. A szülők persze igyekeztek, sajnálkoztak, de láthatóan nem mindenkinek volt annyira fontos az óvónők szabadideje, mint a saját elfoglaltsága. A szerzők jól viselkedő közgazdászként azt javasolták, hogy a nem mindennapos, de a könnyen elviselhetőnél gyakoribb késések megszüntetése érdekében az óvónők tisztességesen árazzák be a késést. Ennek eredménye az lett, hogy megszaporodtak a késések – a lelkiismeret-furdalástól megmenekült szülők örömmel fizettek az ügyeleti szolgáltatásért, ha fontosabb dolguk akadt még délután. A pénzbüntetés megszüntetése ellenben nem hozta vissza a korábbi (óvónők által akkor már visszaírt) normát. Az ismételt játékok elméletével összhangban, a szülők már más ösvényen jártak, és nem tudták a nagy késési arányra vonatkozó várakozásokat módosítani („ha mindenki késik, akkor nekem sem ciki”).

#### III.2.2.4. Normafenntartás nagycsoportban: kapcsolatok hálózata

A hálózati paradigma szociológus képviselői némileg eltérően látják a többszemélyes csoportokba ágyazott, a normák fenntartásában szerepet kapó kétszemélyes interakciókat, mint a matching game modelljét bevezető közgazdászok. Arra mutatnak rá, hogy az emberi közösségeket behálózó társas kapcsolatokban hosszabb, ismételt játszmáknak is tekinthető együttműködések zajlanak. Ezekben a közeli ismerősök által játszott kétszemélyes játszmákban a társak megfigyelése is megoldott. Ennek következtében megfelelő kapcsolatháló struktúra mellett nem jelent problémát a csoport mérete: mindenki a jól megfigyelhető ismerősei körében ellenőrzi, hogy betartják-e a nagycsoport együttműködési szabályait. Ez a hálózati paradigma nagyon régi alaptétele (Granovetter 1973), amelynek játékelméleti leírására már csak jóval Kandoriék eredményei után került sor (Janky és Takács 2010, Takács, Janky és Flache 2008). Az általam is jegyzett modellben a kétszereplős játszmák elkülönülnek a teljes csoport  $n$ -szereplős fogolydilemmájától. E kétszereplős helyzetekben – a bizonytalan végű kétszereplős fogolydilemmára vonatkozó népi tétel (és némi optimizmus) segítségével – az együttműködés megoldott, és az együttműködés/potyautazás a nagycsoport fogolydilemmájában választott viselkedés jutalmazására/büntetésére szolgál. E későbbi modell fényében úgy tűnik, hogy Kandori (1992) kapcsolatokra vonatkozó gyengébb feltevései a hálózatkutatók sokszereplős együttműködésre vonatkozó tételeinek kiterjesztését szolgálják. Azt mutatják meg, hogy a tartós kétszemélyes társas kapcsolatok nem is feltétlenül szükségesek a közösségi normák fenntartásához (Kandori elég explicit módon utal modelljének erre az értelmezésére).

Ugyanakkor a hálózati paradigma játékelméleti modellezői szinte már a Kandori-féle modell megjelenésekor figyelmeztettek arra, hogy a laza kapcsolatok Kandori-féle feltevése nem elégséges, hanem éppen hogy szükséges feltétele a nagyobb közösség együttműködésének (Flache 1996, Flache és Macy 1996). Ugyanis ha feltételezzük, hogy kétszemélyes kapcsolatokban a nagycsoport normáinak betartásá-

ra irányuló játszmákon túlmutató interakciók is zajlanak, akkor az erős személyközi kapcsolatok nem feltétlenül erősítik, hanem lehetséges, hogy éppen gyengítik a közösségi normákat. Szerelmünkkel, barátainkkal, rokonainkkal és üzletfeleinkkel is rengeteg ügyletünk van, amelyek közül csak néhány vonatkozik arra, hogy a nagyobb közösségünk érdekében cselekszünk-e a közösség valamely fogolydilemmájában. Ha a közösség fogolydilemmájában az ellenőrzés a potyautazáson gondolkodó egyének legszorosabb szövetségeseire hárul, akkor megnő az esélye annak, hogy más, a párok számára fontosabb ügyletek sikere érdekében, a felek megbocsátóak a tágabb közösséget károsító potyautazással szemben. A különböző játszmák ilyen összefonódására a VII. fejezetben még visszatérek. Fontos szerep jut nekik a normák megértésében.

Kandori (1992) modellje, Flache (1996) és szerzőtársainak megszorító megjegyzéseivel együtt demonstrálja, hogy kényszerítő állami intézmények nélkül meddig juthat az önző és racionális emberek együttműködése. Az eredmények egyrészt megmutatták, hogy bármilyen társas preferencia (és erőszakszervezetek) nélkül is létrejöhet az emberi közösségek együttműködése. De azt is megmutatták, hogy az egyének által belátható csoportméretet elhagyva, a racionális együttműködés esélyei jelentősen csökkennek.

Nyitott maradt tehát a kérdés, hogy mi motiválja a modern nagytársadalmak állam- és szavazópolgárait arra, hogy idegen polgártársaikat érintő döntéseikben önző érdekeikkel esetleg ütköző erkölcsi megfontolásokat vegyenek figyelembe? A következő alfejezetben azt vizsgálom meg, hogy képes-e az önző polgár modellje a szavazói viselkedés magyarázatára.

### III.3. Nyitva maradt kérdés: ki jön az én utcámba?

#### Az önző szavazó

##### III.3.1. EGY KIHÍVÁS

Az önző polgár modelljét többé-kevésbé sikerrel alkalmazták a gazdagok és szegények közötti újraelosztás támogatottságának magyarázatára (például Alesina és La Ferrara 2005, Meltzer és Richard 1981). A magyarázatok sikerességének értékét óhatatlanul csökkenti, hogy ezekben a vizsgálatokban az osztályszavazás szociológiai modelljei is hasonló előrejelzéseket adnak, mint az önző szavazó modellje. Ugyanakkor nehéz a magyarázatokkal szemben kritikát megfogalmazni, mert az egyes szavazók szubjektív mobilitási perspektívái jelentősen befolyásolhatják az eredményeket. Emiatt elégséges kicsiny magyarázóerőt kimutatni, mert a nem magyarázott részt a szubjektív perspektívák idioszinkratikus hatásainak lehet betudni. Azaz olyan helyzeteket enged meg a modell, hogy például egy objektíve jó perspektívákkal rendelkező ember fél a munkanélküliségtől, és ezért lesz az újraelosztást támogató szavazó.

Némileg tisztább tesztje lehet az önző szavazó modelljének az időskori támogatások iránti attitűdök vizsgálata, mivel a felnőtt népesség jelentős része számolhat azzal, hogy nagy eséllyel megöregszik. Ugyanakkor ennek menetrendje sokkal jobban előrejelezhető, mint például a munkanélküliségé. És a perspektívákat meghatározó életkor könnyen mérhető változó.

A téma vizsgálatát különösen izgalmassá teszi, hogy az utóbbi néhány évtizedben a nyugati jóléti államokban egyre nőtt az egyenlőtlenség a jelenlegi és a jövőbeni nyugdíjasok anyagi kilátásai között (Preston 1984).<sup>\*</sup> Ennek fő oka, hogy az adók és jóléti intézmények erőforrásokat csoportosítanak át a fiataloktól és a jövő generációtól az idősebbek felé (Auerbach és Kotlikoff 1987, Gál et al. 2001). Ugyanakkor a szavazók körében nagy az ellenállás a fenntarthatóságot növelő reformokkal szemben, miközben a szakértők már régóta

<sup>\*</sup> Részletesebben lásd Janky (2018).



az állami „piramisjáték” összeomlásának veszélyére figyelmeztetnek (Boeri et al. 2002a,b, Boeri 2004). Ráadásul nem csupán az idősek, hanem a fiatalok is támogatják a generációk közötti újraelosztást végző társadalombiztosítási rendszereket – egyes amerikai vizsgálatok szerint még az idősebbeknél is nagyobb arányban (Silverstein et al. 2000, Hamil-Luker 2001).

Az önző polgár modellje számára kihívás az a tény, hogy a fiatalok támogatnak egy olyan rendszert, amely pénzt von el tőlük a jelenlegi nyugdíjasok támogatására, miközben az elvonásnál kevesebbet kínál saját öregkorukra. Az állami társadalombiztosítási rendszerek sajátosága ugyanis, hogy nem a saját korábbi befizetéseinket kapjuk vissza, hanem az aktuális aktív korosztályok befizetéseit kapjuk meg majd. A demográfiai trendek alapján pedig adott járulékszinthez egyre kevesebb nyugdíjkifizetés tartozik majd a jövőben.

Három megközelítést mutatok be az alábbiakban, amelyek háromféleképpen próbálják feloldani az ellentmondást a nyers adatok és az önző szavazópolgár egyszerű modellje között. Ezek a nyugdíjellátás népszerűsége racionális döntéelméleti vizsgálatainak legfontosabb megközelítései.

### III.3.2. SOK MATEK ÉS KEVÉS EREDMÉNY

Az első közelítés az intézmények szilárdságát építi be korlátozó feltevésként. Az ötletet Browning (1975) vetette fel (vö. Aaron 1966). Browning azt mutatta meg, hogy az intézményekre vonatkozó néhány szigorú feltevés segítségével értelmezhetőek azok az eredmények, melyek szerint a választók többsége olyan rendszert támogat, amely minden időszakban a választók egy kisebbségéhez csoportosít át erőforrásokat, és amellyel – a legelső belépő kohorsz kivételével –, jól láthatóan mindenki rosszul jár. Browning feltette, hogy a választók egy örökké kötelező erejű szavazáson döntenek a nyugdíjrendszer bevezetéséről.

A modell egyszerűen megérthető. Tegyük fel, hogy a társadalom három, egyforma méretű kohorszból áll (fiatalok, középkorúak, idő-

sek). A kiinduló időpillanatban az emberek egy új nyugdíjrendszer bevezetéséről szavaznak. A népszavazás eredménye kötelező érvényű a rákövetkező generációk számára. Ez a kulcsfeltevés. A javaslat  $T$  nagyságú adó kivetését jelenti a fiatalokra és a középkorúakra (egy emberöltőnyi időszakban), és ennek megfelelően  $2T$  nagyságú öregségi ellátás szétosztását az idősek között. Az idősek támogatják a javaslatot, hiszen adóbefizetés nélkül kaphatnak  $2T$  nagyságú ellátást. A középkorúak is támogatják a javaslatot, hiszen  $T$  befizetése után a következő időszakban  $2T$  nagyságú bevételük lesz. Ha eltekintünk a kamatokról és diszkontrátától, akkor a fiatal kohorszoknak mindegy, hogy mi lesz az eredmény. Ha a jövőbeni jövedelmeket diszkontálják a fiatalok, akkor ellenezni fogják a javaslatot. Ám az ellenzők így is csak a szavazók harmadát teszik ki.

Később igazolták, hogy a modern jóléti államokban az életkor szerint tekintett medián szavazó érdekelt a (hosszú távon fenntarthatatlan) járulék- és járadékmértékek fenntartásában (Galasso 2002). Ugyanakkor a modell azt nem tudta megmagyarázni, hogy a nyugdíjrendszerek támogatottságának mértéke miért korrelál nagyon gyengén az életkorral.

A későbbi, komplexebb modellek kicsit lazították a fenti irreális intézményi feltevéseket (Hu 1982, Verbon 1987, Boadway és Wildasin 1989), azonban ezek a finomítások nem győzték meg a kételkedőket az örökre fixált (vagy korlátozottan változtatható, lásd Hu 1982, Verbon 1987) intézményekre vonatkozó feltevés realitásáról.

Egy másik megközelítés egy képzelt játszma-ba helyezte a szavazókat. Sjoblom (1985) azt feltételezte, hogy a generációk között egy implicit megállapodás jött létre, melynek részeként a fiatalok megígérik, hogy támogatni fogják a jelenlegi középkorúakat akkor, amikor azok nyugdíjba mennek, feltéve, ha most a középkorúak is támogatják a jelenlegi időseket (vö. Kotlikoff et al. 1987). Ebből Boadway és Wildasin (1989) továbbá Cooley és Soares (1999a) olyan játékelméleti modelleket építettek, amelyek az előző alfejezetben ismertetett normafenntartó stratégiák racionalitását mutatták meg (Kandori 1992). A fiatal generációk a jövőbeni nyugdíjkifizetések megtagadásával fenyegetik a középkorúakat arra az esetre, ha utóbbiak elmulasztanák a jelenlegi

idősek támogatását. A fenyegetések generációkon átnyúló láncolata azért hiteles, mert a játszmát generációk végtelen sora játssza egymással.

Arról ezek a modellek nem mondanak semmit, hogy miért is képzelnék bele magukat az emberek ebbe a játszmába. Az még megmagyarázható, hogy miért kezdenék el fenyegetni a fiatalok a középkorúakat. De miért lenne ez hiteles fenyegetés? Ráadásul a modell nem szól arról, hogy miképpen szervezik meg magukat a generációs csoportok. Az attitűdök kifejezésénél és a választásoknál hogyan nyilvánul meg ez a fenyegetés? Az osztályszavazás strukturalista modelljei felteszik, hogy az érdekcsoportok valamiképpen megszervezik a szavazói magatartásukat. Talán az előző alfejezetben kifejtett normafenntartási modellek segíthetnek ebben. A vonatkozó szakirodalom nem foglalkozik ezzel a kérdéssel.

A harmadik közelítés a választók valós játszmáikhoz próbál visszatérni. Tabellini (2000) kilép a tisztán önző preferenciák világából, és a már Becker (1974) által megengedett családon belüli altruizmust emeli be szavazási modelljébe. Emellett a jövedelemegyenlőtlenségeket is figyelembe veszi, és nem nukleáris családot feltételez, hanem több generáció altruista szövetségét. E feltevéseknek köszönhetően ebben a modellben nincs szükség Browning szigorú feltevésére az örökre szóló népszavazásról.

A modellben az átlagos, vagy annál magasabb jövedelemmel rendelkező középkorúak egyáltalán nem támogatják a nyugdíjellátást. Ehelyett a privát ellátást preferálják, mivel így több jut el a szüleikhez, mint az adókon keresztül. Az átlagnál kevesebbet kereső aktívak szülei ellenben a dolgozók teljes fizetésének elvonását is megszavaznák. A többség akkor szavaz egy, a fiatalokat és a jövő generációit kizsákmányoló nyugdíjrendszerre, ha sok idős van a társadalomban, és nagyok a jövedelemkülönbségek. Az idősek magasabb aránya segíti elő, hogy a mediánszavazó egy átlagosnál szegényebb, ezért adópárti dolgozó legyen. A nagyobb egyenlőtlenségek pedig azt jelentik, hogy a fenti szegény mediánszavazó az átlagnál jelentősen szegényebb, ezért magas adókat preferál.

Ez a közelítés kiinduló feltevéseiben reális, ugyanakkor – hasonlóan Browning (1975) modelljéhez –, nem képes a szavazói prefe-

renciák megoszlásának magyarázatára. A modell szélsőséges előrejelzéseket ad: a választók egy jelentős része egyáltalán nem akar állami nyugdíjrendszert, egy másik jelentős része ellenben akár minden jövedelmet elvenne a dolgozóktól a nyugdíjak kifizetése érdekében. Ez köszönőviszonyban sincs a közvéleménykutatási eredményekkel. Azt említettem, hogy nincs idős-fiatal törésvonal a kérdésben, és a jövedelem szerint is csak nagyon gyenge (Silverstein et al. 2000, Hamil-Luker 2001). Gelissen (2001) ugyanakkor kétségkívül kimutatta, hogy a magasabb státuszú szavazók az átlagosnál jobban idegenkednek a túlságosan egyenlősítő nyugdíjrendszerektől.

Az önző szavazó elméletének hiányosságait az intergenerációs újraelosztást végző intézmények iránti preferenciák modelljei mutatják meg. Így az önző szavazó elmélete az önző polgár koncepciójának Achilles-sarka.

A következő fejezetben bemutatásra kerülő elmélet olyan kiegészítést emel be a polgárok motivációi közé, amely számos területen könnyíti meg a megfigyelhető közpolitikai preferenciák magyarázatát. Kérdés, hogy antropológiailag helytálló kiegészítésről van-e szó.

## IV. 1984\* – A tisztességes polgár erkölcsse

### IV.1. Kezdetben vala az szemet szemért parancsa

#### IV.1.1. VÁLASZTÁS AZ EGYENSÚLYOK KÖZÜL

##### IV.1.1.1. Az Axelrod játék

A II. fejezetben megmutattam, hogy a neoklasszikus közgazdaságtan paradigmájában is gyorsan megjelent egyes modellekben az önérdékkövetés és egyéb motivációk szétválasztása a hasznosságfüggvényen belül (Becker 1957). Az önérdékkövetésen túli motivációk ebben a korai modellben (mai értékrendünk szerint) negatív érzelmekhez kötődtek: Becker a diszkrimináció iránti preferenciákat vizsgálta. Ez termékeny kiegészítésnek bizonyult. A diszkrimináció közgazdaságtana fontos eredményeket produkált a rákövetkező évtizedekben (Arrow 1973, 1998, Cain 1986).

A hetvenes évektől azonban már a jóindulat modellezése került a kutatók figyelmének a középpontjába. Ekkoriban nem csupán elméleti közgazdász kutatók (Hochman és Rodgers 1969), de egyes biológusok is vizsgálták az altruizmus gyökereit és következményeit (Hamilton 1964, Trivers 1971). A témával foglalkozó kisebb körből is a diszkriminációkutatót megalapozó Gary Becker munkássága

\* Axelrod, R. (1984): *The evolution of cooperation*. New York: Basic Books.

érte el a legnagyobb hatást (Becker 1974, 1981, az elméletttörténeti áttekintéshez lásd Fontaine 2006).

Látni kell ugyanakkor a jóindulat kutatásának irányába tett fordulat korlátait. Becker modelljeit a család jólétére koncentrálnak viselkedés leírására szánta. Nem feltételezte, hogy érdemes az emberi lényekkel, vagy legalább a polgártársainkkal szembeni általános jóindulatot (áldozatkészséget) feltételezni. Ez a szűklátókörűség pedig magában hordozza a diszkriminációt is (figyelem: a diszkriminációs preferencia is magában hordozza a saját anyagi jólétről történő lemondást – például hozzánk közelebb álló, de rosszabb munkavállaló felvételével). A biológiai altruizmusmodellek esetében még erősebb alapfeltételezés volt a diszkriminatív jóindulat feltételezése, amely csak közeli kapcsolatokra terjed ki.

A fent említett modellek megszorításai ellenére az önérdékkövetés modelljének határait feszegető közgazdász diskurzusban gyakran az általános, idegenekkel szemben is érvényes jóindulat elméleti és empirikus alapjait tárgyalták (Phelps 1975). Elméletileg legtisztábban játékelméleti keretekben lehet látni, hogy nagytársadalmakban az altruizmus önmagában nem elegendő a sokszereplős társadalmi csapdahelyzetek megoldására (erről már volt szó). A nyolcvanas években többek között az adományozási tevékenység vizsgálatával mutatták ki empirikusan is az általános jóindulat feltételezésének terméketlenségét (Andreoni 1988). Andreoni altruizmus helyett csak azzal az ad hoc feltételezéssel tudta magyarázni az adományozók megfigyelt viselkedését, hogy önértéke van az adományozásnak („warm glow giving”).

Az előző fejezetben megmutattam, hogy bizonyos strukturális korlátok között az önző emberek is fenn tudják tartani a jóindulatú(nak érzékelhető) együttműködés normáját. A racionális együttműködés kutatási programja az ismételt fogolydilemma játszma elemzésére épült. Azonban az önzőség és önzetlenség dilemmájának e modellje egy másik irányba is elterelte a kutatók érdeklődését: az altruizmust vizsgáló biológusokhoz hasonlóan a közgazdászok, és a társtudományok képviselői is elkezdtek érdeklődni aziránt, hogy milyen viselkedési modulok épülhettek belénk az evolúciós erők hatására. A köz-

gazdaságtan, a pszichológia és a politikatudomány (majd a kulturális antropológia) egyes képviselőit viszont a biológusokkal szemben a rokonságon túlmutató kapcsolatokat szabályozó modulok érdekelték. A nagyobb társadalomtudományi diszciplínák közül csak a szociológusokat hagyta (kevés kivétellel) hidegen az egész problémakör. A szociológia, és az általam vizsgált kutatási program viszonyáról már beszéltem, és a könyv utolsó fejezeteiben újra beszélek róla.

Robert Axelrodot, a Michigani Egyetem politikatudósát is az emberi együttműködés belénk kódolt szabályai érdekelték a hetvenes évek második felében. Az előző fejezetben beszámoltam az észak-amerikai egyetemi körökben meghirdetett programozási versenyéről, ahol a kódok bizonytalan sokszor ismételt, kétszemélyes fogolydilemma „meccseket” játszottak egymással. A kódokat a versenyt megnyerni kívánó racionális emberek írták, azonban menet közben ezek az emberek már nem reagálhattak a fejleményekre, csak előre kódolt módon lehettek reaktívak is. Axelrod már ezzel is jelezte, hogy letér az előző fejezetben tárgyalt, tisztán racionális döntéseméleti kutatási ösvényről. A verseny eredményét követő analitikus elemzése pedig már nyíltan egy túlélési, azaz evolúciós modell.

Az Axelrod-féle (1984), mintegy hatvan program részvételével zajló verseny győztese a fogolydilemmával és más, ún. alapjátékokkal régóta foglalkozó kanadai pszichológus, Anatol Rapoport végtelenül egyszerű programja, a „szemet szemért” (tit-for-tat) névre elkeresztelt, pár soros kód lett. A tit-for-tat győzelméhez az kellett, hogy a versenyben saját magukkal is játszanak a programok, ami nem igazán tett jót a többé vagy kevésbé bonyolult aljaskodó versenyzőknek (programoknak). Ez Axelrod szemében semmit nem vont le a program értékéből, sőt tanulságos előnyként értékelte.

A „szemet szemért” kód szerint az ismételt fogolydilemma első fordulójában kooperálni kell, utána pedig minden fordulóban le kell másolni a partner előző fordulóban tett lépését. Rapoport e kóddal a lehető legegyszerűbben fogalmazta meg a játék keretein belül az emberi interakciók egyik alapvetőnek tűnő szabályát: „amilyen az adjonisten, olyan a fogadjisten”. Ez a morális kód láthatóan középutat képvisel a diszkrimináció negatív és az (általános) altruizmus pozitív

viszonyulása között. A feltételesen együttműködő ember tud jót is, és rosszat is tenni társával.

Axelrod (1984) a verseny értékelésében azt a – politikatudományban korábban Taylor (1976) által megfogalmazott, és később a közgazdászok által formalizált (lásd előző fejezet) –, gondolatot vitte tovább, hogy a közösségi érdek érvényesülésének alapja nem a feltétel nélküli jóindulat, hanem valamiféle feltételes együttműködés: a méltányos, tisztességes magatartás. Axelrod azzal a sejtéssel egészítette ki korábbi tézisét, hogy ez a fajta morál nem csupán racionális interakciók instant következménye, hanem évszázatzredek alatt zajló sikeres és sikertelen emberi együttműködések nyomán belénk kódolt viselkedési mintázat. Ebben az időszakban már egyes közgazdászok is megfogalmazták ezt a feltételezést; köztük a később Nobel-díjjal jutalmazott George Akerlof (1982) is (lásd még például Sugden 1984).

#### IV.1.1.2. Evolúciós érvek

Axelrod (1984) egy evolúciós játékelméleti modellel kívánta bizonyítani, hogy a „szemet szemért” stratégia speciális túlélési előnyökkel rendelkezik, és képes ellenállni másféle stratégiák evolúciós „támadásának”. Erről az eredeti bizonyításról hamar kiderült, hogy bizonyos technikai részleteiben hibás, és a tit-for-tat kiemelt szerepére vonatkozó állítás elhamarkodott (Binmore 1998). Azonban az eredmények – megerősítve az evolúciós biológia egyik legjelesebb képviselője által (lásd Hamilton és Axelrod 1981) – népszerűsége és egyszerűsítései komoly erőfeszítésekre sarkallták az evolúciós biológia és a játékelmélet számos kutatóját. Axelrod népszerű egyszerűsítései termékeny egyszerűsítések voltak, és egy máig tartó színvonalas kutatási programot indítottak el.

Az első jelentősebb pontosítások Martin Nowak és Karl Sigmund bécsi matematikai zoológusok nevéhez köthetőek, akik megmutatták, hogy a generációkon átnyúló, ismételt fogolydilemmákban többféle stratégia tud életképes maradni, ráadásul más stratégiák, például a tanító jellegű tit-for-tattal szemben a tanuló jellegű, „győztes csapa-



ton ne változtass, de a vesztesen igen” („win-stay loose-shift”) stratégiák adott esetben evolúciósan sikeresebbek tudnak lenni (Nowak és Sigmund 1993). Maga a „szemet szemért” elv is többféleképpen megfogalmazható még ebben az egyszerű játékban is. A tit-for-tat szélsőségesen türelmetlen, amely hiányos információs környezetben veszélyes lehet, ugyanakkor szélsőségesen gyorsan felejt, ami nem mindig elég elrettentő a rafináltabb stratégiák számára.

Az ismételt fogolydilemma evolúciós játékelméleti irodalma nagyon gazdag, és az értő tárgyalásához az evolúciós játékelmélet racionális stratégiai gondolkodástól részben eltérő döntési logikáját is meg kellene ismerni. Azonban ez az egyébként fontos kutatási program a mi történetünkben mellékszál, ezért nem vágok bele. A könyv utolsó két fejezetében röviden utalni fogok arra, hogy hol hasznosulnak az evolúciós modellek eredményei – és akkor már látható lesz, miért nem relevánsak a könyvben vizsgált fő kérdés szempontjából.

A fogolydilemma evolúciós játékelméleti kutatásának eredményeiből annyit mindenképpen érdemes leszűrni, hogy az Axelrod tézisének korrigáló későbbi modellek nem cáfolták azt a sejtést, hogy evolúciósan sikeresek lehetnek a mások irányában pozitív és negatív irányú viselkedésre egyaránt hajlandó, a másik szándékaira figyelő, feltételesen együttműködő stratégiák. A következő évtizedekben a neoklasszikus közgazdaságtan, valamint a játékelmélet által képviselt racionális döntéselméleti paradigmában az empirikus érvényesség és a pontos racionális döntéselméleti modell körül folyt komoly tudományos vita.

A magyar szakirodalomban leginkább a méltányosság kifejezést használjuk (Gulyás 2007, Janky 2005) annak a motivációs tényezőnek (preferenciának) a jelölésére, amelyet az imént ismertetett evolúciós sejtés nyomán kibontakozó angol nyelvű irodalomban a *fairness* terminussal jelölnek. A következőkben én is így fogok rá hivatkozni.

## IV.1.2. IRÁNY A LABOR!

### IV.1.2.1. A kezdetek

Az emberek méltányossággal és igazságossággal kapcsolatos elképzeléseivel – Axelrodékkal párhuzamosan, illetve őket megelőzve – sokat foglalkoztak a pszichológusok, a szociológusok és a filozófusok is (Deutsch 1975, Friedell 1968, Jasso 1980, Rawls 1971). Ennek a következő fejezetben lesz jelentősége. Azonban itt ezzel az irodalommal nem foglalkozunk. Azon a ponton kapcsolódunk be a tudományos diskurzusba, ahol a fenti evolúciós sejtés nyomán, a korábban a foglydilemmára és egyes piaci játékokra koncentráló laboratóriumi kísérletekben elkezdtek vizsgálni a méltányosság motivációját is.

Ezek a kísérletek a neoklasszikus közgazdaságtan kinyilvánított preferencia tézisére (Samuelson 1948) építenek, amikor a kutatók – szemben a szociálpszichológusokkal és szociológusokkal –, nem kérdezősködnek az igazságosságról, hanem valós pénzbeli téttel rendelkező döntési dilemmák elé állítják a kísérleti alanyokat. A feltételezés az, hogy mi hazudhatunk, de döntéseink nem hazudnak valós értékeinkről (visszatérünk nemsokára arra, hogy milyen preferenciákat mérnek a laboratóriumi játékokban).

Kevesen tudják, hogy a tisztességtelenséggel szembeni bosszú első, pénzes játszmákon nyugvó laboratóriumi vizsgálatait nem közgazdász, nem is pszichológus, hanem egy szociológus, az akkoriban a Michigani Egyetemen, később Santa Barbarában, a Kaliforniai Egyetemen tanító Morris Friedell (1968) végezte. Az általa kitalált játék két lépéses. Az első lépésben az egyik játékos dönt, hogy egyenlő, vagy saját maga számára kedvezően egyenlőtlen módon oszt ki magának és egy másik játékosnak pénzt. A második lépésben a másik játékos egyenlőtlen felosztás esetén döntés előtt áll: elfogadhatja a döntést, vagy megbosszulhatja azt. Az utóbbi esetben neki is kevesebb pénze lesz, de még inkább csökken az első játékos nyeresége. Friedell vizsgálata – bár egy rangos folyóiratban jelent meg –, később feledésbe merült.

A pszichológusok, a kérdezősködés mellett, már a kezdetektől fogva részt vettek a valós téttel bíró játszmák laboratóriumi vizsgálá-

latában (lásd például Rapoport és Dale 1966). A méltányosság témájában sem történt másképp. Kahneman és szerzőtársai (1986) vezették be például az igazságosság kérdéseire legjobban illeszkedő, legtisztábbnak tűnő döntési helyzet laboratóriumi vizsgálatát. Az ún. diktátorjátékban a kísérleti alany kap valamennyi valós pénzt (utalványt), és az a feladata, hogy ossza el saját maga és egy másik, számára ismeretlen kísérleti alany között. Az utóbbi személy nem befolyásolhatja a döntést, és utólag sem reagálhat rá. Ebben a játékban (ami tiszta formájában nem játék, azaz nem stratégiai döntés) az anyagi önérdéküket követő kísérleti alanyoknak semmilyen okuk sincs a kapott pénz megosztására. Ez a játék vált a méltányossági motivációt vizsgáló kísérletek egyik fő eszközévé.

A diktátorjátéknál még prominensebb szerepet játszott az ún. ultimátumjáték. A játékelmélet Nobel-díjas német kutatója, a százlábú játék paradoxonját fontosnak tartó Reinhard Selten javasolta svájci kollégájának és társszerzőjének, Werner Güthnek, hogy foglalkozzon kísérleteiben a visszafelé göngyölítés problémájával (Binmore és Shaked 2010), amelyről a III.2.1.2. alfejezetben hosszabban írtam. Güth és szerzőtársai (1982) egy olyan játékot kezdtek el játszani diákjaikkal, amely egyesíti a Friedell (1968), valamint Kahneman és szerzőtársai (1986) által vizsgált szituációkat. A játékban két, egymás számára ismeretlen kísérleti alany hoz egymás után egy-egy döntést. Az első játékos, a diktátorjátékhoz hasonlóan, kap egy fix összeget, és azt saját belátása szerint eloszthatja saját maga és a másik (ismeretlen) alany között. Névtelen ajánlatát a kísérletvezetők átküldik a másik kísérleti személynek, aki dönt az „ultimátum” elfogadásáról, vagy elutasításáról. Elfogadás esetén mindkét résztvevő megkapja az első játékos által megítélt összeget. Elutasítás esetén egyik játékos sem kap semmit. Az elutasítás végleges, a játékot a két összesorolt szereplő egyszer játssza egymással, és az ott szerzett információk (ajánlatok és válaszok reputációja) sem vihetők át a kísérletben esetleg még soron következő játékokra. A visszafelé göngyölítés logikája alapján az első játékosnak minimális összeget érdemes felajánlania a másiknak, mivel tudja, hogy a második játékos, ha pénzt akar, akkor kiszolgáltatott: a nagyon kevés is több a semminél (amivel a vissza-

utasítás jár), sőt, (ha a játékelmélet szigorú logikáját tekintjük), még a semmi sem kevesebb a semminél, tehát még a nullát sem *kell* feltétlenül elutasítania.

A méltányosság kísérleti empirikus kutatásai is kapcsolódnak Akerlof (1982) munkahelyi kapcsolatokkal foglalkozó, sokat idézett munkáihoz. A nyolcvanas-kilencvenes évek fordulóján pedig egyes európai munkagazdászok voltak a témában folytatott kísérletek úttörői – közülük Ernst Fehr, a később Svájcba letelepült osztrák közgazdász vált a legbefolyásosabb tudóssá. Ugyanis a munkahelyi viszonyokban válik látványossá, hogy nagy gazdasági tétek esetén is fontos a jogi eszközökkel ki nem kényszeríthető becsület: nagyon sok gazdasági értéket állítanak elő nagyon nehezen (költéségesen) monitorozható munkakörökben, ahol a tapasztalat szerint a költséges monitorozást nem is végzik el a cégek.

#### IV.1.2.2. A fénykor

Az ezredforduló utáni évekre az idegenek közötti együttműködést és szolidaritást segítő emberi motivációk kutatása már rendkívül nagy, és egyre növekvő népszerűségnek örvendett. Érdekes módon az önérdekkövető egyén legsikeresebb modelljét kidolgozó közgazdaságtan művelői körében növekedett látványosan e kutatási téma iránti érdeklődés. A társas preferenciák (ahogy e közgazdászok óvatosabban nevezték az őket érdeklő motivációs tényezőket) kutatása elsősorban Európában, és nem a legmagasabbra rangsorolt kutatási központokban vált különösen népszerűvé (lásd például Kolm és Ythier 2006). Ugyanakkor a szakma elitjébe sorolt kutatók között is voltak követői a programnak. Például Matthew Rabin, a Berkeley egyetem professzora, aki a legsikeresebb fiatal amerikai közgazdász kutatónak járó, és gyakran „elő-Nobel-díjként” emlegetett John Bates Clark-érem 2001. évi kitüntetésétje volt. Ernst Fehrt, a téma legnagyobb hatású kutatóját pedig a világ legnagyobb presztízsű egyetemei közé sorolt Massachusettsi Műszaki Egyetem (MIT) is meghívta vendégoktatónak.

E kutatási terület sikereit nem lehet eléggé túlbecsülni. Fehrt a német nyelvterület, és vele Európa egyik legbefolyásosabb közgazdászának tartják jelenleg (Bernau 2016). De például Rabin egyik legfontosabb kollégája, Kőszegi Botond is pár éve kapta meg a John Bates Clark-érem európai verzióját. Az erkölcsös polgári viselkedés kutatásának áttekintéséhez ezért az az értelmezés a kiindulópontom, amely alapján a terület alapító-vezetője, Fehr és legközelebbi munkatársai bemutatják a kutatási program kialakulását.

Fehr a kétezres évek közepén, már az MIT oktatójaként foglalta össze néhány kollégájával, hogy miként látja a közgazdaságtan viszonyát a morális viselkedés modellezéséhez (Henrich et al. 2004, Gintis et al. 2005). Fehr és kollégái szerint a modern közgazdaságtant megalapozó klasszikus szerzők tisztában voltak az ember kettős természetével: részben önző, racionális, részben viszont morális lények vagyunk. A modern közgazdaságtan alapítójának tekintett Adam Smith a mai közgazdaságtudomány fontos alaptéziseit lefektető klasszikus műve, *A nemzetek gazdagsága* (Smith 1992 [1776]) előtt már kötetet írt *The Theory of Moral Sentiments* címmel (Smith 2010 [1759]). Fehr és munkatársai e két kötetet – amelyek egyrészt az önző, másrészt a társas preferenciák következményeiről szólnak –, együtt tekintik a modern, axiomatikus viselkedéstudomány megalapozásának. Értelmezésük szerint azonban a modern közgazdaságtudomány elhagyta az ember kettős természetére vonatkozó klasszikus feltevést. Úgy látják, hogy a második világháború után uralkodóvá váló neoklasszikus közgazdaságtan kizárólag az ember önző, racionális énjének modellezésére törekedett, és kizárólag e motiváció alapján próbálta magyarázni a gazdasági viselkedést. Fehr meggyőződése, hogy – az ezredfordulón lendületet szerző – társas preferenciák kutatási programja visszatért a klasszikusokhoz, így korrigálva a 20. századi neoklasszikus közgazdaságtan emberi motivációkra vonatkozó alaptéziseinek hiányosságait.

Felhívnám a figyelmet arra, hogy korábban (II. fejezet) amellet érveltem, hogy a közgazdaságtan más úton jutott el a társas preferenciák kutatási programjának elindulásához, azonban a fenti interpretáció ismerete segít a tisztességes polgár modelljének megértésében.

## IV.2. Makacs mítoszok és makacs tények

### IV.2.1. AZ EGALITARIANIZMUS MÍTOSZA

#### IV.2.1.1. Az egyenlőtlenségaverzió modellje

Ahhoz, hogy a neoklasszikus közgazdaságtan kutatói közössége befogadja a méltányosságra vonatkozó hipotézist, szükség volt – a verbális narratíván, és néhány kísérleti eredményen túl – a hasznosságmaximalizáló modellbe illeszthető formális leírására is. Elsőként Gary Bolton (1991) dolgozott ilyen modellen, de később a téma kutatói Bolton és Ockenfels (2000), vagy még inkább Fehr és Schmidt (1999) modelljét alkalmazták. Közel két évtizednyi előkészítés után a két utóbbi cikk megjelenése indította útjára az ún. társas preferenciák kutatási programját a közgazdaságtanban.

Fehr és Schmidt (1999) arra törekedtek, hogy minél egyszerűbb, a tényeket még kielégítő módon magyarázó modellt alkossanak a társas laboratóriumi kísérletek széles körére vonatkozóan (bővebb magyar nyelvű összefoglaláshoz lásd Gulyás 2007, Janky 2005). Kiindulópontjuk a neoklasszikus közgazdaságtan következménycentrikus cselekvésmo­dellje (Smith 2015) volt. Feltételezéseik szerint a cselekvők preferenciáit a méltánytalansággal szembeni egocentrikus ellenérzés jellemzi (self-centered inequity aversion). A következménycentrikus modellben a méltányosság mértékét az interakciók következményeinek felek közötti elosztása határozza meg: ne legyen egyiküknek se sokkal jobb, mint a másoknak. Az anonimitást biztosító laboratóriumi kísérletek letisztult világában csak az alanyoknak juttatott pénz tekintetében manipulálják a kísérletvezetők a kimeneteleket, semmi más nem nyerhetnek/veszíthetnek a résztvevők. Erre alapozva lehetséges egy olyan modell kidolgozása, amely a méltányosság iránti igényt a felek közötti pénzelosztásra vonatkozó preferenciákkal azonosítja.

Ezt a motivációs elemet azért tekintik egocentrikusnak, mert nem általában az elosztás igazságtalanságaival szembeni ellenérzés jelenti, hanem az egyén és mások helyzetének viszonyára vonatko-

zó preferenciákat érthetünk rajta. Az egocentrizmus révén a Fehr és Schmidt-féle meghatározás kapcsolatba hozható a cselekvés alapú méltányossági elvekkel (Fehr és Schmidt 1999). Ez a modell anynyiban is kötődik a korábbi eredményekhez, hogy nem kíván állást foglalni abban, hogy egy-egy konkrét társadalmi helyzetben mi tekinthető méltányos elosztásnak, és mi nem. Mindazonáltal a laboratóriumi szituációkra vonatkozóan – ahol modelljüket tesztelték –, felteszik, hogy a kifizetések egyenlő elosztása a leginkább méltányos megoldás a szereplők számára.

A motivációs modellben, amely magában foglalja az önérdék követés feltételezését is, három alapvető tényező határozza meg az egyén által realizált hasznosságot: 1. A számára rendelkezésre álló anyagi kifizetés haszna. 2. A számára kedvezőtlen elosztással kapcsolatos ellenérzés. 3. A számára kedvező (ám így az egyenlőtlenégtől eltérő) elosztási eredménnyel szembeni averzió. Feltevésük szerint az emberek méltánytalanabbnak érzik a számukra kedvezőtlen egyenlőtlenégeket, mint azokat, amelyeknek ők az anyagi haszonélvezői. Fehr és Schmidt (1999) saját beszámolója szerint a korábbi kísérletek eredményei nagyjából megerősítik feltevéseiket.

#### IV.2.1.2. A modell tesztelése

Nehéz túlbecsülni a 21. század eleji közgazdaságtanra gyakorolt hatását azoknak a kutatásoknak, amelyeket Fehr és kutatócsoportja végzett a méltányosság témájában. Az itt ismertetett cikkét a Google Tudós nyilvános publikációarchívuma szerint több mint tízezer tanulmány idézte 2018 őszéig. Nagyjából kétszer annyian, mint a 2018-ban Nobel-díjjal kitüntetett William Nordhaus legnépszerűbb cikkét (amely a hetvenes évektől gyűjti a hivatkozásokat az üzleti ciklusok témájában, amely központi jelentőségű a makroökonomia számára). Becker (1974) altruizmusmodelljét is csak ötezren hivatkozták eddig; a megelőző fejezetünk kulcsát jelentő Kandori-féle (1992) racionális norma modellt pedig alig kétezren. Az általuk javasolt egyenlőtlenésgaverzió modell iránti érdeklődés

pedig – a hivatkozások számának trendjeiből ítélve –, jelenleg sem csökken.

A méltányossági motiváció empirikus kutatásának jóval nagyobb hatása volt a diszciplína egészére, mint korábban az altruizmusmodell bevezetésének. A társas preferenciák kutatásának térnyerésével párhuzamosan, és részben e kutatási programnak köszönhetően nőtt meg a kísérleti vizsgálatok mennyisége, presztízse és jelentősége a 21. századi diszciplínában. A legnagyobb presztízssű folyóiratok cikkeiben a neoklasszikus közgazdasági elmélet matematikai és analitikus filozófiai gyökerű érvelési logikája mellé felzárkóztak a kísérletek köré szerveződő, pszichologizáló-biológizáló magyarázatok. A társas preferenciák kutatási program vezető tudósainak eredményei a korábbi divatos területek képviselőinél gyakrabban jelentek meg a vezető természettudományos lapokban (például Bernhard et al. 2006; Camerer és Fehr 2006; Fehr és Fischbacher 2003; Fehr és Gächter 2002, Fehr et al. 2008; Knoch et al. 2006).

Érdeemes rámutatni arra, hogy a 21. században, Axelrod szimplifikáló modellje után a méltányosság kutatása ismét egy túlzottan leegyszerűsítő modell nyomán kapott új erőre. Úgy tűnik, erős intuitív ereje van a méltányosság naturalizálásának, nevezetesen annak a hipotézisnek, hogy az ismeretlen emberek közötti interakciók méltányos rendezésének elvei a történelem kezdete óta, univerzálisan velünk élnek – és biológiai eredetük miatt örökké velünk is maradnak.

A nyolcvanas évek első kísérletei után számos helyen, sokféle formában rekonstruálták azokat a kezdeti eredményeket, miszerint a diktátorjátékban az emberek többsége nem tartja meg a teljes kézhez kapott összeget, hanem ennek nem triviális hányadát (de többnyire kevesebb mint a felét) megosztja a tétlen (és ismeretlen) társával. Az ultimátumjátékban pedig az elosztási javaslatot tevők többsége kifejezetten gáláns ajánlatot tesz, miközben az ultimátumot kapó fél néha inkább lemond a felajánlott pénzről, ha az a teljes felajánlható összegnek csak kisebb része (Camerer 2003).

Az elmúlt két évtizedben nagyon sokan használták a laboratóriumi játszmák, a terepkísérletek, a kérdőíves felmérések és a piaci fo-



lyamatok értelmezésére az egyenlőtlenségaverzió modelljét, ugyanakkor nagyon hamar megjelentek azok az eredmények is, amelyek a modell korlátaira mutattak rá. A keretezési kísérletek (Cookson 2000; Meier 2006; lásd még Andreoni 1995) még nem feltétlenül cáfolják a modell relevanciáját. A racionális döntéelmélet alapvető, és termékenységét sok évtizeden át bizonyító modelljeit is meg lehet „viccelni” a laboratóriumi döntési dilemmák ügyes prezentálásával (Ellsberg 1961, Tversky és Kahneman 1974, 1981). Ráadásul az elosztási alapú méltányosságmodell támogatói hangsúlyozták, hogy a praktikus okokból egyszerűsített egyenlőtlenségaverzió modell mögötti elmélet összetettebb értékeléseket feltételez a méltányos döntéshozóról. Például Konow (2003) a méltányosságelmélet kiváló áttekintésében visszacsempészi a szándékok figyelembevételét és a viszonyosság elvét.

Nem cáfolja még az egyenlőtlenségaverzió létezését Charness és Rabin (2002) elemzése sem, azt azonban kétségbe vonják az eredményeik, hogy a méltányosság (mint egyenlőtlenségaverzió) az egyetlen jelentős társas preferencia. Bemutatnak egy olyan típusú diktátorjátékot, amelyben a diktátorok méltányossági preferenciái nem saját önzőségükkel viaskodnak, hanem a hatékonyság elvével: az egyik lehetőség az, hogy diktátor önmagának és a másiknak egyaránt 400-400 utalványt oszt ki, míg a másik opció az, hogy önmagának 400-at, míg a másiknak 750-et ad (tehát a diktátor mindenképpen 400-at kap, az ismeretlen másik viszont jobban jár, ha a végeredmény egyenlőtlen). Ebben a játékban sokan a másiknak kedvező, kollektív szempontból hatékony (legtöbb össznereményt nyújtó) megoldást választják a méltányosnak tekintett, egyenlő elosztással szemben. Nem sokkal később Engelmann és Strobel (2004) háromfős diktátorjátékokban szisztematikusan megvizsgálták több elosztási elv relatív súlyát, és arra jutottak, hogy bár a kísérleti alanyaik nem csupán saját pénznereményüket maximalizálják, és a méltányosság is szerepet játszik a döntéseikben, de vannak a Fehr és Schmidt-féle (1999) egyenlőtlenségaverziónál fontosabb elvek is számukra – például a Charness és Rabin (2002) által is bemutatott kollektív hatékonyság.

A könyv utolsó fejezetében még lesz szó a fent tárgyalt cselekvés-elméleti viták lehetséges konklúzióinak eltérő társadalmpolitikai üzeneteiről. Most csak felhívnam a figyelmet az egyik kisebb, politikai téttel rendelkező vitára: a 20. század második felének politikai gazdaságtani gondolkodásában nagy szerepet kapott a társadalmi egyenlőtlenségek csökkentésének, és a gazdasági növekedés támogatásának inkongruenciája. A fenti elméleti vita tétje, hogy a tanácsadók és gazdaságpolitikai megmondóemberek hivatkozhatnak-e (a tudomány tekintélyére támaszkodva) arra, hogy „az emberek belső igényei” miatt elfogadható az egyenlőség (növekedés) feladása a növekedés (egyenlőség) érdekében.

Az egyenlőtlenségaverzió kiemelt szerepét cáfoló eredmények sok kérdést vetnek fel a modell mögött húzódó evolúciós érvelés érvényességével kapcsolatban. Ugyanis az idézett (és a hozzájuk hasonló) kísérletek eredményei sokkal inkább szétfeszítik az emberi együttműködés méltányossági preferenciákra építő modelljét, mint Axelrod elméletének pontosításai. Azonban az eredeti elméleti alapok aláásásának ellenére ezek a tapasztalatok továbbra is a társas preferenciák tágabb hipotézisét gazdagítják – és nem cáfolják az egyenlőtlenséget csökkenteni akaró szándékok létét feltételező modellt sem.

Azonban nem kellett sokat várni azokra a kísérleti eredményekre sem, amelyek megkérdőjelezzik, hogy a diktátorjátékban az egyenlőtlenségaverzió mozgatja döntéseinket az egyenlő elosztás irányába. Dana és szerzőtársai (2006) a méltányosságpreferencia hipotézissel szembeni szkepszisük igazolására egy olyan diktátorjáték kísérletet terveztek, ahol az eredeti diktátorjáték imént bemutatott szabályainak ismertetése után megadták a lehetőséget a diktátor pozíciójában lévő kísérleti alanyoknak, hogy a kezében lévő pénz egy kis részének feláldozásáért cserébe „kilépjen” a játékból. A kilépés azt jelenti, hogy a másik kísérleti alany nem kap semmit, és nem is értesül róla, hogy egy ismeretlen egy diktátorjátékban döntött úgy, hogy nem kap semmit. Ez a dizájn tekinthető valamiféle ekvivalencia keretezésnek, amelyben a kísérletvezetők más köntösbe bújtatják a gazdasági tartalmukat tekintve azonos kimeneteleket. Ugyanakkor mégiscsak több egyszerű megfogalmazási ravaszkodásnál. A szabályok ismeretében a kísérleti

alanyok számára egyértelmű és világos, hogy senkinek sem javítják a helyzetét azzal, ha kilépnek – a sajátjukat viszont rontják a kilépési költség megfizetésével. Sokan mégis a költséges kilépést választották ebben a helyzetben. Anyagi értelemben annyi történt, hogy ahelyett, hogy megtartották volna a teljes összeget, leadtak egy kisebb részt a kísérletvezetőknek – elosztási szempontból teljesen szükségtelenül. Tehát úgy csökkentették anyagi hasznukat, hogy kihasználatlanul hagyták az ezzel járó egyenlőtenségcsökkentési lehetőséget. Dana és szerzőtársai (2007) egy újabb kísérletsorozatukban tették még kétségesebbé a korábbi, diktátorjátékkal kapcsolatos eredményeket. Itt a legérdekesebb példa az volt, amikor a diktátornak az 5-5-ös, illetve a 6-1-es felosztás (magának hat, a másiknak egy) közül kellett választania. A méltányosságelmélettel összhangban az alanyok többsége nem kérte a kicsit nagyobb összeget, inkább az egyenlő felosztás preferálta. A játék módosított verziójában azonban csak azt árulták el a diktátoroknak, hogy a döntésük nyomán mennyi lesz a saját maguk kifizetése (5, illetve 6 egység). Ugyanakkor azt mondták nekik, hogy azt egy gép sorsolja ki, hogy melyik opcióhoz kerül a másik játékosnál az 1-es, illetve 5-ös kifizetés. Ha kérik, a gép megmutatja nekik a sorsolás eredményét, azaz láthatják, hogy az 5-5 és 6-1, vagy az 5-1 és 6-5 variáció között döntenek. Ha nem kérik, akkor nem tudják meg, hogy mit kap a másik. Elég sok játékos volt, aki (a korábbi kísérleti helyzethez hasonlóan) „kilépett” a játékból, és nem volt kíváncsi arra, hogy mit dobott a gép. Inkább egyszerűen meghozta az egyszerű döntést a 6 és az 5 között (a 6-ot preferálták).

A fenti, és a hozzájuk hasonló (lásd még például Bardsley 2008, List 2007) ravasz kísérletek nem azt mutatták meg, hogy – szemben a korai eredményekkel – a közgazdasági kísérletek alanyai a mesterséges szituációkban a kutatók által definiált anyagi önérdékeiket veszik figyelembe. Sokkal inkább arra utalnak, hogy az embereket anyagi önérdékeiktől nem egy egyenlőségpreferáló gén téríti el, hanem valami más.

Ezek után a téma vezető kutatói, beleértve Fehr kutatócsoportját is, sokkal óvatosabbak lettek az egyenlőtenségaverzió modelljével kapcsolatban, miközben az alkalmazott közgazdasági kutatásokban

– valamint olyan kísérletekben, ahol például magyarázó változóként szerepel –, feltevésként továbbra is sokszor alkalmazták és alkalmazzák az egyenlőtlenségaverziót, mint a társas preferenciák kiemelt típusát. Kétségtelen ugyanakkor, hogy e cáfolatok jelentős hatása is kimutatható a tágabb kutatói közösségre. Igaz ez Binmore és Shaked (2010) polemikus tanulmányára is, amelyben nem újabb bizonyítékokat hoznak az egyenlőtlenségaverzió modelljével szemben, hanem Fehr és Schmidt (1999) eredeti prezentációjának hiányosságaira irányítják rá a figyelmet.

#### IV.2.2. ERŐS RECIPROCITÁS

##### IV.2.2.1. Az ultimátumjáték

Látszólag ellentmondásos, hogy az egyenlőtlenségaverzió eredeti modelljének látványos cáfolatai ellenére továbbra is sokan hisznek abban, hogy az emberek méltányosság alapú társas preferenciákkal születnek. Ugyanakkor a fenti cáfolatok nem a méltányosságelméletet cáfolják, hanem egy konkrét modellt, amely a lehető legegyszerűbb módon illesztette a neoklasszikus közgazdaságtan döntési modelljéhez a méltányosságelmélet feltevéseit. A torták méltányosság alapú elosztására vonatkozó elméletek is gazdagabbak annál (Konow 2003), mint amit az egyenlőtlenségaverzió eredeti modellje kezelni tud. Emlékezzünk az ismételt fogolydilemma paradigmára épülő racionális döntéseméleti modellekre, és a belőlük kinövő, Axelrod által népszerűvé tett evolúciós hipotézisre: az emberek a másik jó- és rosszindulatára reagálnak („amilyen az adjonisten, olyan a fogadjisten”). Figyelemre méltó, hogy a cáfolatok a diktátorjátékot célozták. Ez érthető, hiszen az egyenlőtlenségaverziót ebben a játékban lehet a legtisztábban tesztelni. A szándékokra való reagálást magában foglaló tágabb méltányosságelméleti keretek vizsgálatára alkalmasabb az imént ismertetett ultimátumjáték. Ez valódi játszma, és a méltányosságelmélet érvényességének vizsgálatában valójában ez a társas döntési dilemma játszotta a főszerepet.

Ahogy említettem, a játék empirikus vizsgálata a korábban elemzett visszagöngyöltési paradoxon kapcsán merült fel a játékelmélet egyensúllyal és racionalitással foglalkozó kutatóinak körében (Binmore és Shaked 2010). Azt is említettem, hogy itt a visszagöngyöltés alkalmazása sokkal intuitívabb, mint a kifejezetten provokatív százlábú játékokban: mivel az ultimátumot kapó második játékosnak érdeke elfogadni a legkisebb alamizsnát is, az ultimátumot adó játékosnak érdeke a legkisebb alamizsnát felajánlania. Ezt diktálja az önérdek, és minden más döntés az önérdektől való eltérésre való hajlandóságra utal (legalább az egyik félnél). Ebből az intuitív állításból indultak ki mindazok, akik az emberben rejlő méltányossági motivációk nyomát keresték e játszma empirikus vizsgálatának segítségével.

Több mint három évtized tapasztalatai alapján az ultimátumjáték laboratóriumi eredményei robusztus módon térnek el a visszafelé göngyöltéssel reprodukálható, intuitív pénzhaszon-maximalizáló stratégiakombinációktól. Egyrészt az ultimátumot adó játékosok többsége jelentős összegről mond le a másik javára. Másrészt az ultimátumról döntő, második játékosok gyakran utasítják el az önző megoldáshoz közeli ajánlatokat (Güth és Kocher 2014). Ezt már az első, nyugat-európai egyetemi hallgatókkal végzett kísérletek megmutatták (Güth et al. 1982, vö. Friedell 1968). A diktátorjáték könnyű manipulálhatósága ugyanakkor arra figyelmeztethet, hogy ezekben az első felmérésekben esetleg a tisztesség speciális kulturális normáját mérték a kutatók. Azonban a nyugat-európai eredményeket megismételték azok a tesztek is, amelyek az északi félteke legkülönbözőbb kultúrájú országában vizsgálták az egyetemi hallgatók reakcióját e játék kihívásaira. Oosterbeek és szerzőtársai (2004) metaelemzése szerint az első játékosok által adott ajánlatokban nincsenek érdemi eltérések a különböző országok diákjai között, de az elutasítási ráta például magasabb az ázsiaiak körében mint az európaiak között.

A diákminták alkalmazását sokan a szociálpszichológia és a kísérleti közgazdaságtan legfontosabb problémái közé sorolják. Ezért fontos, hogy a felnőtt népesség körében végzett felmérések sem vezettek érdemben eltérő eredményre, mint a diákok vizsgálata (Exadaktulos

et al. 2013). Terepkísérletek is megerősítik a laboratóriumi eredményeket (Meier 2006).

A viselkedési közgazdaságtan pénzes játékaiban a hallgatók egyikét mozijegy, vagy néhány ebéd árát nyerhetik meg (és alapesetben a felnőtt kísérleti alanyok sem tudnak nagy tétben játszani). Ez is felvetheti az eredmények külső érvényességének kérdését: valóban olyan fontosak az önérdeken kívül álló motivációk a valós élet döntéseiben is, mint a laboratóriumi játékokban hozott döntésekben? Az ultimátumjátékkal végzett kísérletek tapasztalatai ezekre a kétségekre is válaszolnak. Ugyanis néhány módosabb országban dolgozó kutatónak módja nyílt arra is, hogy relatíve szegényebb országok polgárai körében, nagyobb tétekkel végezze el a teszteket (Andersen et al. 2011, Cameron 1999). A nagy tétek esetében érezhető a változás, de egyáltalán nem dominál az anyagi motiváció. Például az ultimátumot fogadó játékos, a méltányosságelmélettel összhangban, a tétek növekedésével az össznyeremény egyre kisebb arányát hajlandó elfogadni, ugyanakkor abszolút értékben egyre nagyobb összegről hajlandóak lemondani, ha azt arányaiban kicsinek tartják (Eckel és Gintis 2010).

#### IV.2.2.2. Univerzális preferenciák?

Az ultimátumjátékban vizsgált alkuviselkedés univerzalitásával kapcsolatban a legnagyobb hatású vizsgálat kezdeményezése Joseph Henrich amerikai repülőgépmérnök nevéhez fűződik. Amikor 1993-ban Ernst Fehr, a társas preferenciák kísérleti kutatási programjának későbbi vezetője publikálta első jelentősebb hatású cikkét az ultimátumjátékról, és annak méltányossági alapú értelmezéséről (Fehr et al. 1993), Henrich még Springfieldben, a General Electric alkalmazásában, légi járművek és földi berendezések számítógépes kontrollfolyamatait ellenőrizte. Aztán hamarosan úgy döntött, hogy mesterfokozatot, majd doktori címet szerez kulturális antropológiából, a korábbi hobbiszakirányából. A mesterképzésen aztán a matematikai kultúraelmélet első számú kutatóinak, Robert Boydnak és Peter Ri-

chersonnak a hatására (lásd például Boyd és Richerson 1988) az emberi döntések kontrolljának tesztjei kezdtek érdekelni.

Henrich úgy döntött, hogy a Boydék kurzusán megismert ultimátumjátékot kipróbálja a nyári terepmunkáján, a perui Amazonas vidékén élő machiguenga indiánok körében, majd összevetette eredményeit a kaliforniai fiatalok alkuviselkedésével (Henrich 2000). Első tapasztalatai alapján kulturális antropológusként a két terep közötti jelentős különbségeket hangsúlyozta (ezekre még visszatérek), azonban a Fehr mellett eltöltött európai ösztöndíja, és egy amerikai magánalapítvány támogatásának elnyerése után egy, a társas preferenciák univerzalitásának bizonyítását célzó, nagyszabású és nagy hatású viselkedési közgazdaságtani és kulturális antropológiai projekt koordinátora lett. A projektben a téma néhány vezető elméleti kutatója maga köré gyűjtött közel húsz antropológust, akik jártasak voltak több, a világ legkülönbélebb tájain élő, és a modernizáció által mérsékelten érintett népcsoport kultúrájában. Az elméleti oktatás után a terepkutatók meglátogatták alanyaikat többek között a Góbi sivatagban, az afrikai szavannákon és az Amazonas mentén, összesen tizenöt helyen, és kb. egynapi jövedelemnek megfelelő összegért játszották a helyiekkel az ultimátumjátékot (földre helyezett korongokkal, és hasonló számítógépmentes eljárásokkal).

Az eredmények megerősítették a vezető kutatók várakozásait abban a tekintetben, hogy a visszafelé göngyölítéssel nyerhető megoldás egyik kultúrában sem jellemző kimenetele ennek az egyszerű alkunak (Henrich et al. 2001, 2005). Az átlagos ajánlatok helyszínek közötti szórása még nagyobb, mint a különböző fejlett országok diákjait összehasonlító vizsgálatokban; ugyanakkor mindenhol szignifikánsan különbözött nullától az az ajánlati érték, amelynek környékén a második játékosok nem jelentéktelen hányada saját anyagi érdekeivel szembeszállva elutasította az alkut.

A nagy ajánlati különbségek és a visszafelé göngyölítés alternatívái még fontossá válnak a kutatási programban, főleg, hogy egybecsengenek az elosztási preferenciák diktátorjátékokban tapasztalt könnyű manipulálhatóságával. Ugyanakkor világossá vált, hogy bárhol keressük is a vizsgálati alanyainkat a Földön, mindenhol lesznek

olyanok, akik inkább lemondanak a felajánlott pénzről, és ezzel a másiktól is elveszik a pénzt, ha nem tetszik nekik az ajánlat. És ez az elutasító magatartás nem random módon jelenik meg, hanem minden egyes terepen erősen függ az ajánlat nagyságától.

#### IV.2.2.3. Az altruista büntetés

Az ultimátumjátékban a második játékos elutasító magatartása kongruens az egyenlőtlenségaverzió modelljével (az ultimátumot elutasító fél 0/0 egyenlő elosztást preferál mondjuk egy 8/2-vel szemben). Ugyanakkor éppen Fehr kutatócsoportja mutatta ki (fogolydilemma játékokban), hogy a kísérleti alanyok olyankor is hajlandók, akár költségek árán is büntetni a potyautasokat, amikor a büntetéssel nem tudják csökkenteni a felek közötti egyenlőtlenségeket (Falk et al. 2005). Az pedig már korábban is köztudott volt, hogy az elutasítás nem magára a végeredményre, hanem kifejezetten az ajánlattevő tisztességtelen szándékára koncentrál. Ugyanis ha a második játékos úgy informálják, hogy egy gép dobta neki az „ajánlatot”, akkor sokkal kisebb arányban méltatlankodnak egy-egy alacsony ajánlat miatt (Blunt 1995). Idővel azok az eredmények is megjelentek, amelyek szerint egyéni szinten a pozitív gesztusok viszonzására való hajlandóság nem kapcsolódik a büntetési hajlandósághoz (Yamagishi et al. 2012). Ezek szerint nem beszélhetünk általános, „amilyen az adjonisten...” preferenciákról.

A fenti kutatásokkal párhuzamosan önmagában is vizsgálták a büntetési hajlandóságot. Bizonyos kísérletekben a korábban tárgyalt fogolydilemma játékot úgy módosították, hogy az egyszeri lejátéshoz egy plusz döntés, a büntető számára ingyenes, vagy költséges büntetési lehetőség is járt. A laboratóriumi fogolydilemma játékok ilyen típusú módosításának azért is volt jelentősége, mert a hagyományos lejátéshoz az volt a jellemző, hogy a játékosok tapasztalatának növekedésével elég gyorsan közelített a zéróhoz (az önérdékkövető stratégiához) az együttműködési hajlandóság (a kísérletekben ugyanis az a jellemző, hogy az ismeretlen partnerek változtatásával



ugyan, de többször is szembesítik ugyanazzal a dilemmával az alanyokat – így tudják a tanulási folyamatot kiszűrni és a döntések robusztusságát tesztelni).

Fehr és munkatársai ezt a csökkenést is meg tudták ugyan magyarázni méltányossági motivációkra hivatkozva, de a racionális tanulási hipotézis épp ugyanilyen trendet jósol. A büntetés lehetőségének bevezetése (amely akár – nem túl nagy – költséggel is járhat) meg tudta állítani az együttműködés hanyatlását, és a kezdeti (jellemzően 50%-os) szint fennmaradását eredményezte a későbbi fordulókban (Fehr és Gächter 2000).

Ha azt feltételezzük, hogy az emberi együttműködés alapja nem a jószándék, és még csak nem is a középutas tisztesség önmagában, hanem a méltánytalanságra adott ösztönös büntető reakció, akkor érthetőbbé válnak a diktátorjátékkal végzett kísérletek eredményeinek bizonytalanságai. Az ultimátumjáték jellemzően méltányos ajánlatai is jól levezethetők az altruista büntetéstől való félelemmel (amitől a diktátorjáték alanyainak nem kell tartaniuk).

A viselkedési közgazdászok célja ekkor már a következményeiben jól beazonosított altruista büntetés ösztönének fiziológiai beazonosítása volt. A cél közelében érezték magukat: úgy vélték, hogy több évtized elméleti spekulációkon, és közvetett empirikus bizonyítékokon nyugvó nyomkeresése után végre meglelték az erkölcsös viselkedés emberi természetben rejlő gyökereit. A bizonyítás érdekében Fehr kutatócsoportja pszichológusokkal fogott össze. Alanyaik pozitronemissziós tomográfiával a fejükön játszottak egy büntetési lehetőséggel bővített együttműködési játékot (de Quervain et al. 2004). Azt találták, hogy amikor a büntetésnek mindkét fél részéről tétje volt, akkor az altruista büntetés mellett döntő fél agyi jutalomközpontja működésbe lépett – a pénzügyi veszteség ellenére is. A *Science* nevű, rangos természettudományos folyóiratban közölt elemzésben Fehrék megmutatták, hogy sok nyelvben miért „édes a bosszú” (de Quervain et al. 2004: 1254). Az idézett nagy hatású kutatást számos további, agyi aktivitásra koncentráló kísérlet követte. Ezek azt az ígéretet hordozták magukban, hogy fizikailag is beazonosíthatóvá (mintegy láthatóvá) válnak az emberi együttműködés agyi mechanizmusai.

A kognitív tudományi előrehaladás részben a racionális közgazdasági modellezés korlátait ellensúlyozta. Ugyanis a neoklasszikus közgazdaságtan hagyományos modellezési keretében (amelyben az altruizmus elmélete és később az egyenlőtlenségaverzió elmélete is megszületett) alapvetően a célszemélyt érintő döntési következményeket lehet beépíteni, és nehéz kezelni más emberek szándékait, illetve a nem következménycentrikus reakciókat (Smith 2015). Tehát a kutatók a társas preferenciák kutatásának fellendülésekor szembesültek azzal a problémával, hogy ami a hagyományos keretekben könnyen modellezhető, azt nem igazolják az empirikus eredmények; amit viszont ki tudnak mutatni, azt nehezen tudják beépíteni a szokásos hasznosságmaximalizáló modellekbe. Lássuk azt is, hogy az altruista büntetést nemcsak könnyebb empirikusan kimutatni, de elméletileg is közelebb áll az együttműködési motivációk evolúciójának Axelrod-féle hipotéziséhez.

A modellezés nehézsége nem lehetetlenséget jelent. Komplikáltabb játékelméleti érveléssel kezelhető a „jó- és rossz cselekedetekre” reagáló magatartás. Rabin (1993) már nagyon korán megmutatta egy egyszerű, két játékost és két-két stratégiát tartalmazó szituációban, hogy miképpen kereshető egyensúly egy egymás szándékaira érzékeny szereplőket feltételező, ún. pszichológiai játékban. Dufwenberg és Kirchsteiger (2004), továbbá Falk és Fischbacher (2006) az egymásra történő reagálás időbeliségét modellező játékformákra (például a százlábú játék) alkalmazhatóan dolgoztak ki a méltányosságot feltételező és elváró egyensúlykonceptiót. Ezek nem a tankönyvek elejére kívánczó konceptiók, de bizonyították, hogy a racionális döntésemélet keretein belül elvileg kezelhető a szándékokra reagáló altruista büntetés viselkedésmintája.

Ezen a ponton érdemes egy pillanatra visszatekinteni az ívre, amelyet az erkölcs kutatásának individualista-racionalista programja befutott a 20. század közepétől. Becker (1957) korai diszkriminációs modellje a gazdasági racionalitástól eltérő negatív érzelmek negatív jóléti következményeire hívta fel a figyelmet a hazájában kibontakozó fekete polgárjogi mozgalom időszakában. Később, a hatvanas évek végétől több vezető amerikai közgazdász – köztük Becker –, a

racionalitástól eltérő jóindulat pozitív társadalmi következményeiről értekezett (például Becker 1974, Hochman és Rodgers 1969, elméletörténeti áttekintéshez lásd Fontaine 2006) egy olyan időszakban, amikor a posztindusztriális átalakulás kezdte látványosan kihúzni talajt a munkásosztály jelentős tömegei, köztük a polgárjogot nyert feketék nagy részének talpa alól, teret engedve a kilátástalanabb, ún. új szegénység terjedésének (Wilson 1987). Nem sokkal később egyre több nyugati kutató a feltétlen segítség motivációja helyett a feltételekhez kötött együttműködésben látta a közjó létrehozásának eszközt (ahogy e fejezetben megmutattam). Végül, az új évezred hajnalán, a közjóért tenni nem akaró bosszúállók váltak az új hőökké (magyar nyelvű összefoglalóhoz lásd Marosán 2008), egy olyan időszakot követően, amikor a szovjet modell összeomlása, egy tartós növekedési periódus és a tudáselit javuló relatív helyzete mind hozzájárulhatott a piacgazdaság új egyenlőtlenségeit legitimáló közbeszédhez, és a státuszukat generációkon átörökítő új szegényekkel szembeni, növekvő türelmetlenséghez (Gilens 1999).

Az altruista büntetés fogalma – a hozzá kapcsolódó empirikus eredményekkel együtt –, némileg eltávolodik a viszonyossági alapú modellek logikájától. Nem zárja ki az ilyen típusú büntetést (például a „szemet szemért” stratégiát), de nem is kötődik már hozzá erősen. Herbert Gintis az önérdéken túlmutató büntetési hajlandóságot mégis az erős reciprocitás fogalmával írta le (magyarul lásd Syi 2008), hangsúlyozva, hogy a pozitív gesztusok viszonzása is hasonló beépült ösztönző lehet, mint a méltánytalanul viselkedők büntetése. Ezzel visszakötötte az eredményeket a méltányossági hipotézis gyökereihez – beleértve Akerlof (1982) adok-kapok játék (gift-exchange) modelljét is. A reciprocitás „erős” jelzője az ismételt játékok racionálisan viszonzó magatartásától történő megkülönböztetést szolgálja. Mindazonáltal a kísérletekben jóval nehezebb kimutatni jelentős mértékű pozitív viszonzást, mint negatívát (ami az altruista büntetés).

### IV.2.3. A MÉLTÁNYOS SZAVAZÓ

#### IV.2.3.1. Az „érdemesség”

Az egyenlőtlenségaverzió alkalmazása magától értetődő a közpolitikai kérdésekben kialakuló álláspontokra. Ugyanakkor a kétséges kísérleti eredmények miatt soha nem merült fel komolyan a társas preferenciák kutatási programjában, hogy a szavazói viselkedést velünk született, nyers, egyenlőségpárti preferenciák bevonásával vizsgálják. Ehelyett az érdemesség szociológiában meghonosodott koncepcióját alkalmazták az erkölcsös polgár kutatási programjában is. Az összetettebb elosztási elméletek lehetővé tették a méltányossági érdemesség elvének kezelését (Konow 2003). Az elv arról szól, hogy aki megteszi a magáét (például a boldogulásáért), annak a javára, a méltányosság elve alapján, szívesen áldoznak az erkölcsös polgárok. A másik oldalról ez azt jelenti, hogy akinek a jelenlegi intézményi szabályok szerint az államon keresztül közösen javakat juttatunk, attól viszonzást várunk el – mindenekelőtt erőfeszítéseket (Bowles et al. 2001).

A társas preferenciák iránt érdeklődő közgazdászok közül Christina Fong (2001) kérdőíves felmérésben tesztelte a szegényeknek juttatandó segítség iránti attitűdök szerkezetét, megerősítve, hogy működik az érdemesség méltányossággal összeegyeztethető elve az emberek fejében. Fong az eredmények fényében, elméleti irányultságú szerzőtársakkal együtt, az erős reciprocitás modelljének közpolitikai preferenciák területén történő alkalmazhatósága mellett érvelt (Bowles et al. 2001).

Az okság keresési mechanizmusára építő, szociálpszichológiai gyökerű értelmezést már a kilencvenes években alkalmazták a politikai gazdaságtan diszkriminációpreferencia elméletének (például Alesina et al. 2001, Luttmer 2001) alternatívájaként az amerikai jóléti intézményrendszer megértéséhez (Gilens 1999). Ezekben a tanulmányokban az volt a kiinduló kérdés, hogy az interkontinentális kulturális, és egyéb intézményi hasonlóságok, és a(z európai) kontinensen belüli jelentős különbségek ellenére miért tér el radikálisan az észak-

amerikai és az európai jóléti intézményrendszer logikája? Míg Európa nagy részén, a háború utáni jóléti rendszerek különböző szervezési elvek alapján, de egységesen széles körben, és többnyire automatikus kiosztási eljárással segítettek a munka nélkül maradó családokon, addig az Egyesült Államokban szűk körben és kemény feltételek mellett zajlott a nem túlságosan nagyvonalú segélyezés (Albrekt Laresen 2008). A magyarázat előkészítéseként Alesina és Glaser (2004) alapos elemzésben mutatta meg az intézményi sajátosságok együttjárását az államok etnikai heterogenitásával. A társas preferenciák programjától teljesen függetlenül jelent meg a neoklasszikus közgazdaságtan diszkriminációspreferencia-alapú magyarázatával szemben álló értelmezés, amely etnikai preferenciák helyett a szorgalomra (érdemességre) vonatkozó etnikai sztereotípiákkal magyarázza az etnikai heterogenitás erodáló hatását a jóléti rendszerre (Gilens 1999). Másként fogalmazva, Amerikában sok fehér szavazó szívesen támogatná az önhibájukon kívül elszegényedett, szorgalmas embereket. Azonban egy jelentős részük lustának tartja a feketéket, és mivel azt látják (részben torzítva), hogy a segélyek főként a feketéknek mennek, ebből azt a következtetést vonják le, hogy a segélyek jelentős része lusta emberekhez kerül, akik ezt nem érdemlik meg.

Ezzel a hipotézissel összhangban vannak azok a vizsgálatok, amelyek a szegénység médiareprezentációjának etnikai hatástól független befolyására világítanak rá (Albrekt Laresen és Dejgaard 2013): az is szűkmarkúbbá teheti az államokat és polgárokat, ha a fehér szegényeket mutatják be semmirekellőnek a tömegkommunikációban. Az érdemesség koncepció alkalmazhatóságát mutatják azok az elemzések is, amelyek a szegénységi ellátások intézményi keretezésének közvéleményt befolyásoló szerepét hangsúlyozzák (Albrekt Laresen 2008): az emberek nagyobb eséllyel gondolkoznak el azon, hogy hibásak-e a szegény polgártársaik a sorsukért, ha a szegényeket ellátó intézmények működésébe bele van kódolva ennek a kérdésnek a folyamatos napirenden tartása.

A fentiekben csak felvillantottam néhány fontos állítást egy nagyon gazdag politikai-szociológiai kutatási területről, amely jelenleg is új eredményekkel bővíti a szegénységgel kapcsolatos polgári atti-

tűdök megértését (van Oorschot és Roosma 2017). Csupán azt akartam érzékeltetni, hogy a társas preferenciák elmélete nem hozott új hipotéziseket a közpolitikai preferenciák vizsgálatába, mert az elmélet közpolitikai alkalmazásának időszakában, más elméleti alapokon, ugyanazokat a hipotéziseket fogalmazták meg.

Jelentős különbség az érdemesség politikai szociológiai, illetve társas preferenciákon nyugvó cselekvéseméleti megközelítése között, hogy az előbbi ambíciója kifejezetten a jóléti ellátásokkal kapcsolatos vélemények (és az intézmények) magyarázata, míg az utóbbi fókusza elvileg szélesebb, a gazdagok adóit is érinthetné, és alkalmas lehetne az általánosabb gazdaságpolitikai kérdésekben elfoglalt vélemények értelmezésére is. Azonban az erős reciprocitás közpolitikai attitűdökkel kapcsolatos alkalmazása sokáig leragadt a jóléti rendszer meglehetősen elnagyolt (és főleg a szegényellátásokra koncentrált) elemzésénél; ahol nem kínált új hipotéziseket. Az utóbbi néhány évben, miután a tézis beszivárgott a politikatudományba, megnőtt az alkalmazások száma, szélesítve a vizsgált területek körét (például egészségügy, lásd Jensen és Petersen 2017 – bár ez az alkalmazás is jelen volt a korábbi szociálpszichológiai irodalomban: Weiner 1995).

Az alternatív értelmezési keret és annak gazdagabb irodalma nem vonja kétségbe azokat az eredményeket, amelyek összhangban vannak az erős reciprocitás modelljével. A nyugati demokráciákban, a jóléti ellátásokkal kapcsolatos közvélemény-kutatásokban megkérdezett polgárok érzékenyek arra, hogy a különböző állami ellátásokban (potenciálisan) részesülő polgártársaik mennyit tettek le az asztalra, mennyire törekedtek arra, hogy rossz sorsukat elkerüljék (van Oorschot és Roosma 2017). Ennek az erős reciprocitással kongruens polgári érzületnek a létezését ugyanúgy megerősítik az elosztási elvekre vonatkozó direkt kérdésekre adott válaszok, mint a különböző csoportok támogatása iránti attitűdök és a csoportokról alkotott sztereotípiák összevetése (Gilens 1999). A kutatási területen olyan, ún. vignettás vizsgálatokat is végeznek, amelyekben hipotetikus kedvezményezettek érdemességéről kell a válaszadóknak nyilatkozniuk, miközben a kutatók a fiktív kedvezményezettek szándékait (például erőfeszítéseit) manipulálják. Ezek a vizsgálatok is megerősítik a más

módon szerzett ismereteket: a válaszadók a saját sorsukért tenni akaró embereket tartják állami támogatásra méltónak (Janky et al. 2014, Petersen 2012, Petersen et al. 2012).

#### IV.2.3.2. Globális preferenciák vagy lokális attitűdök?

Az erős reciprocitás közpolitikai attitűdök vizsgálatára történő alkalmazásának szempontjából érdekesek azok a kísérletek, amelyekben megpróbálták megtalálni az univerzális társas preferenciákat a különböző kultúrákban élő, és a jóléti rendszerről különbözőképpen vélekedő emberek divergáló véleménye mögött. A feltevés az volt, hogy ha a társas preferenciák univerzális emberi motivációk, akkor markáns egyéni különbségek előfordulhatnak ugyan, de a nagyobb populációk átlagában a társas preferenciák hasonlóan befolyásolják a közpolitikai véleményeket a világ különböző részein. A kinyilvánított attitűdökben, a társadalmak között mutatkozó különbségek hátterében, a szegényekre vonatkozó (kultúrától függő) eltérő vélekedéseknek kell állniuk.

Aarøe és Petersen (2014) az Egyesült Államokban és Dániában végzett survey-kísérletek eredményeit vetették össze, hogy teszteljék a fenti hipotézist. Azt találták, hogy – összhangban az erős reciprocitás modelljével – a nyers közvélemény-kutatási adatokban mutatkozó különbségek eltűnnek, ha a vinyették a prezentált fiktív kedvezményezett jelleméről egyértelmű üzeneteket küldenek a válaszadónak. Az érdemesség egyértelmű megfogalmazása esetén a dán szavazók ugyanannyira támogatják vagy utasítják el a bemutatott fiktív kedvezményezett állami támogatását, mint az amerikaiak. Ugyanakor azt is látni kell, hogy a kutatásban meglehetősen szélsőségesen rajzolták körül az érdemes és érdemtelen támogatottakat (amit az elmélet megenged, és a cél épp minden bizonytalanság kiszűrése volt).

Almás és szerzőtársai (2016) közelmúltbeli kutatásukban az egyenlőtlenségaverzió kibővített modelljét (Konow 2003) porolták le, és elvégeztek egy, a fentihez hasonló norvég-amerikai összehasonlító survey-kísérletet, hogy teszteljék az elméletet. A vizsgálatban egy

diktátorjátékot követően egy semleges harmadik fél végzett valós pénzbeli újraelosztást (a játékban a második játékosok nem nyereményként kapták a pénzt az első játékostól, hanem a munkájuk ellenszolgáltatásaként). Fontos, hogy az eredeti diktátorjátékokkal szemben – és közelítve az erős reciprocitás érdemesség koncepciójához –, egyértelműen meghatározták, hogy ki mennyit tett le az asztalra (egy online piactéren végzett munkával), valamint sokkal nagyobb fizikai távolságot iktattak be a pénzosztók és kedvezményezettek közé, ami sokkal egyértelműbb volt, mint a hagyományos laboratóriumi kísérletekben (amelyekben az adakozók és a kedvezményezettek sokszor egy teremben ülnek, legfeljebb kisebb paravánokkal elválasztva). Tehát megtisztították az elosztási játszmát sok szabad (félre)értelmezési lehetőségtől.

Az eredményeik újabb szeget vertek az egyenlőtlenségaverzió modelljének koporsójába, mivel az érdemesség egyértelművé tétele ellenére is jelentős különbség állt fenn a két ország között abban a tekintetben, hogy mekkora egyenlőtlenséget fogadnak el (Amerikában nagyobb). Ez a negatív eredmény nem csupán a korai manipulált diktátorjátékok tapasztalataival egyeztethető össze, de közvetlenül kapcsolódik Henrich és szerzőtársai (2001, 2004) előző alfejezetben bemutatott nemzetközi ultimátumjáték kísérletének eredményeihez is.

Összességében látható, hogy a méltányosságelmélet lazán körvonalazott eredeti modelljéből elsősorban az mutatható ki robosztusan, hogy létezik az altruista büntetésre való hajlandóság a (nem tudni, hogyan meghatározott) méltánytalansággal szemben. Ahogy haladunk a méltányos következmények kvantitatív meghatározása felé, úgy csökken az elmélet érvényessége. A méltányossági motiváció velünkszületett elosztási preferenciaként való értelmezését elég egyértelműen cáfolják az adatok.



### IV.3. Nyitva maradt kérdés: méltányos döntések, vagy tisztességtelen értelmezések?

#### IV.3.1. RACIONÁLIS-E A MÉLTÁNYOS?

##### IV.3.1.1. Viselkedés vs. stratégia az ultimátumjátékban

A társas preferenciák kutatói által végzett ultimátumjátékok, büntetési lehetőséggel kiegészített fogolydilemmák, és hasonló szituációk segítségével végzett kísérletek látványosnak tűnő sikerei ellenére sok kísérleti közgazdász tekintett – és tekint továbbra is – kétségekkel a felhalmozódott empirikus eredményekre. Nem az eredményeket vonják kétségbe, hanem azt, hogy ezek társas preferenciák létezését bizonyítanák.

A Fehr-vezette kutatási program képviselőinek egyértelmű álláspontja, hogy a fenti játszmák jellemző kimenetelei ellentmondanak a neoklasszikus közgazdaságtan egoista embermodelljéből levezethető eredményeknek. A kételkedők azonban arra hívják fel a figyelmet, hogy ezekben a játékokban a résztvevők többsége Nash-egyensúlyi stratégiakombinációkat játszik (Binmore és Shaked 2010). Ugyanis azok a játszmák, amelyeket a legtöbbször használnak a társas preferenciák létezésének bizonyítására, nem rendelkeznek domináns stratégiákkal, azaz teljesen más ösztönzési struktúrájuk van, mint az egyszerű fogolydilemmának.

Az alábbi ábrán ezt nagyon egyszerűen be lehet mutatni. Az átláthatóság kedvéért vegyünk egy olyan egyszerűsített ultimátumjátékot, amelyet egyébként gyakran használnak a laboratóriumi kísérletekben. Itt az első játékosnak csak két opciója van: vagy megtart nyolc egység pénzt, és kettőt átad, vagy elfelezi a tizes pénzcsoportot, és ötöt-ötöt ad magának és a társának.

Az ábrán jól látszik, hogy noha a játék szabályai szerint az ultimátumot fogadó (második) játékosnak kétféle válaszra van módja, ez nem jelenti azt, hogy csak kétféle stratégiát tud kialakítani magában – emlékezzünk a II. fejezetben tárgyalt Nash/Crowe-féle „csajozási játékra”, ahol a felületes interpretációk szintén megfeledeztek erről

7. ábra. Az egyszerűsített ultimátumjáték stratégiai formája

		Második játékos (aki dönt az ultimátum elfogadásáról vagy elvetéséről)		
		Ajánlat	bármit elfogad	csak az egyenlőt fogadja el
Első játékos (aki ultimátumot ad)	8 – 2	Első: 8, Második: 2	Első: 0, Második: 0	Első: 0, Második: 0
	5 – 5	Első: 5, Második: 5	Első: 5, Második: 5	Első: 0, Második: 0

(lásd Binmore és Shaked 2010, Rubinstein 1991). Ugyanis a válasza lehet feltételes is: az igen és a nem mellett dönthet úgy is, hogy „attól függ”. Példánkban konkrétan elhatározhatja, hogy csak azt fogadja el, ha egyenlő felosztást kap. Így a külső szemlélő által megfigyelhető két válaszlehetőség mögött valójában három lehetséges stratégia rejlik. Egy olyan hagyományosabb ultimátumjátékban, ahol egységenként lépdelve nullától tízig bárhova beállíthatja a második játékos a feltételt, nyilván még több (12) stratégiája van. Kérdés, hogy racionális választás-e bármelyik feltételes stratégia a „legalább egyet kapnom kell” stratégián túl?

A mátrixból látható, hogy ha az első játékos csak kettőt ajánl fel, azt a másodiknak el kell fogadnia. Ez a visszafelé göngyölítéssel kapott stratégiakombináció. Ugyanakkor, ha az első úgy dönt, hogy az összeg felét (ötöt) felajánl a másodiknak, akkor a második játékos számára a kívülről megfigyelhető racionális viselkedés (elfogadás) két stratégiát is takarhat: az egyik, a „bármit elfogadok, hisz több mint a semmi” stratégia, a másik viszont a „csak a tisztességes, felezős ajánlatot fogadom el” stratégia. Most képzeljük magunkat az első játékos fejébe! Ha ő (az első játékos) bármilyen okból feltételezi, hogy a második játékos a „csak az egyenlőt fogadom el” stratégiát alkalmazza, akkor ötöt kell felajánlania. A feltételes („csak az egyenlőt”) elfogadás stratégiájára az első játékos legjobb reakciója, ha egyenlő felosztást javasol. És az egyenlő felosztási javaslatra a második játékos (egyik) optimális válasza az, hogy csak az egyenlőt fogadja el. Tehát

ebben a játszmában két stratégiakombináció is Nash-egyensúly (a stratégiakeverési lehetőséget nem számolva). Azaz, az önérdekkövetés játékelméleti modellje nem zárja ki, hogy méltányos elosztási javaslatot tegyen az első játékos az ultimátumjátékban.

#### IV.3.1.2. A képzelt fenyegetés

Ugyanakkor a visszafelé göngyölítés, mint kiemelt megoldási módszer alkalmazásának hívei szerint egy racionális emberekből álló világban hiteltelen fenyegetés a második játékos részéről (irracionális várakozás az első részéről), hogy olyan stratégiát választ, amellyel bármilyen pozitív ajánlatot visszautasít. Ugyanakkor emlékezzünk a visszafelé göngyölítés korábbi tárgyalására a III. fejezetben: Binmore (1997) értelmezése szerint megfosztjuk a racionális egyént egy fontos döntési opciótól, ha nem adjuk meg neki azt a lehetőséget, hogy élve a játék által adott döntési opciókkal, irracionálisnak tüntesse fel magát – illetve megpróbálja átdefiniálni a játékot, és vele a másik fél racionális reakcióját. Binmore ezt egy olyan szituációban (százlábú játék) vetette fel, amelyben egyetlen Nash-egyensúly van. Az érvelése még érvényesebb lehet egy olyan játszmában, amely több egyensúlyi kombinációt tartalmaz.

Másrészt viszont azt is látni kell, hogy Binmore (1997) eredeti érvelése egy olyan, időben elhúzódó játékra vonatkozott, amelyben a többlépcsős lejátás adott lehetőséget a játékosoknak arra, hogy megzavarják a másikat. Az ultimátumjátékban ilyen zavarásra nincs lehetőség, az első játékos nem tudja, és nem is érdeke olyan irányban manipulálni a játékot, hogy elterelje a végeredményt a visszafelé göngyölítéssel járó végkimeneteltől. A Nash-egyensúlyok egyenrangúságában hívó szakemberek válasza erre az, hogy Aumann (1995, 1998) eredeti levezetése nagyon érzékeny arra, hogy a racionális világ zárt és hibamentes legyen, ahol mindenki *tudja* mindenkiről, hogy mindenki tudja mindenkiről a végtelenségig, hogy nincs világ a tökéletes racionalitás univerzumán kívül. Éppen ezért maga Aumann sem ajánlja, hogy a sterilitásuk ellenére mégis kicsit poros laborató-

riumi kísérletek világában alkalmazzák a visszafelé göngyölítésre vonatkozó levezetését (Aumann 2000). Ne feledjük azt sem, hogy az egyensúly koncepciójának finomításában szintén érdekelt, és szintén Nobel-díjas Selten maga ajánlotta az ultimátumjátékot hódító útjára indító Güthnek, hogy kísérletezzen vele – és ő akkor még nem a társas preferenciák tesztelésén, hanem a racionalitás megfelelő modellezésén gondolkodott. Egyébként az egyensúlyok versenyét vizsgáló evolúciós játékelméleti modellekben sem emelkedik ki a visszafelé göngyölítéssel kapható egyensúly a többi Nash-egyensúlyi kombináció közül (Samuelson 1994). Az ismételt fogolydilemma paradigma kísérleteinek egy friss metaelemzése is arra a következtetésre jutott, hogy miközben az emberek a racionális megoldásokat keresik, nem jellemző a konvergencia a visszafelé göngyölítés irányába (Embrey et al. 2017).

Ebben a könyvben a Nash-egyensúly további finomítása ellen érvelők gondolatmenetét követem. Ahogy nemsokára látni fogjuk, van empirikusan védhető ráció abban, ha az első játékos tart attól, hogy a második játékos értetlenségéből kiszúr vele. Ilyen félelmek esetén a méltányos megoldás irányába *érdemes* elmozdulni.

A Nash-egyensúlyok pluralitásának hatása látványosabb a fogolydilemma kísérletek esetében. Az eredeti játékban egy Nash-egyensúly van, és kísérletekben a játékosok jellemzően – az egyensúlytól távolabbról indulva – idővel az egyensúly felé konvergálnak (ahogy újra és újra, mindig új partnerekkel játszanak egy-egy véges játékot). A büntetés lehetőségének bevezetésével – és a fentiekkel analóg módon képződő újabb Nash-egyensúlyok létrejöttével – a játékosok nagyobb eséllyel közelítik meg a mindenkinek kedvezőbb, kooperatív egyensúlyt, ahol – emlékezzünk a racionális normafenntartás logikájára! – a játékosok büntetnék a potyautazást, és ettől való félelmükben az emberek nem is potyautaznak. Figyelem: ehhez nem kell végtelen sok ismétlés! Ha elegendően nagy a büntetés mértéke, akkor a kooperálásra (az egyik) legjobb válasz a feltételes büntetés, a feltételesen büntető stratégiára pedig a legjobb válasz a kooperálás.

Összességében azt mondhatjuk, hogy a Nash-egyensúly bevonásával történő racionális érvelés nem cáfolja a társas preferenciák létezésére vonatkozó hipotézist, azonban a társas preferenciák proponen-

seinek állításával szemben (Gintis et al. 2005) fenntart egy ortodox nullhipotézist, miszerint társas preferenciák nélkül is létrejöhetnek a fentiekben bemutatott, méltányos viselkedésre utaló kísérleti eredmények.

#### IV.3.2. AMIKOR AZ ÉLET BETÖR A LABORBA

##### IV.3.2.1. Konfúzió a laborban

A játékelméleti egyensúlykonceptiók filozófiájával nagyobb barátságban lévő kutatók szívesebben koncentrálnak az eseteknek arra a kisebb hányadára, amikor a játékosok letérnek valamely Nash-egyensúlyi pályáról. Az ultimátumjátékban ez akkor következik be, amikor a második játékos elutasít egy pozitív ajánlatot (Eckel és Gintis 2010). Korábban említettük, hogy a társas preferenciák kutatása eleve a büntető viselkedés értelmezésének irányába indult el az első évek után, amikor az elosztási-preferenciákkal kapcsolatos modellt nem igazolták a kísérletek. Az egyensúlyon kívüli kimenetek a játékok kisebb arányában fordulnak elő (hiszen az első játékosok félnek a büntetéstől), de előfordulásuk az egyenlőtlen ajánlatok esetében viszonylag gyakori. Ez mindenképpen ellentmond a Nash-egyensúlyon nyugvó magyarázatoknak.

A vitának ezen a pontján az ortodox játékelmélet méltányosság-elméletet kritizáló hívei látszólag kilépnek az egzakt érvelés keretei közül. Azonban érveik érdemesek a megfontolásra. Azt állítják, hogy a kísérleti alanyok a laboratóriumban, korlátozott kognitív képességeik és a szűkös idő miatt, a játékelmélet szabályrendszerének instant megtanulása helyett jellemzően valós életük játékaival keresik az analógiákat (Bicchieri et al. 2018, Binmore 1998, 2010). Ez első hallásra azt is sugallhatja, hogy a játékelméletek által oly erősen promótt Nash-egyensúly nem is annyira intuitív; egy egyszerű játékban sem ismerjük fel, azaz nem a hétköznapi emberek döntési logikáját, hanem egy elvarázsolt matematikusnak és követőinek a gondolkodásmódját tükrözi.

Intuitíve talán jobban elfogadjuk az instant tanulás korlátaira vonatkozó érvelést, ha nézünk egy másik példát a modern tudományos szemlélet, illetve az ember által története során megszokott gondolkodási logika között. A nyugati társadalmakban ma már mindannyiunkat megtanítottak jól vagy rosszul térképet olvasni. Sokszor a térkép madártávlati szemszögéből tervezzük meg útjainkat. Azonban a térkép egy racionális civilizációs képződmény, mi földi halandók minden bizonnyal az „egyenesen előre, aztán a fánál jobbra” típusú tájékozódásra vagyunk kalibrálva. Bárki megtapasztalhatja ezt, aki az egyébként velejéig modern, ám a térképkultúrához kevésbé szokott Japánban próbál helyiektől térkép segítségével útbaigazítást kérni. Tehát az, hogy az evolúció megtanított minket a Nash-egyensúly keresésére a valós interakcióinkban, még nem feltétlenül jelenti azt, hogy azonnal olvasni tudjuk a Thomas Schelling által kitalált mátrixokat (bár minden bizonnyal sokkal hamarabb tanuljuk meg azokat olvasni, mint a játékok képletek segítségével történő, „hivatalos” leírását).

Az élet-játék analógia laboratóriumi felhasználásának hipotézisét éppen az a kutatás támogatja a leglátványosabban, amelyet a társas preferenciák univerzalitásának bizonyítására szerveztek meg (Henrich et al. 2001, 2004). A fejezet első felében ismertetett nemzetközi kulturális antropológiai ultimátumjáték projektben Fehr és közgazdász kollégái kiképezték, majd szétküldték a terepismeretekkel rendelkező antropológusokat a szélrózsa minden irányába; az eredmények begyűjtése után pedig összeállítottak egy kötetet, amelynek bevezető fejezeteiben a társas preferenciák elméletének dicsőségét hirdették. Ez a kötet az elmélet egyik kanonizálási kísérlete. Mindezek után a könyv második felében a hálátlan antropológusok, egyenként, saját történetüket elmesélve, mind azt bizonygatják, hogy eredményeiket milyen nagymértékben befolyásolta az, hogy miképp értelmezték (félre) alanyaik a kutatók számára egyszerűnek tűnő ultimátumjátékot. Például Henrich hagyományos perui terepén a vadászok kis zsákmányokat ejtenek, amelyeket nem szívesen osztanak meg másokkal – itt a játékosok a szabadtéri „laborban” is szűkmarkúak voltak (Henrich 2000). Ugyanakkor egy indonéz bálnavadász törzs tagjai közül sokan úgy érezték, hogy tisztességtelen akár a „zsákmány” felét is megtar-

tani. A „hiperaltruista” ajánlatokat viszont jó néhányan visszautasították – ahogy a résztvevők utólagos értelmezéseiből kiderült, nem azért, mert még altruistább döntést vártak, hanem inkább azért, mert felháborodtak a pofátlan tukmáláson (Alvard 2004). A terepen ott-honos kutató megmutatta, hogy a ritka nagy zsákmánnyal hazatérő bálnavadászok jó szokása társaik, különösen a földműveléssel is foglalkozók tukmálása a szűkebb körben elfogyaszthatatlan zsákmánnyal. Soványabb időben azonban a tukmálók visszatérnek a megajándékozott társaikhoz, hogy részt kérjenek a szűkös mezőgazdasági terméskből. Tehát óvatosnak kell lenni, hogy mennyit fogad el az ember a tukmált bálnahúsból. A kulturális különbségekre érzékeny Henrich és szerzőtársai (2001) egyébként kvantitatív elemzéssel is bemutatják, hogy az átlagos ajánlatok összefüggenek a piaci tranzakciók kiterjedtségével: ahol hozzá vannak szokva a sok csereberéhez és alkudozáshoz, ott nagyobb eséllyel hagynak az első játékosok a másik félnek is jelentős részt.

Továbbá vannak arra vonatkozó jelek is, hogy az életben sokszor ismétlődő, a világból való kimenekülést ritkán biztosító interakciók stratégiáitól nehéz szabadulni a rövid laboratóriumi látogatás alatt. Az anonimitás feloldása az együttműködési kísérletekben direkt következmények nélkül is jelentősen megnöveli a kooperációs hajlandóságot (Rege és Telle 2004). Érdekes, de már az az is apróság is jelentősen megnöveli az alanyok nagyvonalúságát a diktátorjátékban, ha döntés közben a számítógép képernyőjén látják mások tekintetét (Haley és Fessler 2005). Én magam is szervezője voltam egy olyan fogolydilemma kísérletnek, ahol az egymást nem ismerő, de számítógépes hálózaton összekötött résztvevők pozitív és negatív szmájlikat küldhettek egymásnak, és már ez az inger is elég volt ahhoz, hogy nagyobb együttműködésre bírjon embereket (Takács és Janky 2007). Ide kapcsolódnak Dana és szerzőtársai (2006, 2007) már említett diktátorjáték kísérletei is, ahol már a fiktív interakciókat is kerülte a résztvevők jelentős hányada, hogy több pénzt tarthasson meg magának.

Mindezek az eredmények arra utalnak, hogy a játékosok vonzódnak az ismétlődő játszmák Kandori (1992) és társai által bemutatott

normafentartó stratégiáihoz azokban a helyzetekben, amikor a laboratóriumban akaratlanul előhívják az attól való (nem feltétlenül tudatosuló) félelmet, hogy bármelyik embertársunkkal összefuthatunk később, vagy legalábbis rossz hírünket kelthetik a jövőben.

#### IV.3.2.2. Tanulás a laborban

Az élet-játékok és a laboratóriumi játékok összefonódását másik irányból még régebben és még többször tesztelték: vajon idővel megtanuljuk-e a laboratóriumban, hogy egy új játszmában vagyunk, amelynek sajátos, az életünkben ritkán előforduló kifizetési struktúrája van? Ha nyomát találjuk ennek a tanulási folyamatnak, az két dolgot is megmutat. Egyrészt igazolja, hogy ha lassan is, de a racionális gondolkodás a Nash-egyensúlyok felé viszi az emberek stratégiai döntéseit. Másrészt óvatosan kell bánni a laboratóriumi eredményekkel, hiszen az induló stratégiák bizonyítják, hogy azok nem jól informált döntésekből erednek. A tanulási kísérletek arra utalnak, hogy az emberek bizonyos dilemmákban rossz modellel kezdenek, de lassan a Nash-egyensúlyi stratégiák felé terelődnek döntéseik (Binmore és Shaked 2010).

A stratégiákkal kapcsolatos informáltságot még közvetlenebb módon is lehet mérni. Bicchieri és szerzőtársai (2016) túlléptek a játékelméleti kísérletekben alkalmazott egyszerű figyelem- és értelmezéskontroll kérdéseken, és részletesebben kérdezték alanyaikat arról, hogy miképpen látják az eléjük tett stratégiai döntési dilemmákat. Kutatásukban csupán a játékelméleti kísérleteikben résztvevők ötöde ismerte fel, hogy az általuk pénzért játszott fogolydilemmában a potyautazás lenne a domináns stratégiájuk, azaz az optimális döntésük nem függhet a többi játékos döntésétől. Ráadásul a többség a számítógép elleni fogolydilemma-játékokban is igazodott a másik (gépi) játékoshoz (gyengítve a feltételes viselkedés méltányossági motivációjára vonatkozó hipotézist).

Ahogy a nyolcvanas években Axelrod méltányossággal kapcsolatos egyszerű feltevéseit bécsi zoológusok szedték ízekre a modell-



jeikben (Nowak és Sigmund 1993), úgy mostanában is a zoológusok kezdték – az ortodox játékelméleti gondolkodók tanulási hipotéziseit empirikusan megtámogatva – a legérzékenyebb pontokon támadni a társas preferenciákra vonatkozó eredményeket. Az oxfordi egyetem zoológiai tanszékének kutatói, Maxwell Burton-Chellew és szerzőtársai (2016), az együttműködési játékokra vonatkozó kísérleteikben végzett részletes kérdezősködés nyomán megmutatják, hogy a fogolydilemmában együttműködők közül elég sokan azt gondolják, hogy nem jótékonykodnak, hanem saját nyeresiményük maximalizálásának az érdekében választják az együttműködést a játék mátrixa alapján. Azt feltételezik, hogy a nyeresiményüket maximalizáló optimális döntésükhöz figyelniük kell a többiek lehetséges döntéseire, (vagy előzetesen gondolkozniuk kell róla), azt remélve, hogy a többiek, szintén az eszükre hallgatva, az együttes kooperációt választják. Pár évvel korábbi, szintén nagy visszhangot kiváltó vizsgálatukban (Burton-Chellew és West 2013) az alanyok egy részének külön felhívták a figyelmét arra, hogy a fogolydilemmában az együttműködő viselkedéssel jót tesznek másokkal. Ennek az információnak a kiemelése kifejezetten csökkentette az együttműködési hajlandóságot. Burton-Chellew és West (2017) legújabb (oxfordi diákok körében végzett kismintás) vizsgálatukban részletesebben is megvizsgálták a többiekre vonatkozó információk hatásmechanizmusait a fogolydilemma-kísérletekben. Eredményeik arra utalnak, hogy a diákok a sikeresnek bizonyuló résztvevő-társak viselkedését próbálták utánózni, és ez gyorsította az együttműködés hanyatlását. Ez összhangban van Bicchieri és Xiao (2009), a diktátor-játékosok motivációira vonatkozó korábbi vizsgálatának eredményeivel, ahol a másokra vonatkozó információkkal ellátott alanyok kevésbé reagáltak döntéseikben arra, hogy mások szerint mit kéne tenniük, annál inkább arra, hogy mások mit csinálnak valójában.

A legutóbbi kritikus eredményekkel kapcsolatban meg kell jegyezni, hogy ezek a fogolydilemma játékot alkalmazó kísérletekben születtek, és még hiányoznak a többnyire kismintás eredmények – független (vagy kifejezetten „ellenérdekelt”) kutatócsoportok által végzett – replikálási kísérletei. A társas preferenciák modelljeit

meggyőzően kétségbe vonó újabb kritikák – a korábbiakhoz hasonlóan – kevésbé érintették az ultimátumjátékok második játékosainak büntető magatartását. Az erős reciprocitás hívei szerint ebben a viselkedésmintában nincs nyoma a Nash-egyensúly irányába történő sodródásnak (Eckel és Gintis 2010). Továbbá, amikor Lamba és Mace (2010) a korábban ismertetett, anonimitást manipuláló kísérletekhez hasonló kutatásukban az ultimátumjáték első és második játékosának viselkedését vizsgálták, azt találták, hogy a második játékos (és csak ő) rezisztens minden, az anonimitást befolyásoló kísérleti kezelésre – azaz nem mutatható ki, hogy viselkedése valamilyen illúzió eredménye lenne (ez összhangban van a tanulási hatás előbb említett hiányával is).

#### IV.3.3. ALTRUISTA BÜNTETÉS, VÉGÁLLOMÁS?

A méltányossági preferenciák több évtizedes kutatásának eddigi konklúziója szerint sok jel utal arra, hogy az emberek költségeket is vállalva, „zsigerből” hajlandóak büntetni a számukra károsnak tűnő viselkedésformákat. Azonban a méltányossági preferenciák elmélete kudarcot vallott abban a tekintetben, hogy megvédjen olyan pozitív módon megfogalmazott univerzális elveket, amelyek megsértésére egyesek zsigeri büntetéssel reagálnak. A negatív eredményekből biztosnak tűnik, hogy az emberek nem született igazságosságelméleti filozófusok, akik a társas élet minden aspektusát tarták felosztásáról szóló alkunak tekintik, és számszerű válaszuk is van a helyes tortaelosztásokra. Azt ugyan elmondhatnánk, hogy az emberek jószándékot várnak el a velük kapcsolatba lépő embertársaiktól, azonban a társas preferenciák vezető kutatóinak optimizmusa ellenére sincs semmilyen mankónk abban a tekintetben, hogy mivel és mihez képest lehet a jószándékot megmérni.

A fenti pozitív és negatív eredmények két irányba csábították az erkölcsös polgári viselkedés titka után nyomozó tudósokat. Egyrészt, a méltányosságelmélet kutatásának felfutása után felépülő egyik kutatási program teljes mértékben elhagyta a racionális szituációs logi-

kai elemzéseket, és metodológiájában a hagyományos pszichológiai megközelítést tartotta irányadónak. Az irányzat képviselői abban látták a kudarc okát, hogy a kialakuló, majd érett kapitalista társadalmat vizsgáló elit-értelmiség az elmúlt kétszáz évben a racionális igazságosságkonstrukciókkal volt elfoglalva, és azt feltételezte, hogy (csak) ezek a konstrukciók irányítják, alaptermészetük részeként, a mindennapi emberek erkölcsi ítéleteit is (Haidt 2012). Ez a pszichológiai paradigma a modern értelmiségi diskurzus (és a rájuk építő racionális döntéseméleti modellek) szűk kereteitől megszabadulva keresi az emberek valós erkölcsi ösztöneit, így adva választ arra is, hogy milyen erkölcsi elvek megsértéséért jár a zsigeri büntetés.

A másik, az utóbbi évtizedben élénkülő kutatási program a méltányossági preferenciák elméletének kudarcából azt a következtetést vonta le, hogyha olyan alapvetőnek tűnő dilemmákban sem kódoltak belénk kész megoldásokat, mint a közjószág-előállítás vagy a tortaszeletelés, akkor meg kell barátkozni azzal a gondolattal, hogy az emberek közötti tranzakciók morálisan „nyitott végűek”, azaz számos döntés tekinthető erkölcsösnek, és számos erkölcstelennek. Mi, gyarló, önző emberek döntjük el – az öntudatlan alkalmazkodás révén megtalált egyensúly rögzülése után, vagy valamilyen közös térben folytatott diskusszió nyomán –, hogy mi a követendő erkölcsös cselekedet, és mi a büntetendő, erkölcstelen tett. A következő két fejezet ezt a két újabb, egymással ellentétes orientációjú, de a méltányosságelméletre egyaránt explicit módon építkező kutatási irányt mutatja be.

## V. 2001\* – Az ösztönös polgár erkölcs

### V.1. Az erkölcs új pszichológiája

#### V.1.1. A MORÁLIS ALAPOK ELMÉLETE

##### V.1.1.1. Néhány mondat a(z elmélet) történeti kontextusról

A morális alapok elmélete (MAE) azt feltételezi, hogy az élet specifikus fenyegetéseire és lehetőségeire specializálódott érzelmi modulok fejlődtek ki az emberekben (Haidt 2012, Smith et al. 2017). A morális alapok elméletének kidolgozása, és vele a mai politikatudományra nagy hatást gyakorló új morális pszichológia megalapozása Jonathan Haidt virginiai pszichológus és munkatársai undorral kapcsolatos kutatásaiból indult ki a kilencvenes években (Haidt et al. 1993, 1994). Haidtet az zavarta, hogy a társadalmi morálról szóló tudományos diskurzus egyrészt a tudatos, racionális megfontolásokra koncentrált, másrészt az emberi társadalmak hosszú története során fontos erkölcsi megfontolásokból csak néhányat emelt be a tudományos vizsgálódás körébe. Ezek az értékek elsősorban a baloldali-liberális gondolkozás számára legkedvesebb értékek voltak. Nem csoda, hogy az ilyen korlátok között építkező kutatások azt mutatták, hogy a baloldali választók inkább elvekkel rendelkező, racionális (erkölcsös) emberek,

\* Haidt, J. (2001): The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.

míg a jobboldaliak inkább ösztönlények (Emler et al. 1983). Az undorral kapcsolatban például a társadalmi erkölcs tudományos kutatói és a politikai gondolkodók úgy vélték, hogy az undorító, de mások számára ártalmatlan (fizikai fenyegetést nem okozó) viselkedés nem ítéltető el morálisan (Haidt 1993, Inbar et al. 2009). Haidt és követőinek álláspontja ezzel szemben az, hogy az undor kizárása a morális világból új jelenség, leginkább a felvilágosodáshoz köthető. Az emberi történelem túlnyomó részében az undorhoz kapcsolódó tisztasági megfontolások a társadalmi erkölcs alapvető pillérei voltak.

Haidt kutatási programja explicit módon kapcsolódott a politikai élet átalakulásához. A posztindusztriális demokráciákban először a baloldalon, majd a jobboldalon is megjelentek az osztálypártok ideológiai logikáját elvető mozgalmak. A szavazói döntésekben fokozatosan csökkent az osztálypozíció hatása, és megnövekedett (és mindmáig növekszik) a gazdaságpolitikától független társadalmi-kulturális kérdésekben elfoglalt álláspontok szerepe. A gazdasági részérdekek becsatornázását végző, közvetítő intézmények hanyatlása elősegítette a kulturális kérdések középpontba kerülését (Csigó 2016). A gazdasági ellentétek helyett egyre inkább a kulturális ellentétek kerültek a közéleti diskurzusok középpontjába. Haidt kutatási programjának kibontakozása egybeesik azzal az időszakkal, amikor szembetűnővé vált, hogy az emberek különböző csoportjai más és más tartanak erkölcsösnek. Csigó (2016) a közelmúltban megjelent tanulmányában azt állítja, hogy nemcsak a létező ellentétek válnak láthatóbbá, hanem valójában a politikai elitek nagyítanak fel, vagy akár kreálnak erősen morális tartalmú ideológiai ellentéteket. Ráadásul mindez egy olyan társadalmi közegben történik, amelyben az erkölcsileg releváns viselkedésmintázatok növekvő pluralizmusát tapasztalhatják meg a polgárok. Az erkölcs kutatásának egyféle, és kifejezetten gazdasági orientációjú igazságfogalma egyre kevésbé reflektált a hétköznapi életben és a közéleti diskurzusokban megélt, sokféle, egymással versengő nem-gazdasági igazságkonceptiók vitájára.

### V.1.1.2. A morális modulok

A kétezres évekre kidolgozott MAE szerint öt alapvető megfontolás, és ehhez kapcsolódó, erkölcsi ítéleteinket meghatározó kognitív modul határozza meg erkölcsi gondolkozásunkat (Haidt 2012).

- Gondoskodás/ártalom: ez az erkölcsi megfontolás más emberek szenvedéseire és szükségleteire érzékeny. Az elmélet szerint az utódok gondozásához szükséges ösztönökre vezethető vissza ez a modul.
- Méltányosság/csalás: az emberi közösségek túlélése szempontjából nélkülözhetetlen interperszonális együttműködési szabályok betartásáról szól.
- Hűség/árulás: nem egy-egy másik ember, hanem az egész csoporthoz való kötődés kötelessége.
- Tekintélytisztelet/behódolás: A társadalmi rend és hierarchia tiszteletének erkölcsi kötelessége. Szintén a nagyobb közösségek fennmaradásának egyik sarokköve.
- Undor/tisztaság/szentség: Ebben a fejezetben ezt az elvet tárgyaljuk részletesen. Az elmélet szerint a paraziták távoltartására kialakult ösztön később a társadalmi rend „tisztaságának” őrzőjévé is vált.

A gondoskodás és a tisztesség megfontolásai az individualista, illetve individualizáló erkölcs alapjai. A másik három a közösségeket összekötő erkölcs alappillére (Haidt 2012). A társadalmak, és a társadalmakon belül az egyének jelentősen különbözhetnek abban, hogy mekkora súllyal veszik figyelembe az egyes megfontolásokat az emberi viselkedés erkölcsi megítélésekor

A morális alapok elméletének egyik legfontosabb tézise, hogy az erkölcs tanulmányozásának korábbi alaptételével szemben az erkölcs nem tudatos megfontolásokon alapul, hanem érzelmeken, illetve morális intuíciókon. Ez az intuitív etika genetikai alapokon, valamint a korai szocializáció során rögzült reakciókon nyugszik (Haidt és Joseph 2004). Eszerint az erkölcsi ítélet érzelmi folyamat, amely intuitívan és automatikusan reagál a környezet eseményeire (Wage-

mans et al. 2018). Haidt és szerzőtársai téziséből empirikusan tesztelhető előrejelzések következnek. Nevezetesen: a tézis alapján ezek a modulok *a*) a korai szocializációt követő életszakaszokban stabilak, továbbá *b*) az összetettebb és/vagy konkrétabb ideológiai, politikai nézeteket okként magyarázzák, végül pedig *c*) öröklődnek (Smith et al. 2017)

Az undor-vágyakozás tengelyt sokan az összes érzelem legalapvetőbb építőkövének tartják (Smith et al. 2011). Garvey és Ford (2014) vizsgálata szerint a gondoskodás és méltányosság megfontolásai erősebben kötődnek a tudatos, racionális gondolkodáshoz, mint a tisztaság (szentség) undoron nyugvó érzülete. Ez az egyik oka, hogy az alábbi áttekintésben az undorral fogunk foglalkozni. Ez az érzület nemcsak negatív irányultságában, hanem ösztönösségében is kontrasztban áll a méltányossági elvvel. Az erkölcs új pszichológiáját megalapozó Haidt és munkatársai (1993, 1994) is az undor vizsgálatával indították el a kutatási programot, és később is ezzel az érzelmmel kapcsolatban végezték a legtöbb vizsgálatot.

## V.1.2. AZ UNDOR

### V.1.2.1. Miről beszélünk?

Az undor negatív érzelem, averzív ingerre reagáló ellenézés (Wagemans et al. 2018). Undor érzése esetén az egyén elutasítja az érintkezést a tisztátalannak ítélt objektumokkal, illetve tartózkodik tisztátalannak ítélt cselekvésektől (Rozin 2008). A kutatók szétválasztják a testi és nem-testi (társadalmi, erkölcsi) undort (Kam és Estes 2016). Az utóbbi évtized kutatásai azt állítják, hogy mindkét típus hasonló agytevékenységekkel jár együtt, hasonló arckifejezést indukál, hasonló kognitív reakciókat eredményez, és hasonló viselkedésre sarkall (Borg et al. 2008, Chan et al. 2014, Chapman et al. 2009, Jones és Fitness 2014, Kam és Estes 2016).

Az interperszonális viszonyokban az undor illetve annak hiánya nem egyszerűen az elutasításról vagy kedvelésről szól. A csoportközi

viszonyok hagyományos pszichológiai modelljeiben kedvelt és diszpreferált csoportokra oszlik a külvilág. A kedvelt csoportok mindenféle támogatást megérdemelnek, a diszpreferált csoportoknak nem szívesen juttatunk bármit is (Nelson és Kinder 1996). Az undor morális pszichológián, és a viselkedési immunrendszer hipotézisének alapuló elmélete ennél specifikusabb reakciókat feltételez, és a mások és másság elutasításának összetettebb modelljeire épít, azokat módosítja.

A sztereotípa-tartalommodell szerint embertársainkat barátságosság és kompetencia szempontjából értékeljük. Vannak, akiket nagyra tartunk, de ellenségesnek is, és ezért diszkrimináljuk őket, míg másoktól nem félünk, de nem is becsüljük őket sokra. Ebben a modellben a barátságosság és kompetencia tekintetében is alacsonyra értékelt személyekkel szemben érzünk undort (Fiske 2010). Azonban az undor ilyen interpretációja ellentétes a MAE feltételezéseivel, amelynek követői szerint Fiske keveri az undort a megvetéssel, amely szerintük más gyökerű érzelem (Clifford és Piston 2016). Fiske modellje az undorítóknak tekintett csoportokkal szembeni támadó magatartást is lehetségesnek tartja, míg az undor fent ismertetett evolúciós elmélete alapján ez az érzelem kifejezetten elkerülő magatartásra ösztönöz.

Wagemans és szerzőtársai (2018) megpróbálták szétválasztani a fertőzési ingert a gusztustalan viselkedés ingerétől. Például kérdőíves kísérleteikben úgy variálták az alanyok elé tárt hipotetikus szituációkat, hogy az egyik kondícióban a kérdeztettek arról olvastak, hogy egy betörő tintát önt szőnyegre, a másikban pedig rávizel. Adataik alapján a fertőzési inger nem befolyásolta az önbevalláson alapuló gusztustalanságérzékenység és a társadalmi tisztátalansággal kapcsolatos averzió kapcsolatát. Azaz a tintázás ugyanolyan undorító, mint a vizelés.

A legutóbbi időkig váratott magára a gusztustalanságérzékenység direkt mértékei, valamint az erkölcsi tisztaság iránti preferenciák MAE által definiált indikátorai közötti kapcsolat tisztázása. Van Leeuwen és munkatársai (2017) mutatták ki a kapcsolatot az ún. *Erkölcsei alapok kérdőíve* (Moral Foundation Questionnaire), és az egyes undorérzékenység indikátorok között. Ez a kérdőív közvetlenül kérdez



rá az emberek egyes erkölcsi megfontolásokkal kapcsolatos, általános véleményére. Wagemans és munkatársai (2018) ezt kiterjesztették az ún. *Morális alapok vignetták* kérdéssorára is. Ezek a vignetták nem általános attitűdkérdéseket tesznek fel, hanem konkrét cselekedetek erkölcsi megítélését kérik a válaszadóktól.

#### V.1.2.2. A viselkedési immunrendszer

Az undor/tisztaság feltételezett moduljához önálló evolúciós modell kapcsolódik. Nevezetesen a viselkedési immunrendszer elmélete (Tybur 2016). A viselkedési immunrendszer az étkezést, a párválasztást, és bármely egyén percepcióját befolyásolja (Murray és Schaller 2016). A viselkedési immunrendszer a biológiai immunrendszert egészíti ki. Akkor aktiválódik, amikor az egyén valamilyen fertőzésveszélyről szerez információt – szaglás, ízlelés, hallás, látás vagy tapintás útján (Terrizzi et al. 2013). Az információra adott érzelmi reakció az undor, a hozzá kapcsolódó kognitív reakció a fertőzés (hátsó) gondolata, az undor által ösztönzött viselkedési reakció pedig a feltételezett fertőzésveszély forrásának elkerülése. Mindehhez nincs szükség tudatos döntésre (Aarøe 2017).

A morális alapok elmélete és a viselkedési immunrendszer modellje evidensnek tekinti az undor evolúciós eredetét: a fertőzések elkerülése növeli a túlélési esélyeket (Smith et al. 2011). Ugyanakkor, ahogy minden evolúciós „mesével” kapcsolatban, itt is könnyen bonyolítható a modell, az empirikus ellenőrzés lehetősége nélkül. Ugyanis az új dolgokkal és élőlényekkel való találkozás hasznot is hozhat. Tehát nem biztos, hogy a minden furcsától való menekülés maximalizálja a genetikai fitnesset, és inkább átváltás (trade-off) van a fertőzés várható költsége, és az újjal való érintkezés várható hasznai között (Aarøe 2017). Ez a két ellentétes erő jelenléte azonban jól magyarázhatja az egyének gusztustalanságérzékenységének magas szórását (Aarøe 2017).

Az undor társadalmi attitűdökre gyakorolt hatásával foglalkozó társadalomtudósok körében elfogadott nézet, hogy genetikai fakto-

rok játszanak szerepet az undorra való hajlandóság kialakításában (például Clifford és Piston 2017). Mendez (2017) friss szakirodalmi áttekintésében az undor neurológiai és genetikai alapjai mellett érvel. A vonatkozó szakirodalom talán leglátványosabb eredményei agykárosodást szenvedett emberek politikai attitűdjeinek változásához kapcsolódnak (kismintás vizsgálatokon, esetenként egytucatnál kevesebb alany adataira építve). Látnunk kell ugyanakkor, hogy az undorhoz kapcsolódó agyi tevékenységek lokalizálása önmagában nem érv sem a genetikai alapok mellett, sem a kulturális folyamatok befolyásoló ereje ellen. Kétségtelen, hogy vannak génekhez kötött eredmények is (Mendez 2017). Ugyanakkor ez sem bizonyíték a kulturális hatások ellen. A feltételezett genetikai hatásokat kimutató kutatások szerint is befolyásolja a baráti kör a gének undorérzékenységre gyakorolt hatását (Settle et al. 2010). Haidt és munkatársai (2009) is második szintű lelki konstrukciónak tekintik a morális intuíciókat, köztük az undort. Ezek a konstrukciók kontextusfüggőek. Tehát az, hogy mi és ki számít undorítósnak, nemcsak az egyén genetikailag meghatározott érzékenységtől függ, hanem kulturálisan befolyásolt (Haidt 2012, Kam és Estes 2016). A kisebbségek fenyegetett identitásának megnyilvánulása lehet például, hogy az amerikai feketék érzékenyebbnak mutatkoznak az önbevalláson alapuló undorskálán (Haidt et al. 1994). A nők kérdőíves válaszaiból kirajzolódó magasabb undorérzékenysége az evolúciós pszichológusok tetszetős elméleteket prezentálhatnának, ugyanakkor a fiziológiai undorérzékenységi vizsgálatokban nincs különbség a nemek érzékenysége között (Rohrman et al. 2008).

### V.1.2.3. Az undor mérése

Az undor társadalmi attitűdökre gyakorolt hatásának vizsgálatakor elsőként azt kell eldönteni, hogy miképpen mérjük az undort. Erre vonatkozóan – a társadalmi attitűdök méréséhez hasonlóan – többféle eljárást dolgoztak ki. Mindenekelőtt az állapotalapú, illetve személyiségalapú undort kell megkülönböztetni (Kam és Estes 2016).

A személyiségalapú undor mérések elvileg stabil gusztustalanságérzékenységet mérnek a kutatók. Ilyenkor a különböző érzékenyséű emberek közötti eltéréseket figyelik meg (például politikai attitűdök tekintetében). Ha az undor-állapotot vizsgálja a kutatás, akkor a kutatók az alanyokból különböző ingerekkel undorérzetet akarnak kiváltani. Ezek olyan kísérletek, ahol a válaszadók kezelt csoportjai politikai attitűdjeik kifejtése előtt undorérzetet keltő ingereket kapnak. Ilyen kutatásokban a kísérleti kezelés következtében undort érző válaszadók attitűdjeit vetik össze a kontrollcsoportba tartozókéval.

A személyiségalapú undorérzékenység mérésének két alapvető módszere van. Az egyik a kérdőív alapú, önbevalláson nyugvó módszer, a másik a fiziológiai jellemzők mérésén alapul (Kam és Estes 2016).

A kérdőíves módszerek közül a társadalmi attitűdök vizsgálataiban gyakran használt itemsor a 25 kérdésből álló *DS-R blokk* (Olatunji et al. 2007), amely magundort, fertőzésundort és állatimaradványundort különböztet meg. Általában az első kettőt használják a társadalmi-politikai véleményekkel összefüggésben, és egyszerűen összeadják a válaszokat. Egy másik gyakran használt skála a *TDDS* (az Undor három területe-skála), amelynek 3-szor 7 iteme erkölcsi-, szexuális-, és kórokozóundort különböztet meg (Tybur et al. 2009).

A fiziológiai mérésnél tipikus esetben undorító dolgokról mutatnak képeket, és eközben például a bőr elektromos reakciójának intenzitását figyelik (Smith 2011). Egy másik eljárás az undor elkerülő magatartásra ösztönző jellegén alapul. Ebben az eljárásban szemmozgást követő módszerrel mérik, hogy mennyi időt tölt valaki különböző érzelmeket kiváltó képek nézésével (Oosterhoff 2018). Az undorító képekről az erre érzékenyek hamarabb elkapják a tekintetüket.

Az önbevallás sokkal egyszerűbb, azonban a társadalmi elvárások befolyásolhatják a válaszokat. Például, ahogy már említettem, a nők saját bevallásuk alapján érzékenyebbek az undorító dolgokra, mint a férfiak. Ugyanakkor gusztustalan dolgokra adott fizikai reakcióik nem erősebbek, mint a férfiakéi (Rohrman 2008, Smith et al. 2011).

Az undorodás létrehozásának egyik bevált módszere az érzelmhez köthető emlékek előhozása. Eszerint megkérlik az undort előhozó kí-

sérlet kezelt csoportjába tartozó válaszadókat, hogy idézzenek fel egy olyan eseményt, amely undort váltott ki belőlük. Bizonyos esetekben eközben az adott érzelmet kifejező arcot is látnak maguk előtt a képernyőn (Banks és Valentino 2012). Undor kiváltható gusztustalan képek bemutatásával is (lásd fenn). Egy speciális eljárás hipnózist használ, hogy aztán bizonyos szavak használata undort hívjon elő az emberekből (Wheatley és Haidt 2005).

## V.2. Ösztönök és társadalmi attitűdök

### V.2.1. KILÉPÉS A HÉTKÖZNAPI INTERAKCIÓK VILÁGÁBÓL: ÖSZTÖNÖK ÉS POLITIKAI IDEOLÓGIA

#### V.2.1.1. Az undor–konzervativizmus kapcsolat tézise

A morális alapok elméletének legnagyobb visszhangot kiváltó alkalmazása a morális intuíciók és a politikai ideológiák kapcsolatának vizsgálata (Smith et al. 2017). A kutatási programot megalapozó pszichológusok a kilencvenes évek óta foglalkoznak a kérdéssel. A politikatudósok figyelme az utóbbi évtizedben fordult a téma felé (Hatemí és McDermott 2012, Kam és Estes 2016, Smith et al. 2008).

A morális intuíciók és a politikai ideológiák kapcsolatával foglalkozó kutatók alaptézise, hogy az erkölcsi tisztaság megfontolásai, illetve az undor iránti érzékenység a (társadalmi) konzervativizmussal áll kapcsolatban. A konzervativizmus a hagyományokhoz ragaszkodást, a normák és a csoporthatárok fontosságát hangsúlyozza. Mindezek illeszkednek a tisztaság elveihez (Terrizzi et al. 2013). Ooserhoff és munkatársai (2018) ehhez hozzáteszik, hogy az undor, mint elkerülésre irányuló ösztönző, csökkenti az érdeklődést az undorra hajlamos emberek körében a társadalom perifériájára szoruló csoportok sorsa iránt – és ez is erősíti a kapcsolatát a konzervatív, a marginális csoportokkal kevésbé szolidáris preferenciákkal.

Az undor és ideológiák kapcsolatára vonatkozó szakirodalomban a legtöbben úgy vélik, hogy a morális intuíciók magyarázzák a politikai

ideológiákat. Azonban vannak példák az ellentétes megközelítésre is: az ideológiák a világ működésére vonatkozó hiteket jelentenek, amelyekből levezethetők a morális ítéletek (Federico et al. 2013). Ennek a fordított megközelítésnek a követői hivatkozhatnak a morális pszichológia megalapozóinak egyes vizsgálataira, ahol a regressziós modellekben magyarázóként szerepel az ideológia, és függőként a morális intuíciók (Graham et al. 2009). Azonban Haidt és kollégái nem rendelnek elméleti mechanizmusokat ehhez az elrendezéshez. Úgy tűnik, az ideológiai változók inkább kontrollként szerepelnek a magyarázó tényezők mellett.

Az undor és a politikai ideológiák kapcsolatát vizsgáló kutatások nagyobb része az egyének undorérzékenysége és a politikai nézeteik közötti összefüggést vizsgálja. Ezek ún. megfigyeléses vizsgálatok, amelyek technikailag csak korrelációkat tudnak megállapítani a vizsgált változók között. Azonban e vizsgálatok többsége feltételezi, hogy az alapvető morális intuíciók hatnak a politikai ideológiákra (Inbar et al. 2009).

#### V.2.1.2. A tézis tesztjei megfigyeléses vizsgálatokban

Az első, jelentős visszhangot kiváltó politikatudományi vizsgálat Inbar és munkatársainak (2009) ilyen irányú elemzése (melyre eddig több mint 400 tanulmány hivatkozott a világban). Vizsgálatuk két kismintás felmérésen nyugszik (egy 181 fős, nem reprezentatív online minta, és egy 87 fős diákminta, mindkettő az Egyesült Államokból). Az undorérzékenység mérésére a Haidt és szerzőtársai által ajánlott klasszikus *DSS skálát* használták. Az első vizsgálatukban az egyének liberális-konzervatív önidentifikációját magyarázták az undorérzékenységgel. A második felmérésben egy előteszt segítségével kiválasztott, az USA-ban akkor politikailag relevánsnak tartott témákról kérdezték a diákokat (abortusz, meglegházasság, szakszervezetek, fegyvertartás, Irán, Irak, segélyek, adózás, pozitív diszkrimináció, halálbüntetés). Mindkét vizsgálatban szignifikánsnak bizonyult az undor politikai nézetekre gyakorolt hatása (Inbar et al. 2009): az

undorra hajlamosabbak konzervatívabb nézeteket vallottak (demográfiai tényezők mellett a vallásosságra is kontrollálva), mint az undorra kevésbé hajlamosak. A konkrét témák esetében erősebb, mint az általános világnézet tekintetében (más-más mintán tesztelték e kettőt!). A témák közül, az elméleti feltevéseknek megfelelően, a szexuális morálhoz köthető abortusz és melegházasság témájában volt erős és szignifikáns a kapcsolat. Ugyanakkor az adózás kérdésében is szignifikáns volt a hatás: az undorra érzékenyebbek a konzervatívok körében népszerű adópolitikát támogatták inkább. A többi területen nem.

Tybur és szerzőtársai (2010) megpróbálták némileg nagyobb mintán ellenőrizni Inbar és szerzőtársai (2009) eredményeit. Ők kizárólag diákmintákkal dolgoztak, de három felmérésükben összesen 1100 egyetemi hallgató vett részt. A legfontosabb kritikájuk Inbar és szerzőtársai kutatásával szemben az volt, hogy a felmérésben használt, Haidt és szerzőtársai (1994) tanulmányában ajánlott undorskála nem eléggé megalapozott a viselkedési immunrendszer elmélete szempontjából. Direkt módon tartalmaz morális ítéleteket (például a szexuális viselkedésről), ezeket kapcsolja össze hasonló morális ítéletekkel (például az abortuszéval). Ha sarkítani akarjuk, a problémájuk az, hogy a később klasszikus hivatkozássá vált Inbar és szerzőtársai (2009) kutatásban előbb megkérdezték, hogy a válaszadó undorítósnak tartja-e a deviáns szexuális magatartást, majd ezt összevetették azzal, hogy mit gondol a melegházasságról. A klasszikus undorérzékenység-skála helyett ezért a más csoportosítást alkalmazó *TDDS skálát* használtak. Ez explicite elkülöníti a szexuális és az erkölcsi undort a kórokozóundortól. A viselkedési immunrendszer modelljében az undor ez utóbbi típusra vonatkozik. Tybur és szerzőtársai (2010) ezt tartják az elmélet igazi próbájának. Az elmélet igazi tesztjének szánt első felmérésükben cáfolták Inbar és szerzőtársai (2009) konklúzióját: A kórokozóundor-érzékenységnek a mintájukban gyakorlatilag nulla hatása volt a liberális-konzervatív önidentifikációra. A szexuális undorral viszont szignifikáns kapcsolatban volt az ideológiai alapállás (megerősítve tézisüket arról, hogy miért található szignifikáns összefüggéseket Inbar csapata).

Második felmérésükben megpróbálták megismételni Inbar és szerzőtársai (2009) elemzését annyi eltéréssel, hogy az Inbarék által használt klasszikus *DS skála* helyett annak a szakirodalomban – a fenti problémák miatt javasolt – redukált változatát (DS-R, Olatunji et al. 2007) használták, amelyben nem szerepelnek direkt szexuális kérdések. E módosított skálával már nem sikerült replikálni Inbar és szerzőtársai (2009) pozitív eredményeit: így sem találtak kapcsolatot az undorérzékenység és a konzervativizmus között. Harmadik felmérésükben a függő változót próbálták felbontani a konzervativizmussal korábban stabilan összefüggésbe hozott érték-, illetve beállítódáskonstrukciókkal (társadalmi dominancia orientáció, jobboldali tekintélyelvűség és vallási fundamentalizmus). E skálák közül a kórokozóundor csak a társadalmi dominancia orientációval volt szignifikáns összefüggésben.

Inbar és szerzőtársai (2012) nem sokkal később nagyobb mintán is megismélték korábbi vizsgálataikat. A morális alapok elméletének mérésére létrehozott YourMoral.org oldalon a látogatók által kitöltött mintegy huszonötezer, USA-ból beérkezett kérdőívet elemezték. Ezen a felületen az undorérzékenységet a – Tybur és szerzőtársai (2010) által is replikációs céllal használt – *DS-R kérdés-blokkal* mérték (amely Tybur és szerzőtársai szerint nem jó mérőeszköz, de legalább a szexuális undor kérdései kikerültek belőle), és ezt vetették össze az ideológiai identifikációval (konzervatív-liberális-skála). Eredményeik szerint az undorérzékenység a konzervatív irányba tolja el az ideológiai beállítottságot (Inbar et al. 2012). A kontrollváltozók hatásaival összevetve az undorérzékenység szerepe viszonylag jelentős mértékű (Inbar et al. 2012, 2. táblázat). Vizsgálatukat megismélték az oldalra érkezett mintegy ötezer külföldi kitöltő adatainak elemzésével. Hasonló eredményeket kaptak, és a világ minden régiójára kimutatták (a többnyire képzett, a kérdőívre önállóan rátaláló válaszadók körében), hogy az általuk mért undorérzékenység kapcsolatban van az ideológiai identifikációval. Inbar és szerzőtársai (2012) külön hangsúlyozzák, hogy az általuk mért fertőzésundor áll a legerősebb kapcsolatban az ideológiával, és ez a viselkedési immunrendszer elméletének megfelelő eredmény.

Ezekben az években több más kutatás próbálta vizsgálni az undorérzékenység és az ideológiai beállítottság közötti kapcsolatot. Terrizzi és szerzőtársai (2013) foglalták össze a vizsgálatok első hullámában született 24 tanulmány tanulságait. Metaelemzésük szerint az undorérzékenység empirikusan összefügg a tekintélyelvűséggel, a társadalmi dominanciaorientációval, a vallási fundamentalizmussal, a kollektívizmussal és a konzervatívizmussal. Az utóbbi terminusnál politikai- vagy társadalmi jelzővel illetik azokat a preferenciákat, amelyek nem a gazdasági preferenciákra vonatkoznak. Az eredmények ezekre a nem gazdasági attitűdökre vonatkoznak. Azonban Terrizzi és munkatársai (2013) elemzése nem kezelte azokat a problémákat, melyekre Tybur és szerzőtársai (2010) hívta fel a figyelmet. Ráadásul metaelemzésük arra utal, hogy az ideológiai preferenciák varianciájának alig tizedét magyarázza a különféleképpen mért undorérzékenység.

Mindeközben Tybur és szerzőtársai (2016) is kísérletet tettek arra, hogy nagymintás elemzésben ismételjék meg saját vizsgálataikat. Nem reprezentatív, ellenben nagy és diverz lakossági mintákon, harminc országban végeztek online felméréseket (Tybur et al. 2016). A kiindulópontjuk továbbra is a viselkedési immunrendszer elmélete, és a kórokozók szembeni averzió szerepe volt. Ezt tekintették az evolúciós elméleti alapok és az oksági viszonyok tekintetében megfelelőnek. Első körben országszintű elemzéseket végeztek, ahol kérdőíves adatok helyett földrajzi-biológiai adatok alapján definiálták egy-egy ország kórokozó fenyegetettségének a mértékét. Ezeknek az adatoknak a függvényében vizsgálták a kérdőívekben mért átlagos ideológiai beállítottságot. Azt találták, hogy a kórokozók által jobban fenyegetett országokban népszerűbbek a tradicionalista értékek (de a fenyegetettség nem magyarázza a társadalmi dominanciaorientáció vagy az undorérzékenység átlagos mértékét). Bár a szerzők konklúziója az elmélettel kapcsolatban alapvetően pozitív volt, és a mintegy tízezer válaszadó átlagában egyéni szinten is találtak összefüggést a tradicionalizmus és (a hagyományosan mért) undorérzékenység között, de ez viszonylag kicsi, és jelentős a szórás a nemzetek között. Elsősorban egyes skandináv, továbbá izraeli adatok húzták



fel az átlagot. Indiától kezdve Dánián keresztül Brazíliáig jónéhány országban nullához közeli az undorérzékenység és tradicionalizmus közötti összefüggés (Tybur et al. 2016, 1. táblázat), Lengyelországban pedig a várttal ellentétes irányú a kapcsolat.

2017-ig kellett várni, hogy az első olyan kutatás eredményei megjelenjenek, amely nemcsak nagy, de ezzel együtt országosan reprezentatív minták adataiból dolgozott (Aarøe 2017, Dánia és USA). Ez a kutatás azonban egy területre, a bevándorlásra koncentrált, itt mutatott ki kapcsolatot az undorérzékenységgel.

Smith és szerzőtársai (2011) az eredeti feltevésekkel, de más mérési paradigmában végezték kutatásukat. Az undorérzékenység mérésének problémáját úgy oldották meg, hogy egy innovatív dizájn keretében a hagyományos kérdőíves módszer mellett a válaszadók galváns bőrreakcióját is mérték az undorítóknak tekintett képek mutatók. A reakció intenzitásának mértéke határozta meg az undorérzékenység mértékét (a képeket a felmérés előtt egy külön vizsgálatban tesztelték). A speciális dizájn hátulütője, hogy csak kisebb elemszámon tudták felmérésüket elvégezni. Összesen 46 válaszadóról szereztek érvényes adatokat. Kismintás vizsgálatuk Inbar és szerzőtársai (2009) kutatásához hasonlóan a feltevéseknek megfelelő pozitív eredményt hozott: a fiziológiai undor reakciók intenzitása szignifikáns mértékben növelte annak esélyét, hogy valaki konzervatívnak vallja magát. Konkrét policy-témák közül (ahogy más vizsgálatokban is) a melegházasság kérdésében mutatkozott szignifikáns kapcsolat. Eredményeiknek az undorra vonatkozó kérdésekkel kapcsolatos fontos tanulsága, hogy a fiziológiai undorérzékenység és az ideológiai beállítottság közötti kapcsolat a kérdőíves undorérzékenység mértékére kontrollálva is szignifikáns maradt (Smith et al. 2011). Ahn és szerzőtársai (2014) követték Smith et al. (2011) megközelítését, és fizikai reakciók mérésével definiálták az undorérzékenységet. Ők az agyi aktivitást figyelték fMRI vizsgálattal, és szintén pozitív összefüggést találtak undorérzékenység és konzervatív beállítottság között.

#### V.2.1.4. Kísérleti dizájnok

Bár igazán minőségi reprezentatív mintákon még nem végeztek átfogó elemzéseket, az eddigi eredmények – korlátaik ellenére is – arra utalnak, hogy az undorérzékenység bizonyos mértékei többé-kevésbé korrelálnak a konzervativizmus bizonyos indikátoraival. Azonban az is látható, hogy ahol az undorérzékenység indikátorai közelebb állnak az elméleti megfontolások elvárásaihoz, ott ez a kapcsolat bizonytalanabb. A hagyományos *DS skála* esetében ráadásul nehezen zárható ki a látszólagos okság gyanúja. Ezekre a problémákra az kínálhat megoldást, ha kísérleti úton indukálunk undort, és ennek az állapotnak a véleményekre gyakorolt hatását vizsgáljuk (Petrescu és Parkinson 2014).

Az undor és a morális ítéletek kapcsolatának korábbi vizsgálataira építve Petrescu és Parkinson (2014) azt elemezte, hogy miként hat a kísérleti kezelésben indukált undor a politikai attitűdökre. Kísérletükben undort keltő képekkel befolyásolták az emberek érzéseit (és más jellegű negatív képekkel kontrolláltak a negatív érzelmek általános hatását, amelyet egyesek szintén összekapcsolnak a konzervativizmussal). Három kismintás (néhány tucat oxfordi hallgatót magában foglaló) felmérésük nem mutatott kapcsolatot a kísérletileg létrehozott undor, és a társadalmi-politikai konzervatív önidentifikáció között.

A kísérleti kezelés megoldja az oksági viszonyok azonosításának problémáját, azonban nem tudjuk, hogy elég erős volt-e az undort keltő inger ahhoz, hogy véleményváltozást idézzon elő. Az előbbi kutatással nagyjából egy időben Adams és szerzőtársai (2014) is kísérletet tettek az undorérzet felkeltésére – egy igazán innovatív (és valóban undorító) kezelést találtak ki annak érdekében, hogy valóban jelentősen növeljék az undorérzetet: kutatók a kezelt csoport érkezése előtt undorító szaggal telítették a laboratóriumot, ahol az alanyok a kérdőívet kitöltötték. A kontroll- és kezelt csoportban összesen 57 kitöltőjük volt, sokféle policy-kérdésben. Szignifikáns különbséget a kezelt és kontrollcsoport között a szexuális normákkal és jogokkal kapcsolatban találtak, máshol nem (Adams et al. 2014).

Ez a módszer is problematikusnak tűnhet, azonban látni kell, hogy más érzellemmel, például a dühvel kapcsolatban (megfelelő megszorításokkal) sikerült az elméleti feltevéseknek megfelelő összefüggéseket találni (Banks 2014: a feldühített amerikai liberális fehérek a kontrollcsoporthoz képest nagyobb eséllyel támogatták az egészségügy szegények számára kedvező reformját).

#### V.2.2. VISSZAVONULÁS A VÉDELMI VONALAK MÖGÉ: UNDOR ÉS ETNIKAI DISZKRIMINÁCIÓ

Nem csupán az undorérzékenység mérése, és az undor kiváltásának módja okozhat problémát. A konzervativizmus indikátorai is kritizálhatók elméleti szempontból (Terrizzi et al. 2013). Felvethető például, hogy a konzervativizmust nem feltétlenül az általános önidentifikációval kell megragadni. Látható, hogy a különböző vizsgálatok átlagában, a konzervativizmuson belül a szexuális morál kérdéseivel áll legstabilabb kapcsolatban az undor. A morális alapok elmélete és a viselkedési immunrendszer elmélete azonban ennél sokkal szélesebben értelmezi az undorérzékeny társadalmi-politikai kérdések körét. Ugyanakkor nem feltétlenül olyan szélesen, hogy abba a konzervativizmus általában beletartozzon. Az eddig bemutatott vizsgálatok többségében, ahol mérték a gazdasági konzervativizmust, az – ritka kivételektől eltekintve – nem mutatott összefüggést az undor különböző indikátoraival.

Az egyik ilyen, az undor összefüggésében mostanában is sokak által vizsgált speciális társadalmi-politikai kérdés a bevándorlók és az őshonos kisebbségek iránti attitűdök kérdése. Az amerikai szociológiában és politikatudományban régóta okoz fejtörést az, hogy míg a rasszizmus indikátorai az idegenellenesség folyamatos csökkenését mutatják az elmúlt fél évszázad folyamán, addig nem növekedett ezzel párhuzamosan a feketék és más etnikai kisebbségek társadalmi integrációját segítő politikák támogatottsága (Banks és Valentino 2012).

A viselkedési immunrendszer elméletének politikatudományi követői szerint a tartós idegenellenesség egyik gyökere abban rejlik,

hogy a statisztikai értelemben vett normális kinézettől való mindenfajta eltérést betegségre utalhat, és ez undort, vele pedig elkerülő magatartást implikálhat (Clifford és Piston 2017, Petersen 2017). Az undorra hajlamos emberek a sötétebb bőrszínt tudattalanul fertőző betegségek jelének láthatják (Aarøe et al. 2017). Ne feledkezzünk meg arról a széles körben dokumentált tényről se, hogy az uralkodó rasszista diskurzus bizonyos kisebbségeket – például az USA-ban a feketéket –, sokáig a kosszal és a fertőzésveszéllyel kötötte össze (Aarøe et al. 2017).

A politikai pszichológia ezt a tézist tovább finomította az elmúlt évtizedben. Újabban a mássággal kapcsolatos undor két dimenzióját különítik el: az egyik a más csoportoktól való undoralapú idegenkedés, a másik a csoporton belüli viselkedési-alapú deviancia elutasítása (Terrizzi et al. 2013, Tybur et al. 2016). Az első a társadalmi dominanciaorientációval hozzák összefüggésbe, a másodikat pedig a tekintélyelvűséggel. A csoportközi undor a máshogy kinéző emberekkel való érintkezés kerülésére ösztönöz, a csoporton belüli undor a deviáns viselkedést utasítja el. Mindkét beállítódás pozitívan korrelál a konzervativizmussal, és mindkét beállítódásnak lehet kapcsolata a bevándorlók iránti attitűdökkel – csak másképpen. A tekintélyelvű polgárok akkor ítélik el a bevándorlókat, amikor úgy érzik, hogy a bevándorlók nem kívánnak beilleszkedni a társadalomba. A csoportközi undoron alapuló társadalmi dominanciabeállítódással jellemezhető polgárok akkor is elutasíthatják a bevándorlókat, ha azok asszimilálódni akarnak. Sőt, az elmélet alapján épp specifikusan ebben az esetben lesznek elutasítóak, ugyanis az érintkezést akarják elkerülni velük, mert tudat alatt fertőzésforrásnak tekintik őket (Aarøe 2017), tehát a gettósodás számukra jobban megfelelő kisebbségi stratégia (Tybur et al. 2016). Ez utóbbi mechanizmus miatt pesszimista néhány jeles politikatudós a bevándorlók integrációját elősegítő erőfeszítések potenciális sikerével kapcsolatban (Aarøe 2017).

Az idegenellenesség vizsgálata első látásra megfelelő szűkebb fókuszra tűnik a viselkedési immunrendszer és a politikai attitűdök összefüggéseinek zavaró tényezőktől mentes feltárására. Ugyanakkor

látni kell, hogy a csoportközi viszonyok érzelmi szempontú elemzései már korábban felhívták a figyelmet a düh szerepére, és ezzel az érzellemmel kapcsolatban már születtek az undoréhoz hasonló eredmények az érzelmek és a politikai attitűdök (konkrétan az idegenellenesség) tekintetében. Néhány kutató például azt feltételezi, hogy a tekintélyelvűség nemcsak a csoporton belüli viselkedésre vonatkozó preferenciákkal áll összefüggésben, hanem a más csoportokkal szembeni düh is növeli ezt a beállítódást (Terrizzi et al. 2013). A társadalmi domancia orientáció ebben a közelítésben kifejezetten a marginalizált, és nem fenyegető csoportokkal szembeni hozzáállást befolyásolja. Vannak olyan amerikai társadalomkutatók, akik úgy gondolják, hogy míg az utóbbi évtizedekben az USA-ban visszaszoruló klasszikus rasszizmus valóban undoralapú volt, addig a mai szimbolikus rasszizmus (Kinder és Sanders 1996) inkább a társadalmi szabályokat állítólag be nem tartó kisebbségekkel szembeni dühből táplálkozik (Banks és Valentino 2012). Az idegenellenesség motiválója az a hiedelem, hogy az egyes kisebbségek tagjai nem igyekeznek eléggé beilleszkedni, és emiatt dühösek rájuk, miközben nem undorodnak tőlük úgy, ahogyan korábban sokan tették. Banks és Valentino (2012) azt is megmutatja, hogy a legkevésbé idegenellenes, az amerikai kisebbségekhez leginkább pozitívan viszonyuló fehéreknek a dühe a többségi fehér társadalom ellen irányul, mivel (spekulálnak a kutatók) ők nem tettek meg mindent a marginalizált kisebbségek integrációjáért. Ha a düh szerepét valahogy el is különítettük, nem oldottuk meg az undor tiszta elemzésének problémáját. Az idegenekkel kapcsolatban a félelem legalább annyira releváns érzelem lehet, mint az undor és a düh (Banks és Valentino 2012).

Mindezeket a zavaró tényezőket észben tartva lehet értékelni azokat az újabb vizsgálatokat, amelyek az undor és a bevándorlók iránti attitűdök kapcsolatára fókuszálnak. A korábbi vizsgálatok eredményei arra utalnak, hogy van érdemi kapcsolat az undorérzékenység és az idegenellenesség között, azonban Aarøe, Petersen és Arcenau (2017) közelmúltbeli vizsgálata az első, amely ebben a témakörben nagy és reprezentatív mintákon végzett elemzéseket. Négy felmérésben összesen több mint négyezer embert kérdeztek meg Dániában

és az Egyesült Államokban. A reprezentatív mintán végzett kérdőíves felméréseket egy kis elemszámú, dániai laboratóriumi vizsgálat egészíti ki. Többféle kérdéssel mérték az undorérzékenységet, és a kismintás vizsgálatban ezt kiegészítették fiziológiai reakciók mérésével is. Továbbá a fertőzésveszély/undor erejét is manipulálták, és vignetták segítségével mérni tudták a különböző típusú bevándorlók iránti preferenciákat is. A nagyobb mintáknak köszönhetően a politikai beállítottság moderáló hatását is vizsgálni tudták (azaz a politikai identitás itt nem volt függő változó). A vignetták mellett attitűdöket is mértek a bevándorlókkal való érintkezés témájában. Ez eddig a legátfogóbb vizsgálat a viselkedési immunrendszer és a politikai attitűdök témájában. Az eredményeik megerősítik a viselkedési immunrendszer elméletére alapozott várakozásokat. Ráadásul az undorérzékenység hatását jelentősnek értékelik, amely összemérhető az iskolázottság és a jövedelem hatásával. Bár az első felméréseikben a jobban kritizálható hagyományos undorérzékenység skálákat használták (Haidt et al. 1994, Inbar et al. 2009), utolsó nagymintás (bár nem reprezentatív) felmérésükben a Tybur és munkatársai (2010) által javasolt, elméletileg megalapozottabb itemeket rakták a kérdőívbe – hasonló eredménnyel (Aarøe et al. 2017, 2. táblázat). A vinyettás vizsgálat eredményei szerint a kulturálisan nagyobb távolságra lévő migráns elutasításában nagyobb szerepet játszik az undorérzékenység, azonban a bevándorló integrációs törekvései nem befolyásolják az undorérzékenység hatását. Az elméleti várakozásaikkal összhangban az undorérzékenység szerepét csökkenteni tudták egy olyan kísérleti kezeléssel, amely csökkentette a kórokozófenyegetettség (kérdőívben mesterségesen megnövelt) érzetét a válaszadók körében. Figyelemre méltó az is, hogy baloldaliak körében erősebb az undorérzékenység hatása: a jobboldali gondolkodásúak amúgy is ellenzik a bevándorlást, a baloldaliak körében az undorra kevésbé hajlamosaknak kevesebb problémájuk van a migránsokkal (Aarøe et al. 2017).

Az undorra való hajlam általános konzervatív beállítottságot erősítő hatásának korábban ismertetett, némileg bizonytalan eredményeihez képest Aarøe, Petersen és Acrenaux (2017) eredményei kétségtelenül robusztusnak és meggyőzőnek tűnnek. A vinyettás

vizsgálatok eredményei (nevezetesen, hogy a bevándorlók eltérő kulturális háttere felerősíti az undor szerepét a bevándorlás megítélésében, azonban a bevándorlók integrációs törekvései nem), a szerzők szerint arra utalnak, hogy az undor elsősorban az idegenektől való ösztönös fertőzési félelem felerősítésében játszik szerepet. Azonban Tybur és szerzőtársai (2016) szintén új, nemzetközi, nagymintás, bár nem reprezentatív felmérései ennek ellentmondanak. A korábban már említett, harminc országra kiterjedő, mintegy tízezer embert elérő elemzésükben ők is megvizsgálták – bár más módszerekkel – az undor csoporton belüli és csoportok közötti viszonyok megítélésére gyakorolt hatását. Eredményeik pedig Aarøe és szerzőtársai (2017) konklúziójával szemben azt mutatják, hogy a (szerintük) csoporton belüli normákhoz kapcsolódó tradicionalizmus függ össze az undorérzékenységgel, míg a csoportközi hierarchiához kötődő társadalmi dominanciaorientáció nem (Tybur et al. 2016).

Petersen (2017) egy kevésbé nagyszabású (bár közel kétezer válaszadót megmozgató), ám igen innovatív kutatásában szintén azt a kérdést feszegeti, hogy miért félünk az idegenektől, milyen mechanizmuson keresztül befolyásolja ezt az undor. A neves dán kutató a „Ki mit mondott?” módszert alkalmazta annak vizsgálatára, hogy miként kategorizálunk különböző embereket. Az eljárás során először nyolc portrét mutatott a kérdezeteknek, amelyek mindegyikéhez tartozott egy-egy állítás a fotón szereplő személytől. Egy kizökkentő feladat után a válaszadóknak a portrékat és állításokat kellett párosítaniuk („Ki mit mondott?”). Az elemzés során a hibákat vizsgálta: kinek az állításait keverik össze kiével? Például ha egy ilyen kérdésben a hajszínt és a nemet manipuláljuk, megvizsgálhatjuk, hogy melyiket tartják (tudattalanul) fontosabb ismérvnek az emberek: ha a nemi hovatartozás a fontosabb, akkor gyakrabban keverik a szőkék állításait a barnáékkal, mint a fiúkét a lányokéval. A (Dániában végzett) kísérlet kontrollcsoportjában négy egészséges fehér, és négy egészséges kelet-indiai ember portréját mutatta meg. A kezelt csoportban a négy kelet-indiai mellett négy ragyás arcú fehér embert mutatott (az egészségesek photoshoppolt változatát). A bőrbetegséget implikáló portrék bemutatásakor megnőtt az esélye a különböző bőrszínű em-

berek összekeverésének: az egészséges kelet-indiaiak jónéhány válszadónál összemósódtak a beteg fehérekkel (Petersen 2017). Petersen (2017) konklúziója, hogy az emberek a fizikai másságra érzékenyek általában, de ez nem jelent az idegenekhez köthető kórokozók miatti célzott etnikai érzékenységet. Az idegenekkel szembeni tartózkodás csak „nem szándékolt” mellékkövetkezménye a betegségtől való félelemnek. Szerintük az idegeneket nem idegen kórokozók hordozóinak tartják, hanem egyszerűen – a fura kinézetük miatt – önkéntelenül betegnek nézik őket. A Petersen által feltárt mechanizmus közelebb áll a csoporton belüli deviancia miatti undor téziséhez (lásd Tybur et al. 2016), azonban megmagyarázza azt is, hogy miért érzékenyek az etnikai másságra az emberek (lásd Aarøe et al. 2017).

A korábbi kutatások megmutatták, hogy nem csak az undor válthat ki ellenérzést az idegenekkel és a marginális kisebbségekkel szemben. Néhány közelmúltbeli vizsgálat rámutat arra is, hogy az undor más mechanizmusokon keresztül is befolyásolhatja a diszkriminációs preferenciákat. Clifford és Piston (2017) hajléktalanokkal kapcsolatos intézkedések iránti attitűdök vizsgálatával mutatja ki (összesen közel ezeröttszáz fős, részben reprezentatív mintákon), hogy az undor elsősorban a probléma láthatatlanná tételét segítő politikák támogatását növeli – függetlenül a hajléktalanokat segítő politikák preferálásától. Az undorérzékeny emberek közül sokan egyszerre támogatják, hogy adjanak több segítséget a hajléktalanoknak, és azt, hogy tiltsák ki őket a közterekről. Clifford és Jerit (2018) pedig nemrég megjelent elemzésükben azt mutatják meg, hogy az undor a problémáktól való elfordulást ösztönzi oly módon is, hogy egy veszélyhelyzetben csökkenti az új információk iránti érdeklődést. Plauzibilis feltevés az eredmények alapján, hogy az ilyen-olyan okból undorítóknak tekintett csoportoktól való elfordulás csökkenti a problémáik, szükségleteik iránti ismereteket, és vele az empátiát is. Az undor így nem direkt módon befolyásolja az idegenellenességet, hanem a tudatlanságon keresztül.



### V.3. Nyitva maradt kérdés: az ösztön a végső mozgató?

#### V.3.1. MOST AKKOR JOBBRA VAGY BALRA?

##### V.3.1.1. Undor a baloldalon

Az erkölcspszichológia Haidt-féle elméletének népszerűsége egyelőre töretlen. Ugyanakkor folyamatosan, és a legutóbbi időben egyre gyakrabban érkeznek olyan jól megalapozott kutatási eredmények, amelyek alapjaiban kérdőjelezik meg az elmélet által felállított ok-sági lánc érvényességét. Az alfejezet végén (és a VII. fejezet zárásaként) szó lesz arról is, hogy ezek a friss eredmények, és a nyomukban kirajzolódó mechanizmusok amúgy kongruensek a pszichológia főáramának modern elméleteivel.

Első lépésként, az undor és konkrét policy preferenciák kapcsolatának vizsgálata segít elszakadni attól az általánosítástól, hogy az undor csak jobboldali, konzervatív politikai nézetekkel függhet össze, azaz mindenképpen ilyen nézeteket implikál. Már említettem Clifford és Piston (2017) közelmúltbeli vizsgálatát, amelyben a hajléktalanokkal kapcsolatos intézkedések támogatottságát vizsgálták. Kutatásukban el tudták választani a viselkedési immunrendszer elmélete alapján várható elkerülő attitűdöt a segítségnyújtásra való hajlandóságtól. Az undorérzékenység vizsgálatukban növelte az olyan politikák támogatottságát, amelyek a hajléktalanok közterekről való kizárását célozzák. Nem volt összefüggésben azonban a hajléktalanok általános megítélésével, és a számukra nyújtandó állami/önkormányzati segítség támogatásával.

Konkrét politikák és ügyek esetében már korábban kimutatták, hogy az ügyek mentén baloldalinak minősülő polgárok az átlagosnál érzékenyebbek az undort keltő ingerekre. Ugyanez igaz az állatok jogaiért küzdő aktivistákra (Herzog és Golden 2009), és az erkölcsi alapú vegetáriánusokra is (Rozin et al. 1997).

Petrescu és Parkinson (2014) pár évvel ezelőtti vizsgálatukban már ennél tovább mentek: kifejezetten általános baloldali prefe-

renciákkal próbálták összefüggésbe hozni az undorra való hajlamosságot. Elvi szinten ez már korábban felmerült, hiszen az általunk is emlegetett *TDDS* undorérzékenység-skála tartalmaz becsületes-ségre és reciprocitásra vonatkozó itemeket, amelyek a morális alapok elmélete szerint liberális értékek is számítanak (Tybur et al. 2010). Petrescu és Parkinson (2014) azonban az undor általános morális ítéletekre gyakorolt hatását vizsgáló egyes kísérleteket tekintette kiindulópontnak. Nevezetesen azokat, amelyek a pénzügyi elosztás méltánytalanságával szemben érzett undorra vonatkoztak (például Chapman et al. 2009). Ha ezek az eredmények megállják a helyüket – érvelnek a szerzők –, akkor a modern társadalmakban létrejövő jövedelemeloszlási igazságtalanságokkal szemben is érzékenyebbnak kell lenniük azoknak, akik undor-érzékenyebbek. A jövedelmi igazságosság illetve egyenlőség pedig klasszikus baloldali téma. Hipotézisük teszteléséhez Evans és szerzőtársai (1996) itemsorát használták, amelyet a baloldali gazdasági nézetek mérésére dolgoztak ki (ilyen kérdésekkel, mint például: „a kormány feladata a jövedelmek újraelosztása”). Az általuk hivatkozott irodalom a kísérletileg indukált undor erkölcsi ítéletekre gyakorolt hatását vizsgálta, így ők is ezt csinálták: undort keltő képeket mutattak a kezelt csoport tagjainak. Kísletszámú diákmintákon (összesen alig kétszáz embert kérdeztek meg) végzett felmérésük eredményei szerint az undort keltő képekkel szembesülő diákok némileg baloldaltlibb gazdasági nézeteket fejtettek ki (Petrescu és Parkinson 2014). Azonban a hatás mértéke csekély volt. A kicsi statisztikai erő, és a kis hatás mellett visszafogottságra int az, hogy az oxfordi diákmintájuk erősen baloldali volt, és az általuk hivatkozott szakirodalom azt hangsúlyozza, hogy az undor felpiszkálása általában szigorítja a morális ítéleteket (és erre hozzák példának az elosztási méltányosság kísérletekben jól bejáratott játszmat). A szerzők is elismerik, hogy egy baloldali mintában ez az általános szigorodás is vezethet az újraelosztási preferenciák erősödéséhez, és a kis szórású, kis létszámú mintában az ideológiát nem lehet kontrollálni (Petrescu és Parkinson 2014). Az empirikusan gyenge lábakon álló eredmények nem vonnak le a megközelítés értékéből, és ez az irány további

vizsgálatokra érdemes. Főleg, mivel az undor és a konzervativizmus kapcsolat tézise, és a tézist támogató eredmények a társadalmi-politikai konzervativizmusra vonatkoznak, és nem a gazdasági konzervativizmusra (amelyhez például a piaci intézmények és folyamatok tiszteletének értéke sorolható); míg Petrescuék kutatása a gazdasági baloldaliságra irányul. Jelen áttekintés írásáig azonban nagyobb és diverz mintán még nem ismételték meg ezt kísérletet.

Kam és Estes (2016) átfogóan próbálja meg tisztázni, a konkrét szakpolitikák iránti attitűdök szintjén, hogy mely területeket érintheti egyáltalán, és – ebből következően – mely területeken járhat inkább jobb-, és mely területeken inkább baloldali preferenciákkal az undor, illetve az undorra való hajlamosság. A viselkedési immunrendszer logikájából levezethetően, a fertőzésvédelemmel hozzák összefüggésbe az undort. Kutatásukban a már korábban azonosított területek (bevándorlás, melegházasság, abortusz) mellett az ételminőség-biztonság területén is találtak (a pártpreferenciák és az ideológiák befolyásához hasonló mértékű, azokra kontrollálva is létező) hatásokat. Más policy-területeken nem találtak összefüggést undor és társadalmpolitikai attitűdök között.

Tehát a kilencvenes években felvetett klasszikus hipotézis, miszerint az undor a konzervatív ideológiához kötődik, átalakult. Ma már inkább azt feltételezik a téma kutatói, hogy az óvó-védő (protective) politikák támogatásában van az undornak szerepe (Aarøe 2017).

A morális alapok elmélete szerint az undor az erkölcsi tisztaságon nyugvó ítéleteket megalapozó érzelm. A MAE mérésére kidolgozott kérdőívek megfelelő itemjeinek egy része kifejezetten reflektál az undorra. Ugyanakkor egyes, az undor attitűdökre gyakorolt hatását fontosnak tartó társadalmtudósok úgy gondolják, hogy valójában más erkölcsi területekhez is kapcsolódhat undor jellegű érzelm (Cameron et al. 2015). A MAE feltételezéseivel szemben álló, legerőteljesebb alternatív hipotézis szerint minden terület erkölcsi ítéleteit egyforma mértékben befolyásolhatja az undor. Egy mérsékelt alternatív hipotézis szerint az undor nemcsak a tisztasághoz, hanem általában a csoportkötődést erősítő erkölcsi területekhez kapcsolódik (például a hűséghez és a tekintélytisztelethez). A MAE szigorú fel-

tevéseit kissé lazító feltevés szerint az undorérzékenység szignifikánsan erősebben kapcsolódik a tisztaság megfontolásaihoz, mint a többi megfontoláshoz (Wagemans et al. 2018). Wagemans és szerzőtársai (2018) öt kisebb és közepes méretű, nem reprezentatív mintán végzett vizsgálatukban megmutatták, hogy az undorra való hajlandóság valamivel erősebben kötődik az erkölcsi tisztaság megfontolásaihoz, mint a MAE által feltételezett többi erkölcsi dimenzióhoz. De érdeemi eltérés mutatkozott a holland (többnyire diákokból álló) minták, és az utolsó felmérésükben használt amerikai, nem reprezentatív online minta adatai között: a hollandok körében a tisztaságon kívüli erkölcsi aspektusok gyengébben kötődtek az undorhoz, mint az amerikaiaknál. Adataik alapján a csoportkötődéshez tartozó hűség és tekintélytisztelet szubjektív fontossága nem kötődik jobban az egyéni undorérzékenységhez mint az individualizálónak tekintett gondoskodás jelentősége.

#### V.3.1.2. Fizikai undor és erkölcs: létezik-e egyáltalán a kapcsolat?

A morális pszichológia és az azt alkalmazó politikai tudományok alapfeltevése, hogy az emberek nemcsak fertőzésveszélyes objektumok közelében, de társadalmi erkölcsi szabályok áthágását látva is érezhetnek undort (Kam és Estes 2016, Wagemans et al. 2018). A kétezres években ezzel kapcsolatban születtek a tézist megerősítő kutatási eredmények. Sanfey és szerzőtársai (2003) neurológiai vizsgálata azt mutatta, hogy az előző fejezetben részletesen tárgyalt ultimátumjátékban méltánytalan ajánlatot kapó alanyok agyának azon része aktiválódik, amely testi undor esetén is. A kérdésben a legtöbbet (több mint kilencszázszor) hivatkozott vizsgálat Wheatley és Haidt (2005) kísérlete, amelyben csekély számú kísérleti alanyaik egy részét hipnózissal tették érzékennyé arra, hogy bizonyos, egyébként semleges szavak említése esetén undort érezzenek (tehát undorrállapotot mértek). Azt vizsgálták, hogy az így keltett undor miként befolyásolja bizonyos viselkedések megítélését. Ezek között voltak

erkölcsileg semlegesnek tekintett cselekmények, továbbá különféle szabályok áthágását jelentő aktusok (vérfertőzés, saját kutya megévése, megvesztegetés, lopás). Konklúziójuk szerint az undort érző válaszadók élesebben ítélték meg a különböző erkölcsi szabályok áthágását. Ugyanakkor adataik azt mutatják, hogy a vérfertőzés és a megvesztegetés kivételével az undornak csekély hatása van az erkölcsi ítéletekre (ezekre is csak az egyik felmérésükben a kettőből). Ráadásul második felmérésükben az undorérzet erkölcsi szabályszegésekre gyakorolt gyenge hatása párosult a nem erkölcsi cselekmények megítélésére gyakorolt szignifikáns hatással. A vizsgálat a kis elemszám, és a bizonytalan eredmények ellenére elsődleges hivatkozási forrásá vált az undor szakirodalmában.

Voltak ugyanakkor későbbi vizsgálatok is, amelyek megerősítették az undorérzet és az általános erkölcsi ítélkezés kapcsolatát. Chapman és szerzőtársai (2009) például első körben undorító képeket nézegető kísérleti alanyaik arckifejezését vizsgálták, majd ezután leírták velük az érzéseiket is. Ezeket a reakciókat vetették össze azzal, ahogyan az alanyok a méltánytalan pénzügyi ajánlatra reagáltak. A kutatás azt mutatta, hogy a fizikai gusztustalanságra ugyanúgy reagálnak az alanyok, ahogy a méltánytalanságra (bár utóbbi esetben dühöt is éreztek).

Kérdés, hogy van-e ennek hatása például a fenti alkujátékokban mutatott viselkedésre. Moretti és Pellegrino (2010) azt állítja, hogy igen. Kutatásukban az undort keltő ingerrel szembesülő válaszadók magasabb arányban utasították vissza a méltánytalan ajánlatokat, mint a kontrollcsoport tagjai. Ugyanakkor a közelmúltban Landy és Goodvin (2015) áttekintette a tézist vizsgáló összes elérhető kutatást, és összességében csekély hatásokat és ambivalens, nem stabil eredményeket találtak. A közelmúltban a korábbi vizsgálatokat nagyobb mintán megismétlő kutatócsoport hasonló konklúzióra jutott (Johnson et al. 2016).

### V.3.2. AZ OKSÁGI IRÁNY MEGKÉRDŐJELEZÉSE

Haidt kutatócsoportja (Graham et al. 2009) nagymintás (bár nem reprezentatív) mintán kimutatták az öt modul öt önálló faktorként való elkülönülését – ez volt a MAE predikciója. Azonban látni kell, hogy ezt az eredményt később nem sikerült megismételni (Graham 2011). Az újabb vizsgálatok egyrészt az individualizáló, másrészt a csoportkötődést erősítő modulok együttmozgását mutatják, azaz a morális megfontolások kétfaktoros modellje működik. A kutatócsoporttól független friss vizsgálatok is – az elmélet előrejelzésével szemben –, kétfaktoros modellt találtak (Smith et al. 2017).

Néhány kutatás esetében említettem, hogy az undorérzékenység a vallásosságra kontrollálva is befolyásolja az ideológiai identifikációt. Azonban például a szexuális vonatkozású undor nagyon szorosan kapcsolódik a vallásossághoz (Tybur et al. 2010). A vallásosság esetében viszont plauzibilis az a feltevés, hogy a vallásos neveltetés és elvárások növelik annak esélyét, hogy valaki negatívan nyilatkozik az abortuszról és az átlagostól eltérő szexuális viselkedésről (és nem az undorérzékenysége miatt kap vallásos neveltetést). És az is széles körben elfogadott feltevés, hogy az emberek inkább válnak vallásosként a konzervatív politikai pártok támogatóivá, mint fordítva (jobboldali pártok híveiből válnak istenhívókké).

Kérdés, hogy tartható-e a viselkedési immunrendszer elméletére alapozott modell, miszerint az undorra való hajlamunk befolyásolja ideológiai beállítottságunkat. A Nebraskai Egyetem politikatudósa, Kevin Smith és szerzőtársai (2017) az *American Journal of Political Science* című folyóiratban nemrég megjelent, és rögtön nagy figyelmet keltő tanulmányukban ennek próbáltak alaposan utánajárni. Mindenekelőtt azt vizsgálták, hogy az adataik miképp illeszkednek a morális alapok elméletének fentírt erősebb állításához, miszerint a különböző érzelmek olyan modulok, amelyek egy-egy erkölcsi területhez (gondoskodás-bántás, méltányosság-csalás stb.) kapcsolódnak. Az elméleti modellhez leginkább az erkölcsi ítéletek és preferenciák terének ötdimenziós (ötfaktoros) leírása illik. Azonban faktorelemzésük, az elmélet feltevéseitől eltérően, kétfaktoros meg-

oldást mutatott (Smith et al. 2017). De elemzésüknek nem ez volt a lényege.

A kutatócsoport másfél-két éves periódusban, két hullámban vizsgált egy viszonylag nagy, ausztrál ikertestvérekből álló mintát. Azt találták, hogy az erkölcsi ösztönök indikátorai jelentős volatilitást mutatnak a két adatfelvétel között, és ezek a változások nem magyarázzák a politikai attitűdök két hullám közötti változásait. Az erkölcspszichológia elméletéből következő öröklődési hipotézist is el kellett vetniük az ikrekből álló minta adatai alapján.

Ciuk (2018) e könyv írásakor megjelent politikatudományi indítatású tanulmányában olyan kísérletet végzett, amelyben a MAE feltevéseivel szembehelyezkedve, a politikai identitás előhívásával próbálta befolyásolni az alapösztönök sztenderd indikátorainak értékeit. Amerikai mintájának eredményei azt mutatják, hogy a fordított irányú oksági kapcsolat működik: a republikánus/demokrata szavazók vélt morális beállítottságára vonatkozó információk befolyásolták a válaszadók undorral és egyéb ösztönökkel kapcsolatos beállítódásait.

A morális alapok elmélete azonban a legerősebb, legátfogóbb, és elméletileg is megalapozott kritikát a pszichológián belülről kapta. Cameron és szerzőtársai (2015) átfogó irodalmi áttekintésükben pszichologiaelméleti perspektívából, ám a kísérleti módszertan részleteire is figyelve tekintették át az erkölcsi ítéletek és az érzelmi ösztönök kapcsolatát. Konklúziójuk az, hogy a meglévő empirikus eredmények nem támasztják megfelelően alá az ösztönös érzelmek, és erkölcsi ítéletek/ideológiai beállítottság közötti oksági kapcsolatokra vonatkozó téziseket.

Az eredmények sokkal inkább kongruensek az érzelmek a fentiekhez hasonlóan régi, ám csak az utóbbi időben teret nyerő elméleteivel, amelyek arra hívják fel a figyelmet, hogy a nyelvileg megragadott érzelmi koncepciók a mélyből fakadó és nehezen megfogható lelkiállapotok társadalmilag konstruált modelljei (Cameron et al. 2015, Lindquist 2013).

Ez a könyv a racionális döntéelmélet erkölcsi modelljeiről szól, ezért nem szánok egy külön fejezetet annak ismertetésére, hogy milyen elméletek állnak az új erkölcspszichológia téziseit cáfoló ered-

mények háttérében. Az utolsó fejezet utolsó alfejezeteiben azonban érintőlegesen visszatérek még ezekre az elméletekre. Röviden álljon itt annyi: míg a racionális döntésmélet keretei között (némi pongyolasággal) a pszichologizálás és a szituációs logika szembenállása határozza meg az erkölcs forrásával kapcsolatos vitákat, addig a pszichológián belül a naturalizáló és a konstruktivista elképzelések vetélkedése figyelhető meg. Az – általam – pszichologizáló modelleknek nevezett irányzatok a naturalista pszichológiai elméleteket használják, míg, mint látni fogjuk, a szituációs logikához ragaszkodó racionális döntésméleti irányzat a konstruktivista vonulathoz közelebb álló elméletek eredményeit képes jobban értelmezni.



## VI. 2006\* – A normavezérelt polgár erkölcsse

### VI.1. Vissza a Nash-egyensúlyhoz

#### VI.1.1. MIT TANULTUNK EDDIG?

Ez a fejezet az erkölcs forrásainak egy új, a korábbiakra reflektáló kutatási irányát foglalja össze. Ebben az irányban folytatom kutatásaimat én is. A szóban forgó kutatási program megközelítése szemben áll az előző fejezetben bemutatott pszichológiai irányzat szemléletével, azaz a pszichologizálás kiteljesítése helyett (a könyvben bemutatott kutatási program kezdetén kidolgozott) szituációs logikai elemzés irányában folytatja az erkölcsi viselkedés megértését célzó kutatási folyamatot.

Mindenekelőtt azonban érdemes összefoglalni, hogy milyen tanulságokkal szolgáltak a témában a második világháború után elindult, racionális individualista megközelítést alkalmazó elméleti és empirikus kutatások.

A racionális döntéselmélet klasszikus axiómáira építő (II. fejezet), és alapjában az önérdékből (bármit is értsünk rajta) kiinduló játékt-

\* Bicchieri, C. (2006): *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press. Lásd még Binmore, K. (2005): *Natural justice*. Oxford: Oxford University Press. Továbbá: Gintis, H. (2009): *The bounds of reason: Game theory and the unification of the social sciences*. Princeton: Princeton University Press.

elméleti modellek megmutatták, hogy az élet ismétlődő játékaiban pusztán egyéni érdekekre támaszkodva is fenn lehet tartani a közérdeket szolgáló, az egyének rövid távú érdekeivel szemben álló viselkedésformákat (III. fejezet). Utaltam arra is, hogy egyes racionális döntéelméleti szociológusok azt is megpróbálták megmutatni, hogy ez elvileg nagylétszámú közösségekben is lehetséges, a társas kapcsolatok egész társadalmat összefogó hálózatán keresztül. Az erkölcs fenntartása racionális büntetések fenyegetéseinek láncolatán keresztül történik (ahol a büntetők is a büntetéstől félve teszik a dolgukat). Az utóbbi évtizedekben empirikus vizsgálatok sora mutatta meg, hogy az emberek erkölcsös viselkedését gyakran tényleg az ismétlődő interakciókban várható büntetéstől való félelem tartja fenn.

Ugyanakkor ezek a modellek arra is egyértelműen rávilágítottak, hogy a közjó létrehozását segítő mechanizmusok ugyanúgy támogatják a közjót nem szolgáló kollektív viselkedési minták fenntartását is. A szóban forgó játékelméleti modellek nem mondanak semmit azokról a mechanizmusokról, amelyek garantálják, hogy a racionális emberek közössége az erkölcsösnek tekintett viselkedésformákat tartsa fenn.

A fenti modellekre nyíltan reagáló méltányosságelméletek hipotézise az, hogy az emberekben kódolva van az együttműködés egy bizonyos szabályrendszere iránti vágy. Ez a vágy arra irányul, hogy az emberek a másokat segítő erőfeszítéseik arányában részesüljenek a közösség tagjai által nyújtott segítségből; és e vágy olyan öncél, amely arra készíti az embereket, hogy akár egyéb (például anyagi) céljaik feláldozása árán is, *a*) törekedjenek a tisztességes viselkedésre, és *b*) büntessék azokat, de csak azokat, akik megsértik a méltányosság elvét (IV. fejezet). Tehát az ismétlődő emberi játszmákban elképzelhető, sokféle racionális egyensúlyi kimenetel közötti választást egy olyan genetikai (vagy esetleg nagyon ősi kulturális) kód határozza meg, amely a tisztesség-méltányosság-egyenlőség elvei mentén biztosítja, hogy az emberek tenni akarjanak a közjóért.

Hatalmas mennyiségű empiria halmozódott fel az elmúlt két-három évtizedben e tézis tesztelésével kapcsolatban. Az adatok azt mutatják, hogy tényleg hajlamosak vagyunk rövid távú érdekeinkkel szembeállítva büntetni a valamilyen szabálytól eltérő embertársain-

kat. E hajlamra vonatkozó hipotézis eddig robusztusnak mutatkozott azokra a tesztekre, amelyek megpróbálták kizárni a racionális büntetési motivációkat. Ugyanakkor azt több évtized után sem tudjuk egyértelműen, hogy pontosan milyen méltányossági elveket próbálunk betartani és betartatni. Minél jobban konkretizáljuk, és minél közelebb visszük a méltányos preferenciákra vonatkozó szabályokat a racionális döntésmélet modellezési keretéhez, az adatok annál határozottabban cáfolják a bennünk élő, méltányosságra irányuló vágyakra vonatkozó feltevéseket.

Pszichológusok egy kis csoportja már két évtizede felhívta a figyelmet arra, hogy a méltányosságelméletek kizárólag néhány, az elmúlt kétszáz év politikai-filozófiai diskurzusában vizsgált erkölcsi szempontra fókuszálnak, és nem veszik figyelembe, hogy az emberi társadalmak együttélési problémái sokrétűbbek annál, hogy e néhány szempont kezelni tudja őket. Ha azt feltételezzük, hogy az erkölcsi elvek genetikailag kódolva vannak, akkor más, a humanoid közösségek évszázerezdes (vagy hosszabb) története során szerepet játszó, ma már kevésbé feltűnő erkölcsi kihívásokkal is foglalkoznunk kell. Ez az irányzat racionális döntésméleti modellezés nélkül, evolúciós spekulációk nyomán keresi az alapvető erkölcsi ösztöneinket a viselkedési kísérletekben és kérdőíves vizsgálatokban. E kötetben az undorra koncentráltam, mert egyrészt ennek kutatásával indult ez az irányzat, másrészt ennek az evolúciós narratívája a legerősebb (és ezt feltételezik az erkölcsi tartalommal bíró érzelmek közül a legősibb reakciónak), továbbá e tényező „erkölcsi pantheonba” helyezése mutat rá a leglátványosabban arra, hogy az objektív evolúcióbizológiai alapokra hivatkozó méltányosságelméletek valójában társadalmi értékek által telítettek (V. fejezet).

Az elmélet és empiria gazdagodásával azonban az erkölcsünket meghatározó vágyak és érzelmek listája (szinte már) esetlegesen bővült, és egyre szembetűnőbb, hogy milyen könnyű egy-egy újonnan felmerülő ösztönhöz verbális-anekdotalisztikus, evolúciós narratívát alkotni. Továbbá az érzelmek folyamatosan fejlődő pszichológiai elmélete is erős alternatív modelleket kínál az erkölcspszichológia végső mozgatóként értelmezett ösztönmoduljaival szemben.

Az erkölcspszichológia kutatási programja az elnagyolt elméleti alapok finomítása helyett a megalkotott konstrukciók empirikus koherenciájára, és a társadalmi attitűdökkel kapcsolatos magyarázó erejére koncentrált. Más pszichológiai irányzatok képviselői azonban képesek kimutatni az erkölcspszichológia szerint évszázadok (vagy akár milliók) alatt belénk kódolódott modulok társadalmi konstruáltságát, és egyre gyűlnek azok a bizonyítékok is, amelyek az ősiinek tekintett ösztönök empirikus modellekben használt indikátorairól mutatják be, hogy azok értékei valójában függvényei azoknak a tényezőknek (például a pártállásnak), amelyeket magyarázniuk kellene (Ciuk 2018).

A társadalmi normák racionális, játékelméleti alapokon nyugvó, 21. században fejlődésnek induló elméletei abból a tételből indulnak ki, hogy nem feltétlenül léteznek a lelkünkben olyan velünk született erkölcsi modulok, amelyek egyértelmű irányt mutatnak közös ügyeink erkölcsös rendezése felé. Erre építve vizsgálják, hogy a racionális emberek miképpen találhatják meg a társadalmi térben az erkölcsösnek minősülő viselkedési formákat, amelyeket – a korábban megismert racionális vagy altruista büntetések láncolatának mentén – a közösség tagjai követnek. A következőkben ezt a kutatási irányt mutatom be.

### VI.1.2. MIT ADTAK NEKÜNK A MATEMATIKUSOK?

Az alábbiakban tárgyalt, a normák racionális játékelméleti elemzésén nyugvó elméletek elutasítják azt a feltételezést, hogy velünk született modulok irányítanak minket az erkölcsös viselkedés irányába. Kiindulópontjuk az erkölcs játékelméleti kutatásának egyik konklúziója, miszerint a racionális emberek játszmáiban a Nash-egyensúlyok közül kerül ki a végeredmény, és a számos próbálkozás ellenére (amelyeket nem fejtettem ki teljes egészében) nincs olyan univerzális varázsalgoritmus, amely több egyensúlyi pálya esetén mindig meg tudna nevezni egyetlen olyan cselekvéskombinációt, amely a többi racionális megoldásnál is racionálisabb szupermegoldás.

Az elmúlt évtizedekben az élethelyzetek ilyen kevésbé determinált leírása miatt egyébként sok kritika érte a játékelméletet (Hedoin 2015). Éppen a játékelmélet antropológiai feltevéseitől és modellezési eszköztárától nem idegenkedő, a játékelmélet kutatóit tanzségein befogadó neoklasszikus közgazdaságtan képviselői hozták fel a játékelmélet alkalmazásával szemben ezt az érvet. Erre a kritikára a következő fejezetben visszatérek.

Mielőtt áttekinteném a társadalmi normák szerepét a racionális egyensúlyok közötti választásban, szeretném röviden érzékeltetni, hogy mennyiben segíti az emberi interakciók vizsgálatát a Neumann tézisével általánosított Nash-féle egyensúlykonceptió, és hol vannak a határai. A tankönyvekben (ebben a könyvben is) bemutatott és társadalomtudományi alkalmazásokban használt  $2 \times 2$ -es márixjátékok végtelenül egyszerű struktúrában elemeznek meglehetősen komplex élethelyzeteket. Ezekben a helyzetekben egyensúlyi és egyensúlyon kívüli kimenetelekből is kevés van, mivel konstruálójuk már eleve a legértelmesebb döntési alternatívák (egyszerűsített) leírására szorítkoznak. Az ilyen szituációk könnyen átláthatóak, ám nem tudják érzékeltetni a Nash-egyensúly konceptiójának erejét és korlátait.

Egy ilyen tipikus tankönyvi szituáció helyett vegyünk például egy valóságban nagyon egyszerű (legalábbis az itt tárgyalt komolyabb erkölcsi dilemmáknál sokkalta egyszerűbb) élethelyzetet – a leírását jelentősen leegyszerűsítettem, de azért nem olyan végletes mértékben, mint a könyvben tárgyalt modellek a mintául szolgáló erkölcsi dilemmákat.

Ebben az egyszerű példában a menzán egy négyfős társaság tagjai a pénztárnál jelzik egymásnak, hogy látnak egy asztalnál négy szabad helyet egymás mellett. Van tétje annak, hogy ki hova ül: egyik helyen szembe sül a nap, a másik közelében egy udvariatlanul szétterpeszkedő ember ül stb. Tehát nem teljesen közömbös a társaság tagjainak, hogy melyik széken tudják elkölteni az ebédjüket. De ezeknél a szempontoknál fontosabb, hogy a leüléshez készülődés közben ne ütközzenek össze – és még sokkal fontosabb, hogy nem akarnak megküzdeni egymással az egyik vagy másik helyért. Most azt a pillanatot kapjuk el, amikor az asztalt már kellően megközelítve mindenkinek

gyorsan döntést kellene hoznia arról, hogy melyik szék irányába indul el. A társaság tagjainak helyzetét ábrázolhatjuk térben úgy is, hogy elhelyezkedésük erős fogódzót nyújt a választáshoz, de ha ügyesek vagyunk, akkor létrehozhatunk olyan helyzetet is, amikor igazából bárki dönthet ép ésszel bármelyik szék mellett. Most az utóbbi verziót választjuk, és feltételezzük, hogy döntésük (első körben mindenképpen) szimultán születik meg.

Ha az egyéni kényeskedés (tűző nap, terpeszkező asztaltárs stb.) szempontjait felülírja az ütközés és verekedés elkerülésének vágya, akkor egyszerű meghatározni az összes kimenetel mellett a Nash-egyensúlyokat is. Tekintsük először az összes lehetséges kimenetelt: négyen négy hely közül  $4^4 = 256$ -féleképpen választhatnak (beleértve azokat az eseteket is, amikor többen választották ki ugyanazt a széket). Fontos látni, hogy ez a szituáció egészének korrekt leírása (amely nagyon sok tankönyvi játékmódelellől már kimarad). E választások többsége viszont ütközéssel járna (amikor többen nézték ki ugyanazt a széket), és ezt mindenképpen elkerülnék az érintettek. Ha az egyik szék (képzeletben) már foglalt, akkor oda elvileg megpróbálhatna leülni mindhárom másik társa, azonban azt feltételezzük, hogy bár megtehetnék, ezt nem akarják. Ezért ha valaki foglaltnak lát/gondol egy széket, oda már nem akar leülni, és inkább a másik hátról választ. De ha a maradék hátról foglaltanak véli a preferált széket, akkor értelemszerűen, az ütközés elkerülése érdekében, már csak kettőből *érdemes* választania (noha elvileg továbbra is négyből lehetne!). Az iskolában tanult egyszerű kombinatorikai algoritmussal kiszámolhatjuk, hogy ennek a játszmának  $4!$ , azaz összesen  $4 \times 3 \times 2 \times 1 = 24$  Nash-egyensúlya van.

Egyrésztől azt látjuk, hogy a Nash-egyensúly értelmes és jelentős szűkítését adta a lehetséges kimeneteknek a résztvevők preferenciáiról tett egyszerű feltevések alapján is (emlékezzünk rá, hogy nem is kellett pontosan megrajzolni, hogy ki mit és mennyire szeretne). Ez számunkra triviális, hiszen ebben élünk. Ám ugyanezt a folyamatot el lehet játszani számunkra rejtélyesebb szituációkkal is. A fogalom kidolgozásakor Nash rögtön be is bizonyította, hogy az élet játékaiknak bármely hasonló kis modelljében fogunk találni ilyen egyensúlyt

(lehet, hogy éppen csak egyet). A legrosszabb esetben azt kell mondanunk, hogy az emberek indifferensen fognak viselkedni – de a határozatlanságra vonatkozó állítás is egy határozott előrejelzés!

Ugyanakkor azt is látni kell, hogy ez az egyensúlykonceptió csak speciális esetekben kínál egyetlen várható kimenetelre vonatkozó előrejelzést (az egyszerű fogolydilemma egy ilyen speciális eset). Egy oly triviális probléma, mint az asztal köré ülés is rengeteg bizonytalanságot rejt magában. Ezért is törekedtek arra sokáig a játékelméletet megalapozó kutatók, hogy a tudományos elvárásoknak megfelelően növeljék a modell előrejelző képességét azáltal, hogy a racionális döntés valamilyen magasabb szintjét tudják meghatározni. Nagyon fontos és hasznos koncepciók születtek, de egyik sem bizonyult a racionalitásból egyértelműen levezethető univerzális koncepciónak (ezek közül a visszafelé göngyölítés eljárásáról hosszabban írtam a III. fejezetben). Jogos tehát az egyén tudatos döntésén túlmutató mechanizmusok keresése. Ilyen kutakodás az erkölcsi döntésekkel kapcsolatban is folyik. Ezeknek két iránya van. Az egyik a tudattalan alkalmazkodást vizsgálja, a másik pedig a cselekvést megelőző kommunikáció lehetőségeit. Az alábbiakban e két megközelítés tárgyalása következik.

### VI.1.3. EGYENSÚLY ELÉRÉSE CSENDES ADAPTÁCIÓ RÉVÉN

Az erkölcsös viselkedés gyökereinek vizsgálatával kapcsolatban a Nash-konceptió fáradhatatlan védelmezője, a könyv korábbi fejezeteiben is sokszor idézett brit matematikus-közgazdász (és ma már hivatalos státusza szerint filozófus) Ken Binmore (1998, 2005, 2010, 2015) képviseli azt az álláspontot, hogy azért is felesleges a Nash-egyensúlyok közötti választást megoldó szuperracionális döntés keresése, mert életünk játékait eleve nem szuperracionális okoskodással oldjuk meg, csak éppen a végeredmény fog úgy kinézni, mintha a Nash-féle racionális mérlegelés logikáját követtük volna (vö. Friedman 1953).

Binmore szerint a döntéseink túlnyomó részét kitevő életjátékaink sokszor és bizonytalan számossággal ismétlődnek. Igazából észre sem vesszük, hogy lejátszunk ezeket a játszmákat (lásd a fenti székfoglalós dilemmát). Sokszor hibázunk is, vagy a rossz irányba fordulunk. A menzán, a tálcával a kezünkben, néha épp arra indulunk, amerre a társunk is menni akar. Nash logikája azt jelenti, hogy mindenféle szuperracionalitás nélkül korrigálunk, ha látjuk, hogy mások viselkedésére nem az optimális választ adtuk. És mindenki ezt teszi. És a korrekciók – bizonyos interakciónál akár másodpercekben sem mérhetően rövid idő alatt –, a Nash-egyensúly irányába terelnek mindenkit. Nagyobb formátumú társadalmi folyamatoknál persze az időléptékek is változhatnak. De Binmore figyelmének fókuszában az erkölcs kapcsán is olyan helyzetek állnak, mint például, hogy ki lépjen be előbb az ajtón, vagy nyújtsak-e kezét ilyen-olyan szituációkban stb.

A Nash-egyensúlyhoz közeledést tehát nem úgy kell elképzelni, hogy a játékelméleti tankönyvek mátrixaiban értelmezzük helyzetünket, és kikalkuláljuk az optimális lépéseket. Az evolúció nem a Thomas Schelling által kitalált mátrixok olvasására készített fel bennünket (még kevésbé a kutatók egymás közötti kommunikációjában használt képletek értelmezésére), hanem a fent említett korrekciókra. A mátrixokban és a képletekben csak racionális döntésként modellezzük adaptációnk végeredményét. Ugyanúgy, ahogy minket, földi halandókat az evolúció nem a madártávlatú térképek alapján történő tájékozódásra készített fel. Erről bárki meggyőződhet, aki olyan országban kérne a helyiektől térkép segítségével útmutatást, ahol nem verik az iskolások fejébe a térképhasználatot. Viszont a térképek alkalmasak a terepviszonyok és közlekedés észszerű modellezésére. Ez az érvelés egybevág azokkal a válaszokkal, amelyeket a neoklasszikus közgazdaságtan korai proponensei adtak a kezdeti kritikákra (Alchian 1947, Friedman 1953). A racionális döntésselméleti modellezés korai védelmezőivel szemben Binmore és az evolúciós játékelmélet képviselői konkrétan megfogalmazták, és formális modellekben is vizsgálták azokat a mechanizmusokat, amelyek révén a hibákra reagáló adaptív viselkedés a Nash-egyensúly felé tereli



a döntéseket (Samuelson 1994). Az adaptív döntések dinamikáját elemző evolúciós modellek mellett e könyvben is sokat emlegetett játékelméleti kísérletek dinamikájának elemzése is abba az irányba mutat, hogy az emberek kudarcuk nyomán tanulnak és adaptálódnak a Nash-egyensúly által rájuk osztott szerephez (Binmore és Shaked 2010).

Ezekben az evolúciós jellegű modellekben az alapjátékok bizonytalanságával szemben valamit lejátszanak az adaptálódó játékosok, tehát a modell választ a lehetséges Nash-egyensúlyok közül. Az útfüggőség jelentős lehet az adaptációs folyamatban (az egyén észlelése arról, hogy merre indul először a másik, aki a következő pillanatban már az egyén reakciójára reagál). Ugyanakkor Binmore (2005) például jelentős erőfeszítéseket tett azért, hogy a tortafelosztási alkukkal kapcsolatos elvekre vonatkozóan megalapozzon egyféle (egalitáriánus, vagy arra visszavezethető) megoldást. De mint láttuk, empirikusan nem sikerült bizonyítani, hogy az emberek automatikusan tendálnának az egyenlő elosztást kínáló megoldások felé. Ráadásul – anélkül, hogy a részletekbe belemerültem volna –, arra is utaltam, hogy az alapjátékban az egyetlen Nash-egyensúlyt tartalmazó, ismételt fogolydilemma evolúciós modelljei (amelyek a genetikai evolúció modelljei mellett magukban foglalják az adaptációs modelleket is) nem emelnek ki egyetlen specifikus stratégiakombinációt, ahova az adaptív döntések az idővel konvergálnának (Nowak és Sigmund 1993).

Nagyobb létszámú csoportokban – például egymás korlátozott észlelése miatt – ezek az adaptációs folyamatok problémásak is lehetnek. Néhány új evolúciós (adaptációs) modell kétségbe vonja azt is, hogy ilyen nagy csoportban az egyének mindig eljutnak a Nash-egyensúlyi stratégiakombinációk közelébe (Morsky és Akcay 2018). Sok ember viselkedésének szimultán koordinációja esetén csak speciális esetekben lehet hinni a csendes adaptáció gyors koordináló erejében. És ott vannak még az erkölcs olyan, a politikai filozófia középpontjában álló dilemmái, amelyek nagyobb súlyú és tudatosabb döntéseket igényelnek, és sokszor tényleg inkább egyedi játékként modellezendők, mint sok hasonló interakció sűrű sorozataként. A normák játékelmé-

leti elemzői e helyzetek megértéséhez a Nash-egyensúly egyik korai, speciális módosításához nyúltak vissza a koncepcionális alapokért (Aumann 1974). Most erről lesz szó.

## VI.2. A normák koreográfiája

### VI.2.1. KOREOGRÁFUSOK ÉS KORRELÁLT EGYENSÚLYOK

Az erős reciprocitás koncepcióját kidolgozó Herbert Gintis amerikai alternatív közgazdász a társadalom rendjének játékelméleti alapú, interdiszciplináris kutatására tette fel életének elmúlt két évtizedét (Gintis 2000, 2009, 2016). Ő volt talán az első, aki az erkölcs játékelméleti kutatói közül megfogalmazta, hogy a társadalmi rend megértéséhez ki kell lépni az egyéni preferenciák világából. Gintis (2009) a normák kapcsán a játékelmélet metodológiai individualizmusát is megkérdőjelezte.

Gintis (2009) tézise, hogy a Nash-egyensúlyok sokasága közül – különösen az előző alfejezet végén említett esetekben –, a társadalmi térben megtalálható közös tudatunkban élő erkölcsi szabályok alapján választunk. E szabályokat koreográfusoknak nevezte, amelyek megakadályozzák, hogy a sokszereplős, sok Nash-egyensúlyt tartalmazó emberi játszmákban a bizonytalan emberek nem megfelelően koordinált döntései káoszt okozzanak. Binmore-hoz hasonlóan Gintis is úgy gondolja, hogy az erkölcs vizsgálatának fogolydilemma szituációkra való szűkítése nem megalapozott, az élet sokkal többféle erkölcsi dilemmát tartogat. Ezen felül osztja Binmore szkepszisét azon közgazdászok törekvéseivel kapcsolatban, akik a Nash-egyensúlyok közül egy szuperracionális algoritmus segítségével próbálnak választani (Eckel és Gintis 2010). Ezért is van szükség szerinte koreográfusokra.

A koreográfia mechanizmusának leírására a korrelált egyensúly koncepcióját használja (Aumann 1974), amely a játékelmélet korábbi, az egyensúly finomítására irányuló törekvésekben gazdag időszakában született. Gintis nem az egyetlen, aki e régi koncepciót

szélesebb körben alkalmazza a gyakorlatban. A társadalmi normák játékelméleti alapokon nyugvó értelmezésének jeles képviselője, Francisco Guala (2016, lásd még Hendriks és Guala 2015) olasz közgazdász-filozófus szintén a korrelált egyensúly fogalmát használja a társadalmi intézmények működési mechanizmusainak megértéséhez. A mindenkivel kritikus Binmore (2015) sem idegenkedik (teljesen) ettől az interpretációtól.

Egy játék korrelált egyensúlya olyan stratégiakombináció, amelyet a szereplők a döntésük előtt közösen megfigyelt jelzés (utasítás) követésével hoznak létre. A koncepció elmozdulás a könyvben nem tárgyalt, ún. kooperatív játékelmélet megközelítése felé. A kooperatív játékelméletben feltételezik, hogy a játékosok nem önállóan hozzák meg a döntésüket, hanem egy tárgyalás során olyan megállapodást köthetnek, amelynek végrehajtását aztán ki tudják kényszeríteni. Ám a korrelált egyensúly koncepciója nem megy ilyen messzire. Aumann (1974) koncepciója szerint az emberek nem kötelezhetők a szignál követésére, azt saját belátásuk szerint, a Nash (és igazából már Neumann) által lefektetett, racionális stratégiai viselkedésre vonatkozó szabályok szerint követik: egy szignál akkor lesz korrelált egyensúlyra vonatkozó utasítás, ha minden résztvevőre igaz az, hogy a többiek engedelmeskedése esetén neki is megéri követnie a szabályt. A korrelált egyensúly tulajdonképpen egy, a közös jelre vonatkozó információval kiterjesztett játék Nash-egyensúlya (Binmore 2015).

Először is látni kell, hogy a szereplők közötti térben elhangzó, Nash-egyensúlyi stratégiakombinációkat javasoló utasítások mindig korrelált egyensúlyt hoznak létre, azaz a Nash-egyensúlyok egyben korrelált egyensúlyok is. De ha több Nash-egyensúlya is van az adott játszmának, akkor a Nash-egyensúlyok közötti térben megjelennek olyan stratégiakombinációk is, amelyek korrelált egyensúlyok, de nem Nash-egyensúlyok.

A normák racionális követésének megértése szempontjából alapvető koncepcióról van szó, ezért az alábbiakban, egy egyszerű példán keresztül, részletesebben is megmutatom a működését és a hatásait.

## VI.2.2. HARMADIK PÉLDÁZAT: NYULAK A DELPHOI JÓSDÁBAN

Vegyük a korrelált egyensúly egy tankönyvekben gyakran alkalmazott (és talán sok olvasó számára ismerős) illusztrációját, amely egyúttal bemutatja azt is, hogy miért lehet szerepe ezeknek a közös szignáloknak a társadalmi káosz elkerülésében.

8.a ábra. A gyáva nyúl játszma

		Te	
		mész előre	kitérsz
Én	megyek előre	összeütközünk, mindkettőnknek a legrosszabb	nyertem, vesztedél, nekem nagyon jó, neked rossz
	kitérek	vesztettem, te nyertél, nekem rossz, neked nagyon jó	döntetlen lett, egyikünknek se jó, de nem is rossz

8.b ábra. A gyáva nyúl játszma – boldogságértékekkel kifejezve

		Te	
		mész előre	kitérsz
Én	megyek előre	Én: 0, Te: 0	Én: 3, Te: 1
	kitérek	Én: 1, Te: 3	Én: 2, Te: 2

8.c ábra. A gyáva nyúl játszma boldogságértékekkel – és az egyensúly felé ható „erőkkel”

		Te	
		mész előre	kitérsz
Én	megyek előre	Én: 0 ↓, Te: 0 →	Én: 3, Te: 1
	kitérek	Én: 1, Te: 3	Én: 2 ↑, ← Te: 2

A játék tankönyvi alaptörténete két heves fiatal férfiról szól, akik bátorságukat akarják demonstrálni egy párbajban, ezért egy üres útszakaszon elindulnak egymással szemben az autóikkal. A párbajt az nyeri, aki az utolsó lehetséges pillanatban (amelyet a fenti mátrix

modellez) sem rántja félre a kormányt, hanem vállalja az ütközés kockázatát. A párbaj leegyszerűsített sémája számos személyközi és csoportközi konfliktusnak. Talán azért ez a játszma az első számú illusztráció a korrelált egyensúllyal kapcsolatban, mert az emberi közösségekben e konfliktusokat láthatóan nem mindig sikerül harmonikusan, külső kényszerítő erő nélkül megoldani. A mátrix úgy van leegyszerűsítve, hogy a méltányos kompromisszum lehetőségét kizárja.

Bár rendkívül nehéz döntési dilemmákat modellez, a fenti  $2 \times 2$ -es mátrixjáték struktúrája végtelenül egyszerű. Ha valahogy megtudnám, hogy te semmiképpen nem fogod félrerántani a kormányt, akkor a legrosszabb elkerülése érdekében én mindenképpen kitérek. Ha viszont megneszelem, hogy te végül ki fogsz tért, akkor bátran megyek előre, és győzök. A játék alapleírása azonban semmilyen, a felek elszántságára vonatkozó információt nem tartalmaz, ezért is fordulhatnak rosszra a dolgok.

A II. fejezetben már utaltam rá, hogy minden ilyen, véges formában felírt játszmahoz létre lehet hozni a stratégiák keverésének egyensúlyi kombinációját. Azaz a játékosok kockáztatnak, figyelembe véve azt, hogy mennyire szeretnék a győzelmet, és mennyire félnek az ütközéstől. A Nash/Crowe-féle „csajozási játékban” már bemutattam a randomizált döntés logikáját: akkor randomizálok, ha a másik úgy keveri a lapokat, hogy nem tudok dönteni egyértelműen egyik stratégia mellett sem. És a másik fél akkor ragaszkodik az engem elbizonytalanító hozzáállásához, ha én is úgy keverem a lapokat, hogy ő se tudjon dönteni egyértelműen. Ebben a játékban akkor bizonytalanodom el, ha úgy érzékelem (tippelem), hogy a másik fél hozzáállása a következő egyenletnek megfelelő:  $0p + 3(1-p) = p + 2(1-p)$ , ahol  $p$  annak az (érezkelt) esélye, hogy a másik fél kitart. Az egyenlet baloldalán a boldogságom várható értékét számítom ki abban az esetben ha én megyek előre, a jobb oldalán ugyanezt arra az esetre, ha kitérek. Könnyen kiszámolható, hogy ilyen boldogságértékek mellett mindkettőnknek egy pénzérmét kell feldobnunk, és így az esetek felében a kitartás mellett döntünk. Ekkor egyenlő eséllyel valósul meg mind a négy lehetséges kimenetel, és várhatóan 1,5 bol-

dogságértéket nyerünk, ha ilyen játszmáink felében kitartunk, másik felében pedig kitérünk a konfliktus elől.

A keverés racionalitása még kézzelfoghatóbbá válik, ha például két, egymással csatára készülő, nagyjából egyenlő erőket felvonultató görög városállam dilemmáját képzeljük el. A dilemma a vakmerő támadás és a rugalmas elszakadásra (megfutamodás) való hajlandóság között merül fel, egy olyan helyzetben, amikor a vezetők számára a győzelem mellett az emberi élet védelme is fontos. Ismereteink szerint ezekben a helyzetekben gyakran fordultak a vezetők az isteneikhez, nem csupán segítségért, de információkért is. Kíváncsiak voltak arra, hogy az égi jelek mire utalnak: segíteni fogják-e őket az istenek vagy sem. Ha már sok emberről van szó, akkor a Nash/Crowe játszma módosításához hasonlóan (II.2.3.2. alfejezet) húzzuk kicsit szét a boldogságértékeket, hogy jobban érzékeljük a káosz elkerülésének nyereségét! A halálos összeütközést jelentő nulla boldogsághoz képest tegyük százas értékre a bátor győztesek jutalmát. A győzelemnél sokkal rosszabb, de a halálnál sokkal jobb vereség értéke legyen 48, a gyávaság szégyenének megosztása ennél kicsivel jobb, 52 boldogságegységet ér. A csata előtti (külsőleg randomizálásnak tűnő) döntésünket meghatározó bátorságfaktor értékének számolása a másik hasonló ösztönzői esetén a következő:  $0p + 100(1-p) = 48p + 52(1-p)$ . Könnyen belátható, hogy ilyen boldogságértékek mellett továbbra is a pénzfeldobás mellett maradunk (hiszen  $p=0,5$ ), és minden második csatánkba felbátorodva indulunk, a többi esetben pedig visszavonulásra készülünk. Várható boldogságértékünk pedig 50 boldogságegység: az egyenlő arányban előforduló 0, 100, 48 illetve 52 boldogságértékek átlaga.

Most tételezzük fel, hogy a csata előtt a felek annyira félnek a piruszi (sok halálos áldozattal járó) győzelemtől (döntetlentől), hogy megállapodnak a következőben: mindkét fél kiválasztja a legjobb harcosát, akik megküzdének egymással. Aki győz, annak a városa nyeri a csatát. Előre egyik fél sem tudhatja, hogy az ő harcosa fog-e győzni, így ez is valamiféle kockadobás. Azonban ezt, a csata előtti jóslásokkal szemben, a két fél közösen végzi – ugyanis nem független egymástól az, amit a két harcos küzdelmének az eredménye a két

fél számára üzen. Ha az egyik fél harcosa nyert, az azt jelenti, hogy ők nyertek, és nem hátrálnak meg. Ugyanakkor a vesztes félnek is világossá vált, hogy a győztes fél nem hátrál meg. Az optimális választuk erre az, hogy elfogadják a számukra kedvezőtlen végeredményt. A győztes fél persze annak a tudatában nem hátrál meg, hogy a párbajban vesztes fél ilyenkor meghátrál.

Ennek az eljárásnak az eredménye, hogy az egymástól független információforrásokra hagyatkozó döntések esetével szemben, a felek mindenképpen elkerülik a vérontást. Az 50-es átlagos boldogságszint helyett, egyenlő erők párbaja esetén, egyforma eséllyel nyernek 100-at illetve 48-at, ami 74-es átlagot jelent.

Kritikaként merülhet fel a fenti példával szemben az, hogy előzetes megállapodásokat feltételez, így kilép az eddigi keretek közül, ahol egymástól független döntések racionalitását vizsgáltuk. Képzeljünk hát el, hogy nem párbajban dől el a csata győztese, hanem az alkalmazott (támadó vs. védekező) stratégiáról továbbra is az istenek hangulatára vonatkozó információk alapján döntenek a felek. Azonban a csaták előtt mindkét fél Delphoiba jár kifürkészni az istenek szándékát. Amennyiben egy-egy csatával kapcsolatban a jósök következtetések, és mindenki tudja, hogy mindenki oda jár jóslatolni, akkor nem szükséges külön megállapodás ahhoz, hogy a közös szignálra hagyatkozva csökkentsék a felek a kockázataikat (elkerüljék a túlzott vérontást).

A delphoi jósdá tehát mindenki számára közös útmutatást ad az istenek hangulatáról, amely ebben a játszmában az Aumann (1974) által definiált korrelált egyensúlyhoz vezet. Ahogy jeleztem, azért nevezhetjük ezt egyensúlynak, mert nem vakon követik a jósök útmutatását, hanem saját jól felfogott érdekükben. Ez nagyon fontos! A felek a jóslattal kiegészített játszmában döntenek arról, hogy kövessék-e a jóslat indikációit. Ezt a döntési dilemmát ugyanúgy fel lehet írni egy egyszerű egyenletben. Ha például a spártaiak úgy érzik, hogy a jósök egyszer így, másszor úgy jósolnak (azaz fele-fele arányban), akkor azért döntenek a jóslatra való figyelés mellett, mert – ha a másik fél követi a jóslat utasításait –, akkor számukra  $100 \times 0,5 + 48 \times 0,5 > 52 \times 0,5 + 0 \times 0,5$ . Az egyenlőtlenség bal oldalán látható a jóslat követésének

következménye akkor, amikor mindkét fél a jóslat szerint jár el, a jobb oldalon pedig a jóslattal való szembeszegülés következménye akkor, ha a másik fél hallgat a jósda javaslatára (az utóbbi esetben mi is visszavonulunk, amikor a másikat visszavonulásra ösztönzik; és előretörünk, amikor a másikat ösztönzik előretörésre).

A jósda feles-feles lapkeverése annyit mutat meg, hogy a mi szempontunkból a Nash-egyensúlynak megfelelő random viselkedés (az esetek felében támadunk, a másik felében visszavonulunk) a közös koreografáló erőnek köszönhetően jobb állapotba juttatja mindkét felet annál, mintha mindketten más-más, egymástól független helyen firtatnák az istenek hangulatát. Ebből látszik a koreográfia közös haszna, miközben követésének egyéni racionalitása is kimutatható.

Ugyanakkor azt is látni kell, hogy a fenti számok többet is mondanak annál, mint hogy a közösség koreográfusai, ha akarnak, jobb helyzetbe hozhatnak minket. Ugyanis, ha ebben a játszmában mindenki tudja, hogy mindenki elfogadja a delphoi jósok koreográfus szerepét, akkor nekik nagyobb hatalom van a kezükben, mint pusztán feldobni egy pénzérmét, és a szerint jósolni mindkét félnek a vérontás elkerülésének érdekében. Ugyanis a fenti képletből látszik, hogy a másik fél szabálykövetése esetén a delphoi jósok bármilyen lapkeverési stratégiáját érdemes elfogadni, bármennyire is kedvezőtlen az számunkra. Ha tehát a jósdában nem kedvelik a spártaiakat, és mindig ellenük jósolnak, azok nem sokat tehetnek alárendelt szerepük ellen (csak azt, hogy ha van elég erőforrásuk, akkor fegyverkezés és/vagy önfeláldozás révén átalakítják a hadi játszmáikat, és domináns stratégiává teszik maguk számára a támadást).

Ebben a játszmában – a fent megadott boldogságértékek mellett –, három Nash-egyensúly van: 1. az egyik támad és a másik elfut, 2. az egyik elfut és a másik támad, és 3. mindketten egy érme feldobásával döntenek el, hogy támadnak-e. Azaz a támadás valószínűsége lehet 100%, 50% vagy 0%. A korrelált egyensúlyi mechanizmus bevezetésével azonban végtelenre nő a lehetséges egyensúlyok száma, 0% és 100% között bármekkora lehet a támadás esélye, a jósok kártyakeverési politikájának függvényében. Tehát ebben a játszmában az az eljárás, ami egyértelmű útmutatót ad a játékosnak egy kockázatos



helyzetben, az a külső megfigyelő számára teljesen előrejelezhetetlenné teszi a történéseket: bármi megtörténhet.

Ugyanakkor ez a leegyszerűsített játék megmutatja, hogy milyen korlátok állnak a Nash-egyensúlyok halmazának közös útmutató általi kiterjesztése előtt. Ugyanis a korrelált egyensúly eléréséhez a koreográfus csak „hozott anyagból” dolgozhat: nem terjeszkedhet túl a Nash-egyensúlyok által kifizített stratégiatéren, csak a köztük lévő „lyukakat” tömheti be. Például egy, a fentihez hasonló alkuhelyzetben a koreográfusok nem tudnak szélsőségesebb (egyenlőtlenebb) megoldást ránk erőltetni annál, mint amit magunk is elfogadnánk egy spontán alkufolyamat eredményeként. A koreográfus csak a független döntések nyomán elképzelhető egyensúlyok közötti tetszőleges keverést teszi lehetővé. Leginkább azért, mert a koreográfus dobókockáján csak Nash-egyensúlyok szerepelhetnek. Mégpedig azért, mert ő egyszerre ad mindenkinek utasítást, és ha egy dobás eredménye nem egyensúly, azt legalább egyvalakinek nem érdemes végrehajtania, bármennyire is előnyösnek tűnt számára a koreográfia (egy-egy utasítások esélyeinek listája) a dobás előtt. Egyébként, ha biztosak vagyunk benne, hogy a koreográfus tényleg csak a Nash-egyensúlyok közül sorsol, akkor már azt sem kell tudnunk, hogy konkrétan milyen utasítást ad másoknak. De ez már a vizsgálat tárgyának szempontjából kevésbé lényeges részletkérdés.

### VI.2.3. A SZOCIOLÓGIAI KONCEPCIÓK FELÉ

#### VI.2.3.1. A korrelált egyensúly jelentősége

A játékelméletet régóta alkalmazó közgazdaságtanban a korrelált egyensúly ötletét a vele rokon (Forges és Peck 1995) napfolt-egyensúly koncepciójának (Cass és Shell 1983) alkalmazásánál használták fel. A közgazdászokat az érdekelte, hogy miképp befolyásolhatják a gazdasági szereplők viselkedését olyan információk, amelyeknek elvileg nincs közük a gazdaságot valójában meghatározó tényezőkhöz. Például a rossz hatással dolgozó hitelminősítő- és tanácsadó cégek

hogyan képesek önbeteljesítő jóslattá változtatni a megalapozatlan előrejelzéseiket azáltal, hogy minden gazdasági szereplő úgy számol, hogy minden gazdasági szereplő úgy véli, hogy minden gazdasági szereplő hallgat a hitelminősítőkre (Ferri et al. 1999).

A társadalmi normákkal játékelméleti alapokon foglalkozó új elméletek (Gintis 2009, Gintis és Helbing 2015, Guala 2016) a korrelált egyensúly koncepciójának jóval nagyobb, általánosabb jelentőséget tulajdonítanak. Úgy vélik, hogy minden „társadalmi elvárások” által befolyásolt döntésünk a koreográfusként működő társadalmi normák/intézmények utasításának követésével létrejövő, korrelált egyensúlyokhoz igazodik.

A korrelált egyensúly létrejöttét segítő mechanizmus előnye, hogy nagyméretű csoportokban nem kell mindenkit egyenként figyelni, nem kell mások racionalitásáról annyit pszichologizálni, mint az egyes játzmák Nash-egyensúlyai esetében. És legfőképp, a szereplők számára megoldja a sok egyensúly jelenlétéből eredő bizonytalanságot. A játékosoknak azt kell figyelniük, hogy melyek azok a játékos-társaik által is látható-hallható utasítások, amelyekre a többiek is figyelnek, és ezek mit tartalmaznak az épp soron lévő játzmára vonatkozóan.

Amennyiben a társadalmi normák és intézmények imént idézett új elméletei jól értelmezik a korrelált egyensúly szerepét, az a játékelmélet régi nagy elméleti kérdésére, (az egyensúlyok közötti választás dilemmájára) is adhat egy, a korábbiaktól eltérő szemléletű választ. A játékelméleti kutatások főáramához tartozó közgazdászok között is vannak olyanok, akik így gondolják. Például a játékelmélet eddig e könyvben nem említett Nobel-díjas kutatója, Roger Myerson (Klarreich 2017, Myerson 1999), Gintishez (2009) hasonlóan, a Nash-egyensúlynál természetesebb koncepciónak tartja a korrelált egyensúlyt a játékosok informáltságának és kalkulációs képességének realistább modellje miatt.

A normák játékelméleti alapú elemzése során a korrelált egyensúly fogalmától függetlenül is megjelent a gondolat, miszerint a játzmák résztvevői számára kulcsfontosságú annak az információnak a figyelése, hogy milyen szabályok vonatkoznak az épp aktuális, er-

kölcsileg releváns élethelyzetre. Bicchieri (2006, lásd még Bicchieri és Sontuoso 2015) a személyközi kommunikáció normákra fókuszáló szerepéről beszél (erre még visszatérek). Az általa leírt egyensúlykeresési mechanizmus kompatibilis a korrelált egyensúlyhoz vezető úttal.

#### VI.2.3.2. Közös tudat

A könyv előző fejezeteiben is hangsúlyoztam annak a jelentőségét, hogy az emberi játzmák szereplői mit gondolnak arról, hogy mit gondol a többi szereplő az adott játzmáról, és arról, hogy a többiek mit gondolnak. A társadalmi normák játékelméleti elemzése kapcsán nem halogatható tovább ennek a kérdésnek a rövid, de szisztematikus megtárgyalása. Ugyanis a sok ember viselkedését koordináló utasítások kapcsán felmerül, hogy a közös társadalmi térben észlelhető ideák tőlünk független entitások, amelyeket nem lehet kezelni az erkölcs kutatásának eddig bemutatott módszertani individualista eszközrendszerével. Ez az elgondolás explicit módon fel is merül a normák játékelméleti irodalmában (Gintis 2009, Hedoin 2015). A helyzet azonban ennél egy fokkal összetettebb.

A vonatkozó irodalomban David Lewis filozófus (1969) nevéhez fűződik a társadalmi normák, illetve konvenciók értelmezéséhez szükséges interaktív ismeretelmélet fogalmi alapjainak lerakása. Az egymás gondolataira vonatkozó hiteink formális modelljét a matematikus-közgazdász-filozófus Aumann (1976) nevéhez kötik. Bár ahogy Binmore (2015) rámutatott, az ultimátumjáték előzményei kapcsán a IV. fejezetben már említett Morris Friedell (1969) szociológus, Aumannéhoz hasonló eszköztárral, már korábban definiálta formálisan is az interaktív ismeretelmélet kulcskonceptióját, amelyet köztudott tudásnak hívunk. Ennek van némi elméletörténeti jelentősége, hiszen Aumann részben épp az interaktív ismeretelmélet formális alapjainak lerakásáért kapott közgazdasági Nobel-díjat 2005-ben.

Friedell a koncepció explikálásakor épít Thomas Scheff (1967) szociológusnak az *American Sociological Review*-ban a konszenzusról

megjelent tanulmányára, amelyben a mások hiteire vonatkozó hitekről is értekeznek. Egyébiránt a Friedell által elhagyott, világhírű michigani survey-kutatási központ munkatársait is érdekelték ekkoriban a mások hiteire vonatkozó hitek (Fields és Schuman 1976). A koncepcióval kapcsolatos szociológiai érdeklődésre mindjárt vizsgálatok születtek.

De miről is van szó? Lewis (1969) a konvenciók természetével foglalkozó elemzésében megkülönböztette azt a helyzetet, amikor egy esemény bekövetkeztéről mindenki tud, attól az állapottól, amikor ezen túlmenően mindenki tudja, hogy mindenki tudja, hogy mindenki tudja stb. a végtelenségig, hogy az esemény bekövetkezett. Az előbbi kölcsönös tudásnak nevezte el, az utóbbit pedig közös- vagy köztudott tudásnak (common knowledge). Friedell (1969) és Aumann (1976) annak lehatárolásához nyújtott matematikai eszköztárat, hogy egy adott helyzetben miként viszonyulhatnak egymáshoz a kölcsönösen tudott és a köztudott tények. Nyilvánvaló, hogy többnyire sokkal több dolgot tudunk mindannyian, mint amennyiről tudjuk, hogy mindenki tudja, hogy mindenki tudja stb., hogy mindenki tud róla.

A köztudott tudás valódi „game changer”. A fogalom használata nélkül a visszafelé göngyölítés észszerűségének feltételezéséhez nem elegendő (III. fejezet), hogy mindkét fél felismerje, hogy a százlábú játékban már el sem érdemes indulni. Az egyik félnek tudnia kell a másikról, hogy az tudja róla, hogy ő tudja a másikról stb. A Lewis (1969) által tárgyalt konvencióknál maradványként, tegyük fel, hogy az intenzív kamionforgalommal terhelt budapesti Hungária körúton vágok át keresztbe gépjárművel, egy üres keresztútca felől, lendületből, az épp zöldre váltó közlekedési lámpa bátorítására. (Ez is egy jólétnövelő, korrelált szignál! Képzeljük el, ha két külön cég egymástól függetlenül írná a két irány lámpáinak programjait!) Én akkor megyek be nagy lendülettel a keresztirányban közlekedő kamionok, és a mellettük lavírozó száguldozók között a nekünk adott zöld (keresztforgalomnak adott piros) jelzés miatt épp szabaddá váló sávba, ha tudni vélem, hogy az oldalról érkezők is tisztában vannak a közlekedési lámpa színeinek jelentésével; sőt, úgy gondolják, hogy én is tisztában vagyok vele, és azzal is, hogy ők tisztában vannak vele

ezért bátran behajtok, tudva, hogy onnan nem akar senki az utolsó utáni pillanatban átslisszolni. Ha nincsen elegendően magas szintű (tudom, hogy tudják, hogy tudom...) magabiztos hitem, akkor félhetek attól, hogy a keresztbe közlekedők egyike a félelmemet feltételezve, azt kihasználva, még átslisszol, mert úgymint körülnézek és fékezek, hogy meggyőződjek, hogy minden keresztbe közlekedő megállt-e már, és vajon nem akar-e még valaki átslisszolni (és igen, ilyen hitek mellett a vállalható kockázatnál nagyobb eséllyel lesz olyan, aki még áthajt a friss piroson).

Ugyanakkor a köztudott tudás egy elméleti konstrukció. A mások hiteire vonatkozó hiteink egy elméleti ideáltípusa. Az imént emlegetett visszafelé göngyölítés kapcsán ezért is merült fel, hogy nehéz egy olyan, a játzsmák racionális lejátszására vonatkozó megoldást laboratóriumban tesztelni, amely feltételezi, hogy például egy játzsmalejátszásának észszerű lépései a köztudott tudás részét képezik (Aumann 2000, Binmore és Shaked 2010). Azt feltételezhetjük, hogy mindenki tudja, hogy mindenki tudja, stb., hogy mindenki látja, hogy miképpen néz ki a kísérleti alanyok elé tett döntési fa. Az a probléma, hogyha ezen túlmenően azt is feltételeznünk kell, hogy mindenki képes levonni a döntési fából a helyes következtetést, ezt mindenki feltételezi a többiekéről, és azt is, hogy mindenki feltételezi a többiekéről, hogy feltételezik a többiekéről stb.

Bicchieri (2006) a normák játékelméleti értelmezéséhez kapcsolódva vizsgálta, hogy az emberek milyen szintig képesek tudatosan kezelni a mások gondolkodására vonatkozó hiteiket. Eredményei szerint az emberek két-három szintig képesek gondolkodni. Sokszor ez is elegendő magabiztosságot ad koordinációt igénylő helyzetekben. Ráadásul vannak olyan kommunikációs struktúrák, amelyek – amennyiben érzékeink és a fizika alapvető szabályait mindenki által elfogadottnak tekintjük –, minden, mások gondolataival kapcsolatos további morfondírozás nélkül biztosíthatják bizonyos tények kvázi köztudott tudássá válását. Ilyen helyzet, amikor körbeállva látjuk és halljuk valakinek az állításait (Binmore 1991, Chwe 2001). Az utóbbi néhány évben zajlottak a köztudott tudás első empirikus pszichológiai vizsgálatait, amelyekben a kutatók épp arra voltak kíváncsiak,

hogy a játékelmélet előrejelzéseinek megfelelően magabiztosabban koordinálják-e viselkedésüket az emberek, ha kölcsönös tudás helyett a közös kommunikációs térben köztudott tudássá tesznek a koordináció szempontjából fontos tényeket (Thomas et al. 2014). Az első eredmények szerint az emberek többsége valóban reagált arra, hogy valamilyen koordinációval kapcsolatos ismeret (potenciálisan) köztudott tudássá vált.

A korrelált egyensúly fogalmának használata magával hozta a szignálokra vonatkozó magasabb rendű hitek jelentőségét. Gintis (2009) hangsúlyozza, hogy nagy csoportokban a korrelált egyensúlyok létrejöttéhez a játék szereplőinek kevesebb információra van szüksége mint a szignál nélküli Nash-egyensúlyok eléréséhez. Ez azért van, mert az egyes egyénekre vonatkozó pszichologizáló spekulációk helyett a szereplőknek a koreográfusok mindenki által megfigyelhető utasításaira kell figyelniük. Azokat viszont nem önmagukban kell figyelniük, hanem azt is nézniük kell, hogy mások is „veszik-e az adást” – és a többiek is tisztában vannak-e azzal, hogy mindenki veszi-e az adást. Az eddigiek alapján ez azt jelenti, hogy azoknak, akik azt gondolják, hogy az erkölcsös (állampolgári) magatartás fő mozgatórugói a korrelált egyensúlyok utasításai, elegendő csak a „józan paraszti ész” feltételeznük az egyénekről, és vizsgálataikat az emberek közötti kommunikáció struktúrájára (Chwe 2001), és a különféle módokon kialakult közös normák tartalmára koncentrálhatják.

Úgy tűnik, ezzel visszaérkeztünk a könyv elején megidézett Émile Durkheim szemléletmódjához. A játékelmélet keretei között született interaktív ismeretelmélet, és Durkheim közös tudatra vonatkozó elméletének összekötése nem új ötlet. A fent idézett Scheff (1967) és Friedell (1969) is sokat foglalkozik Durkheimmel. A racionális döntésemélet szociológiai alkalmazásának egyik hazai úttörője, a tragikusan korán elhunyt Csontos László ezt több mint három évtizede megtette (Csontos 1986). Durkheim észlelte azt az ellentmondást, hogy az egyes emberek fejében összeálló gondolatok kényszerítő erővel bíró közös gondolatokká állnak össze. Nem voltak meg azonban az eszközei arra, hogy ezt a folyamatot egzakt módon megragadja.

Mindazonáltal nem ezen a ponton szeretném befejezni az erkölcs forrásának vizsgálatát. A következtetések levonása előtt szeretném még egyszer megvizsgálni a korrelált egyensúlyok természetét az egyéni ösztönzők szempontjából. Elemzésem ezért továbblép a társadalmi normák játékelméleti alapokon nyugvó újabb elméleteinek (Gintis, 2009, 2016, Gintis és Helbing 2015, Guala 2016) kollektivisták konklúzióin. A továbblépés Binmore (2015) figyelmeztetésének szellemében történik: a korrelált egyensúlyt ne tekintsük önmagától működő csodaszernek, hanem vessük alá racionális analízisnek. Ám előtte még megvizsgálom, hogyan keletkezhetnek a korrelált egyensúlyokhoz vezető utasítások.

#### VI.2.4. A KOREOGRÁFIÁK KOREOGRÁFIÁJA

##### VI.2.4.1. Hogyan keletkeznek a szabályok?

A társadalmi normákat értelmező szakirodalomban egymással szemben álló elméletként tűnik fel a Nash-egyensúly apró korrekciók során történő elérését feltételező elmélet (Binmore 2005, 2010), és a korrelált egyensúly koncepciójára építő koreográfuselmélet (Gintis 2009, 2016, Guala 2016). Én úgy vélem, hogy ezeket az egyensúly-elérési mechanizmusokat nem szabad egymást kizáró modelleknek tekinteni. A mindennapi, apró-cseprő interakciók normáinak kialakítása nem feltétlenül ugyanúgy zajlik, mint a klímaváltozásra adott erkölcsös polgári reakciók megvitatása. Az egyensúly elérésnek több módja lehet, és az idézett elméletek segítségével könnyű azonosítani az alapvető típusokat.

A Binmore (2005) által feltételezett korrekciós mechanizmus egy szélső ideáltípus. Ezt nevezhetjük hallgatólagos adaptációnak is. Az ilyen típusú alkalmazkodás nem igényli a normák diszkusszióját. Egymást figyeljük, hogy ki miként reagál egy-egy szituációban, és a sikeresebb, magasabb státuszú egyének reakciója elterjedhet a releváns csoportban. Ezzel párhuzamosan, az egyes játszmákban a partnereink – a téma explicit tárgyalását nélkülöző –, reakciói igazítanak útba

minket. Egy-egy fizikai ütközés, egy-egy csodálkozó, megdőbönt, mérges vagy undorodó tekintet, vagy az elítélő hanghordozás elegendő, hogy értésünkre adja, hogy letértünk az egyensúlyi pályáról. Binmore példáiból kiderül, hogy nem akarja a filozófusokat követni abban, hogy különbséget tesz normatív erővel nem rendelkező konvenciók (Lewis 1969), és kollektív problémákat megoldó normák (Bicchieri 2006, Ullmann-Margalit 1977) között. Ezért hivatkozik többek között a nyelvire (Binmore és Shaked 2010). A következők alfejezetben részletesen tárgyalni fogom a normativitás kérdését. Itt csak annyit jegyzek meg, hogy azok a racionális döntésemélet felől közelítő elméletek, amelyek a nyelvi konvenciót elemzik, valóban normatív erő nélküli konvencióként tekintenek a nyelvire (Mühlenbernd és Franke 2012, Nowak et al. 1999, Trapa és Nowak 2000). Nagy a csábítás arra, hogy a hallgatólagos adaptáció egyéb példáit is normatív erő nélküli konvenciónak tekintsük. Ugyanakkor minden szociológus és szociolingvisztikus számára nyilvánvaló kell legyen, hogy a nyelvi kifejezésformáknak, a kiejtéstől a szóhasználaton keresztül a mondatok összeállításáig, erős normatív tartalma lehet (Bourdieu 1991 [1982]). A koreográfusok kétségtelenül megvitatják ugyan a nyelvi kifejezésmódok közötti normatív választások egy részét a nyilvános térben, de nem lehet tagadni a hallgatólagos szabályok létezését sem. Ez talán az akcentusok, és az egyazon objektumokat jelölő szinonimának tekinthető szavak közötti választás esetében a legnyilvánvalóbb.

A másik szélsőséget a Gintis (2009) által bemutatott koreográfia centralizált formája képviseli. Ebben az esetben egy adott csoport elitjének tagjai, a csoport közös kommunikációs terében nyíltan tárgyalják meg, hogy egy-egy adott szituációban melyek az elvárt viselkedési formák. Az ilyen normameghatározó, illetve -megerősítő aktusok jelentőségét egyes kulturális antropológusok már régen felismerték (Bloch 1986, Mauss 1954 [1923], Turner et al. 1967). A köztudott tudással foglalkozó kortárs társadalomtudósok is felfedezték ennek jelentőségét. A közgazdász Michael Suk-Yung Cwhe (2001) részletes tanulmányban szedte össze a vonatkozó kulturális antropológiai irodalmat, és helyezte a normamegerősítő aktusokat racioná-



lis döntéseméleti (konkrétan interaktív ismeretelméleti) keretbe. A közösségek spirituális vezetőinek a közösség színe előtt tartott példabeszédei a Gintis (2009, 2016) által meghatározott koreografálás ideáltipikus példái. Nagyon fontos, hogy e beszédek nem külön-külön hallgatják meg a közösség tagjai, hanem együtt, egy helyen: így nyilvánvaló mindenkinek, hogy mindenki hallja a beszédet, és mindeközben láthatják a is többiek reakcióit. Tehát mindenkinek nyilvánvaló, hogy mindenkinek nyilvánvaló hogy..., és így tovább, a végtelenségig. A normasértések nyilvános büntetése, és az, hogy a közösség tagjai ezt belenyugvással veszik tudomásul, hasonló hatást érhet el. A nagyméretű, modern társadalmakban a fizikai együttlét, és a köztudott tudás elvi koncepcióját megközelítő közös tudat létrehozása nehezebb. Azonban a tömegkommunikáció csatornáiban olyan megszólalási lehetőséget nyújtanak e társadalmak elitjének, és olyan mértékben köztudott, hogy e csatornákat valamilyen szinten követi az emberek jelentős része, hogy az itt elegendő intenzitással megjelenő üzenetek esetében jó közelítéssel tartható a korrelált egyensúlyhoz szükséges feltevés, miszerint „tudom, hogy mindenki tudja stb., hogy mindenki hallotta az üzenetet”.

Úgy vélem, hogy mindkét tiszta modell finomítható, és a koreográfiák közlésének köztes formái is elkülöníthetők. Részletes kommunikációelméleti elemzés helyett legalább egy köztes ideáltípus elkülönítése is elegendő az átmeneti egyensúlykeresési módok természetének megértéséhez. Decentralizált diskurzusnak nevezhetjük a társadalmi normák olyan tárgyalását, amelyet nem az elit irányít, azonban túllép a hallgatólagos megegyezéssel történő korrekciókon. Sok olyan apró szabály létezik, amelyről nem feltétlenül beszélnek korunk felkent papjai a tömegkommunikáció elektronikus szószékeiről. Ugyanakkor ezek a szabályok felmerülnek a mindennapi beszélgetéseink során, és megtárgyaljuk, hogy így vagy úgy cselekedni normális-e (vö. Bicchieri 2006, Bicchieri és Sontuoso 2015). E diskurzusoknak nem kell feltétlenül kifinomult normaértékeléseknek lenniük, és az értékelt szituációtól függ, hogy mennyire tartoznak hozzá a bántó normasértéseket kísérő indulatok (ra vonatkozó állítások). Szabadon cseveghetünk arról, ki hogyan szereti inni kedvenc italát. Saját döntésem

miatt pironkodva elárulom a társaságban, hogy jégkockával iszom a sört. A Binmore-féle (2005) hallgatólagos adaptációt súroló esetként elegendő lehet egy megfelelő hangsúlyozással odavetett megjegyzés: „Tényleg?” – és már tudom, hogy ez nem kifejezetten menő dolog, pedig többet nem esik szó erről a normasértéséről.

A decentralizált diskurzus másik szélső formája, amikor egy társaságban részletesen pletykálunk távol lévő személyek cselekedeteiről, és megpróbálunk konszenzusra jutni abban, hogy a helyzetükben mi a helyes, és mi a helytelen cselekvés. A pletyka racionális döntéseméleti szakirodalma (Greif 1994, Piazza és Bering 2008, Sommerfeld et al. 2007) elsősorban a pletyka azon lehetséges funkcióját emeli ki, hogy információt nyújt fixnek tekintett normák megsértéséről. Erről a könyv III. fejezetében szó volt. Azonban a koreográfia elméletet átgondolva logikusan felmerül, hogy a pletykálás normameghatározó szerepet is játszhat. Pletykálkodásunk egy része nem csupán monoton szörnyülködés (ilyen is van, persze), hanem olyan vita, ahol a terítékre kerülő szereplők és tettek mellett is elhangoznak érvek. „Meg kell érteni Tamást, nem rosszindulatból csinálta ezt, hanem...”; „Nagyon nehéz ügy ez, magam sem tudom, hogy mit csináltam volna ilyen helyzetben”. Az ilyen típusú pletykálkodások centralizált papoláshoz leginkább hasonló formája az, amikor a pletykálkodó szereplők elszakadnak a pletyka terítékére került cselekményektől, és már nyíltan határozzák meg a hasonló szituációkban követendő erkölcsös viselkedés szabályait – a jeges-sörös példához hasonlóan tesztelve közönségüket a szabály érvényességére vonatkozóan: „Én úgy gondolom, hogy az alap, hogy beteg gyermeket nevelő anyát nem hagyunk magára, bármilyen csábításnak is vagyunk kitéve.” Lényegében a társas kommunikáció ilyen funkciójáról beszél Bicchieri (2006, lásd még Bicchieri és Sontuoso 2015).

Azért hoztam ezt az utóbbi példát, mert ismét hangsúlyozni akarom, hogy az egyensúlyhoz vezető diskurzus típusa nem determinálja az adott cselekvéshez tartozó normatív nyomást. Tehát, ahogyan a hallgatólagos adaptációhoz is tartozhat erős érzelmi reakciókon keresztül történő ösztönzés, úgy a diskurzus része is lehet leíró normák közvetítése. Azaz, ahogy akár a jégkockás sörözés (szerintem ártal-

matlan!) szokását is súlyosan büntetheti egy megsemmisítő pillantás, úgy a legutoljára idézett, a pletykakörben részletesen diszkusszált normatív kijelentést is ellenpontoszhatja valaki leíró normával. Így jelzi, hogy távolra kerültünk a pletykakörben kíváncsnak tartott egyensúlytól, és a realitás egy másik egyensúly – tekintettel arra, hogy a diskurálók decentralizált köre nem képes központi koreográfusként meghirdetni az erkölcsi üzenetét a korrelált egyensúly reményében a teljes csoportban: „Figyelj, a férfiak már csak ilyenek, nincs bennük tisztesség, mindegyik a micsodája után megy”. E leíró kijelentésnek következménye van arra vonatkozóan, hogy a pletykálásban résztvevők milyen stratégiát tarthatnak optimálisnak a saját életükben. Ha hisznek a leíró kijelentésnek, akkor úgy találhatják, hogy a fenti, erkölcsi érzéküket sértő viselkedés esetén az általuk amúgy kívánatosnak tartott kiközösítésnek nincsenek jó esélyei a tágabb közösségben, ezért erre a fenyegetésre nem építhetik fel a saját párkapcsolati stratégiájukat.

#### VI.2.4.2. Honnan jönnek a szabályok?

Már most látszik a normák játékelméleti elemzésének egyik fontos tanulsága (amelyre a könyv végén még visszatérek), miszerint a korrelált egyensúlyokat vezénylő koreográfusok játéktere meglehetősen nagy – és az eddigi elemzés még távolról sem merítette ki az összes lehetőséget. Fontos kérdés, hogy a külső szemlélő felfedezhet-e valamilyen logikát abban, hogy milyen elvek válnak normává.

A IV. és V. fejezetben bemutatott pszichológizáló elméletek egyértelmű választ adnak erre a kérdésre. Normatív hajlamaink viszonylag pontosan meghatározzák, hogy mit várunk el egy-egy szituációban magunktól és embertársainktól. Ám ebben a fejezetben éppen azt az elméleti megközelítést tárgyaljuk, amely tagadja, hogy – a kézfogás szabályaitól kezdve a közpolitikai preferenciáig bezárólag velünk született utasítások határozzák meg erkölcsös polgári viselkedésünket.

Az élet apró játszmáira vonatkozóan erősek Binmore (2005) többször említett feltevései a folyamatos korrekcióról. E korrekciókra el-

vileg építhetőek adaptív cselekvőket feltételező modellek, amelyek elvezethetik a kutatót a specifikus egyensúlyokhoz. A tortaelosztási alkukra vonatkozóan Binmore (2005) kísérletet is tett az egalitárista elosztás kialakulásának racionális megértésére.

A dán politikatudós Michael Bang Petersen és szerzőtársai (2012) nem közvetlenül Binmore-ból kiindulva, hanem a méltányos elosztásra vonatkozó preferenciák egyéb elméleteire építkezve feltételezik, hogy a mikroszintű interakciókban használt stratégiáinkat átvisszük makrotársadalmi problémák értelmezésére is (ez összhangban van Binmore korlátozott észlelési-tanulási képességre vonatkozó feltevéseivel). Petersenék méltányosságra vonatkozó heurisztikákról beszélnek, amelyek peer-to-peer kapcsolatainkat irányítják. Ezeket a heurisztikákat visszük át komplexebb társadalmi problémák értelmezésére. Konkrétan a különböző rászoruló csoportok állami megsegítésére vonatkozó attitűdöket vizsgálták, nagyjából a Fong és szerzőtársai (Fong 2001, Fong et al. 2005) által kijelölt ösvényen haladva. Azt találták, hogy a korábban keményen dolgozó támogatottak juttatásait inkább emelnék az emberek, míg a jóléti rendszeren élőködőnek tűnő munkanélküliektől sajnálják a szavazók a közpénzt.

Nem tudok olyan empirikus eredményekről beszámolni, amelyek cáfolnák a mikrointerakciókra alapozott heurisztikák normaképző erejét. Azonban még a tézist támogató eredményekből is kevés van. További tesztekre van szükség ahhoz, hogy bátrabb kijelentéseket tegyünk erről a feltételezett mechanizmusról. Kézenfekvő kritika a tézissel szemben, hogy azok a preferenciák, amelyek különbséget tesznek a jóléti rendszer érdemes és érdemtelen kedvezményezettjei között, valójában nem általános méltányossági heurisztikákat tükröznek, hanem csak konkrét viselkedésre vonatkozó társadalomfüggő normák betartását (megsértését) jutalmazza (bünteti). A peer-to-peer interakciókra koncentráló, a IV. fejezetben ismertetett laboratóriumi kísérletek e könyvben is hangsúlyozott konklúziója az, hogy leginkább normasértést büntetnek az emberek, és nem univerzális méltányossági elveket követnek.

Aarøe és Petersen (2014) egy fontos közelmúltbeli eredménye azt mutatja, hogy az USA és Dánia polgárai hiába élnek olyan or-

szágokban, amelyek radikálisan eltérően kezelik a szegénységet, és ahol a választópolgárok is radikálisan eltérően ítélik meg a szegények helyzetét, méltányossági attitűdjeik mégis közös nevezőre hozhatók (lásd IV.2.3.2.). A vizsgálatuk mögött álló elmélet azt mondja, hogy a két ország polgárainak eltérő jóléti attitűdjei mögött a szegényekre vonatkozó, eltérő sztereotípiák állnak. Az amerikaiak azt gondolják, hogy a szegények lusták, és ezért a méltányossági heusztikák alapján nem méltóak támogatásra (vö. Gilens 1999). Dániában kevésbé elterjedtek ezek a negatív sztereotípiák, és inkább külső okokkal magyarázzák a szegénységet. Tehát ugyanazokból a méltányossági elvekből vezethetők le a dánok és az amerikaiak egymáséttól jelentősen eltérő közpolitikai preferenciái. Kísérletükben fiktív helyzetek bemutatásával megszüntették a bizonytalanságot: nem hagyták, hogy a válaszolók sztereotípiái dolgozzanak, hanem pontosan leírták (meglehetősen szélsőséges módon), hogy mennyire lusta vagy szorgalmas az az illető, akinek az államtól kapott juttatásairól kérdezik a válaszadókat. A konkrét személyiségleírások segítségével sikerült eltüntetni minden különbséget az amerikaiak és a dánok véleménye között.

Látnunk kell azonban azt, hogy bár ezek az eredmények összhangban vannak az univerzális heusztikák elméletével, de nem cáfolják a minden univerzalitástól mentes normák tézisének sem. Lehet, hogy a megfelelően szélsőséges példák bemutatásával egy platformra hozták két kultúra tagjait, akik egy adott korszak globális nézőpontjából viszonylag közel állnak egymáshoz, ám változhat a helyzet, ha időben és térben kilépünk ebből a körből. Például több országban indult el intenzív gondolkodás az univerzális alapjövedelem bevezetéséről. Azt tudjuk, hogy a jóléti intézmények korrelálnak a jóléti politikára vonatkozó attitűdökkel (Albrekt Laresen 2006). Nem tudhatjuk, hogy az alapjövedelem nem forgatná-e fel teljesen az érdemesség jelenlegi nyugati jóléti rendszerekhez igazodó koncepcióját.

Továbbá ismét szeretnék visszautalni a IV. fejezet tanulságaira. Nem igazán bizonyítható az sem, hogy a tortafelosztás jól definiált helyzetében az evolúciós érvekkel megerősített egalitarianizmus akár velünk született, akár tanult döntést befolyásoló elv lenne. Ráadá-

sul a társas interakciók általános szabályozó mechanizmusait vizsgáló kutatók előszeretettel koncentrálnak olyan helyzetekre illetve elvekre, ahol a szereplők közötti szimmetria adott, vagy létrehozható. Mindeközben akármerre nézünk időben és térben, a társadalmak tele vannak a javak tartósan egyenlőtlen elosztását eredményező egyensúlyi helyzetekkel. Ezekhez persze meg lehet találni az egyenlő csere ideológiáját – ez sok esetben létezik is az adott kultúrában a szituáció értelmezésére. De a normák elveinek gyökereit kereső kutató számára nem hangzik túlságosan megnyugtatóan az olyan érvelés, miszerint minden társadalmi norma, általában a hozzájárulások figyelembevételével, méltányos/egyenlő jutalmak kiosztására törekszik, azonban az egyes társadalmak különbözhetnek abban a tekintetben, hogy különböző rétegek konkrét hozzájárulását mennyire értékelik. Így bármilyen konkrét normára ráfoghathatjuk, hogy méltányosság elveit követi. Például a szülésznők a miénkhez hasonló, mai európai kultúrában komoly megbecsülésre méltó munkát végeznek, míg az indiai civilizáció egyes kultúráiban a prostituáltakkal azonos, „tisztátalan” kategóriába esik a tevékenységük társadalmi értéke (Minturn 1984).

Megfelelő empirikus tesztek hiányában általánosabb szinten is megfogalmazható elméleti kritika a mikro-interakciókhoz kötődő ösztönök és heurisztikák komplex normákba épülésével kapcsolatban. A 21. század elejének társadalomtudományos gondolkodásában kétségkívül befolyásos közelítés az, hogy a civilizációs normáink gyökereit az emberiség történetének hosszú hajnalán elménkbe égett ősi elvekben, vagy kifejezetten a „zsigereinkben” keressük (lásd a IV. és V. fejezetet). Ugyanakkor más korszakokban sokkal népszerűbb volt ennek a nézetnek az ellenkezője: a civilizáció normái éppen az állatias ösztöneink elnyomására születtek. Mai nyelvhasználatunk és a civilizáció hétköznapi értelmezése egyébként ez utóbbi értelmezést támogatja.

Láttuk, hogy a modern társadalmak normáit nem egyszerű közvetlenül visszavezetni mikro-interakciók „törvényszerű” szabályaira. Gintis (2016, lásd még Gintis és Helbing 2015) azonban hisz abban, hogy a bármilyen módon kialakult szabályok, a parsonsi elméletnek megfelelően, generációról generációra enyhén módosulva, lassan

változnak. Kivételes esetekben megindul a korrelált egyensúlyokra vonatkozó utasítások újratárgyalása, amely gyors változásokat eredményezhet. Azonban egyrészt látni kell, hogy a normák eredetének ilyen típusú vizsgálatához sokat nem ad hozzá a játékelméleti megközelítés. Másrészt, a szociológiában és a kulturális tanulmányok területén több évtizede gyűlnek a bizonyítékok a Parsons-féle fokozatos érték- és kultúraváltozás modellel szemben (Lamont et al. 2014, Patterson 2014, 2015).

A kortárs kultúraelmélet (Patterson 2014) azt mondja, hogy a napi interakciók során a meglévő tudáskészletek folyamatosan újraértékelődnek, csiszolódnak. Sokkszerű környezeti változások nyomán a kollektív tudáskészlet is gyorsan változhat – bár nem szükségszerű, hogy mindenfajta mintázat esetében történjen ilyen gyors változás. A klasszikus elmélet szerint például az etnikumspecifikus kulturális mintázatok az etnikai csoport kollektív történetével magyarázhatóak, ezért ezek az egész csoportra jellemző sajátosságok. Mára elég sok kutatási eredmény bizonyítja, hogy az etnicizált szegénység által érintett társadalmakban nem határolható le valamiféle kisebbségi szegénykultúra. Sokféle kulturális miliő él egymás mellett (Patterson és Fosse 2015). Nincs egységes cigánykultúra, szegénykultúra, vagy szegénycigány-kultúra. A kultúra kortárs kutatóinak többsége tagadja, hogy minden kulturális folyamat a körülményekhez történő kollektív alkalmazkodást segítené elő. De az széles körben elfogadott, hogy egyes kulturális folyamatok részei a kollektív (és gyors) alkalmazkodásnak. Az is igaz, hogy más folyamatok erősebben állnak ellen a környezeti változásoknak.

Amit a kultúra új elmélete hangsúlyoz: a tudáskészletünk, cselekvési repertoárunk lesz jellemzően „ragadósabb”, hiszen annak határai nem is tudatosulnak bennünk (mert definíciószerűen úgy tudjuk, hogy tényszerűen azok a cselekvési határaink). A tudatosult értékeinket a parsonsi elmélet feltételezésével szemben könnyebben megváltoztatjuk, hiszen azokról tudjuk, hogy választások, tehát létezik alternatíva.

A kultúra új elméletének eredményei alapján az a sejtés körvonalazódik, hogy a hallgatólágos adaptációra építő egyensúlyépítés

követhet leginkább racionálisan rekonstruálható dinamikát. Egy jó példa lehet erre a nyelv és nyelvhasználat fejlődése, amely többé vagy kevésbé normatív konvenciókból épül fel, de nem teljesen kaotikusan. A nyelvészet képes jól meghatározni a dinamikára és a belső szerkezetre vonatkozó törvényszerűségeket (Jackendoff 2003, Reali és Griffiths 2010). A hallgatóságos adaptációhoz közeli peer-to-peer diskusszió alapuló koreográfiák is viszonylag nagy arányban követhetnek bizonyos evolúciósan stabilizálódott döntési reflexeket, heurisztikákat. A koreográfiák létrehozásának másik szélső típusa – sejtésem szerint –, azonban meglehetősen függetlenné tud válni a mikro-interakciók logikájától. Ha egy szűk körű, belső kooperációra képes (mert például belül is erősen hierarchizált) elite odafigyelnek, amely számos egyensúly közül választhat, akkor miért pont az általa nem is ismert kulturális-genetikai koevolúció által diktált egyensúly felé vezetné a népet? Miért is nem arra, amerre a saját érdekei kívánják? És az érdekei milyen logika alapján kell, hogy egybeessenek az útfüggőségre és a lokális optimalizációra építő, hallgatóságos adaptációs logika által kiválasztódó társadalmi egyensúlyi pályával?

A centralizált koreográfia látványos példái a modern demokráciák politikai közösségeiben hirdetett értékek. A III. fejezetben írtam arról, hogy az egyének hétköznapi interakcióikra jól alkalmazható önérdékalapú döntési modellek nehezen magyarázzák az erkölcsös megoldások választását akkor, amikor az emberek a modern nagytársadalmak szavazóiként teszik le a voksukat a tágabb közösséget befolyásoló egyik vagy másik stratégia mellett. Az alábbi alfejezetben azt mutatom be, hogy modern demokráciák választóinak pusztán önérdéken túlmutatón viselkedése könnyen és empirikusan alátámaszthatóan értelmezhető úgy, mint a politikai elitek által meghatározott normák követése. Ezen a területen a normakövető szavazó modellje versenyképes a méltányos szavazó és az ösztönös szavazó modelljével – sőt, a szavazói preferenciákra vonatkozó empirikus eredmények bizonyos kérdésekben épp e modellt erősítik, cáfolva a másik kettőt.



### VI.2.5. NEM MEGYEK A TE UTCÁDBA! A NORMAVEZÉRELT SZAVAZÓ

#### VI.2.5.1. Pártpolitikai identitás és közpolitikai attitűdök

A saját célokkal (érdekekkel és értékekkel) szemben társadalmi normákat követő szavazói attitűdök megértéséhez a politikai elitek véleménybefolyásoló szerepének megértésén keresztül vezet az út. Nem csak a politikai közösségekben kialakuló normákat követhetik a szavazók, és a politikai identitáshoz kötődő normák sem csak a politikai elit által koreografált vélemények lehetnek. Azonban a normakövetésnek és -meghatározásnak ez a mechanizmusa a leginkább alkalmas arra, hogy elválasszuk az attitűdökben a normák hatását a szavazók belső motivációinak következményeitől.

Mára már jól dokumentált jelenség, hogy a politikai eliteknek milyen hatása van az átpolitizált kérdésekre vonatkozó szavazói attitűdökre (Bullock 2011, Carsey és Layman 2006, Highton és Kam 2011). Az utóbbi évtizedben kísérleti módszertant alkalmazó vizsgálatok új hullámát indította el az attitűdpolarizációval kapcsolatos megnövekedett érdeklődés (Slothuus 2010). Ebben az évtizedben számos kísérlet eredmény erősítette meg a tézist, miszerint a pártok vezetői által hangoztatott álláspontoknak más tényezőktől független, önálló hatása van a szavazók attitűdjeire (Bisgaard és Slothuus 2018, Bolsen et al. 2014, Boudreau és MacKenzie 2014, Bullock 2011, Ciuk és Yost 2016, Druckman et al. 2013, Leeper és Slothuus 2015).

Ennek a kutatási programnak néhány kutatója azt is vizsgálta, hogy mi motiválja a szavazókat arra, hogy a számukra kedves párt elitjének véleményét kövessék (Han és Wackman 2017, Malka és Lelkes 2010, Petersen et al. 2013).

A politikatudományban a pártelit véleménybefolyásoló szerepét hagyományosan a már említett Downs-féle (1990 [1957]) racionális szavazói modell keretei között értelmezték (Leeper és Slothuus 2014). A racionális szavazók bizonyos ügyekben azért követik a pár-

tok álláspontjait, mert más ügyekben megbizonyosodtak arról, hogy az adott párt az ő érdekeik és értékeik szerint alakítja ki álláspontját. A szavazók ezzel időt és energiát takarítanak meg (Bullock 2011, Slothuus 2010, Zaller 1996). Értelemszerűen, a racionális választópolgárok a számukra fontos ügyekben saját álláspontjuk alapján választanak pártot. A kevésbé fontos ügyekben ugyanakkor hajlamosak a választott párt álláspontját elfogadni (Ciuk és Yost 2016).

Az utóbbi időben azonban néhány politikatudós elkezdte hangsúlyozni, hogy egy másik motivációs mechanizmus, a pártpolitikai identitás is érdemi szerepet játszhat a politikai elit véleményformáló hatásában (Malka és Lelkes 2010, Iyengar et al. 2012). Szerintük a szavazók érzelmileg kötődnek az általuk preferált pártpolitikai közösséghez, és ennek közvetlen hatása van az attitűdjük kialakítására. A csoportkötődés hatására az emberek önsztereotipizálásba kezdenek, amelynek eredményeképp a közpolitikai véleményeik az általuk érzékelt csoportnormákhoz igazodnak. Praktikusan az érzékelt tipikus pártpolitikai álláspontokhoz közelítik a véleményeiket (Han és Wackman 2017, Hogg 2003).

A politikatudományban ezt a véleményalakító mechanizmust az identitásra vonatkozó pszichológiai elméletekhez kötik (Turner et al. 1987). Ezekben az elméletekben különös szerepet kap egy-egy csoporthoz való kötődés (csoportidentitás) szembetűnősége (Iyengar et al. 2012, Oakes 1987, Turner et al. 1994). Ez az empirikus tesztek miatt fontos.

Az empirikus tesztek miatt fontos továbbá az a széles körű feltételezés is, hogy a pártpolitikai konfliktusok következtében szembetűnőbbé válik a pártidentitás (Druckman et al. 2013, Han és Wackman 2017). Leeper és Slothuus (2014) egyenesen amellett érvel, hogy a pártelitiek között zajló közpolitikai viták nemcsak azért hatnak az emberek véleményére, mert konkrét álláspontok hangzanak el, hanem magának a vitának is van véleménybefolyásoló szerepe. Tehát a vita által felerősített csoportkötődés is a tipikus csoportvélemény irányába tolja az egyéni attitűdöket.

### VI.2.5.2. Az identitás hatásának mérése

A legutóbbi évek kísérleti módszertannal nyert eredményei arra utalnak, hogy a pártidentitás valóban önálló motivációs szerepet játszik a közpolitikai attitűdök kialakításában. Erre sok, többé-kevésbé közvetett eredmény is utal.

Egyrészt a nyugati demokráciákra vonatkozóan meggyőző eredmények bizonyítják, hogy a pártidentitás fontos a szavazóknak, akik bíznak is az általuk preferált politikai közösségben (Huddy et al. 2018, Huddy et al. 2015, Iyengar és Krupenkin 2018).

Másrészt a racionális tanulás modellje nem képes teljes mértékben megmagyarázni az elitek véleménybefolyásoló szerepével kapcsolatos jelenségeket. Például egyes kísérletek eredményei szerint a szavazók nem mindig időtakarékoságból követik a pártok véleményeit; a pártok véleményének megismerése után nem csökken az érdeklődésük a szubsztantív információk iránt (Boudreau és MacKenzie 2014, Ciuk és Yost 2016). Ráadásul vannak olyan eredmények, amelyek szerint a fontosabbnak tűnő ügyekben erősebb a pártok befolyása, mint marginális kérdésekben (Slothuus és Vreese 2010, de vö. Ciuk és Yost 2016).

Végül, de nem utolsósorban, vannak közvetlenebb eredmények is a pártpolitikai normák befolyására vonatkozóan. Bár látni kell, hogy a normák és az identitás befolyásának azonosítása különösen nehéz feladat, ha az elit által promotált vélemények megtanulásának folyamatát vizsgáljuk. A korábbi fejezetekben már idézett Petersen és szerzőtársai (2013) például kísérleti alanyaik választási idejéből próbáltak messzemenő következtetéseket levonni. Úgy vélték, hogy a válaszadók a pártidentitás direkt hatására hezitálnak akkor, amikor a kísérletben egy atipikus véleményt olvasnak kedvenc pártjuktól (például jobboldali pártokra jellemző közpolitikai álláspontot egy baloldali párttól). Ugyanis azzal a dilemmával szembesülnek, hogy az értékeik, vagy az elit által promotált norma szerint válaszoljanak-e a survey-ben. Ugyanakkor könnyen lehet érvelni amellett, hogy a hezitálás a racionális tanulási folyamat része, ugyanis a meglepő infor-

mációk feldolgozásával a racionális tanulók több időt töltenek, mint a kevésbé meglepőkével.

Han és Federico (2017) egy friss tanulmányukban szintén a párt-identitás közvetlen hatását kívánták kimutatni. Hírkeretelési kísérletükben a kiválasztott közpolitikai kérdésben zajló pártpolitikai vitát tették többé, illetve kevésbé szembetűnővé a válaszadók számára. Azt találták, hogy amikor a vita szembetűnőbb, a válaszadók véleménye jobban közelít a preferált pártok véleményéhez. De ez a kísérleti eredmény nagyjából egybevág a racionális tanulási modell előrejelzéseivel is. Ugyanis ez utóbbi modell is azt feltételezi, hogy a pártálláspontok feltűnőbbé válásával erősödik a pártelitek véleménybefolyásoló hatása.

Mindazonáltal két – e könyv írásakor –, egészen friss tanulmány a fentieknél meggyőzőbb eredményeket szolgáltat a normák és identitás direkt hatására vonatkozóan. Toff és Suhay (2018) azt hasonlították össze a kísérletükben, hogy a szavazói véleményekre a párteliteknek, vagy a többi szavazó véleményének volt-e nagyobb hatása. Azt találták, hogy a szavazótársak befolyása erősebb, mint az eliteké. Meggyőzően érvelnek amellet, hogy a racionális tanulási modell a pártoktól való tanulást magyarázza jól, míg a társak álláspontjának figyelembevétele az identitáselméletekkel áll leginkább összhangban. Azonban a szkeptikusok a racionális hordamagatartás (Chamley 2004) jeleként is tekinthetnek ezekre az eredményekre. Ciuk (2018) már az V. fejezetben is említett tanulmánya az erkölcspszichológia által elkülönített alapösztonokra vonatkozó kérdéseket manipulált pártos álláspontokra vonatkozó információkkal. A válaszadók „általános ösztönei” változtak a megismert pártos vélemények függvényében. Ez nem csupán az erkölcspszichológia alapfeltevéseivel áll szemben, hanem általában a racionális tanulás modelljével is (amelyet többféle motivációs elméletre is lehet alkalmazni). Ez ugyan indirekt, de mégis elég erős eredmény a normák befolyására vonatkozóan.

A normavezérelt szavazó modellje empirikusan egyre több megerősítést kap. De erre a modellre is érvényesek a normavezérelt viselkedéssel kapcsolatos elméleti kérdések, amelyeket a következő alfejezetben tárgyalok.

### VI.3. Nyitva maradt kérdés: a normakövetés rejtélye

#### VI.3.1. A KORRELÁLT EGYENSÚLY ALKALMAZÁSÁNAK PROBLÉMÁI

##### VI.3.1.1. Amit a korrelált egyensúly tud

E fejezet utolsó részében azt szeretném megmutatni, hogy a normák játékelméleti alapokon nyugvó modelljei nem teljesek. A hiány abból adódik, hogy e modellek nem tudják elkerülni, hogy – a pszichologizáló modellekhez hasonlóan – normatív preferenciákat építsenek be az emberi motivációk közé. Tehát végső soron visszatérnek egy olyan magyarázathoz, miszerint azért (is) követünk normákat, mert *szeretünk* normákat követni. Az alábbiakban megmutatom, miért szükséges, hogy a jelenlegi modellek keretein belül nem nélkülözhetjük a normakövetésre vonatkozó társas preferenciák feltevését.

Mindenekelőtt azonban tisztázni kell néhány – részben már felmerült – kérdést a korrelált egyensúllyal kapcsolatban! Az egyensúlyhoz vezető jelzés a játékosnak egyértelmű utasítást ad, elvileg megoldva a tipikus életjátszmákban előálló sok Nash-egyensúly közötti választás dilemmáját. Azonban a Nash-egyensúlyok azonosítása után nem nyújt további segítséget az emberi játszmák modellezőinek. Tehát magából a játszmák szerkezetéből (az ott azonosítható egyéni érdekekből) nem jobban előrejelezhető e játszmák kimenetele, ha a Nash-egyensúly helyett a korrelált egyensúly fogalmára támaszkodunk. Sőt, ha eleve több Nash-egyensúly azonosítható az eredeti szituációban, a korrelált egyensúly alkalmazásával még növekszik is a lehetséges kimenetek halmaza.

Azt is látni kell, hogy diszkrét választási alternatívákkal modellezett játszmákban a potenciális egyensúlyi eredmények bővítése az alternatívák közötti random döntések lehetőségeinek bővítéséből adódik. Az egyensúly bemutatásához használt korábbi példában a randomizált viselkedést előíró normák életszerűnek tűnnek: „Támadj akkor, ha isteneid jókedvükben vannak!” Az egyenlő felek közötti, ismétlődő szituációkra vonatkozó alkukban szintén gyakori az ilyen típusú elvárás: „A házaspár tagjai közül az menjen a gyerekért, aki

aznap épp hamarabb végzett a munkájával!” Azonban az erkölcsös, jó cselekedetekre vonatkozó normatív kijelentések tipikus esetben mégis valamilyen tiszta stratégiát várnak el a szereplőktől: „A családból háborúba harcolni mindig a férfiak menjenek, viszont kisgyerekekkel otthon mindig a nők maradjanak!” A korrelált egyensúly nagy varázslata ez utóbbi esetekben a Nash-egyensúlyok közötti választás. Ismétlem: a szereplőknek ez óriási jelentőségű. Ugyanakkor a játzmák elemzői akkor lettek okosabbak, ha már megfigyelték az adott játzmákra vonatkozó, a résztvevők által koreográfiának tekintett utasításokat. A társas cselekvés normáinak ilyen jellegű megfigyelését sok-sok évtizede végzik szociológusok, kulturális antropológusok, és a rokon diszciplínák képviselői.

A normák ilyen megfigyeléséhez és az eredmények értékeléséhez a játékelméleti eszköztár semmit sem tesz hozzá. Önmagában nagyon fontos következtetés ugyanakkor a normák kortárs játékelméleti elemzésének az a konklúziója, miszerint ezek a vizsgálatok közelebb visznek a társadalmi viselkedés megértéséhez és előrejelzéséhez, mint a pszichologizálás (Binmore 2010). Ez minden olyan vizsgálat implicit kiindulópontja, amely a normatív viselkedések vizsgálatokor, a velünk született lelki hajlamok modelljeire építő elemzések helyett például szellemtörténeti, vagy kulturális antropológiai vizsgálódásba kezd. A normák meglehetősen absztrakt és általános játékelméleti elemzésének egy fontos eredménye tehát a különböző társadalomkutatási hagyományokkal kapcsolatban megfogalmazott ítélet. A játékelméleti elemzés, és konkrétan a korrelált egyensúly koncepciója tehát fontos metaelméleti állítások megalapozásához járul hozzá.

Ugyanakkor a korrelált egyensúly normákra vonatkozó alkalmazása maga is tartalmaz belső ellentmondásokat. Ugyanis ha az egyének racionális döntéseire alapozó játékelméleti logikát követjük – és most éppen a normák játékelméleti alapokon nyugvó elméleteiről beszélünk –, akkor nem tekinthetjük vulgár-durkheimista módon adotttnak a közös kommunikációs térben elhangzó, normatív státuszra pályázó utasítások kényszerítő erejét.

## VI.3.1.2. A korrelált szignál hitelessége

Szigorúan követve a racionális döntéelmélet – vagy konkrétan a Nash-egyensúly – logikáját, egy közösen megfigyelt utasítás elhangzásakor nemcsak azt a kérdést kell feltennünk, hogy mit tegyünk, ha a többiek követik az utasítást, hanem azt is, hogy mi történik egyéb esetekben. Az világos, hogy a koncepció alkalmazásának haszna épp az, hogy kikerüljük ezt a bonyolult morfondírozást azzal a természetes feltevessel, hogy a szereplők első észszerű gondolata az, hogy a többiek követik a közös térben elhangzó utasítást, és annak alapján hoznak józan döntést (Myerson 1999).

Bizonyos utasítások (és utasítók) esetében azonban éppoly természetes, (és a bonyolult morfondírozást elkerülő) feltevés lehet az, hogy a szóban forgó játszma résztvevői közül senki sem követi a közös térben elhangzó utasítást. Vegyük csak a bináris döntésekkel kapcsolatban elhangzó utasításokat! Például: „Zöldnél haladj, pirosnál állj!” (A könyvben, ahogy a játékelméleti tankönyvekben és szakirodalomban általában, ilyen egyszerű a modellezése a társadalmi élet jelentős dilemmáinak.) Ilyen helyzetekben – a korrelált egyensúlyhoz tartozó kalkulációhoz hasonlóan – az egyénnek egyszerű döntést kell hoznia arról, hogy neki mit kellene tennie. Előfordulhat, hogy az utasítás egységes elutasításának felvétele mellett, minden egyes szereplőnek magának is érdemes elutasítania a szabályt. Azaz az adott szabályhoz tartozó korrelált egyensúly párjaként létezik egy „anti-korrelált egyensúly” is – amely az adott kontextusban éppen olyan „természetesnek” tűnhet a résztvevőknek, mint Myersonnak (1999) az eredeti korrelált egyensúly. Az utasítás ez esetben is kisegíti a résztvevőket, azonban a közös térben elhangzó szabály megfigyelése önmagában nem ad támpontot a szituáció külső, tudós elemzőjének.

Nézzünk egy példát! Részben saját emlékeim alapján rekonstruállok egy történetet. 1997 február végén bankpánik tört ki Magyarországon. A Postabank nevű, egyszerre tekintélyes és tragikomikus pénzügyintézet ügyfelei néhány nap alatt több tízmilliárd forintot vettek ki a számláikról, megingatva ezzel a bank stabilitását (Rovó 2017). A bankfiókok előtti hosszú sorok és a tömegjelenetek tették min-

denki számára látványossá a bankrohamot. Utólag látható, hogy bár a bank felelőtlen gazdálkodása miatt később az állam sok közpénz feláldozásával rendezte a bank veszteségeit, a konkrét helyzetben nem volt veszélyben a pánikba esett betétesek pénze. Így ez klasszikus bankpánik volt, ahol – a bank helyzetére vonatkozó (tév)hitek mellett –, azok a reális(abb) várakozások hajtják a bankfiókokhoz a betéteseket, hogy épp a sok megrémült betétes pénzkivétele teszi tönkre a bankot (Diamond és Dybvig 1983). A bankroham kiindulópontja nem a terrorcselekményekhez hasonló helyzeteknek megfelelő fogolydilemma jellegű pánik (erről nemsokára lesz szó). Ez utóbbiakban bármit tesznek a többiek, nekem rohannom kell. A „tisza” bankroham előtt az optimális döntés a többiek döntésétől függ. Ha az ügyfelek, a szokásos ügymenetet leszámítva, nem veszik ki a a pénzüket, akkor a bank nem fog összeomlani, ezért nekem sem érdemes kivennem. Ha viszont mindenki elindul kivenni a pénzét, akkor össze fog omlani, tehát nekem is sietnem kell, hogy az összeomlás előtt a pénzemhez jussak.

Jómagam az igazi bankroham első napján, február 28-án délután sétáltam el a Postabank budapesti központi fiókja előtt. Rácsodálkozva a nagy tömegre, megkérdeztem az egyik sorbanállót, hogy miért állnak itt. Ő azt felelte, hogy bement a rádió, hogy baj van a Postabankkal, azért sietett ide. Szavainak kontextusa a közzéadóra utalt, ami csodálkozással töltött el: nem gondoltam volna, hogy az állami rádió adásában pánikot keltenek. Az esti híreket nézve és hallgatva rekonstruáltam, hogy mire gondolhatott: kora délután egy sajtótájékoztatón a Pénzügyminisztérium államtitkára, az Állami Pénz- és Tőkepiaci Felügyelet elnöke, valamint a bank elnöke biztosítottak mindenkit, hogy nincs baj a bankkal (Rovó 2017). A példám azoknak a játszójának vonatkozik, akik ennek nyomán keltek útra (én véletlenül egyiküket kérdeztem meg a bankfiók előtt).

Az idézett sajtótájékoztató olyan korrelált egyensúly elérését célozta, amely nagyobb közjót eredményez, mint a koordinálatlan Nash-egyensúlyok közül számos, megvalósulásra esélyes kimenetel. Azonban sokan voltak, akik nem hittek a hivatalos szerveknek. Ahogy az eseményeket feldolgozó, egyik közelmúltbeli cikk is idézi (és címébe



is foglalja) az egyik sorbanállót: „Nézze, annyit hazudtak régen, és annyit hazudnak ma is nekünk, hogy az ember már nem hisz senki-nek” (Rovó 2017).

Egyrészt, a fentiek alapján a rohamozók egy részére vonatkozóan rekonstruálhatunk egy olyan játszmát, ahol *a*) a résztvevők a bankroham előtt azt gondolják, hogy a bankkal minden rendben, azonban *b*) a mozgolódás hatására tartott sajtótájékoztató megváltoztatja a bankkal kapcsolatos ítéletüket, és olyan vészhelyzet jeleként értékelik, ahol domináns stratégia lesz a bankroham. De ami érdekesebb, hogy a bankroham egyes résztvevőire vonatkozóan rekonstruálhatunk egy olyan játszmát, ahol a fenti *a*)-jelű feltevés mellett a sajtótájékoztató nem a bankról közöl információkat, hanem a többi betétes várható viselkedéséről. Erről szól a korrelált egyensúly. Jelen esetben azonban az információ (játékelméleti szaknyelven a „korrelált szignál”) az, hogy ha eddig a bankban bízó emberek otthon maradtak, akkor most, azt feltételezve, hogy mindenki azt feltételezi, hogy senki sem hallgat az utasításra, mindenki el fog indulni. Azt gondolom, hogy ebben, a valóságban sokféle szereplő részvételével, időben is széthúzódó játszmában a bank megrohanásának egyik, sokak számára releváns motivációja modellezhető az „anti-korrelált egyensúly” koncepció segítségével. Azt is gondolom, hogy a kialakult helyzet egy alacsony közbizalommal rendelkező társadalomban nem volt egyedülálló és meglepő. (Az idősebbek és középkorúak gondoljanak vissza az élelmiszerboltok megrohamozására az egy-két napos ünnepek előtt, és az azokat megelőző híradós tudósításokra, amelyekben a jelentős pékipari vállalatok felelős munkatársai nyugtatták meg az embereket, hogy nem lesz kenyérhiány az ünnep előtti napon, és a szükségleteknél jóval többet állítottak elő az ünnepekre készülve.) A korrelált egyensúlyok egyértelmű és természetes sorvezetőjének tekintett szignálok hatásának nehéz előrejelezhetőségét mutatja továbbá az is, hogy pár nappal később a kormánzatnak határozott kiállással sikerült rávennie a megmaradó betéteseket arra, hogy a nyugalmi állapotnak megfelelő egyensúlyi stratégiát kövessék, és hagyják bent a pénzüket (Rovó 2017). Ráadásul – hogy közelítsünk az erkölcsi viselkedéshez –, a szociológiai és kulturális antropológiai

szakirodalom rámutat, hogy a normák nyílt megerősítésének kísérletei tendenciaszerűen kötődnek a szóban forgó normákkal kapcsolatos elbizonytalanodáshoz (Weeks 1995), és e tendencia magában hordozza, hogy a nyílt szignálok önmaguknak ellentmondó, esetleges implicit üzeneteit nem szabad figyelmen kívül hagynia a normák játékelméleti elemzőinek (sem). Azaz a korrelált egyensúly – az eredeti definíciója alapján –, nem tekinthető magától értetődő koreografáló erőnek.

Ezen a ponton egyet kell értenünk Binmore (2015) figyelmeztetésével, miszerint a korrelált utasításokat a vizsgált játék információs kiterjesztésével érdemes bevonni a hagyományos Nash-féle elemzésbe. Ráadásul nem azért, mert a korrelált egyensúly koncepciója lényegtelen lenne, hanem éppen azért, mert az emberi játszmák dilemmáinak ez a megoldása van olyan fontos, hogy alaposabban elemezzük, pontosan milyen feltételek teljesülése esetén alkalmazható.

#### VI.3.1.3. Normativitás az egyensúlyok halmazán kívül

Meggyőződésem, hogy a társadalmi normák játékelméleti alapú elemzése az erkölcs kutatására vonatkozó metateoretikus következtetéseken túl is nyújthat segítséget. Vegyük mindenekelőtt a fogolydilemmát, amelyet a társadalmi együttműködéssel és normákkal foglalkozó közgazdaságtani, politikatudományi és játékelméletet alkalmazó filozófiai kutatások a leggyakrabban elemeztek, és amely egyes közelítésekben maga a normák szükségességét előhívó egyetlen, vagy az egyik legfontosabb társadalmi csapda (például Bicchieri 2006, Taylor 1976, Ulmann-Margalit 1977)! A könyv III. fejezetében ismertetett rögzített, véges lejátszási körből álló fogolydilemmában a korrelált egyensúly koncepciója semmiben nem változtatja meg a sok évtizeddel ezelőtti racionális elemzések konklúzióit: az egyetlen racionális stratégia a potyautazás. Ez a játszma egyetlen korrelált egyensúlya is (ez következik az korrelált egyensúly fent ismertetett sajátosságaiból). Nincs olyan csodatévő koreográfia, amely racioná-

lis egyéneket együttműködésre bírna ebben a helyzetben. Tehát az embereket véges fogolydilemma helyzetekben együttműködésre bíró normákat nem lehet a korrelált egyensúly fogalmának segítségével magyarázni.

Lássunk egy példát! A New York-i Világkereskedelmi Központ toronyházainak számos dolgozóját meggyilkoló, 2001. szeptember 11-i terrortámadás után a magyar sajtóban megszólaltatták a tragikus események néhány magyar tanúját. Az egyik tanú története emlékezetemben a következőképpen él: „A sokadik emeleten lévő irodám előtt, a kávégép mellett várakoztam, amikor jó pár emelettel feljebb a repülő becsapódott az épületbe. Nem sokáig voltam bizonytalanságban, mert a hangosbemondóba bemondták, hogy mi történt. Azt is hozzátette a bemondó, hogy mindenki maradjon nyugodtan az irodájában; és várjon arra, hogy az ő emeletét szólítsák a menekítési tervnek megfelelően.” A hangosbemondó üzenete rendkívül kézzelfogható példája a korrelált szignáloknak. Úgy vélem, a szignál kollektív követésének kollektív jólétre gyakorolt pozitív hatását sem kell külön magyarázni. Gyakran kapunk híreket a világ legkülönbélebb tájairól a valós vagy vélt vészhelyzetekben kitörő tömegpánikok tragikus következményeiről, amikor a koordinálatlan tömegben akaratlanul agyontaposnak vagy megfullasztanak embereket. Az ikertornyok dolgozóinak kimenekítéséről szóló helyi beszámolók szerint az épületek korrelált szignálok által koreografált kiürítése a helyzethez képest sikeres volt: nem alakult ki tolakodás, és ennek köszönhetően a felsőbb emeletlakóinak kivételével a nagy többség kijutott, mielőtt a tornyok összeomlottak volna. (A menekítés legnagyobb hősei azok a tűzoltók voltak, akik, bár végig a kijáratok biztonságos közelében tartózkodtak, az épületek összeomlásáig végezték feladatukat, és együtt haltak meg a túl magasról és/vagy a menekítési tervnek engedelmessé válva túl későn indulókkal).

A történet mesélője azonban így folytatta: „Annak köszönhettem az életem, hogy nem hallgattam a hangosbemondó utasításaira, és az irodám helyett a lépcsőház felé vettem az irányt, és így az irodámnak szánt indulási időpontnál jóval hamarabb elindultam lefelé. Munkatársaim arról a sokadik emeletről mind odavesztek.” Ha

a megtámadott épületekben dolgozók játszmáját plauzibilis módon próbáljuk modellezni, könnyen jutunk arra következtetésre, hogy a magyar sajtónak nyilatkozó túlélő a saját domináns stratégiáját követte. Magyarul az egyetlen észszerű dolgot tette. Ugyanakkor, ha mindenki így tett volna, a lépcsőházban könnyen a menekülést akadályozó tumultus alakulhatott volna ki. A szignál nélküli alapjáték a fogolydilemmhez hasonlít.

Tehát a korrelált szignál a fenti esetben úgy működött, hogy nem a korrelált egyensúlynak megfelelő utasítást adta. Épp ellenkezőleg, a korrelált egyensúly megvalósulását kívánta elkerülni – a tragikus eseményeket követő számos beszámoló szerint nagyjából sikerrel. A magyar sajtóban megszólaltatott szabályszegő túlélő – úgy tűnik –, a kivételek közé tartozott.

### VI.3.2. EGYENSÚLY VAGY NORMATÍV ÖSZTÖNÖK?

#### VI.3.2.1. Az ortodox játékelméleti álláspont

A véges fogolydilemma helyzet fontos példája annak, amikor a korrelált egyensúlyhoz vezető utasítások nem nyújthatnak megfelelő ösztönzőket a racionális emberek számára az erkölcsös viselkedésre. Úgy tűnik tehát, hogy a korrelált egyensúly koncepciója nem oldja meg azt az egyensúly-szelektálási problémát, amelyet a klasszikus racionális döntéseméleti modellek hagytak maguk után, és amelyek megoldására megszülettek a belső erkölcsi késztetések népszerű, ám empirikusan gyenge lábakon álló elméletei. Az alábbiakban a normativitás rejtélyének négy olyan megoldási lehetőségét tekintem át, amelyek anélkül próbálják feloldani a rejtélyt, hogy visszanyarodnának a konkrét tartalommal bíró erkölcsi késztetések elméleteihez.

Binmore (1998, 2005, 2010, 2015) ortodox közelítése következetesen ragaszkodik a Nash-egyensúly koncepciójához. Az élet játékaik szerinte bizonytalan végű ismétlődő interakciók. A legtöbb társas játszma ráadásul gyakran ismétlődik, gyorsan játszódik le, és

közben ritkán tudatosítjuk magunkban a dilemmáinkat. Ennek két fontos következménye van, melyekről már volt szó. Egyrészt, az erkölcsös közjával szemben álló domináns stratégia leginkább elméleti konstrukció. A valóságban, mivel az emberek nagyon sokszor nem tudják, hogy az egyes társas kapcsolataik mikor érnek véget, minden racionalitáson túlmutató feltételezés nélkül előfordulhat, hogy a racionális büntetés fenyegetésének árnyékában egymás, vagy a „társadalom” javát szolgálják. Másrészt a gyakori ismétlődés tudattalan korrekciókra ad lehetőséget, lokális optimumként működő, „közeli” Nash-egyensúlyokba terelve az embereket. Az emberek tehát így oldják meg a választás problémáját. Amit választottak, az lesz az erkölcsös. A normativitás így leginkább elméleti konstrukció. A legtöbb, észrevétlen döntés esetében az emberek nem is foglalkoznak vele.

Vegyünk egy példát, amelyre vissza fogunk térni. Számos mai modern társadalomban, ha egy férfi egy nővel együtt érkezik egy bejáráshoz, akkor a férfi különösebb gondolkodás nélkül lassít, hogy a nő előbb léphessen be – aki valószínűleg még a férfinél is kevésbé foglalkozik a sorrend problémájával, hiszen gondolkodás nélkül is „tudja”, hogy a férfi nem áll az útjába. Ha két másodperccel később, az ajtó túloldalán a férfi belebotlik egy férfi ismerősébe, sok kultúrában önkéntelenül nyújtja a jobb kezét kézfogásra.

A kutatók nézőpontja szempontjából Binmore nem tekinti az általános elmélet problémájának az ismétlődő játszmák számos Nash-egyensúlya közötti választás kérdését. Tehát az elméletnek nem kell azzal foglalkoznia, hogy az „egy férfi és egy nő egyszerre az ajtónál” játszma esetén számos kultúrában miért mindig a nő megy előbb, és miért kéznyújtás a köszöntés, és miért jobbkézzel. Konkrét normák esetében középszintű elméletek, vagy a konkrét helyzetre vonatkozó történeti vizsgálódások adnak választ ezekre a kérdésekre. Binmore (2010) szerint a szociológia feladata lehet a kultúraspecifikus szokások történeti hátterének bemutatása. Középszintű elmélettel viszont maga is foglalkozott, amikor az elosztási normák evolúciójának logikáját próbálta levezetni (Binmore 2005).

### VI.3.2.2. A normatív szituációk elkülönítése

Bicchieri (2006, lásd még Bicchieri et al. 2018) ellenben filozófusként arra figyelmeztet, hogy nem csupán a normakutató filozófusok fejében válnak szét a leíró, illetve előíró szabályok. Az emberek érzékelik, hogy valaki csak a saját kárára tér el egy mindenkinek előnyös konvenciótól, vagy a normakövető emberek kárára potyautazik egy fogolydilemma jellegű helyzetben. Bicchieri és szerzőtársai (2018) empirikus kutatásaikban arra jutottak, hogy az emberek másképp reagálnak arra az információra, hogy mit tesz a többi ember, illetve arra, hogy mit gondol helyesnek a többi ember (az előbbi erősebben befolyásolja az egyének döntéseit).

Térjünk vissza a kézfogás fenti példájára! Azt nyilvánvalónak tekintjük, hogy a mi kultúránkban a kézfogás kezdeményezésének elmulasztása nem feltétlenül normasértés, azonban a másik fél részéről kezdeményezett kézfogás elutasítása az. Sokan emlékeznek talán arra a botrányra a kétezres évekből, amely abból keletkezett, hogy egy állami kitüntetésben részesülő tudós (egy másik kitüntetett díjazásával szembeni tiltakozásként) nem volt hajlandó kezet fogni a miniszterelnökkel egy kitüntetési ceremónián. A köztársasági elnök pedig ugyanazon az ünnepségen annak a kitüntetettnek nem nyújtotta a kezét, akinek a kitüntetése miatt az előbb említett tudós a miniszterelnököt büntette (Megyeri 2006). Az eset tárgyalásakor többször elhangzott az a kifejezés, hogy az egyik fél a jobbját nyújtotta, azt sugallva, hogy maga a jobbkezes kézfogás az erős normatív tartalommal bíró egyensúlyi stratégia. Tény, hogy a „felmentéssel rendelkező” vívókon és a sérült, vagy félkezű embereken kívül nem látunk példát arra, hogy a valaki a balját nyújtja. Bár az gyanúra adhat okot, hogy a balkéz nyújtására nem szoktak kézfogást megtagadni. Egy vezető beosztású tanszéki kollégám mindenestre nem elégedett meg azzal, hogy empirikus teszt nélkül mossuk össze a kézfogást a jobbkezes kézfogással. Ezért egy egyetemi diplomaátadó ünnepségen (a humán objektumokra irányuló vizsgálatok terveit ellenőrző egyetemi kutatásetikai bizottság engedélye nélkül!) kísérletbe kezdett. Az öt-másodpercenként érkező, az összes kari tanszékvezetővel kezet fogó

hallgatók véletlenszerűen kiválasztott, „kezelt” (bocsánat) csoportjának bal kézzel nyújtott kezét. Ez megakasztotta ezeket az alanyokat, hiszen friss diplomájukat a bal kezükbe szorongatva mentek végig a tanszékvezetők előtt. A kollégám a megakadás ellenére nem találkozott megütközéssel. A bal kézzel találkozó hallgatók megálltak, és binmore-i típusú (lokális) korrekciót hajtottak végre: szépen áttették jobb kezükbe a diplomájukat, és büszke fiatal értelmiségiekként fogadták egy egyetemi tanszékvezető megtisztelő balját. Azaz, míg a kinyújtott kéz elfogadásának szokása normatív erővel bír, a jobb kéz nyújtásához – úgy tűnik –, nem kötődik ilyen erő. Az egy egyszerű konvenció. Binmore logikája alapján pedig nem lehet különbséget tenni a két döntés normatív ereje között.

Bicchieri tehát fontos, Binmore által ignorált distinkcióra hívja fel a figyelmet. Ugyanakkor nincs határozott megoldása arra vonatkozóan, hogy mi történik másképp a normatíván terhelt szituációkban. Gintis (2009, 2016) elmélete szerint a normabetartást szolgáló erős reciprocitás ad magyarázatot a normatív erőre. Gintis a korrelált egyensúlyt megerősítő belső hajlamként tekint erre az ösztönző tényezőre. Azonban ennek hatása éppen akkor látványos, ha az erkölcsös megoldást a játszma alapszerkezete nem teszi egyensúlyivá. Például világos, hogy nincs végtelen ismétlési lehetőség, és az önérdék nem támogatja az erkölcsös viselkedést.

Gintis a normatív erő forrására vonatkozó javaslatát beleilleszke-dik a társas preferenciákkal foglalkozó közgazdasági kutatások egy újabb, egyelőre periferikus áramlatába, amely a társas preferenciák korábbi modelljeinek problémáira keresi a választ (Gächter et al. 2017, Kimbrough és Vostroknutov 2016, Kranz 2010, Lopez-Perez 2008, 2010). Ez az irányzat adottnak tekinti, hogy az empirikus tesztek alapján kétséges, hogy léteznek-e egyáltalán méltányossági, vagy egyéb, az emberi játszmák konkrét lejátszási formáit támogató társas preferenciák. Emellett bízunk azokban az egyre gyarapodó empirikus eredményekben, amelyek arra utalnak, hogy az emberek a legkülönb-félőbb normákhoz igazodva viselkednek erkölcsösen. Ezeket a negatív és pozitív eredményeket összegzik olyan modellekben, amelyek az egyenlőségre, méltányosságra, közjóra stb. vonatkozó, velünkszüle-

tett vágyakat a normák betartásának inherens vágyával helyettesítik az ember preferenciarendszerében.

A normavezérelt racionális polgárt leíró modell antropológiai logikája szorosan követi a méltányosságelmélet és az erkölcspszichológia által kijelölt nyomvonalat. Valamilyen, a közvetlen önérdek követésével nehezen magyarázható viselkedési tendencia (vélt) megfigyelése nyomán azt feltételezzük, hogy az embert a veleszületett vágyai térítik el a racionális döntés által kijelölt pályáról a megfigyelt irányba. Azonban e modell társadalomelméleti következtetései jelentősen eltérnek az előzőektől. A belső készítés hatására követett normák tartalmára vonatkozóan nincs előírt recept. Ugyanakkor a normák genetikus-kulturális koevolúciójára vonatkozó elméletek visszacsempészhetik a megfigyelt szabályok szükségszerűségére vonatkozó feltételezéseket (lásd Binmore 2005, Gintis 2016, Gintis és Helbing 2015). Minél visszafogottabb a normák valamely játékelméleti modellje a normák evolúciójára vonatkozó feltevéseit tekintve, annál közelebb kerül a valóság társadalmi konstrukciójának elméleteihez (Berger és Luckmann 1966). A könyv végén ennek metaleméleti és praktikus következményeit is tárgyalni fogom.

A normakövetés ortodox játékelméleti modellje (például Binmore 1998, 2005, 2010, Kandori 1992, Kreps et al. 1982) tehát a bizonytalan ideig ismétlődő interakciókat egyensúlyi pályán tartó racionális erőkkkel azonosítja az erkölcsi erőt. Ez az ortodox modell azonban nem terjeszthető ki azokra a helyzetekre, ahol egyedi, nagy téttel járó döntéseket hozunk, valamint az egyéni és közérdek valamilyen módon ütközik. Például nem minden nap kell döntenünk arról, hogy az életünket kockára téve engedelmessékedünk-e a közösségi elvárásoknak egy terrortámadás célpontjává vált épületben, vagy inkább mentjük a életünket, a normák megsértése árán. De idézzük fel a modern társadalmak mindenkit érintő, III. fejezetben tárgyalt dilemmáját: hogyan foglaljunk állást a generációk közötti állami újraelosztás kérdésében? A normák és intézmények fenntartásának korrelált egyensúly koncepcióján nyugvó modelljei (Gintis 2009, Guala 2016) nem oldják meg az egyéni racionalitás és a közérdek ütközését.



ből eredő problémát, mivel ezek is csak a koordináció megoldásának egy lehetséges mechanizmusát írják le.

A normakövetés új, unortodox játékelméleti modelljei (például Gintis 2009, 2016, Lopez-Perez 2008) kézenfekvő választ nyújtanak az egyéni érdek és a normatív elvárások ütközésének dilemmájára: belső késztetésünk van a normatív elvárások teljesítésére. Döntéseinkben a késztetés relatív súlya a kockán forgó anyagi érdekek nagyságától, és e belső késztetésünk erejétől függ.

#### VI.3.2.3. A normakövetési motiváció feltételezésének kritikája

Egyelőre még kevés empirikus munkában tesztelték a normakövetésre vonatkozó belső késztetések újabb modelljeit. A késztetés egy formája azonban kivételt jelent: az altruista büntetésé. Ez utóbbira vonatkozó eredményekről már volt szó korábban. A normakövetés újabb modelljeinek empirikus erejére vonatkozóan azonban még nem vonható le erős következtetés. A normatív késztetések közvetlen empirikus azonosítását célzó kísérletek kritikája helyett az ortodox játékelméleti közelítés elméleti ellenvetéseire koncentrálok.

A tisztán racionális döntéseméleti megközelítés a szituációs logikán nyugszik, és arra ösztönöz, hogy a döntések környezetének sajátosságaiiban keressük az elsőre irracionálisnak tűnő döntések hátterét. Például beazonosítottunk egy fogolydilemmát, amelyben valaki együttműködően viselkedik? Ne elégedjünk meg az együttműködési hajlandóság feltételezésével, hanem kutassunk tovább, és így észrevehetjük, hogy a helyzet ismétlődik! Máshol ugyanakkor azt látjuk, hogy a fogolydilemma ismétlődése sem bírja együttműködésre embereket egy csoportját? Egy másik csoportot ugyanakkor együttműködésre bír? Vegyük észre, hogy az egyikben a csoporttagok közötti interakciók struktúrája nem oldja meg a büntetők és büntetendők „párosítását”, a másokban meg igen (Kandori 1992)! Aztán egy harmadik csoportban azt látjuk, hogy a büntetési mechanizmust lehe-

tővé tevő interakciós struktúra ellenére sincs együttműködés? Ha az emberek személyiségének vizsgálata helyett a körülményeket vesszük alaposan szemügyre, akkor észrevehetjük, hogy ebben a harmadik csoportban a racionális büntetők nem értesültek megfelelő bizonyossággal a potyautasok személyéről.

Ezen a ponton kanyarodjunk vissza a II. fejezet egy fontos állításához! A neoklasszikus közgazdaságtan és a játékelmélet embermodelljének bemutatásakor hangsúlyoztam, hogy az alapmodellek semlegesek az emberi motivációk tekintetében (lásd Becker és Stigler 1977, Binmore 2010). Tehát, míg a könyvben a társas preferenciákkal és az erkölcsi ösztönök modelljeivel kapcsolatban bőségesen tárgyalt empirikus kritikák állítása valóban az, hogy nem létezik a feltételezett formájában és súlyában ez vagy az az erkölcsi motiváció, addig az elméleti kritika nem állít ilyet. Az ortodox racionális döntés- és játékelméleten nyugvó elméleti kritika mondanivalója az, hogy tegyük meg előre tétjeinket: ágyazzuk be elméletileg vagy empirikusan megalapozott motivációs modellünket a racionális önérdékkövető preferenciákra vonatkozó modellünkbe. Például, ha meggyőződésünk, hogy sok embert valamilyen altruizmus hajt, akkor a beágyazás után az altruizmus legyen az új önérdék, majd ezzel a kibővített modellel végezzünk racionális egyensúlyi elemzést. Ha a lélek titkainak vizsgálatából kiderül, hogy az emberek egy részének veleszületett vágya a tüntetésre járás, akkor ez már az ő önérdékük, és ezt megfelelő módon modellezve kiderülhet, hogy a tüntetési részvétel dilemmája nem fogolydilemma számukra. Következésképpen a részvételt támogató és akadályozó tényezőket a fogolydilemma modell alkalmazásával nem fogjuk megtalálni. Ha pedig utólag, társas preferenciaként, hozzáadjuk a tüntetési vágyat a modellhez, nehezen kerüljük el, hogy ad hoc módon, a kiegészítő vágyakra vonatkozó feltevések manipulálásával tömködjük be a meg nem magyarázott jelenségek keltette réseket ahelyett, hogy a szituáció viselkedést befolyásoló tényezőire figyelnénk.

A méltányosságelmélet empirikus tesztelésének és alakításának története kiváló példa arra, hogy miképpen akadályozza bizonyos élethelyzetek jobb megértését az ellentmondások egyszerű feloldása

a motivációs feltevések foltozgatása révén. Mindeközben a kutatók, akik ragaszkodtak az ortodox racionális döntésméleti modellhez, (amely a konkrét esetekben azt feltételezte, hogy a kísérleti alanyok pénzszerzés céljából nyomogatják a gombokat a laborokban), képtelenek voltak feltárni, hogy milyen környezeti tényezők befolyásolják a laboratóriumi diktátor- és ultimátumjátékokban az emberek viselkedését (tanulási lehetőség saját és mások döntéseiből, laboratórium falain túli reputáció-rombolás esélye stb., lásd a IV. fejezetet).

Mindezeket figyelembe véve tartom elégtelennek a normakövetés belső motivációira vonatkozó kiegészítő feltevéseket még abban az esetben is, ha ezek a feltevések jól illeszkednek az eddigi empirikus eredményekhez (Capraro et al. 2018, Krupka és Weber 2013). Meggyőződésem, hogy az erkölcs forrásának még jobb megértéséhez tovább kell folytatni azt a munkát, amely az ad hoc kiegészítő motivációs feltevések beemelése helyett a szituációk pontosabb leírását célozza. Ehhez a munkához kívánok hozzájárulni egy olyan modell felépítésével, amelynek informális vázlatát az alábbiakban mutatom be.

## VII. Konklúzió: a beágyazott önző polgár erkölcsse

### VII.1. Egy döntés, több játszma: javaslat egy új modellezési keretre

#### VII.1.1. A MODELL ALAPJAI

##### VII.1.1.1. Alapjátszma vs. csoportjátszma

A javasolt modell kidolgozásakor a magam és néhány kollégám által korábban épített kollektív cselekvési modelleket (Flache 1996, Janky 2000, Janky és Takács 2004, 2010, Takács, Janky és Flache 2008) gondoltam tovább, több szempont szerint általánosítva azokat. Az alábbi verbális kifejtés mögötti játékelméleti algebra egy fokkal kifinomultabb, mint a fenti modellekben, azonban még mindig jelentős leegyszerűsítésekkel modellezi az ismételt játszmák stratégiáit. De ez nem érinti a fejezetben kifejtésre kerülő tartalmi tanulságokat.

Az alábbiakban bemutatásra kerülő normakövetési modell alapfeltevése az, hogy az erkölcsileg releváns helyzetekben egy döntéssel valójában mindig két játszmában veszünk részt. Az első játszma, vagy másképpen az *alapjátszma* maga az *erkölcsi dilemma*. Például, férfiként előre engedjem-e az ajtóban a velem együtt odaérő nőt? „Előre engedjem-e” házastársamat ma este a házimunkákban? Előre engedjem-e a tolakodót a bolti sorban? Becsületes amerikai polgárként hallgassak-e a terrortámadást szenvedett épület hangosbemon-

dójának utasításaira? Friss diplomásként fogadjam-e egy tanszékvezető gratuláló balját? Ügyeljek-e a hétvégi takarításkor a veszélyes hulladékok speciális elhelyezésére? Tüntessek-e jövő héten egy egyetem, vagy egy egyetemi szak megmaradásáért? Támogassam-e, hogy több adót szedjen be tőlem az állam a szegények érdekében? És így tovább.

E szituációk egy részében a helyzet gyakori ismétlődése megteremti annak a lehetőségét, hogy az ortodox játékelmélet logikája (Binmore 1998) alapján magában a helyzetben keressük a normatív erőt. Ezt speciális esetnek tekintem. Általános esetben nem feltételezhetjük, hogy magában a szituációban rejlik a normatív erő.

A normatív erő a másik játszmában rejlik. Ez a *csoportjátzsma*, egy olyan *metajátzsma*, amelyben egy (vagy több) csoport normáihoz való igazodásomat jelzem. Ennek a könyvnek nem témája a társadalmi csoportok határainak elemzése. Nem szükséges nagyon komoly szociológiai műveltség annak belátásához, hogy a kortárs társadalmak emberi játszmáinak elemzésekor nagyon kényes feladat a cselekvők számára erkölcsileg releváns csoportok definiálása. Az általam javasolt modell bemutatásakor azonban nem csupán a játékelméleti komplexitásból nyesek le, hanem a csoportidentitás szociálpszichológiájának és a társas élet szociológiájának számos szempontját is mellőzöm.

Egy, a csoport (egyes) tagjaival játszott alapjátzsma normájának betartásával egyúttal igazodom a csoportjátzsma normájához is. Ez az igazodás biztosítja, hogy a későbbi, különböző alapjátzmákban úgy léphetek kapcsolatba e csoport különböző normakövető tagjaival, hogy ők is normakövető csoporttagként tekintenek rám. E kölcsönös várakozások alapján a jövőbeni alapjátzmákban is a kölcsönös normakövetés a várható kimenetel. Erről szól a meta- vagy csoportjátzsma. Ez a csoportjátzsma nagy vonalakban modellezhető az ismételt normabetartási játékok mintájára (III. fejezet). Például egy kisebb közösségben Kandori (1992) modelljének analógiájára, ahol a csoportból bármikor bárkivel összefuthatok. Egy nagyobb csoportokra alkalmazható logikát mutat meg Flache (1996), Janky és Takács (2010), ahol hálózatokba szerveződhetnek az interakciók, és ebben

a metajátszmában sok emberrel csak áttételesen van összekötve az egyén.

A csoport-/metajátszmából nemcsak az alapjátszmáktól elválasztható döntések hiányoznak, hanem az azoktól elválasztható következmények is. A metajátszma kifizetései a különböző, erkölcsileg releváns, egyszeri vagy ismételt alapjátszmák (előreengedés, kézfogás, tüntetés, takarítás stb.) kifizetéseiből származnak. Egy adott döntési helyzetben a csoportjátszma kifizetései úgy kalkulálhatóak, hogy megnézem, mi várhat rám a csoport tagjaként a jövőben normakövetőként, illetve normasértőként.

1. Egyrészt azt vizsgálom meg, hogy a csoport tagjaival várható jövőbeni játszmáimban milyen kifizetéseket kaphatok a játszmák normakövető stratégiáinak választása révén, ha ezekben a játszmákban a partnereim is normakövetők. E kifizetések összesített, ún. jelenértékét kell hozzáadni a normakövető stratégia alapjátszmájának következményeihez.

2. Másrészt azt vizsgálom meg, hogy legjobb esetben mit tudok kihozni abból, ha a csoport többi tagja a jövőben normasértőként/idegenként tekint rám. A normasértés többféle következménnyel járhat. a) Lehet, hogy ugyanazokat az alapjátszmákat játszom majd a jövőben, ugyanazokkal az emberekkel, mint normakövetőként, csak éppen mindenki mindig az adott alapjáték optimális stratégiáját követi majd (azaz nincs tekintettel arra, hogy mi az erkölcsös választás). b) Lehet, hogy ugyanazokat az alapjátszmákat lejátszom majd a jövőben, ugyanazokkal az emberekkel, mint normakövetőként, de ezek az alapjátszmák ki vannak bővítve a normasértéshez tartozó büntetési aktusokkal, és a partnereim büntetnek engem ezekben a játszmákban. A büntetés egyik típusa lehet, hogy megtagadják tőlem az eredeti interakcióban való részvételt. A normasértéshez tartozó jövőbeni kifizetések kalkulációját tovább bonyolíthatja, hogy a csoportban várható játszmáim rossz perspektívái esetén alternatív interakciós partnereket kereshetek más csoportokból (vö. Takács et al. 2008).

## VII.1.1.2. A csoportjátzsma jelentősége és természete

Látható, hogy miközben a viselkedésem objektíve az alapjátzsmaival modellezett interakcióban figyelhető meg, jó esély van arra, hogy a viselkedésemet meghatározó döntésemet főként a metajátzsma messze ható következményeire alapoztam, és nem a konkrét helyzetre. Tehát egyetlen aktuális interakció közvetlen következményei állnak szemben rengeteg jövőbeni interakció lehetséges hatásaival. Ez mutatja a normatív nyomás potenciális erejét. Ez az erő messze túlmutat az adott alapjátzsma keretein. Sokszor még akkor is, ha az alapjátzsma maga is egy hosszan ismétlődő játzsma. Az adott konkrét helyzetben lényegtelené válhat, hogy az önző polgár modellje mit jósol az alapjátzsma várható történéseire vonatkozóan.

Ez a potenciálisan nagy erejű normatív nyomás külső nyomás. Az általam javasolt racionális döntésméleti modell e kérdésben a szociológia domináns álláspontját tükrözi. A játékelméleti megközelítés ugyanakkor segít megérteni, hogy miképpen jöhet létre ilyen (és ennyire erős) külső nyomás. Az ereje abból származik, hogy érdeklünkben áll a nyomást gyakorló csoport tagjaival a későbbiekben is együttműködni, és a csoport többi tagjának is érdekében áll a csoport tagjaival a későbbiekben is együttműködni. És ennek tudatában mindenki azt várja, hogy mindenki azt várja, hogy mindenki elvárja a normakövető viselkedést, amely magában foglalja a norma tárgyát képező viselkedésen túl a normasértés büntetését is, és a büntetés elmaradásának büntetését is, és így tovább.

A külső nyomás a többi polgár *racionális* döntéseinek az eredménye. Sőt, a többi polgár *önző* döntéseinek az eredménye. Sőt, úgy gondolom, hogy jó közelítéssel a többi polgár egyszerű *materiális érdekein* nyugvó önző döntéseinek az eredménye. És a nyomásnak, a fenti mechanizmusokra tekintettel, önző módon, hosszú távú materiális érdekeinket követve engedünk. És a nyomásnak érdekeink miatt engedve mi magunk is önző nyomásgyakorlóvá válunk, és hozzájárulunk a másokra gyakorolt társadalmi nyomás erejéhez.

Az érdekekre vonatkozó hivatkozás rögtön értelmet kap, ha belátjuk, hogy az egyéni ösztönzők játzsma-ról játzsma-ra változnak, és

velük változik a társadalmi nyomás jelentősége is. Vannak például olyan különleges alapjátzmák, amelyekben az egyszeri döntéseknek messzemenő következményeik vannak. Például a házassággal, gyermekvállalással és karrierrel kapcsolatos döntések. Ezekben a helyzetekben az alapjátzmában azonosítható következményeknek súlya a csoportjátzmában összegződő következményekkel azonos súlycsoportba eshet. Továbbá éppen a kisebb jelentőségű játzmákban fordulhat elő, hogy a normaszegés büntetése sem örökre szóló kiközösítés. Az alapjátzma következményei ezekben a helyzetekben is fontosak lehetnek. Ráadásul a csoporton belüli büntetések jelentőségét csökkentheti, ha a csoporton belüli interakciók könnyen behelyettesíthetőek egy másik csoport tagjaival folytatott interakciókkal, ahol nem büntetnek az eredeti csoportommal szembeni engedetlenségemért, és összességében nem követelnek többet tőlem a jövőben, mint amennyit az eredeti csoportom kér a normakövetőtől. Végül itt kell megemlékezni arról is, hogy a normaszegések detektálása sok szituációban nehéz. Az ilyen, rejtőzködést lehetővé tevő helyzetekben is megnő az alapjátzma következményeinek a jelentősége.

A nagy tekintélynek örvendő családokon és gondozó intézményeken belüli rejtett erőszak esetei mutatják a legdrámább (és hadd fogalmazzak így: leggusztustalanabb) módon a személy tulajdonságaként értelmezett általános normakövetési hajlandóság (Bicchieri 2006, Gintis 2009, 2016) magyarázóerejének korlátait. A hivatkozott motivációs modell szerint mindig a felszínen normasértőnek bizonyuló emberek körében kellene keresni a rejtett bűnözőket is. A kettős játzma modellje ezzel szemben magyarázni tudja a normakövetési hajlandóság szituáció- és státuszfüggését, továbbá a kettő interakcióját. Például e modellben érthető egyes nagyon magas és nagyon alacsony státuszú emberek általános (és néha hivalkodó) normatív lazasága. Különösen érthető annak az embernek lazasága, aki azért magas státuszú, mert értékesebb az interakciós partnereinek, mint azok neki – azaz nem ő veszít sokat, ha bosszúálló csoporttagok megtagadják tőle az együttműködést. És annak az embernek az általános normaszegési hajlama is érthető, aki már csak a láncait veszít-



heti. Ugyanakkor az ezekkel szemben álló „kispolgári” képmutatás – beleértve a legdurvább kettős életet is –, szintén jól elhelyezhető a társadalmi struktúrában a kettős játszma modellje alapján.

Nagy a kísértés, hogy a második játszmát úgy egyszerűsítsük le, hogy ismételt fogolydilemmaként modellezzük. Ez akkor lenne jó közelítés, ha a normák vizsgálatának analitikus filozófiai hagyományát követve (Bicchieri 2006, Ullmann-Margalit 1977) azt feltételeznénk, hogy a közösségekben elvárt erkölcsi döntések összességükben minden csoporttag önérdekének erőteljes korlátozásával járnak, és az elvárások teljesülése esetén az egész csoport jobban jár. Azonban ezt a feltevést leegyszerűsítőnek tartom. A társadalmi viszonyokban rengeteg az aszimmetria. Az egyes alapjátszmák normatív lejátszásai különböző mértékű veszteséget okozhatnak a különböző résztvevőknek. Ráadásul a normatív nyomás terheinek egyenlőtlensége gyakran az alapjátszmák összességét tekintve is fennmarad. Egyes emberek számos kulturális kontextusban szisztematikusan rosszabbul járnak a normák betartásával mint mások, és ezek az egyenlőtlenségek az emberek társadalmi struktúrában elfoglalt helye alapján sokszor előre jelezhetőek.

Az egyszerre reális és még kezelhető megoldás az lehet, ha egy-egy csoport különböző szerepeket betöltő tagjaira vonatkozóan jellemezzük a csoporttagság várható előnyeit és költségeit. Például a különböző nemű szereplőket tartalmazó alapjátszmák társadalmi normái gyakran aszimmetrikusak: csak a férfiktól vagy csak a nőktől várnak jelentős áldozatokat. Ugyanakkor sok társadalmi kontextusban a metajátszma is aszimmetrikus: hosszú távon is más a haszna egy nőnek, illetve egy férfinak abból, ha a társadalmi normákat betartva és betartatva vesz részt a közösség játszmáiban. Ez a felvetés egyúttal magával hozza azt is, hogy van értelme szétválasztani az együtt élő emberek számos interakcióját identitás-specifikus játszmatípusokra. Azaz például a ma Magyarországon élő felnőttek erkölcsi játszmáin belül érdemes külön definiálni egy egyedi játszmákból, és egy metajátszmából álló nemek munkamegosztása játszmát; továbbá egy hasonlóan kettős szerkezetű politikai közösségi játszmát, és így tovább, sorra véve a releváns identitásokat.

Ha az erkölcsi döntések mozgatórugóit a maguk teljességében kívánjuk feltárni, akkor Flache (1996) kollektív cselekvési modelljét alkalmazva, a fenti, különböző, kettős szerkezetű erkölcsi játszmákhoz hozzáilleszthetjük az emberi kapcsolataink privát csereügyleteit is. De ennek tárgyalására ebben a monográfiában nem vállalkozom. A fókusz itt a leegyszerűsített alapmodellen marad.

## VII.1.2. ERKÖLCÉS ÉS RACIONÁLIS DÖNTÉS: A DEFINÍCIÓK

### VII.1.2.1. Erkölcsös viselkedés, erkölcsös cselekvés

Ezennel elérkeztünk arra a pontra, hogy hét évtized játékelméleti indíttatású kutatásának áttekintése nyomán meghatározhatóak lettek az erkölcsös viselkedés és cselekvés fogalmi. A viselkedés és cselekvés megkülönböztetése Weber (1987 [1922]) distinkcióján nyugszik. A viselkedés a külsőleg is megfigyelhető magatartás. Cselekvés a cselekvő szándékai révén értelmet nyerő döntés. Ez a játékelméletben is fontos distinkció (vö. Binmore 2010, Rubinstein 1991). Döntés és stratégia megkülönböztetéséről a II. és IV. fejezetben ejtettem szót.

Az idézett kutatási eredmények, és az azokra építő, általam javasolt kettős játszma modellje alapján a következőt mondhatjuk:

1. Az erkölcsös viselkedés és erkölcsös cselekvés mindig egy konkrét kontextushoz – nevezetesen egy konkrét alapjátszmához, konkrét időponthoz, konkrét csoporthoz, és e csoporton belüli szerepekhez – köthető erkölcsi szabályok szempontjából értelmezhető.
2. Egy alapjátszmában az, és csak az a magatartás tekinthető erkölcsös viselkedésnek, amely az adott kontextus erkölcsi szabályai által megkövetelt mintázatot követi – függetlenül az egyén motivációitól valamint a magatartás egyéni és kollektív jóléti következményeitől.
3. Az erkölcsös viselkedés akkor és csak akkor tekinthető egyúttal erkölcsös cselekvésnek is, ha a cselekvő a csoportjátszma követ-

kezményei miatt viselkedett erkölcsösen; azaz, ha figyelmen kívül hagyta volna a csoportjátzsma következményeit, akkor nem viselkedett volna erkölcsösen.

Az erkölcsös cselekvés tehát valamiféle csoportorientált racionális döntés. A csoportorientált döntés koncepciója itt nem a csoport jólétére irányuló cselekvési motivációt tételez fel. A csoportorientáció és a racionalitás együtt jelenti *a)* a csoport tagjaival várható interakciókhoz köthető, hosszú távú egyéni érdekek figyelembevételét, továbbá *b)* a csoport megfigyelt szokásaira, és a csoport koreográfusai felől érkező viselkedési elvárásokra való odafigyelést. Az erkölcsös viselkedés akkor is lehet csoportorientált racionális döntés, ha nem erkölcsös cselekvés. A csoportorientáció ilyen esetben nem szükségszerűen foglalja magában az *a)* pontban meghatározott kalkulációt. Bizonyos esetekben ugyanakkor nemcsak az erkölcsös cselekvés, de a csoportorientáció is hiányozhat az erkölcsi viselkedésből. Ez utóbbi esetben a cselekvő pusztán az adott alapjátzmára, és az ottani partnereire koncentrálna találja meg azt a számára optimális stratégiát, amelyet amúgy látens vagy explicit szabályok elő is írnak számára.

A fenti meghatározások igyekeznek megőrizni az önző polgár koncepciójának materialista szemléletmódját. A materialista nézőpont következtében elkerülheti a figyelmünket néhány valóban inherens emberi szükséglet. Ugyanakkor ez a nézőpont biztosíthatja, hogy elméleteinkben – a piaci viselkedés elemzéséhez hasonlóan –, jól le tudjuk határolni a mérhető környezeti tényezők és egyéni jellemzők azon körét, amelyek befolyásolják az egyén jólétét, és így döntési dilemmáit.

Az erkölcsös polgár itt javasolt koncepciója abban különbözik a III. fejezetben bemutatott önző polgárétól (a homo oeconomicus játékelmélet által okosított verziójától), hogy az erkölcsös cselekvő figyelmének fókusza sokkal tágabb, mint a homo oeconomicusé: túlnyúlik az aktuális egyszeri, valamint az ismételt interakciók és szereplők elemzésén. Az erkölcsös polgár magát és játékostársait társadalmi csoportok tagjaiként értelmezi; az előtte álló konkrét döntési alternatívákat pedig a csoportlojalításra vonatkozó jelzésértékük, és

e jelzések következményeinek szempontjából is elemzi. A fő célja azonban az itt definiált erkölcsös polgárnak is ugyanaz, mint az önző állampolgárnak: minél kevesebb szenvedés árán jusson minél többhöz, mi szem-szájnak ingere.

Látni kell, hogy a fenti meghatározások egy kutatási folyamat eredményeképp születtek. Ez a végeredmény indokolja a könyv elejétől alkalmazott terminológiát. Tehát ez a terminus, és a hozzá tartozó definíció nem a kutatási kérdés megfogalmazásakor született. Erre már a könyv legelején utaltam. És a koncepció további kutatások nyomán még módosulhat is. Ez a fogalomalkotási szemlélet élesen szemben áll a fogalmi kategóriák stabilitására törekvő filozófiai törekvésekkel (vö. Syi 2014); ellenben szorosan következik a Lakatos-féle (1998 [1973]) evolutív tudomány szemléletből – amely e monográfia gondolatmenetét alapvetően meghatározta.

Azt is látni kell, hogy ez a meghatározás nagyon idegen az altruizmusra és a hozzá hasonló elvekre épülő modellek (Bereczkei 2009) szemléletétől. Nagyon közel áll ellenben a Durkheim-féle klasszikus szociológiai szemlélet erkölcsképéhez, amelynek ma is termékeny, és nagyon fontos kutatási iránya például a szerepelvárások és szerepidentitás vizsgálata (Stets és Burke 2000, Stryker 1968, Stryker és Serpe 1982). Ugyanakkor ez a hasonlóság nem jelenti azt, hogy a Durkheim-féle funkcionalista szemléletmód is jellemezné ezt a koncepciót.

Továbbá azt is látni kell, hogy a fenti, Durkheim-féle koncepcióhoz közel álló erkölcsmeghatározás a Durkheim metodológiai nézeteivel kevésbé konguens, Weber által (is) megalapozott, megértő társadalomtudomány metodológiáját alkalmazva született. Azaz a fenti erkölcsdefiníció a durkheimi koncepciók weberiánus alapokon történő megfogalmazása.

#### VII.1.2.2. Erkölcsös egyensúlyok

A fentiek alapján az erkölcsi szituációk néhány alapvető típusát is elkülöníthetjük. Létezhetnek olyan alapjátszmák, amelyekben

- a) a normák univerzális követése akkor és csak akkor valósul meg, ha mindenki erkölcsös döntést hoz;
- b) a normák univerzális követése során az alapjátzsma szereplőinek egy része erkölcsös döntést hoz, míg a többi szereplő, a csoportjátzsma ösztönzőitől függetlenül (a többiek normakövetését természetesen figyelembe véve) is erkölcsösen viselkedik;
- c) a normák univerzális követése a csoportjátzsma ösztönzőitől függetlenül valósul meg, azaz az univerzális normakövetés az alapjátzsma egyensúlyi stratégiakombinációja.

Az első helyzet tipikus példája a véges fogolydilemma (külön büntetési lehetőség, és egyéb ravaszkodások nélkül). A harmadik helyzetbe tartoznak a normák játékelméleti elemzői által vázolt erkölcsi szituációk (VI. fejezet), beleértve Binmore (2005, 2010, 2015) iterált játzmáinak adaptív egyensúlyait, valamint Gintis (2009, 2016) és Guala (2016) koreográfusok által meghatározott korrelált egyensúlyait is.

A normákkal és szerepelvárásokkal kapcsolatos viták struktúrájának és dinamikájának – és vele a kultúra evolúciójának – jobb megértése szempontjából kulcsfontosságúnak tartom a köztes kategóriát (b). Ezekben a helyzetekben a csoporttagok egy része természetesnek tekinti a norma követését, mert az számukra pusztán az értelmes döntésre rámutató figyelmeztetés. A többiek azonban külső nyomásként élik meg az elvárásokat – hiszen az adott szituációban nem érzik a belső késztetést a norma által előírt viselkedésre. Különösen akkor lehet jelentős a két alcsoport közötti feszültség, ha azok, akik a normát erkölcsi ösztönző nélkül is követik, érezhetően jobban járnak az erkölcsösnek tekintett viselkedéssel, mint azok, akiket csak a csoportjátzsma ösztönzői vesznek rá arra, hogy az erkölcshez igazodjanak.

Úgy vélem, hogy például a késő modern társadalmak nemi szerepelvárásaihoz kötődően számos ilyen aszimmetrikus szituáció keletkezik. Jellemzőbb, hogy elsősorban a nők egy része érez külső nyomást ott, ahol a férfiak többsége természetes nemi magatartásmintákat lát. Ugyanakkor fordított helyzetek is vannak.

A következő alfejezetben röviden, de szisztematikusan próbálom áttekinteni, hogy a fenti modellezési keretből milyen alapvető következtetések származnak.

## VII.2. Erkölc alapú érdek, érdek alapú erkölcs

### VII.2.1. FÜGGŐSÉG, IDENTITÁS, ERKÖLCS

#### VII.2.1.1. Érdék vs. beágyazottság

A kettős játszma modellje általános elmélet. A középszintű elméleteket útmutatóval segíti abban, hogy megtalálják azokat a környezeti tényezőket, amelyek befolyásolják, hogy valaki követi-e az erkölcs által előírt magatartási mintákat. Az egyéni viselkedést befolyásoló tényezők felfedésével a kollektív dinamikára is következtethetünk.

A modellre épülő útmutató legfontosabb megkülönböztető sajátossága, hogy felhívja a figyelmet arra kérdésre, hogy mennyire függ – pusztán anyagilag –, a cselekvő az erkölcsi szabályokat meghatározó csoporttól. Konkrétan: a csoport tagjaival várható tranzakciók mit jelentenek neki hosszú távon.

Vegyük a mai magyar, fiatal értelmiségi párok családtervezési alakúját! Azt akarjuk megérteni, hogy milyen társadalmi tényezők járulhattak hozzá, hogy egyes párok korán vállalnak gyereket, mások később, van ahol az anyuka ígéretes karrierjét feladva sokáig otthon marad, míg máshol időigényes állást vállal, és a szülés után gyorsan bölcsődébe kerül a gyerek, stb.

A homo oeconomicus modellje jól körülhatárol néhány fontos befolyásoló környezeti tényezőt. Mindenekelőtt a családalapítók emberi tőkéjének aktuális értékét meghatározó tényezőket kell figyelembe venni: például mennyire piacképes a két fél tudása, stb. Költségvetési korlátok is számíthatnak: például milyen a kiinduló lakáshelyzetük. Leginkább szociológusok foglalkoznak a befolyásoló tényezők egy másik csoportjával: a társadalomban uralkodó erkölcsi szabályok hogyan formálták közvetlenül és közvetve a két fél embe-

ri tőkéjét. A tényezők egy harmadik, az ilyen helyzetek elemzésekor szintén gyakran vizsgált típusát alkotják az (adott anyagi érdekek mellett megfigyelhető) erkölcsi motivációk: a különböző társadalmi csoportok egymástól eltérő elvárásokat fogalmaznak meg a nők család- és karriertervezési mintáival kapcsolatban.

A kettős játszma modellje összeköti a fenti közelítéseket. Fontosnak tartja az anyagi érdekek vizsgálatát, de nem csak a két fél piaci lehetőségeire fókuszál, hanem arra is, hogy az emberi tőke kihasználásának különböző pályái milyen csoportkötődéseket hoznak létre. A másik irányból tekintve: a kettős játszma modellje fontosnak tartja a releváns erkölcsi szabályok vizsgálatát, de nem önmagukban elemzi a társadalmi környezet elvárásait, hanem azt is megnézi, hogy milyen (anyagi) függőségek kötik a különböző elvárásokat megfogalmazó közegekhez a szereplőket.

#### VII.2.1.2. Érdek és beágyazottság: erkölcs

A konkrét példában a szereplőkre a családi szerepmegosztás döntéseiben hatnak „kibocsátó” családi-baráti közegek nemi szerepelvárásai. Tétélezzük fel, hogy ezek erősen szétváló férfi és női szerepeket (ún. hagyományos családmódellet) promotálnak. Ezek az értékek családi ebédeken, és a régi iskolai barátokkal történő kocsimázások alkalmával manifesztálódnak. Direkt módon (például „gyerek nélkül nincs élet”) vagy indirekt módon, mások viselkedésének megítélésekor (például „Manci szerintem egy önző, karrierista néember”). Ezek a szerepelvárások egy (szubjektíve) jó karrieresélyekkel rendelkező nő számára magas-, míg egy jó karrieresélyekkel rendelkező férfi számára alacsony költségeket jelentenek. Erősebb normatív nyomás kell ahhoz, hogy a nő engedjen e szerepelvárásoknak, mint a férfi. A kevésbé perspektivikus kapitalista karrier elé néző nők és férfiak számára fordított a helyzet. A normatív nyomás erőssége e modelltben nem egy exogén változó, hanem az határozza meg, hogy a szerepelvárással szembesülő személy mennyire szorul rá érzelmileg és/vagy anyagilag a szerepelvárásokat megfogalmazó közegek tagjaival folytatott interak-

ciókra. A családi ebédeken elhangzó erkölcsi elvárásoknak akkor van súlya, ha a fiatal még akar valamit szüleitől, vagy a tágabb családjától. A kibocsátó közeg mellé odaképzeltünk egy új, egyetemi, munkahelyi közegből kinövő baráti-ismerősi közösséget is, ahol esetleg épp ellentétes erkölcsi parancsokat osztogatnak egymásnak az emberek. Ha túl akarnám bonyolítani, akkor itt még szétválaszthatnánk a férfiak és nők diszkurzív közegét. Tegyük fel, hogy ez a közeg egyrészt emancipatorikus, másrészt erősen karrierorientált erkölcsi értékeket közvetít a fent leírt direkt és indirekt módokon. E közegben a normatív nyomások nemek és karrieraspirációk szerinti megosztása ellentétes a kibocsátó közegben tapasztaltakkal, de ugyanazon logika alapján működik, és itt is a jövőbeni interakciók perspektívája határozza meg a normatív nyomás mértékét. A privát cserefolyamatok hatásainak Flache (1996) által javasolt bevonása azt jelentené, hogy – immár tényleg a párok magánéletében turkálva – megvizsgáljuk, hogy milyen érzelmi és fizikai boldogságot tudnak nyújtani egymásnak a felek, és ebben mennyire érezhetik magukat monopolhelyzetben. A példának hozott nemek harca játszmákban ezek a privát alkufolyamatok felülírhatják a szociológiai tényezők hatását egyes párok esetében. Hasonló mechanizmust írt le Flache (1996) és Flache és Macy (1996) az erős kapcsolatok gyengeségét vizsgálva: a diádok privát cserefolyamatai alááshatják a közös erkölcsi értékeket (a kollektív cselekvést).

A normakövetésre ösztönző belső késztetéseket feltételező modellel szemben a kettős játszma modellje az erkölcsi viselkedést összeköti a csoporttagsággal. Ráadásul nem binárisan, mivel a csoporton belüli interakciók perspektíváin keresztül értelmezni képes az erősebb és gyengébb kötődéseket is. És ha eljutunk idáig, akkor kezelni tudjuk a versengő identitások problémáját is. Ezt a fenti családtervezési példa is mutatja. A különböző identitások szembetűnőségének hatása tekintetében a modell előrejelzései összhangban vannak a szociálpszichológia régi, gazdag empiriával megtámogatott várakozásaival. De ezen túlmenően a kettős játszma modell segítségével vizsgálni lehet a különböző identitásokhoz tartozó csoportokkal való viszony strukturális jellemzőit is.



A piaci viselkedés elemzése során a középszintű közgazdasági elméletek (beleértve a piacon kialakuló lokális normák vizsgálatait is, lásd Greif 1994, Kandori 1992) azt a kérdést teszik fel, hogy milyen anyagi költségekkel és hasznokkal járnak az elméletek által vizsgált piaci tranzakciók különböző kimenetelei. Az erkölcsös polgár koncepciójával kongruens középszintű szociológiai elméletek a konkrét tranzakciók értelmezésekor azok közvetlen következményei mellett azt vizsgálhatják, hogy hosszú távon milyen anyagi hasznokkal és költségekkel járnak a csoporttagoknak a szerepelvárásoknak megfelelő, illetve az azt elutasító stratégiák. A szociológia kétségtelenül sokat foglalkozik a társadalom normatív rendjének különböző csoportokra gyakorolt anyagi hatásával. Továbbá a szociálpszichológia fontos kutatási területe annak mérése, hogy milyen erővel képes befolyásolni döntéseinket alapjátázmákban az, ha felhívják figyelmünket a csoporttagságunkra és az ahhoz kötődő szerepelvárásokra.

Leginkább egyes gazdaságsszociológiai kutatások próbálják meg a két – önmagában is rendkívül fontos! –, kutatási területet szisztematikusan, racionális értelemzés segítségével összekapcsolni, és a szerepelvárások hatásainak elemzésébe bevonni az alapjátázmák következményeinek, valamint a csoportfüggőség mértékének a vizsgálatát is. Például a bevándorlók emberi- és társadalmi tőkétől függő stratégiáinak kiváló vizsgálatai, a racionális döntésemélet elutasítása ellenére, egész szoroson és eredményesen követik a fenti koncepcióból levezethető kutatási stratégiát (Di Maggio és Garip 2012, Nee és Sanders 2001, Portes 1995). Ez talán nem véletlen, ugyanis egyes bevándorló csoportok etnikai gazdaságainak empirikus kutatása során szembetűnő egyrészt az anyagi javak szerzésére irányuló tevékenység társadalmi beágyazottsága, másrészt az, hogy lehetséges a különböző erkölcsi szabályokat előíró csoportok közötti választás, és e választásnak komoly anyagi következményei vannak.

## VII.2.2. AZ ERKÖLCS HATALMA

### VII.2.2.1. A csoport hatalmából erkölcsi játéktér

A csoportjátzsma létrehozza azt az ösztönzőt, amely normatív erőt ad az egyes alapjátzmákban erkölcsösnek minősülő döntések meghozatalához. És ha az erkölcsös döntéseket az alapjátzsma ösztönzőin túlmutató normatív erő támogatja, akkor egy adott alapjátzmában még több döntéskombináció válhat egyensúlyivá, mint amennyit az alapjátzmában definiált korrelált egyensúly koncepciója megengedne. Ugyanis a normatív erő átalakítja az alapjátzsma eredeti következményeit, pozitív és negatív boldogságértékeket adva egyes alternatívákhoz. Például a VI.3.1.3. alfejezetben bemutatott, terrortámadás utáni menekülés alapjátzmájában hiába domináns stratégia a fejvesztett menekülés, ha a közösség közös tudatában az él, hogy megvetendő dolog a csoport kollektív érdekét segítő menekülési tervtől eltérően előretolakodni. Nem a következő terrortámadásban, hanem az élet más területein korlátozza majd emberi interakcióiban az előretolakodó túlélőt az, ha más túlélők elterjesztik róla, hogy mit tett. Ha elég erős ez az erkölcsi szabály (azaz komoly büntetésre számíthat az, aki megszegi), akkor Nash-egyensúlyivá válhat például az a döntéskombináció (is), amelyben mindenki csak akkor tolakszik, ha a többség ezt teszi (vö. Bicchieri 2006, Bicchieri et al. 2018).

A kettős játszmában létrejövő normatív erő alapvetően különbözik a jelenlegi szakirodalomban feltételezett, velünk született általános normakövetési hajlandóságtól. Az utóbbi az egyént jellemzi, szituációtól függetlenül: vannak erősen, és vannak kevésbé normabetartó emberek. A kettős játszmában létrejövő normatív erő ezzel szemben a szituációt jellemzi – noha, ahogy nemsokára kifejtem, nem zárja ki az egyén pszichológiai beállítottságának figyelembevételét sem. A normatív erő hatásának egyének közötti különbségeit az egyének jövőben várható interakciós struktúrája (a normát kibocsátó csoport tagjaival várható interakciók jelenértéke) határozza meg. Az egyén normatív beállítottságára építő modell a fenti példában szereplő, New Yorkban dolgozó magyar normasértő magatartását az

illető személyiségére vezeti vissza. A kettős játszma modellje ellenben rá tud mutatni arra, hogy egy vendégmunkás számára az esetleges amerikai kiközösítés nem feltétlenül jelenti életlehetőségeinek oly mértékű ellehetetlenítését, mint egy bennszülött számára. Úgy számolhatott, hogy hazájában nem fogják megvetni azért, hogy az alapjáték domináns stratégiáját követte. Tettéről én sem rosszindulatú pletykákból értesültem, hanem onnan, hogy ő maga mesélte el egy magyar újságnak, amelyik leközölte. Egy őshonos amerikai nehezebben vonul ki az őt esetleg megvető amerikai társadalomból, mint egy vendégmunkás.

A fentiekből kirajzolódik az összefüggés az egyének csoportfüggősége, valamint az erkölcsi szabályok korlátai között. Egy olyan csoportban, amelyből komolyabb negatív következmények nélkül lehet távozni, az erkölcsi szabályok az alapjátékmák korrelált egyensúlyai közé vannak szorítva. Mászt nem fogadnak el a csoporttagok. Ezekben a csoportokban a koreográfus leginkább az egyensúlyok közötti választást segítő szolgáltatást nyújt a tagoknak. Ellenben egy olyan csoportban, amelynek tagjaival folytatott interakciókban nehezen helyettesíthető javakhoz jut az egyén, az erkölcsi szabályok elvileg sokkal tágabb keretek között határozhatók meg. Ilyen csoportokban az alapjátékmák egyensúlyai nem nyújtanak értelmes útmutatót az elvileg lehetséges világokról. Ugyanis ilyen csoportokban – némi túlzással – szinte bármi fenntartható erkölcsi nyomással.

A csoportkötődés és erkölcsi szabályok összefüggésére vonatkozó fenti következtetések kongruensek Durkheim meglátásaival, továbbá a normák és a társadalomszerkezet összefüggéseit vizsgáló későbbi irodalommal is. Például Lindenberg (1988) különíti el a csoporttagokat összekötő, erős, illetve gyenge szolidaritás koncepcióját. Az előbbi többet vár el az egyénektől, és többet is tud nyújtani a csoporttagoknak, mint az utóbbi. Az utóbbi ellenben az egyének csoportok közötti mozgásával kompatibilisabb erkölcsi szabályokat hoz magával, mint az előbbi. A kettős játszma modellje segít abban, hogy pontosabban írjuk körül azokat a körülményeket, amelyek erősebb, illetve gyengébb erkölcsi hatalommal rendelkező csoportok létrejöttéhez vezetnek. A korábbi elméletek fókuszában a modernizáció cso-

porterőre gyakorolt hatásának általános trendjei álltak. Azonban a kortárs társadalmakon belül is jelentős különbségek lehetnek abban, hogy különböző emberek milyen költségek árán képesek kiszabadulni a csoporton belüli játszmákból. Erre nemsokára visszatérek.

#### VII.2.2.2. Endogén csoportfüggőség

Egyes csoportok egyes tagjai nemcsak az erkölcsi szabályokat képesek befolyásolni, de a szabályokon keresztül manipulálni tudják a többi csoporttag csoportfüggőségét is. A világ legkülönbözőbb (szub)kulturáiban figyelhetőek meg az egyes csoporttagok menekülési útjait nehezítő, az erkölcsrendszert erősítő erkölcsi szabályok. A normák játékelméleti modellje alapján az ilyen, menekülést gátló szabályok nagyobb eséllyel jelenhetnek meg azokban a helyzetekben, ahol az alapjátszmában hozott döntés hosszú távra, és sok élethelyzetre határozza meg a cselekvő boldogságát, és ezért az alapjátszma nagy tétje miatt az erkölcsi szabályok relatív ereje kisebb. (Például a korábban már említett párválasztási helyzetekben.) Ennek irodalma jelentős, és része a nemek közötti egyenlőtlenség kurrens közéleti diskurzusának is.

A csoportfüggőség manipulálása „kinek az érdeke?” jellegű kérdéseket vet fel az erkölcsi szabályokkal kapcsolatban. Ez átvezet minket az erkölcs jóléti elemzéséhez. Az erkölcs jóléti szempontból semleges definiálása, és az erkölcsi világok sokféleségére vonatkozó következtetés nem zárja ki, hogy a kettős játszma modellje alapján jóléti elemzéseket végezzünk. Úgy vélem, hogy ez a modell a jóléti elemzések finomításában is segít.

Az világos, hogy az elegendően erős függőségben lévő egyénekből álló csoportok tudnak megoldani nagy tétekkal rendelkező fogolydilemmákat. A kollektív cselekvésre vonatkozó szakirodalom ugyanakkor arra is figyelmeztet, hogy adott esetben a normatív nyomás a kollektív rossz fenntartásához is hozzájárulhat. Ez utóbbit a kettős játszma modellje sem zárja ki. Azonban ebben a modellezési keretben fel kell tenni a kérdést: kik erőltethetik, hogy a mindenki számá-

ra káros erkölcsi szabályok kristályosodjanak ki a csoporttagok közötti kommunikációs térben?

Az általam javasolt modellezési útmutató a fenti kérdésre azt a választ adja, hogy érdemes keresni azokat a csoporttagokat, akiknek érdekében állhat a mások számára káros erkölcsi szabályok közös tudatba juttatása és/vagy ott tartása. Az erkölcsös cselekvő számára negatív következményekkel járó erkölcsi szabályok így ráirányítják a figyelmet az aszimmetrikus helyzetekre, amelyeket már említettem korábban. Arra is utaltam, hogy az aszimmetrikus következményekkel járó erkölcsi szabályok által uralt alapjátzmák veszteségei és nyereségei hosszú távon kiegyenlíthetik egymást. Bizonyos csoportokban azonban az erkölcsi szabályok betartásának következményei szisztematikusan és jelentősen különböznek a különböző alcsoportokban. Az ilyen csoportok vizsgálata átvezet minket a csoporttagok közötti hatalmi viszonyok kérdéséhez, és az erkölcsi szabályok tartalmi értékeléséhez.

### VII.2.3. A HATALOM ERKÖLCSE?

#### VII.2.3.1. Erkölcstelen erkölcs

Az előző alfejezetben azt hangsúlyoztam, hogy az erkölcsi szabályok extrém hatalma nem extra erős normatív preferenciák következménye. A súlyos érdeksérellemmel járó erkölcsös döntések extrém kiszolgáltatottságból erednek. A normák játékelméleti modellje alapján azt is állítottam, hogy egyes csoporttagok hosszú távon fennmaradó érdeksérelmei mögött érdemes más csoporttagok érdekeit keresni. Ilyen csoportviszonyok a kényszer és hatalom koncepcióihoz vezetnek a kutatót. Ezek a helyzetek sokunk erkölcsi érzékét sértik, és óhatatlanul megkérdőjelezzik a kettős játszma modelljében definiált erkölcs-koncepció termékenységét. Többek között felmerül az a kérdés, hogy miképp kezeli ez a modell a hatalom és uralom koncepciójának hagyományos distinkcióját (Weber 1987 [1922]), amely megkülönbözteti a nyers erő előtti meghajlást a morális legitimitással rendelkező

parancsnak történő engedelmességtől. Az alábbiakban azt próbálom elmagyarázni, hogy miképpen létezhet erkölcstelen erkölcs, és hol lehet a helye a modellben a hatalom és az uralom koncepcióinak.

Az erkölcstelen erkölcs értelmezésének kulcsa ebben a modellezési keretben az erkölcsi szabályok kontextusfüggősége. A kontextusfüggőség két meghatározó dimenziója az erkölcsi szabályt érvényesítő csoport mérete, valamint a szabály időbeni tartóssága. Az erkölcstelen erkölcs a különböző kontextusban született erkölcsök ütközéséből ered. Ez nagyon régi gondolat. Ahogy – ehhez kapcsolódóan – az is, hogy az idők változásával ugyanabban a csoportban erkölcstelené válhat az, ami korábban erkölcsös volt. A kettős játszma modellje alapján azonban ennél többet is mondhatunk.

#### VII.2.3.2. Személyközi hatalmi játsz máktól az egyetemes erkölcsig

A kettős játszma modellje segít annak megértésében, hogy nemcsak egymástól időben vagy térben szeparált csoportok erkölcs e különbözhet. Kis csoportok egyes erkölcsi szabályai élesen szemben állhatnak a kis csoportot magában foglaló nagyobb csoportok szabályaival. A kis csoport lehet egy két főből álló háztartás is, ahol olyan szabályok érvényesülnek, amelyek a háztartást magában foglaló tágabb társadalmi közegben elfogadhatatlanok. Mindaddig, amíg a kiscsoporton belül az erkölcsös cselekvés itt lefektetett mechanizmusa tartja fenn az egyes belső szabályokat, e szabályok a kiscsoport (erkölcstelen) erkölcsi szabályai. Tehát ha egyes háztartástagok hosszú időn keresztül túri k más tagok erőszakos viselkedését azért, hogy ne kerüljenek utcára, és legyen mit enniük, akkor ők a „családjuk” erkölcstelen lojalitási erkölcs einek engedelmeskednek – anyagi szükségszerűségből. A kettős játszma modelljének útmutatója szerint egy isten háta mögötti faluból nagyvárosba csábított, *látszólag* szabad prostituált „szermes lojalitása” mögött is a túlélés kényszereit kell keresni, és „szermes lojalitása” az itt definiált értelemben (erkölcstelen) erkölcsös cselekvés (az érzelmek értelemezésére nemsokára még visszatérek).

A kettős játszma modelljében ezeknek az egészen kiscsoportos játszmáknak nemcsak a külvilág értékítélete miatt gyengül az erkölcsi tartalmuk. Ezek a játszmák a modellben határesetet képeznek. Minél egyszerűbben leírhatók a felek közötti interakciók, annál erősebben lehet érvelni amellett, hogy a kiscsoportra vonatkozó csoportjátzmát ne kettős játszmának, hanem egyetlen ismételt alapjátzmának tekintsük. De még ha kettős játszmaként modellezük a kiscsoportos hatalmi viszonyokat, a hatalmi kényszer alkalmazó játszma súlya nagyra nőhet a felek összes interakcióján belül. Ha valakit ugyanaz a férfi ver meg minden nap, ugyanazokat a kényszerítő eszközöket használva, és minden egyéb napi tevékenységüket is beárnyékolja a fizikai erőszak fenyegetése, akkor ez egy erkölcsi tartalom nélküli, rendkívül aszimmetrikus játszmának tekinthető. A szeparált csoportjátzma eljelentéktelenedésével pedig az uralmi viszonyok egyszerű hatalmi viszonyokká alakulnak át.

A kiscsoportok erkölcstelen erkölcei nemcsak aszimmetrikus tranzakciókat irányíthatnak. Bonnie és Clyde típusú, erősen lokális „közjavak” (maguknak rabolt pénz) előállítását is segíthetik az ilyen szabályok.

Ugyanakkor számos tragikus példa mutatja, hogy nagyobb közösségekben is észlelhetők erkölcsi legitimitás nélküli hatalmi viszonyok. A modell útmutatásai szerint az ilyen helyzeteknek kétféle értelmezési irány lehetősége. Az egyik a hatalmi kényszer elfogadásához kötődő, a nem elég alapos megfigyelő elől rejtve maradó erkölcsi döntések feltárása. Erre lehetnek példák Kertész Imre *Sorstalanságának* leírásai a lojálisan viselkedő zsidó foglyokról. Kertész számára kifejezetten fontosak az engedelmesség motivációi. A lényeg: a hatalmi kényszer alatt szenvedő közösségekben is működhethetnek hosszabb vagy rövidebb ideig az engedelmesség erkölcsi szabályai. A másik értelmezési irány pedig a szereplők közötti játszmák feltárása. Nagyobb csoportban is előfordulhat, hogy a hatalmi kényszer játszmái kényszerítő és kényszerítettek közötti páros (vagy kiscsoportos) alapjátzmák, amelyekre a fenti, a kiscsoportok hatalmi viszonyaira vonatkozó meglátások érvényesek. Azaz az erkölcs eltűnhet a játszmákból. Ilyenkor nem más elnyomottakkal együtt játsszák a szenvedők az engedelmisségi játsz-

mát, hanem egyes elnyomó ágensekkel. Az engedelmesség az elnyomó ágens közvetlen fenyegetésének eredménye, és nem a szenvedők nagycsoportjához tartozás iránti vágy hajtja.

Az erkölcsi spektrum másik végén találhatóak a robusztus, (kvázi-)univerzális erkölcsi szabályok. A kettős játszma modellje jó alapokat szolgáltat e szabályok sajátosságainak szituációs logikai alapú vizsgálatára is. Egy ilyen vizsgálódás túlmutat e könyv keretein. Nem nélkülözheti a normákra vonatkozó filozófiai és evolúciós irodalom tanulságait sem. Ezzel az irodalommal itt nem foglalkoztam (vö. Binmore 2005, Syi 2014).

### VII.3. Nyitott kérdés (?): a morális indulat

#### VII.3.1. ÉRDEKEK ÉS ÉRZELMEK

##### VII.3.1.1. Az érzelmek kiértékelélmélete

A bennünk élő normatív erőt feltételező elméletek látszólag erős empirikus alapokra építkeznek. A korábbiakban én is megmutattam, hogy a mások normasértését követő indulatok, és az indulatok nyomán alkalmazott altruista büntetések széles körben megfigyelt jelenségek (IV. fejezet). Az altruista büntetésre utaló magatartás sokkal robusztusabban jelenik meg a társas preferenciákat vizsgáló kísérletekben, mint bármilyen konkrét társadalmi rend iránti vágy. Ezért a normák játékelméleti modelljei (Bicchieri 2006, Gintis 2009, 2016) is feltételezik, hogy az ilyen indulatok valamiféle „rásegítő” szerepben terelik a normakonform viselkedés irányába az embereket.

Ugyanakkor az erkölcs új pszichológiáját ismertető V. fejezet egyik fő tanulsága az volt, hogy megkérdőjelezhető az a modell, amely az érzelmeket a viselkedés végső mozgatórugóiként értelmezi. Az undorreakció ideológiai befolyásolhatóságára vonatkozó egyes eredményeken túl (Ciuk 2018) az érzelmek pszichológiájával foglalkozó irodalom mutat egy, a viselkedési közgazdászok által átvett modelltől egészen eltérő kognitív perspektívát.



Az érzelmek kiértékelésmélete (appraisal theory) azt tételezi, hogy érzelmeink nem kognitív folyamatainkat torzító korlátok, hanem éppen a világról szerzett ismereteinket költségtakarékos módon összegző döntéselőkészítő segédeszközök. A külső szemlélő racionális elemzéssel beláthatja, hogy milyen típusú ingerek és perspektívák milyen típusú érzelmeket implikálnak.

Az érzelmek kiértékelésmélete szerint egy kétlépcsős folyamatban keletkeznek az érzelmek (Lazarus 1991). Az elsődleges értékelés során felmérjük, hogy a környezetünk jellemzői összhangban vannak-e a céljainkkal. Ez alapján keletkeznek pozitív, illetve negatív érzelmek bennünk. A másodlagos értékelés során részletesebben felmérjük a helyzetet és a lehetséges kezelési stratégiákat, majd ezek alapján alakulnak ki cselekvésünket ösztönző, konkrét érzelmek (öröm, sajnálat, méreg, undor, félelem stb.). Például ha fenyegetést észlelünk és nem kontrolláljuk a helyzetet, akkor félelemmel reagálunk (amely menekülésre készít minket); ellenben ha a fenyegetettséggel a kontroll érzete és a félelem forrásának ismerete társul, akkor dühösek leszünk, és támadásba lendülhetünk a fenyegetés forrása ellen (Lazarus 1991). A dühvel kapcsolatban például Mackie és szerzőtársai (2000) kimutatták, hogy dühös emberek könnyebben aktivizálódnak ellenségesnek tartott csoportokkal szemben.

Koncentráljunk most a negatív ingerekre! Bizonyos helyzetekben egy számunkra kellemetlen objektum elkerülése a legcélravezetőbb, mindenféle egyéb akció nélkül. Ilyenkor az undorhoz hasonló, elkerülésre sarkalló érzelmek irányítják döntéseinket. Máskor az egyszerű elkerülés még nem oldja meg a problémánkat, mert a veszélyt hordozó objektum aktívan tehet azért, hogy újra a közelünkbe kerüljön. Ilyenkor a félelem lesz rajtunk úrrá – azért, hogy az egyszerű elkerülés ne hozzon megnyugvást számunkra. Van azonban olyan helyzet, amikor a negatív inger okozója nem olyan veszélyes, hogy ne tudnánk megzabolázni, és a megzabolázás inkább érdekünkben áll mint az egyszerű elkerülés. Például azért, mert az objektum megzabolázása vagy elpusztítása nélkül a negatív ingerek nem szűnnek meg. Ilyen helyzetben dühösek leszünk. Hasonlóan reagálunk akkor is, ha a megzabolázott objektummal később még számunkra gyümölcsöző

módon léphetünk interakcióba. Ilyenkor az érzelmek pszichológiája szerint dühösek leszünk, és támadólag lépünk fel az objektummal szemben (Lazarus 1991).

Az érzelmek kavalkádját a IV. fejezetben bemutatott laboratóriumi játékokban is vizsgálták. Chapman és szerzőtársai (2009) ultimátumjátékot játszottak a kísérleti alanyaikkal, akik a kutatók által felajánlott pénzre kaptak elosztási javaslatot egy általuk ismeretlen személytől. A javaslatokat valójában a kutatók manipulálták, hogy figyelhessék a méltánytalan ajánlatokra érkező érzelmi reakciókat. A fele-fele felosztási javaslatokkal a válaszadók döntően elégedettek voltak. Ahogy az ismeretlen ember által ajánlott részesedésük csökkent, úgy lettek egyre szomorúbbak, dühösebbek és éreztek undort is. Kissé méltánytalan ajánlatoknál a szomorúság és az undor dominált, „pofátlan” ajánlatok esetén ezek rovására erősödött a válaszadók dühe.

#### VII.3.1.2. Közgazdaságtani interpretációk vs. pszichológiai elméletek

A düh vizsgálata kapcsán ezen a ponton visszajutunk az érzelmek közgazdaságtani elemzésének érveléséhez (Hirshleifer 1987). Eszerint az érzelmek nem nekünk, hanem partnereinknek nyújtanak hiteles információt az elkötelezettségeinkről. Az érzelmeink kifejezését nehéz kontrollálni, miközben érdemben – és a közgazdaságtani érvelés szerint –, az érdekeinkkel is szembeszállva befolyásolják döntéseinket. Hatásuk lehet pozitív is. Esetünkben a düh fenyegetéseket hitelesítő szerepe az érdekes. A dühünk fejezi ki, hogy egy konfliktushelyzetben olyat is képesek vagyunk megtenni, amit utólag magunk is megbánunk (azaz esetleg nem a Nash-féle racionális útmutató alapján választunk stratégiát). A korrelált stratégia értelmezéséhez bemutatott példában, a VI. fejezetben található „gyáva nyúl” játszmában kiválóan értelmezhető ez a funkionalista érvelés. Ha dühös vagyok ellenfelemre, akkor az indulat tudatszűkítő hatása alatt mindenképpen támadni fogok. Ha ellenfelem racionális, akkor ő meghátrál, és így a „meccset megnyertem”. Az erős reciprocitás elmélete (Gintis 2000)

ilyen automatikus, irracionális, a racionális partnereket a normasértéstől elrettentő érzelmek tekinti az altruista büntetést. És az ultimátumjátékra vonatkozó viselkedési közgazdaságtani kísérletek valóban arra utalnak, hogy az emberek tartanak, és jogosan tartanak az ultimátumot kapó második játékos irracionális dühétől.

Ezt az érzelmeket magyarázó közgazdasági modellt nem szeretném cáfolni. Mindazonáltal arra hívnám fel a figyelmet, hogy az érzelmek kiértékelésmélete nem zárja ki az érzelmek ilyen stratégiai működését, hiszen az alapfeltevése éppen az, hogy az érzelmek racionális döntéseinket segítik. A düh esetében azzal is, hogy bevállaljuk a rövid távú költségeket a hosszú távú sikerek érdekében. Ennek egy verziója, hogy költségvállalási hajlandóságunkat hitelesen közvetítjük a megzabolázandó fél irányába, aki jó esetben meghátrál, és megment minket a költségektől. A két elmélet kompatibilitása visszafelé nem működik. A Hirshleifer-féle (1987) elmélet specifikus érzelmek specifikus funkcióit magyarázza, és nem kezeli az érzelmek sokrétűségét. Ahogy arra az érzelmek egyes kutatói rámutatnak, az egyes érzelmek maguk is sokrétűek lehetnek, és a velük kapcsolatos fogalmi lehatárolásaink is társadalmi konstrukcióknak tekinthetőek (Barrett 2009, Lindquist 2013). Például maga a düh koncepciója is sokféle érzelmi állapotot jelölhet. Van olyan típusú dühöngés, amely alapvetően kifelé szól, és hitelessége épp ezért megkérdőjelezhető – legalábbis a partnerek gyorsan megtanulhatják, ha valaki ugyanabban a körben többször próbál így előnyhöz jutni. De dühként élük át az emberek az olyan csendes érzelmet is, amely a másik ijesztgetése helyett a bátorság összeszedésében segíti a támadni kívánó felet (akit céljai elérésében éppenséggel akadályozhat is, ha előre felhívja saját érzelmeire a megtámadott partner figyelmét). Az érzelmek fenti közgazdasági modellje hallgat a legalább olyan fontos, ám nem az ellenérdekelt feleknek szóló fenyegetés vagy ígéret funkcióját ellátó érzelmekről. A sokat emlegetett undor elkerülő magatartása nem fenyeget senkit. A félelem is a fenyegetés elfogadásáról szól. Ezeknek az érzelmeknek is lehet kifelé szóló üzenete, de elsősorban a velünk közös érdekelt-ségű felek számára nyújt információt, például: „Ti is figyeljtek oda, veszélyes objektum a közelben!”

Térjünk vissza az undorra! Bizonyos objektumok (például fertőzés-veszélyes élőlények) ösztönösen, minden tanulás nélkül kelhetnek undort bennünk. Más objektumokkal (például az adott kor és társadalom normáit sértő szexuális viselkedést folytató emberrel) kapcsolatban meg kell tanulnunk, hogy el kell kerülnünk őket. Az undor kialakulásának e két formája keveredik az erkölcs új pszichológiájának elméletében – amely a tanulási folyamatot is mindenáron ösztönös ingerekre akarja visszavezetni. Az érzelem kiértékeléselmélete szerint ugyanakkor az undor közvetíti a veszélyre vonatkozó információt – szerezzük bárhogyan is azt: génjeinkbe kódolva, vagy racionális tanulás útján.

Ebben a vizsgálati keretben az érzelmek mint információközvetítő mechanizmusok jelenléte – különösen mint a hosszú távú önérdék közvetítő mechanizmusai az egyszeri játszmákban –, nem mond ellent a kettős játszma tisztán racionális, játékelméleti alapú elemzésen nyugvó modelljének. Hiszen az érzelmek részben éppen azokat az információkat foglalják össze, és gyűrik döntésbefolyásoló mechanizmusokká, amelyek a játékelméleti modellekben olyan, bonyolultabb racionális elemzéseket igényelnek, amelyek elvégzését antropológiailag hitelesen nem feltételezhetjük. Ezáltal éppen megnyugtatják az egyre bonyolultabb modelleket építő játékelméleti kutatót: nem kell feltétlenül félni attól, hogy komplex módon modellezzünk egy ismétlődő játszmákat is tartalmazó döntési dilemmát, mert tudjuk, hogy van olyan mechanizmus, amely anélkül közvetíti a modellben szereplő döntésbefolyásoló tényezőkre vonatkozó releváns információkat, hogy a döntéshozónak bonyolult kalkulációkat kellene elvégeznie. Az külön kutatási téma lehet, hogy milyen módon „csaphatóak be” ezek az érzelmek, és ennek következményeként mikor vezetnek valóban irracionális döntéshez.

### VII.3.2. ÉRZELMEK, ÉRDEKEK ÉS NORMATÍV ERŐ

Amennyiben elfogadjuk az érzelmek szerepére vonatkozó fenti értelmezést – és én azt javaslom, hogy fogadjuk el! –, akkor új meg-

világításba kerül a deskriptív és előíró társadalmi normák közötti megkülönböztetés is (Bicchieri 2006). Ugyanis ebben a keretben a normativitás értelmezhető az érzelmi reakciók szempontjából is. A higgadt fejjel hozott kalkulatív döntések arról szólnak, hogy mennyire tűnik hasznosnak egy-egy lépés. Az érzelmekkel kapcsolatba hozható döntések szólnak inkább arról, hogy mit kell, vagy mit tilos megtennie az embernek. Azonban ebben a keretben az érzelmek mögött ott vannak a döntés következményeinek hosszú távú perspektívái is.

A fentiek alapján az előíró normativitás értelmezhető folytonos változóként, amelynek mértékét a normasértés esetén jelentkező érzelmek intenzitása jelzi. Az érzelmek intenzitását ugyanakkor a szituáció racionális döntésemélet kereteiben értelmezhető jellemzői határozzák meg. Négy lehetséges befolyásoló tényezőre hívom fel a figyelmet. Feltételezéseim szerint a normasértésre adott reakciók intenzitását fokozza: *a*) a másik normasértő magatartásának szokatlansága okozta megdöbbenés; *b*) a másik normasértő magatartásának megfigyelőre gyakorolt kellemetlen hatása; *c*) az adott norma betartásának jelentősége (jelzésértéke) a különböző csoporttagságok (identitások) szempontjából; végül *d*) az adott csoportok (identitások) fontossága a megfigyelő számára.

Emlékezzünk kísérletező kedvű tanszéki kollégám balkezes kézfogás kísérletére (VI. fejezet)! A kollégám konklúziója az volt, hogy a kéznyújtás előíró normájával szemben a jobbkezes kéznyújtáshoz nem tartozik külön előírás, az csak egy normatív erő nélküli konvenció. Tegyük még hozzá e két viselkedéshez a társadalmi normák irodalmának egyik kedvenc példáját, a sorbanállás szabályainak betartását egy pénztár előtt (Hindriks és Guala 2015, Smith 2015, Young 2007). A sorbanállás szabályaihoz erős normatív erőt társítanak azok, akik ezt példának hozzák. A balkezes kéznyújtás a diplomaosztón kétségtelenül rendkívül szokatlan (*a*), ellenben legfeljebb epszilon mértékű kellemetlenséget okoz az elszenvedőnek (*b*), miközben nincsenek olyan identitások, amelyekhez a bal- vs. jobbkezes kéznyújtás jelezne az odatartozást, ráadásul az egyetemről épp örömmel kirepülő végzősöket valószínűleg kevésbé érdekli, hogy milyen normákkal (és

milyen normasértő, kelekótya professzorokkal) hagyják maguk mögött az egyetemet. Az adott szituációban a kézfogás teljes elmaradása sem okozott volna feltétlenül hatalmas megrökönyödést a diplomázók körében, azonban a tanszékvezető kollégák részéről már annál inkább. A balkezes kézfogással is leginkább a vele egy csoportba tartozó kollégák rosszallását kockáztatta kísérletező kedvű kollégám – a *d*) pont miatt. A kézfogás elmaradása sem jár kézzelfogható (elnevezést) hátránnyal. Azonban az olvasók számára is nyilvánvaló, hogy annak *jelentése* van.

Modellezési keretünkben ezt a jelentést azzal ragadhatjuk meg, hogy nem mindenhol és nem mindenkinek jár kéznyújtás. Csoport-hovatartozást jelölhetünk azzal, ha nőknek kezet nyújtunk, hiszen ma is sok ember identifikálja magát olyan csoportokkal, ahol nőknek se nők, se férfiak nem nyújtanak kezet. A nemi diszkrimináción túlmenően is szelektáljuk az embereket abból a szempontból, hogy nyújtunk-e nekik kezet. Az együttműködésre méltó csoporttagoknak adjuk meg ezt a gesztust. Ha valakitől megtagadják a kézfogást, akkor annak *jelentése* miatt származhat mégis hátránya az illetőnek (*b*). A sorbanállási szabály megsértését azért érdemes megemlíteni, mert első látásra inkább a balkezes kézfogáshoz hasonlít, és nem a nőknek felajánlott jobbkezes kézfogáshoz. Kifejezetten tolakodást promotáló kultúrák nincsenek, legfeljebb olyanok, amelyek nem tudják elég erős normává tenni a sorbanállást. Abban a társadalmi kontextusban, amelyben e könyv olvasói élnek, ritka a nyilvánvaló tolakodás (*a*). Ezt az olvasó nem így érzi, mert elfeledkezik az egy tolakodóra jutó becsületes sorbanállók arányáról; főleg azt nem számítja bele, hogy a ritka tolakodók is csak különleges eseményekkor jelennek meg, miközben sokszor hónapok vagy évek telnek el, miközben megannyi sort kiállunk az élelmiszerboltokban és máshol anélkül, hogy találkozoznánk ilyen kísérlettel. A sorbanállási rend megsértése viszont azonnali és kézzelfogható hátrányt okoz a megfigyelők (egy részének) (*b*). Ráadásul akár mikro- (a sorbanállók kis ideiglenes sorsközössége), akár makroszinten értelmezzük a sorbanállók csoportját, a csoporttagság fontos a sorbanállóknak (*d*); mikroszinten pedig a sorbanállási szabály betartása elsődleges fontosságú, hiszen ez egy meg-

lehetősen szembetűnő közjószág (c) (vö. Bicchieri 2006). A példát felhozó kutatókhoz hasonlóan én is úgy sejttem, hogy erős normatív tartalommal bír ez a szabály, és felháborodást vált ki a megsértése.

Ebben a modellezési keretben is döntéskorlátozó tényezők az érzelmek. Azonban nem a minden szempontból racionális döntések korlátozói, hanem a rövid távon hasznosnak tűnő opciók választását korlátozzák hosszú távú érdekeink figyelembevételével. És az érzelmek ilyen típusú működésébe belefér a Hirshleifer (1987) által bemutatott fenyegetési funkció is.

Az általam bemutatott kettős játszma modell, kiegészítve az érzelmek kiértékelés elméletének eredményeivel, egy olyan értelmezési keretet ad az erkölcsös viselkedésnek, amelyben egyrészt megkülönböztethetőek a normatív nyomással elért egyensúlyok a morális jelentőséggel nem bíró konvencióktól; ugyanakkor nincsen szükség semmilyen belső erkölcsi motiváció feltételezésére az erkölcsi rend fenntartásához. A modell nem öncélú gondolati konstrukció. A racionális döntésemélet szituációs logikájának fenntartása biztosíthatja, hogy a viselkedési tendenciák különbségeinek okait ne az aggregált szinten amúgy is eltűnő pszichológiai jellemzők különbségeiben keressük, hanem a szituációkban rejlő ösztönzők eltéréseit figyeljük.

## VIII. Epilógus: a fiamról, a norvég íróstárról, és a tudománytalan tudományok fontosságáról

*„Milyen érdekes, gondoltam, a nép életében nincsenek átmenetek. A nép az egyik pillanatban még némán eszi a szart, kibontja a csillogó sztaniolpapírból, és nyeli békésen, egyik kakát a másik után, de a másik pillanatban már belezőkéssel őzjög.”*

Cserna Szabó András (2018). *Az abbé a fejével játszik*. Budapest: Helikon 60. oldal (Free jazz)

Nem siettünk, de szokásunk szerint szaporán szedtük rövid lábainkat Zsomborral. A szentendrei Anyóskonyhára igyekeztünk menzapótló rántothús-ebédre. Valamelyik kedvenc videobloggeréről faggattam. Büszke vagyok rá. Angol nyelvű blogok, és tizenkét évesen bármilyen témáról komoly összefoglalót tud adni. Múltkor a szerzői jog kérdéseiben próbáltam okoskodni, de leiskolázott. Egy másik alkalommal a méhekkal kapcsolatos sztereotípiák ügyében is. Mondjuk az aznapi konkrét témára nem emlékszem. Csak a válaszomra: frappánsan kifejtettem, hogy miközben ilyen hülyeségekkel foglalkozunk, addig csendben élhetetlenné tesszük a bolygót. Szokása szerint pontosította trehányul odavetett mondatomat. Nem biztos, hogy ebben a formában szó szerint igaz az, amit mondtam. Lassan felfogtam. Ami számomra negyvenhat évesen a Facebookon lógó hobbipolitizálók valamely aktuális frázisának leakasztása, az számára húsba vágó kérdés.

Hobbipolitizáló vagyok, követem, ahogy örültekházává válik a nyugati demokráciák közéleti színpada. Nem létező bevándorlókat próbálunk kizárni, miközben leállnak a gyárok, mert sajnos tényleg nem léteznek. A politikai fórumokról a csúnya vélemények miatt befogott füffel kivonuló social justice warriorok legalább messze vannak, lehet rajtuk nevetni. Ugyanakkor közel áll hozzám az a jófej kollégám, aki büszkén mutatja a lányának vett könyvet a mai progresszív kislányok példaképeiről, köztük a világ leghíresebb, büntetlen szexuális bűnözőjének (Bill Clinton) legközelebbi cinkosáról (Cin-



tonné). Ad hoc példák voltak, nem folytatom. Csigó Péter (2016) megírta, mi történt, e könyvben is utaltunk e folyamatokra. Fontos: néhány generációval ezelőtt sokkal nagyobb, igazi, vérre menő örüléthullámok söpörtek végig a Földön. Mégis lehet, hogy ebbe az óvodás hülyéskedésbe, amit ma közéletnek hívunk, végül többen fogunk belehalni.

Ez a könyv a racionális elemzés fontosságáról szól. De nem csak a tudományos problémák racionális elemzéséről. Az általam ismertett racionális modellek konklúziója, hogy életbevágóan fontos a racionális közéleti diskurzus. A húsbavágó materiális érdekekre hivatkozó racionális diskurzus. Magyarul észszerű párbeszéd. Nem vagyunk arra programozva, hogy egyedül megoldjuk a társas élet révén közössé váló problémáinkat. De képességet kaptunk arra, hogy közösen megbeszéljük, és ennek nyomán közösen válaszokat adjunk rájuk (Henrich 2017). Nincs garancia arra, hogy jó válaszokat adunk. Közönnyel, vagy épp a legnagyobb morális elszántsággal gyűjthetjük magunkra a bolygót. De a lehetőség happy endek végtelen sorára is megvan. Racionális párbeszéddel több esély van a happy endre. És a jó és rossz változás is a pillanat műve lehet. Tétje van annak, miről és hogyan beszélgetünk. Igyekeztem elmagyarázni, hogy miért van így.

Látni kell, hogy az egyéni érdekekből kiinduló, húsbavágó kérdésekről folyó észszerű párbeszéd szükségszerűen a közös érdekekről szól. És csak a húsbavágó kérdésekről folyó észszerű párbeszéd szól a közös érdekekről. És ehhez tartozik, hogy az egyéni érdekek értelmezhetetlenek mások érdekei nélkül. Zavarosnak hangzik, de remélem, a könyvben megmutattam, hogy logikus, hogy így van.

A legtöbb társadalomtudós azonban ezt másképp gondolja. Közgazdász, pszichológus, politikatudós és (ma már) kulturális antropológus kutatók tízezrei hisznek abban, hogy a közös érdekek a fejünkben vannak. És az egyéni érdekeink is értelmezhetőek saját magunkban. Az egész világ értelmezhető, ha megfejtjük, mi van a fejünkben. Ezek a tézisek újabb és újabb tudományos szupersztárokat termelnek ki, e könyvben is bemutatattam néhányat közülük. Nem tudnak hipotézisekben és méréseikben akkorát hibázni, hogy elméleteik ne maradjanak tudóstízezrek számára vonzóak. Már a politikatudósok is drága

mérőműszerekkel hatolnak az idegrendszer mélyére, hogy a politikai rendszer működését megfejtsék.

Az egyénben rejlő titkok szakadatlan keresése nem a kortárs tudósok sajátos szenvedélye. Ez a modernitás korszelleme. Műveletlenségem megakadályozta, hogy e könyvben belevágjak a korszellemet jellemző individualista tézisek szellemtörténetének mégoly rövid áttekintésébe. De egy közepesen műveletlen ember is átlátja, ahogy például a képzőművészet a 19. század végétől elfordult a külső világtól, és lelkünk megfestésének látott neki (figyelem, ismerős: racionális konstrukciók és irracionális impressziók irányzatai a lélekbúvár képzőművészek között) (Rothko 2006).\* Egy közepesen művelt ember pedig azt is átlátja, hogy az irodalom miként fordult a bizonytalan külső világ megismerésétől a bizonytalan belső világ megismerése felé.

A képzőművészek modern egotripjét némi fenntartásokkal fogadja a nagyközönség, ám az irodalomban mintha más lenne a helyzet. Például Karl Ove Knausgard, a lehető legszélsőségesebb módon egoista regényfolyam norvég alkotója ma talán a nyugati világ legnépszerűbb kortárs szépírója.

Knausgard a legújabb hírek szerint második felesége összeomlása után teljes kudarcnak ítélte piacilag szupersikeres hiperegoista projektjét (Franklin 2018). Némi problémát okozott, hogy Knausgard ország-világ elé tárta kapcsolatuk egy olyan olvasatát, amelyben lelkileg teljesen leválva tekinti magát közös családtagjairól és közös projektjeikről.

A társadalomtudományokban nem érzem a hiperegoista projektek összeomlásának előszelét. Bár általam okosnak érzett emberek harcát látom, és azt is, hogy vannak, akik hallgatnak rájuk. A könyvben megpróbáltam bemutatni a legújabb eredményeiket.

Létezik ugyanakkor a társadalomtudósok között egy nagy csapat, amelynek legtöbb tagja a lélekbúvárkodás évszázadában is az emberek közötti társadalmi térben kereste a társadalmi jelenségek rejtélyeinek megfejtéseit. Sejtésem szerint a szociológusok azért voltak

\* Köszönet Rajkó Andreának a vonatkozó irodalom bemutatásáért.

képesek kitartani az eredeti megközelítésmódjuk mellett, mert egyre kevésbé figyeltek a tudományos világ véleményére. Az önszelekció (figyelem, racionális döntéselméleti magyarázat!) révén erősödött egy olyan szociológus kultúra, amely nem kapaszkodott a társadalomtudományos valóságkeresés elméleti és módszertani trendjeibe, miközben rendületlenül hitt a saját igazságkereső programjában. És ez a tudományos közeg ténylegesen egyre kevesebb látványosat tud mondani a valóságról. De – ez egyik fő következtetésem –, ezzel együtt az alapkérdésekben igazuk van. Csak nem igazán akarja meghallgatni ma már őket senki (Szelényi 2015). Belülről látom ezt a közeget, és megvallom őszintén, néha nehezemre esik felháborodni e mellőzöttségen. De nem mindig!

A Bertalan László által alapított hazai szociológiai iskola tagja vagyok. Az iskola fogyatkozó tagjai vallják, hogy a játékelméleti szemléletmód a szociológiai megismerés megfelelő eszköze. Nagyjából a világ összes szociológusa meglehetősen irtózik a játékelmélettől. Azt hiszem, úgy látják, hogy annak az igazán tudományos társadalomtudományi programnak egy prominens irányzata, amelynek igazságatlan valóságfeltárásától rettegnek a vallásos igazságkeresésbe zárkózó szociológusok (bocsánat az általánosító és leegyszerűsítő megfogalmazásért). Nem látják, hogy mennyire tudománytalan tudomány a játékelmélet, amelynek egyes alapkutatói modelljei gyakran legalább annyira a társadalmi mítoszok által gyártott, megérthető igazságot keresik a természeti valóság helyett, mint a szociológia tudományos tudománytól leginkább eltávolodott alágai (Roell é.n.). Próbáltam erre is utalni a könyvben. De tudom, hogy nem könnyű bárkit meggyőzni, hiszen a sok alkalmazott matematika és a Nobel-díj (természet) tudományos álruhája ezt jól elfedi.

Könyvem e két tudománytalan tudományterület szövetségéért született (vö. Vaisey és Valentino 2018). Hiszek abban, hogy a 21. században szárba szökkenő új típusú örültségek, és a valódi problémákkal kapcsolatos hagyományos tétlenség ellen tehetnek a társadalom természetét értő, a magatartástudomány és adattudomány korában tudománytalanoknak tartott diszciplínák. Főleg, ha összefognak.

### Az epilógus epilógusa

A radarjelző villogás jelenségét (I.1.1. alfejezet) a IV-től a VI. fejezetig tárgyalt elméletek (tisztességes, ösztönös, illetve normavezérelt polgárok elméletei) könnyedén magyarázni tudják: a bennünk élő erkölcsi motivációkra hivatkoznak. Talán a rejtélyes jelenségekre adott végtelen egyszerű válaszaik (azért csinálják az emberek ezt, mert úgy születtek, hogy ilyet kell csinálniuk) miatt is lehetnek népszerűek ezek az elméletek. Az önző polgár modelljeinek (III. és VII. fejezetek) kognitív korlátokra kell hivatkozniuk a jelenség magyarázatában (például az autósok valamit félreértenek: azt hiszik, hogy szembejövők felismerhetik őket; vagy csak nem gondolnak bele, hogy ez pont nem olyan helyzet, ahol felismerhetőek, és szellemi restségből alkalmazzák a szemtől szembeni interakciókban hasznos stratégiát). E korlátok működésének tesztelése a konkrét szituációban nem is könnyű. De talán nem is lehetetlen. Az ilyen tesztek elemzése nem sokkal Neumann és Morgenstein (1944) könyvének megjelenése után megkezdődött (Allais 1953). A korlátozott racionalitás kurrens elméletei (Gabaix 2014) talán már megfelelő alapot is adnak a radarjelző villogás titkának elemzéséhez szükséges komplexebb kísérletekhez is. A világ bonyolult (bocsánat a közhelyes kijelentésért). A tudósoknak nem szabad megijedni a komplexitástól.

## Hivatkozások

- Aarøe, L., & Petersen, M. B. (2014). Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues. *The Journal of Politics*, 76(3), 684–697.
- Aarøe, L., Petersen, M. B., & Arceneaux, K. (2017). The Behavioral Immune System Shapes Political Intuitions: Why and How Individual Differences in Disgust Sensitivity Underlie Opposition to Immigration. *American Political Science Review*, 111(2), 277–294.
- Aaron, Henry J. (1966). „The Social Insurance Paradox” *Canadian Journal of Economics and Political Science* 32: 371–374.
- Abend, G. (2011). Thick concepts and the moral brain. *European Journal of Sociology/Archives Européennes de Sociologie*, 52(1), 143–172.
- Adams, T. G., Stewart, P. A., & Blanchar, J. C. (2014). Disgust and the politics of sex: Exposure to a disgusting odorant increases politically conservative views on sex and decreases support for gay marriage. *PloS one*, 9(5), e95572.
- Ahn, W. Y. et al. Montague, P. R. (2014). Nonpolitical images evoke neural predictors of political ideology. *Current Biology*, 24(22), 2693–2699.
- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, 97(4), 543–569.
- Albrekt Laresen, C. & Dejgaard, T. E. (2013). The institutional logic of images of the poor and welfare recipients: A comparative study of British, Swedish and Danish newspapers. *Journal of European Social Policy*, 23(3), 287–299.
- Albrekt Laresen, C. (2008). The institutional logic of welfare attitudes: How welfare regimes influence public support. *Comparative political studies*, 41(2), 145–168.
- Albrekt Laresen, C. (2013). *The rise and fall of social cohesion: The construction and de-construction of social trust in the US, UK, Sweden and Denmark*. Oxford University Press.

- Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *Journal of political economy*, 58(3), 211–221.
- Alesina, A., & La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of public Economics*, 89(5-6), 897–931.
- Alesina, A., Glaser, E., & Sacerdoce, B. (2001). Why doesn't the USA have a European type Welfare state. *Harvard Institute for Economic Research, Cambridge/Mass.*
- Alesins, A. & Glaser, E. (2004). *Fighting poverty in the US and Europe*. Oxford: Oxford University Press.
- Alford, R. R. (1962). A suggested index of the association of social class and voting. *Public Opinion Quarterly*, 417–425.
- Alford, R. R. (1963). The role of social class in American voting behavior. *Western Political Quarterly*, 16(1), 180–194.
- Almås, I., Cappelen, A. W., & Tungodden, B. (2016). Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?
- Allais, M. (1953). L'extension des théories de l'équilibre économique général et du rendement social au cas du risque. *Econometrica*, 21(2): 269–290.
- Alvard, M. S. (2004). The ultimatum game, fairness, and cooperation among big game hunters. In: Henrich, J. P., Boyd, R., Bowles, S., Fehr, E., Camerer, C., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press. pp. 413–435.
- Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7), 3427–39.
- Anderson, C. (2008). Wired. The end of theory: the data deluge makes the scientific method obsolete. Retrieved from <http://www.wired.com/2008/06/pb-theory/>; letöltve: 2018.10.15.
- Andreoni, J. (1988). Privately provided public goods in a large economy: the limits of altruism. *Journal of Public Economics*, 35(1), 57–73.
- Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 110(1), 1–21.
- Arceneaux, K. (2017). Anxiety Reduces Empathy Toward Outgroup Members But Not Ingroup Members. *Journal of Experimental Political Science*, 4(1), 68–80.
- Aron, R. (1950). Social structure and the ruling class: part 1. *The British Journal of Sociology*, 1(1), 1–16.

- Arrow, K. (1973). The theory of discrimination. *Discrimination in labor markets*, 3(10), 3–33.
- Arrow, K. J. (1998). What has economics to say about racial discrimination? *Journal of Economic Perspectives*, 12(2), 91–100.
- Auerbach, Alan J. and Kotlikoff, Laurence J. (1987). *Dynamic Fiscal Policy*. Cambridge: Cambridge University Press.
- Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1), 67–96.
- Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics* 4(6): 1236–1239.
- Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1), 6–19.
- Aumann, R. J. (1998). On the centipede game. *Games and Economic Behavior*, 23(1), 97–105.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396.
- Baldwin, P. (1990). *The politics of social solidarity: class bases of the European welfare state, 1875–1975*. Cambridge University Press.
- Banks, A. J. (2014). The public's anger: White racial attitudes and opinions toward health care reform. *Political Behavior*, 36(3), 493–514.
- Banks, A. J. (2016). Are group cues necessary? How anger makes ethnocentrism among whites a stronger predictor of racial and immigration policy opinions. *Political Behavior*, 38(3), 635–657.
- Banks, A. J., & Valentino, N. A. (2012). Emotional substrates of white racial attitudes. *American Journal of Political Science*, 56(2), 286–297.
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Evidence on questionable research practices: The good, the bad, and the ugly.
- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2), 122–133.
- Bargheer, S., & Wilson, N. H. (2018). On the Historical Sociology of Morality: Introduction. *European Journal of Sociology/Archives Européennes de Sociologie*, 59(1), 1–12.
- Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion*, 23(7), 1284–1306.
- Becker, G. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.

- Becker, G. S. (1957). *The economics of discrimination*. University of Chicago.
- Becker, G. S. (1974). A theory of marriage: Part II. *Journal of political Economy*, 82(2, Part 2), S11-S26.
- Becker, Gary S. (1983). „A Theory of Competition among Pressure Groups for Political Influence”, *The Quarterly Journal of Economics* 98: 371–400.
- Becker, Gary S. and Mulligan, Casey B. (1998). „Deadweight Costs and the Size of the Government” *NBER Working Paper* 6789.
- Bereczkei, T. (2009). *Az érény természete. Önzetlenség, együttműködés, nagylelkűség*. Budapest: Typotex.
- Berelson, B. R., Lazarsfeld, P. F., McPhee, W. N., & McPhee, W. N. (1954). *Voting: A study of opinion formation in a presidential campaign*. Chicago: University of Chicago Press.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. New York.
- Bernau, P. (2016). Ernst Fehr an der Spitze der Ökonomen. *Frankfurter Allgemeine Zeitung*, 2016 szeptember 2.
- Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105), 912.
- Bicchieri, C. (1990). Norms of cooperation. *Ethics*, 100(4), 838–861.
- Bicchieri, C. (1997). *Rationality and coordination*. Cambridge: Cambridge University Press.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.
- Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.
- Bicchieri, C., & Sontuoso, A. (2015). I cannot cheat on you after we talk. In Peterson, M (ed.) *The prisoner’s dilemma*. Cambridge: Cambridge University Press, pp. 101–114.
- Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191–208.
- Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). Social norms. forthcoming in. *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/social-norms/>. Accessed, 11.
- Binmore, K. (1991). *Fun and games. A text on game theory*. London: D.C. Heath.
- Binmore, K. (1997). Rationality and backward induction. *Journal of Economic Methodology*, 4(1), 23–41.
- Binmore, K. (2005). *Natural justice*. Oxford University Press.



- Binmore, K. (2015). Institutions, rules and equilibria: a commentary. *Journal of Institutional Economics*, 11(3), 493–496.
- Binmore, K. G. (1998). *Game theory and the social contract*. (Vol 1–2.) Cambridge, MA: MIT press.
- Binmore, K., & Shaked, A. (2010). Experimental economics: Where next? *Journal of Economic Behavior & Organization*, 73(1), 87–100.
- Bisgaard, M., & Slothuus, R. (2018). Partisan elites as culprits? How party cues shape partisan perceptual gaps. *American Journal of Political Science*, 62(2), 456–469.
- Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.
- Bloch, M. (1986). *From blessing to violence: history and ideology in the circumcision ritual of the Merina* Cambridge: Cambridge University Press.
- Boadway, Robin W. and Wildasin David E. (1989). „A Median Voter Model of Social Security” *International Economic Review* 30(2): 307–328.
- Boeri, Tito (2004). „Silence is Golden? Assessing the Public Debate on Pension Reforms in Europe” *Presentation at CEPS 14 September 2004*.
- Boeri, Tito, Börsch-Supan, Axel and Tabellini, Guido (2002a). Would you like to reform the Pension System? The opinion of European Citizens Manuscript University of Mannheim [http://www.mea.uni-Mannheim.de/mea\\_neu/pages/files/nopage\\_pubs/dp07.pdf](http://www.mea.uni-Mannheim.de/mea_neu/pages/files/nopage_pubs/dp07.pdf).
- Boeri, Tito, Börsch-Supan, Axel and Tabellini, Guido (2002b). „Pension Reforms and the Opinions of European Citizens” *American Economic Review* vol 92(2): 396–401.
- Boldrin, Michele and Montes, Ana (1998). *Intergenerational Transfer Institutions: Public Education and Public Pensions*. Mimeo Universidad Carlos III., Madrid.
- Boldrin, Michele and Rustichini, Aldo (2000). „Political Equilibria with Social Security” *Review of Economic Dynamics* 3: 41–78.
- Bolsen, T. – Druckman, J. N. – Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36(2), 235–262.
- Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. *The American Economic Review*, 1096–1136.
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166–193.
- Borg, J. S., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of cognitive neuroscience*, 20(9), 1529–1546.

- Bosker, B. (2018). The nastiest feud in science. *The Atlantic*, September 2018 issue.
- Boudreau, C. – MacKenzie, S. A. (2014). Informing the electorate? How party cues and policy information affect public opinion about initiatives. *American Journal of Political Science*, 58(1), 48–62.
- Bourdieu, P. (1991 [1982]). *Language and symbolic power*. Cambridge, MA: Harvard University Press. [Részben megegyezik: *Ce que Parler Veut Dire: L'Économie des Échanges Linguistiques*. Paris: Librairie Arthème Fayard.]
- Bower, G. H., & Forgas, J. P. (2001). Mood and social memory. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 95–120). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bowles, Samuel and Gintis, Herbert (2000). „Reciprocity, Self Interest, and the Welfare State” *Nordic Journal of Political Economy* 26: 33–53
- Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. Chicago: University of Chicago press.
- Brams, S. J. (2018). *Divine Games: Game Theory and the Undecidability of a Superior Being*. MIT Press.
- Breyer, Friedrich (1994). „The Political Economy of Intergenerational Redistribution” *European Journal of Political Economy* 10: 61–84.
- Breyer, Friedrich and Craig, Ben (1997). „Voting on social Security: Evidence from OECD Countries” *European Journal of Political Economy* 13: 705–724.
- Bridges, Benjamin (1978). „Why The Social Insurance Budget Is Too Large in a Democracy: Comment” *Economic Inquiry* 16(1): 133–138.
- Browning, Edgar K. (1975). „Why The Social Insurance Budget Is Too Large in a Democracy” *Economic Inquiry* 13(3): 373–388.
- Browning, Edgar K. (1978). „Why The Social Insurance Budget Is Too Large in a Democracy: A Reply” *Economic Inquiry* 16(1): 139–142.
- Buchanan, J. (1992 [1969]). A választás tudománya-e a közgazdaságtan? In. *Piac, állam, alkotmányosság*. Budapest: KJK. [Is economics the science of choice? In. *Roads to freedom: Essays in honour of Friedrich A. von Hayek*, 47–64.]
- Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review*, 105(3), 496–515.
- Burton-Chellew, M. N., & West, S. A. (2013). Prosocial preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences*, 110(1), 216–221.
- Burton-Chellew, M. N., El Mouden, C., & West, S. A. (2016). Conditional cooperation and confusion in public-goods experiments. *Proceedings of the National Academy of Sciences*, 113(5), 1291–1296.

- Burton-Chellew, M. N., El Mouden, C., & West, S. A. (2017). Social learning and the demise of costly cooperation in humans. *Proc. R. Soc. B*, 284(1853), 20170067.
- Busby, E., Flynn, D. J., Druckman, J. N., & Hall, S. (2018). Studying Framing Effects: Existing Research and Lingering Questions. In: Paul D'Angelo (ed.). *Doing News Framing Analysis II* New York: Routledge.
- Butler, Monika (2000). „The Political Feasibility of Pension Reform Options: The Case of Switzerland” *Journal of Public Economics* 75: 389–416.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.
- Camerer, C. F., & Fehr, E. (2006). When does” economic man” dominate social behavior? *Science*, 311(5757), 47–52.
- Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, 37(1), 47–59.
- Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*, 19(4), 371–394.
- Camic, C. (1979). The utilitarians revisited. *American Journal of Sociology*, 85(3), 516–550.
- Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3), 818–827.
- Capraro, V., & Rand, D. G. (2018). Do the Right Thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*, Vol. 13, No. 1, January 2018, pp. 99–111.
- Carsey, T. M. – Layman, G. C. (2006). Changing sides or changing minds? Party identification and policy preferences in the American electorate. *American Journal of Political Science*, 50(2), 464–477.
- Cass, D., & Shell, K. (1983). Do sunspots matter? *Journal of political economy*, 91(2), 193–227.
- Chamley, C. (2004). *Rational herds: Economic models of social learning*. Cambridge University Press.
- Chan, C., Van Boven, L., Andrade, E. B., & Ariely, D. (2014). Moral violations reduce oral consumption. *Journal of consumer psychology: the official journal of the Society for Consumer Psychology*, 24(3), 381.
- Chapman, H. A., & Anderson, A. K. (2014). Trait physical disgust is related to moral judgments outside of the purity domain. *Emotion*, 14(2), 341.

- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, 323(5918), 1222–1226.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817–869.
- Chwe, M. S. Y. (2001). *Rational ritual. Culture, coordination, and common knowledge*. Princeton: Princeton University Press.
- Ciuk, D. J. (2018). Assessing the contextual stability of moral foundations: Evidence from a survey experiment. *Research & Politics*, 5(2), 1–9. 2053168018781748.
- Ciuk, D. J. – Yost, B. A. (2016). The effects of issue salience, elite influence, and policy content on public opinion. *Political Communication*, 33(2), 328–345.
- Clifford, S., & Jerit, J. (2018). Disgust, Anxiety, and Political Learning in the Face of Threat. *American Journal of Political Science* forthcoming.
- Clifford, S., & Piston, S. (2017). Explaining public support for counterproductive homelessness policy: The role of disgust. *Political Behavior*, 39(2), 503–525.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Conde-Ruiz, J. Ignacio and Galasso, Vincenzo (1999). „Positive Arithmetic of the Welfare State” *CEPR Discussion Paper Series 2202*.
- Conde-Ruiz, J. Ignacio and Profeta, Paola (2004). *The Redistributive Design of Social Security Systems*. Manuscript, Università Bocconi.
- Conesa, Juan C. and Kruger, Dirk (1999). „Social Security Reform with Heterogeneous Agents” *Review of Economic Dynamics* 2: 757–795.
- Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics*, 3(1), 55–79.
- Cooley, Thomas F. and Soares, Jorge (1999a). „A Positive Theory of Social Security Based on Reputation” *Journal of Political Economy* 107(1): 135–160.
- Cooley, Thomas F. and Soares, Jorge (1999b). „Privatizing Social Security” *Review of Economic Dynamics* 2: 731–755.
- Cox, Donald (1987). „Motives for Private Income Transfers”. *Journal of Political Economy* 95: 508–546.
- Cukierman, Alex and Meltzer, Allan H. (1989). „A Political Theory of Government Debt and Deficits in a Neo-Ricardian Framework” *The American Economic Review* 79(4): 713–732.
- Csontos, L. (1986). Bevezetés. in. Durkheim, É. *A társadalmi munkamegosztásról – Részletek* (ford. Csákó Mihály), Budapest: MTA Szociológiai Kutató Intézet.

- Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes*, 100(2), 193–201.
- Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67–80.
- De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254–1258.
- DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Caidric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science*, 15(5), 319–324.
- Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social issues*, 31(3), 137–149.
- Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of political economy*, 91(3), 401–419.
- DiMaggio, P., & Garip, F. (2012). Network effects and social inequality. *Annual review of sociology*, 38, 93–118.
- Downs, A. (1990 [1957]). Politikai cselekvés a demokráciában: egy racionális modell. *Közgazdasági Szemle*, 38(9), 993–1011. [An economic theory of political action in a democracy. *Journal of political economy*, 65(2), 135–150.]
- Druckman, J. N. – Petersen, E. – Slothuus, R. (2013). How Elite Partisan Polarization Affects Public Opinion Formation. *American Political Science Review*, 107(1), 57–79.
- Duckitt, J., & Sibley, C. G. (2010). Right-Wing Authoritarianism and Social Dominance Orientation differentially moderate intergroup effects on prejudice. *European Journal of Personality*, 24(7), 583–601.
- Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, 47(2), 268–298.
- Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personality and Individual differences*, 47(6), 541–546.
- Durkheim, É. (2001 [1893]). *A társadalmi munkamegosztásról*. Budapest: Osiris.
- Durkheim, É. (2003 [1897]). *Az öngyilkosság*. Budapest: Osiris.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology*, 6, 621.

- Eckel, C., & Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior & Organization*, 73(1), 109–119.
- Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1), 117–130.
- Ellison, G. (1994). Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies*, 61(3), 567–588.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643–669.
- Elster, J. (1995 [1989]). *A társadalom fogaskerekei*. Budapest: Osiris.
- Embrey, M., Fréchette, G. R., & Yuksel, S. (2017). Cooperation in the finitely repeated prisoner's dilemma. *The Quarterly Journal of Economics*, 133(1), 509–551.
- Emler, N., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology*, 45(5), 1073.
- Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review*, 94(4), 857–869.
- England, P. (1989). A feminist critique of rational-choice theories: Implications for sociology. *The American Sociologist*, 20(1), 14–28.
- Evans, G., Heath, A., & Lalljee, M. (1996). Measuring left-right and libertarian-authoritarian values in the British electorate. *British Journal of Sociology*, 93–112.
- Evans, Geoffrey (1993). „Class, Prospects and the Life-Cycle: Explaining the Association between Class Position and Political Preferences.” *Acta Sociologica* 36: 263–276.
- Exadaktylos, F., Espín, A. M., & Brañas-Garza, P. (2013). Experimental subjects are not different. *Scientific reports*, 3, 1213.
- Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, 54(2), 293–315.
- Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73(6), 2017–2030.
- Federico, C. M., Weber, C. R., Ergun, D., & Hunt, C. (2013). Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition. *Political Psychology*, 34(4), 589–610.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785.

- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868.
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079.
- Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The quarterly journal of economics*, 108(2), 437–459.
- Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, 24(1), 56–62.
- Ferge, Z. (1966). Társadalmi rétegződés Magyarországon *Valóság*.
- Ferri, G., Liu, L. G., & Stiglitz, J. E. (1999). The procyclical role of rating agencies: Evidence from the East Asian crisis. *Economic Notes*, 28(3), 335–355.
- Fields, J. M., & Schuman, H. (1976). Public beliefs about the beliefs of the public. *Public Opinion Quarterly*, 40(4), 427–448.
- Flache, A. (1996). *The Double Edge of Networks*. Amsterdam: Thesis Publishers.
- Flache, A., & Macy, M. W. (1996). The weakness of strong ties: Collective action failure in a highly cohesive group. *Journal of Mathematical Sociology*, 21(1–2), 3–28.
- Flood, M. M. (1958). Some experimental games. *Management Science*, 5(1), 5–26.
- Flood, M. M., & Dresher, M. (1952). Research memorandum rm-789-1-pr. *Rand Corporation, Santa-Monica, CA, USA*.
- Foner, Anne (1999). „Age Integration or Age Conflict as Society Ages?” In., White Riley, M – Uhlenberg, P. (szerk.), *Age Integration: Update and Critique*. National Institute of Aging, Bethesda, MD, pp. 23–31.
- Fong, C. M., Bowles, S., & Gintis, H. (2005). 10 Reciprocity and the Welfare State. *Moral sentiments and material interests*, 277.
- Fong, C. M., Bowles, S., & Gintis, H. (2006). Strong reciprocity and the welfare state. *Handbook of the economics of giving, altruism and reciprocity*, 2, 1439–1464.
- Fong, Christina (2001). „Social Preferences, Self-interest, and the Demand for Redistribution” *Journal of Public Economics* 82: 225–246.
- Fong, Christina, Bowles, Samuel and Gintis, Herbert (*forthcoming*) (2006). „Strong Reciprocity and the Welfare State” In. Serge Kolm and Jean Mercier-Ythier (eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*. Amsterdam: Elsevier.

- Fontaine, P. (2006). From philanthropy to “altruism”: incorporating unselfish behavior into economics, 1961–1975. Centre for Philosophy of Natural and Social Science, Discussion Paper Series 81/06.
- Fontaine, P. (2012). Beyond Altruism? Economics and the Minimization of Unselfish Behavior, 1975–93. *History of Political Economy*, 44(2), 195–233.
- Forges, F., & Peck, J. (1995). Correlated equilibrium and sunspot equilibrium. *Economic Theory*, 5(1), 33–50.
- Franklin, M. N. (1985). *The decline of class voting in Britain: changes in the basis of electoral choice, 1964–1983*. Oxford University Press, USA.
- Franklin, R. (2018). How writinig 'My Struggle' undid Knausgard. *The Atlantic Magazine*, November 2018.
- Friedell, M. F. (1968). A laboratory experiment in retaliation. *Journal of Conflict Resolution*, 12(3), 357–373.
- Friedell, M. F. (1969). On the structure of shared awareness. *Behavioral Science*, 14(1), 28–39.
- Friedman, M. (1953). *Essays in positive economics*. Chicago: University of Chicago Press.
- Fudenberg, D., & Maskin, E. (1986). The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3), 533–554.
- Fudenberg, Drew and Tirole, Jean (1991). *Game Theory*. Cambridge, MA: MIT Press.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4), 1661–1710.
- Gächter, S., Gerhards, L., & Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97, 72–86.
- Gächter, S., Nosenzo, D., & Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3), 548–573.
- Gál, Róbert I., Simonovits, András and Tarcali, Géza (2001). „Generational Accounting and Hungarian Pension Reform” *Social Protection Discussion Paper Series 127*.
- Galasso, Vincenzo (1999). „The US Social Security: What does Political Sustainability Imply?” *Review of Economic Dynamics*2: 698–730.
- Galasso, Vincenzo (2002). „The US Social Security: A Financial Appraisal for the Median Voter” *Social Security Bulletin* 64(2): 57–65.



- Galasso, Vincenzo and Profeta, Paola (2000). „Politico-Economic Models of Social Security” *Econpublica Working Papers* 69.
- Galasso, Vincenzo and Profeta, Paola (2002). „The Political Economy of Social Security: A Survey” *European Journal of Political Economy* 18: 1–29.
- Galasso, Vincenzo and Profeta, Paola (2004). „Lessons for an Ageing Society: The Political Sustainability of Social Security Systems” *Economic Policy, April*: 63–115.
- Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 9–15.
- Gelissen, John (2001). „Old-age Pensions: Individual or Collective Responsibility? An Investigation of Public Opinion across European Welfare States.” *European Societies* 3: 495–523.
- Gintis, H. (2000). *Strong reciprocity and human sociality* (No. 2000–02). Working Paper, University of Massachusetts, Department of Economics.
- Gintis, H. (2016). *Individuality and entanglement: the moral and material bases of social life*. Princeton University Press.
- Gintis, H., & Helbing, D. (2015). Homo socialis: An analytical core for sociological theory. *Review of Behavioral Economics*, 2(1–2), 1–59.
- Gintis, H., Bowles, S., Boyd, R. T., & Fehr, E. (Eds.). (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. Cambridge, MA: MIT Press.
- Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, 29(1), 1–17.
- Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5), 1029.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of personality and social psychology*, 101(2), 366.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology*, 91(3), 481–510.
- Granovetter, M. (2017). *Society and economy: Framework and principles*. Cambridge, MA: Harvard University Press.
- Green, D., & Shapiro, I. (1996). *Pathologies of rational choice theory: A critique of applications in political science*. Yale University Press.

- Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy*, 102(5), 912–950.
- Guala, F. (2016). *Understanding institutions: The science and philosophy of living together*. Princeton: Princeton University Press.
- Gulyás, A. (2007). A méltányosságelmélet alapjai: Modellek és nézőpontok. *Közgazdasági Szemle*, 54(2), 167–183.
- Güth, W., & Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization*, 108, 396–409.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Haidt, J. (2008). Morality. *Perspectives on psychological science*, 3(1), 65–72.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon Books.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31(1), 191–221.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of personality and social psychology*, 65(4), 613.
- Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, 16(5), 701–713.
- Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies*, 9(1), 107–131.
- Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, 26(3), 245–256.
- Hamil-Luker, Jenifer (2001). „The Prospects of Age War: Inequality between (and within) Age Groups.” *Social Science Research* 30: 386–400.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of theoretical biology*, 7(1), 17–52.
- Han, J. – Federico, C. M. (2017). Conflict-Framed News, Self-Categorization, and Partisan Polarization. *Mass Communication and Society*, 1–26.

- Han, J. – Wackman, D. B. (2017). Partisan Self-Stereotyping: Testing the Salience Hypothesis in a Prediction of Political Polarization. *International Journal of Communication*, 11, 603–625.
- Hansson, Ingemar and Stuart, Charles (1989). „Social Security as Trade Among Living Generations” *The American Economic Review* 79(5): 1182–1195.
- Hardin, G. (1968). The tragedy of the commons. *science*, 162(3859), 1243–1248.
- Harsanyi, J. C. (1962). Mathematical models for the genesis of war. *World Politics*, 14(4), 687–699.
- Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.
- Hatemi, P., & McDermott, R. (2012). Policing the perimeter: Disgust and purity in democratic debate. *PS: Political Science & Politics*, 45(4), 675–687.
- Hédoin, C. (2015). Accounting for constitutive rules in game theory. *Journal of Economic Methodology*, 22(4), 439–461.
- Heijden, Eline van der, Nelissen, Jan and Verbon, Harrie A. A. (1997). „Altruism and Fairness in a Public Pension System.” *Journal of Economic Behavior and Organization*, 32: 505–518.
- Henrich, J. (2000). Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 90(4), 973–979.
- Henrich, J. (2017). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton: Princeton University Press.
- Henrich, J. P., Boyd, R., Bowles, S., Fehr, E., Camerer, C., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2), 73–78.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367.
- Herzog, H. A., & Golden, L. L. (2009). Moral emotions and social activism: The case of animal rights. *Journal of Social Issues*, 65(3), 485–498.
- Highton, B. – Kam, C. D. (2011). The Long-Term Dynamics of Partisanship and Issue Orientations. *The Journal of Politics*, 73(1), 202–215.

- Hindriks, F., & Guala, F. (2015). Institutions, rules, and equilibria: a unified theory. *Journal of Institutional Economics*, 11(3), 459–480.
- Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. *The Dark Side of the Force*, 198–219.
- Hitlin, S., & Vaisey, S. (Eds.). (2010). *Handbook of the Sociology of Morality*. Springer Science & Business Media.
- Hitlin, S., & Vaisey, S. (2013). The new sociology of morality. *Annual Review of Sociology*, 39, 51–68.
- Hobbes, T. (1970 [1651]). *Leviatán vagy az egyházi és a világi állam anyaga, formája és hatalma*; ford. Vámosi Pál, utószó Heller Ágnes, Budapest: Magyar Helikon–Európa.
- Hochman, H. M., & Rodgers, J. D. (1969). Pareto optimal redistribution. *The American Economic Review*, 59(4), 542–557.
- Hodgson, A. (2017). A beautiful mind got Nash equilibrium wrong. [https://www.youtube.com/watch?v=DTcmmD\\_MWas](https://www.youtube.com/watch?v=DTcmmD_MWas), letöltve 2018.09.10.
- Hogg, M. A. (2003). *Social Identity*. Guilford Press, New York.
- Hu, Sheng-Cheng. (1982). „Social Security, Majority-Voting Equilibrium and Dynamic Efficiency” *International Economic Review* 23: 269–287.
- Huddy, L., Bankert, A., & Davies, C. L. (2018). Expressive vs. Instrumental Partisanship in Multi-Party European Systems. *Advances in Political Psychology* 39(S1), 173–199.
- Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 109(1), 1–17.
- Imai, K. (2017). *Quantitative social science: an introduction*. Princeton University Press.
- Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition and emotion*, 23(4), 714–725.
- Inbar, Y., Pizarro, D., Iyer, R., & Haidt, J. (2012). Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science*, 3(5), 537–544.
- Inglehart, R. (1971). The silent revolution in Europe: Intergenerational change in post-industrial societies. *American Political Science Review*, 65(4), 991–1017.
- Inglehart, R. (1981). Post-materialism in an environment of insecurity. *American Political Science Review*, 75(4), 880–900.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

- Irwin, Sarah (1996). „Age Related Distributive Justice and Claims on Resources.” *British Journal of Sociology* 47: 69–92.
- Iyengar, S., & Krupenkin, M. (2018). Partisanship as Social Identity; Implications for the Study of Party Polarization. In *The Forum* 16(1), 23–45.
- Iyengar, S., Sood, G. & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3), 405–431.
- Jackendoff, R. (2003). Précis of foundations of language: brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, 26(6), 651–665.
- Janky, B. (2001). Rational choice theory in sociology and quantitative empirical research: A survey. *Review of Sociology*, 7(1), 21–34.
- Janky, B. (2005). *Szolidaritás és jóléti preferenciák*. Budapest: Akadémiai Kiadó.
- Janky, B. (2018). A társadalombiztosítás politikai gazdaságtana: A racionális magyarázat kísérletek rövid története. *Politikatudományi Szemle* 27(4): 57–73.
- Janky, B., & Takács, K. (2010). Efficient and inefficient social control in collective action. *CEU Political Science Journal*, 5(3).
- Janky, B., Bako, B., Szilagyi, P., & Bognar, A. (2014). Stigmatising the poor without negative images: images of extreme poverty and the formation of welfare attitudes. *Sociological Research Online*, 19(3), 1–10.
- Jasso, G. (1980). A new theory of distributive justice. *American Sociological Review*, 3–32.
- Jensen, C., & Petersen, M. B. (2017). The deservingness heuristic and the politics of health care. *American Journal of Political Science*, 61(1), 68–83.
- Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., Ebersole, C.R. & Narr, R. K. (2016). The effects of disgust on moral judgments: Testing moderators. *Social Psychological and Personality Science*, 7(7), 640–647.
- Johnson, D., & Krüger, O. (2004). The good of wrath: Supernatural punishment and the evolution of cooperation. *Political theology*, 5(2), 159–176.
- Jones, A., & Fitness, J. (2008). Moral hypervigilance: the influence of disgust sensitivity in the moral domain. *Emotion*, 8(5), 613.
- Kahneman, D. (2011). *Thinking, fast and slow* New York: Farrar, Straus and Giroux.
- Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728–741.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59(4/2): S285-S300.

- Kam, C. D., & Estes, B. A. (2016). Disgust sensitivity and public demand for protection. *The Journal of Politics*, 78(2), 481–496.
- Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies*, 59(1), 63–80.
- Kaplan, M. A. (1960). Some problems of the extreme utilitarian position. *Ethics*, 70(3), 228–232.
- Kertész I. (1975). *Sorstalanság*. Budapest: Magvető.
- Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608–638.
- Kinder, D. R., & Kam, C. D. (2009). *Us against them: Ethnocentric foundations of American opinion*. University of Chicago Press.
- Kinder, D. R., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*. University of Chicago Press.
- Klarreich, E. (2017). In game theory, no clear path to equilibrium. *Quanta Magazine* 18. July 2017. Retrieved from <https://www.quantamagazine.org/in-game-theory-no-clear-path-to-equilibrium-20170718/>, letöltve: 2018.09.20.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *science*, 314 (5800), 829–832.
- Kolm, S. C., & Ythier, J. M. (Eds.). (2006). *Handbook of the economics of giving, altruism and reciprocity: Foundations* (Vol. 1). Elsevier.
- Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of economic literature*, 41(4), 1188–1239.
- Kotlikoff, Laurence J.; Persson, Torsten; Svensson, Lars (1988). „Social Contracts and Assets: A Possible Solution to the Time Inconsistency Problem” *The American Economic Review* 78(3): 662–677.
- Köthenbürger, Marko; Poutvaara, Panu and Profeta, Paola (2005). „Why are More Redistributive Social Security Systems Smaller? A Median Voter Approach” *Econpublica Working Paper n. 103*.
- Kranz, S. (2010). Moral norms in a partly compliant society. *Games and Economic Behavior*, 68(1), 255–274.
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners’ dilemma. *Journal of Economic theory*, 27(2), 245–252.
- Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495–524.
- Kurz, M. (1978). Altruism as an outcome of social interaction. *The American Economic Review*, 68(2), 216–222.

- Lakatos, I. (1998 [1973]). *Bizonyítások és cáfolatok*. Typotex Kft., Budapest. [Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge: Cambridge University Press.]
- Lamba, S., & Mace, R. (2010). People recognise when they are really anonymous in an economic game. *Evolution and Human Behavior*, 31(4), 271–278.
- Lambertini, Luisa and Azariadis, Costas (1998). *The Fiscal Politics of Big Governments: Do Coalitions Matter?* Mimeo, UCLA.
- Lave, L. B. (1960). An Empirical Description of the Prisoner's Dilemma Game. Santa Monica, CA: RAND Corporation, 1960. <https://www.rand.org/pubs/papers/P2091.html>, letöltve: 2018.09.21.
- Lave, L. B. (1960). *Applications of the Theory of Games to Economics* (Doctoral dissertation, Reed College).
- Lave, L. B. (1962). An empirical approach to the prisoners' dilemma game. *The Quarterly Journal of Economics*, 424–436.
- Lave, L. B. (1965). Factors affecting co-operation in the prisoner's dilemma. *Behavioral Science*, 10(1), 26–38.
- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.
- Leeper, T. J. – Slothuus, R. (2014). Political Parties, Motivated Reasoning, and Public Opinion Formation. *Political Psychology*, 35(S1), 129–156.
- Leeper, T. J. – Slothuus, R. (2015). Can Citizens Be Framed? How Information, Not Emphasis, Changes Opinions. *Kézirat*, Aarhus University.
- Lewis, D. (1969). *Convention: A philosophical study*. John Wiley & Sons.
- Lindbeck, A., & Weibull, J. W. (1988). Altruism and time consistency: the economics of fait accompli. *Journal of Political Economy*, 96(6), 1165–1182.
- Lindenberg, S. (1988). Contractual relations and weak solidarity: The behavioral basis of restraints on gain-maximization. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 144(1), 39–58.
- Lindenberg, Siegwart (1992). The Method of Decreasing Abstraction. In: James S. Coleman and Thomas J. Fararo (eds.): *Rational Choice Theory. Advocacy and Critique*. Newbury Park: Sage, pp. 3–20.
- Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, 5(4), 356–368.
- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482–493.
- Liu, J. (2016). How A beautiful mind misrepresented John Nash and game theory? <https://www.youtube.com/watch?v=2EX2WIXDJEG>, letöltve, 2018.09.10.

- López Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic behavior*.
- López-Pérez, R. (2010). Guilt and shame: An axiomatic analysis. *Theory and Decision*, 69(4), 569–586.
- Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. PD: p. 95.
- Luttmer, E. F. (2001). Group loyalty and the taste for redistribution. *Journal of political Economy*, 109(3), 500–528.
- Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: explaining offensive action tendencies in an intergroup context. *Journal of personality and social psychology*, 79(4), 602.
- Macy, M. W. (2015a). An Emerging Trend: Is Big Data the End of Theory? *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–14.
- Macy, M. W. (2015b). Big theory: a trojan horse for economics? *Review of Behavioral Economics*, 2(1–2), 161–166.
- Malka, A. – Lelkes, Y. (2010). More Than Ideology: Conservative–Liberal Identity and Receptivity to Political Cues. *Social Justice Research*, 23(2–3), 156–188.
- Manza, J., & Brooks, C. (1999). *Social cleavages and political change: Voter alignments and US party coalitions*. OUP Oxford.
- Manza, J., Hout, M., & Brooks, C. (1995). Class voting in capitalist democracies since World War II: dealignment, realignment, or trendless fluctuation? *Annual review of sociology*, 21(1), 137–162.
- Marosán, G. (2008). A büntetés dicsérete. *Szociológiai Szemle* 19(3): 109–123.
- Mauss, M. (1954 [1923]). *The gift: The form and reason for exchange in archaic societies*. London: Cohen & West. [Essai sur le don: Forme et raison de l'échange dans les sociétés archaïques. *l'Année Sociologique*, seconde série, 1923–1924].
- Megyeri D. (2006). Egy elmaradt kézfogás. Magyar Nemzet, 2006. március 20. Letöltve: [https://mno.hu/migr\\_1834/egy-elmaradt-kezfogas-530101](https://mno.hu/migr_1834/egy-elmaradt-kezfogas-530101), letöltve, 2018.10.01.
- Meier, S. (2006). A survey of economic theories and field evidence on pro-social behavior. Federal Reserve Bank of Boston Working Paper no. 06–6.
- Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of government. *Journal of political Economy*, 89(5), 914–927.
- Meltzer, Allan H. and Richard, Scott F. (1981). „A Rational Theory of the Size of Government” *Journal of Political Economy* 89(5): 914–927.



- Mendez, M. F. (2017). A neurology of the conservative-liberal dimension of political ideology. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 29(2), 86–94.
- Minturn, L. (1984). Changes in the differential treatment of Rajput girls in Khalapur: 1955–1975. *Medical anthropology*, 8(2), 127–132.
- Morehous, L. G. (1966). One-play, two-play, five-play, and ten-play runs of Prisoner's Dilemma 1. *Journal of Conflict Resolution*, 10(3), 354–362.
- Moretti, L., & Di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion*, 10(2), 169.
- Morsky, B., & Akcay, E. (2018). The blind choreographer: evolution of social norms and correlated equilibria. *bioRxiv*, 435586.
- Mulligan, Casey B. and Sala-i-Martin, Xavier (1999a). „Gerontocracy, Retirement, and Social Security” *NBER Working Paper 7117*.
- Mulligan, Casey B. and Sala-i-Martin, Xavier (1999b). „Social Security in Theory and Practice (I): Facts and Political Theories” *NBER Working Paper 7118*.
- Mulligan, Casey B. and Sala-i-Martin, Xavier (1999c). „Social Security in Theory and Practice (II): Efficiency Theories, Narrative Theories, and Implications for Reform” *NBER Working Paper 7119*.
- Mulligan, Casey B. and Sala-i-Martin, Xavier (2003). „Social Security, Retirement, and the Single-Mindedness of the Electorate” *NBER Working Paper 9691*.
- Mulligan, Casey B., Gil, Ricard and Sala-i-Martin, Xavier (2004). „Do Democracies Have Different Public Policies than Nondemocracies?” *Journal of Economic Perspectives* 18(1): 51–74.
- Mühlenbernd, R., & Franke, M. (2012). Signaling conventions: Who learns what where and when in a social network? In *The Evolution Of Language* (pp. 242–249).
- Myerson, R. B. (1999). Nash equilibrium and the history of economic theory. *Journal of Economic Literature*, 37(3), 1067–1082.
- Nash, J. (1951). Non-cooperative games. *Annals of mathematics*, 286–295.
- Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1), 48–49.
- Navarrete, C. D., & Fessler, D. M. (2006). Disease avoidance and ethnocentrism: The effects of disease vulnerability and disgust sensitivity on intergroup attitudes. *Evolution and Human Behavior*, 27(4), 270–282.
- Nee, V., & Sanders, J. (2001). Understanding the diversity of immigrant incorporation: a forms-of-capital model. *Ethnic and racial studies*, 24(3), 386–411.

- Nelson, T. E., & Kinder, D. R. (1996). Issue frames and group-centrism in American public opinion. *The Journal of Politics*, 58(4), 1055–1078.
- Neumann, J. V. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1), 295–320.
- Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14), 8028–8033.
- Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, 364(6432), 56.
- Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological bulletin*, 135(2), 303.
- Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., & Elwood, L. S. (2007). The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychological assessment*, 19(3), 281.
- Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: theoretical investigations. *American journal of sociology*, 85(6), 1356–1375.
- Olson, M. (1997 [1965]). A kolektív cselekvés logikája. Közjavak és csoportelmélet. Budapest: Osiris [*The Logic of Collective Action: Public Goods and the Theory of Groups*]. Cambridge, MA: Harvard University Press.
- Oosterbeek, H., Sloof, R., & Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, 7(2), 171–188.
- Oosterhoff, B., Shook, N. J., & Ford, C. (2018). Is that disgust I see? Political ideology and biased visual attention. *Behavioural brain research*, 336, 227–235.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Orthmayr, I. (1997). Módszertani individualizmus. *Szociológiai Szemle*, 8(3), 3–31.
- Ostrom, E. (1990). *Governing the commons*. Cambridge: Cambridge University Press.
- Pampel, Fred C. (1994). „Population Aging, Class Context, and Age Inequality in Public Spending.” *American Journal of Sociology* 100: 153–195.
- Pareto, V. (1917). *Traité desociologie générale*. Paris-Genève: Librairie Droz.
- Parsons, T. (1949). Social classes and class conflict in the light of recent sociological theory. *The American Economic Review*, 16–26.
- Parsons, T. (1951). *The Social system*. New York: Routledge.
- Pascal, B. (1670). *Pensées*. Paris.

- Patterson, O & Fosse, E., (eds.) (2015a). *The Cultural Matrix: Understanding Black Youth*. Harvard University Press.
- Patterson, O. & Fosse, E. (2015b). Introduction In. Patterson, O & Fosse, E., (eds.) *The Cultural Matrix: Understanding Black Youth*. Harvard University Press, pp. 1–25.
- Patterson, O. (2014). Making sense of culture. *Annual Review of Sociology*, 40, 1–30.
- Patterson, O. (2015a). The Social and Cultural Matrix of Black Youth. In. Patterson, O & Fosse, E., (eds.) *The Cultural Matrix: Understanding Black Youth*. Harvard University Press, pp. 45–137.
- Patterson, O. (2015b). The Real Problem With America's Inner Cities. The New York Times, Sunday Review, Opinion. 9 May 2015. <http://www.nytimes.com/2015/05/10/opinion/sunday/the-real-problem-with-americas-inner-cities.html>.
- Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological methods*, 19(4), 459.
- Pearl, J. (2018). *Causal and Counterfactual Inference*. Manuscript prepared for the Handbook of rationality by MIT Press, UCLA, Computer Science Department, Los Angeles, October 2., 2018.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: a primer*. John Wiley & Sons.
- Persson, Torsten and Tabellini, Guido (2000). *Political Economics: Explaining economic Policy*. Cambridge, MA: The MIT Press.
- Persson, Torsten and Tabellini, Guido (2002). „Political Economics and Public Finance”, In. Alan J. Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics Volume 3*. Amsterdam etc: Elsevier, pp. 1549–1659.
- Petersen, M. B. (2012). Social welfare as small-scale help: evolutionary psychology and the deservingness heuristic. *American Journal of Political Science*, 56(1), 1–16.
- Petersen, M. B. (2017). Healthy Out-Group Members Are Represented Psychologically as Infected In-Group Members. *Psychological science*, Dec; 28(12): 1857–1863.
- Petersen, M. B., Sznycer, D., Cosmides, L., & Tooby, J. (2012). Who deserves help? evolutionary psychology, social emotions, and public opinion about welfare. *Political psychology*, 33(3), 395–418.
- Petersen, M. B. – Skov, M. – Serritzlew, S. – Ramsøy, T. (2013). Motivated Reasoning and Political Parties: Evidence for Increased Processing in the Face of Party Cues. *Political Behavior*, 35(4), 831–854.

- Petrescu, D. C., & Parkinson, B. (2014). Incidental disgust increases adherence to left-wing economic attitudes. *Social Justice Research*, 27(4), 464–486.
- Phelps, E. S. (Ed.). (1975). *Altruism, morality, and economic theory*. Russell Sage Foundation.
- Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior*, 29(3), 172–178.
- Popper, K. (1989 [1957]). *A historicizmus nyomorúsága*. Budapest: Akadémiai Kiadó [The poverty of historicism. London: Routledge].
- Portes, A. (Ed.). (1995). *The economic sociology of immigration: Essays on networks, ethnicity, and entrepreneurship*. Russell Sage Foundation.
- Preston, Samuel H. (1984). „Children and the Elderly: Divergent Paths for America’s Dependents”, *Demography* 21: 435–457.
- Profeta, Paola (2000). „Demography, Retirement and the Future of Social Security” *Econpubblica Working Paper* 68.
- Quillette Magazine (2018). The grievance studies scandal: Five academics respond. *Quillette Magazine October 1 2018*. Letöltve: <https://quillette.com/2018/10/01/the-grievance-studies-scandal-five-academics-respond/>; letöltve 2018.10.15.
- Quora (é. n.) Why is Nash equilibrium is explained incorrectly in A beautiful mind? <https://www.quora.com/Why-is-Nash-equilibrium-explained-incorrectly-in-A-Beautiful-Mind>; letöltve 2018.09.10.
- Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General*, 145(4), 389.
- Rangel, Antonio (2003). „Forward and Backward Intergenerational Goods: Why Is Social Security Good for the Environment?” *The American Economic Review* 93(3): 813–834.
- Rapoport, A., & Dale, P. S. (1966). The “end” and “start” effects in iterated Prisoner’s Dilemma. *Journal of Conflict Resolution*, 10(3), 363–366.
- Rasmusen, E. B., & Hirshleifer, D. (1989). Cooperation in a repeated prisoner’s dilemma with ostracism. *Journal of economic Behavior and Organization*, 12, 87–106.
- Raub, W., & Weesie, J. (1990). Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, 96(3), 626–654.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

- Real, F., & Griffiths, T. L. (2010). Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1680), 429–436.
- Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of public Economics*, 88(7), 1625–1644.
- Rey, J.-M. (2008). If we all go for the blonde. Plus Magazine, June 1. <https://plus.maths.org/content/if-we-all-go-blonde>; letöltve: 2018.09.10.
- Roell, S. (é. n.). The best books on game theory: Recommended by Ariel Rubinstein. <https://fivebooks.com/best-books/ariel-rubinstein-on-game-theory/>, letöltve: 2018.10.01.
- Rogers, J. (2018). The game theory glitch in A beautiful mind. Law and Liberty, April 27. <https://www.lawliberty.org/2018/04/27/the-game-theory-glitch-in-a-beautiful-mind/>, letöltve: 2018.09.10.
- Rohrmann, S., Hopp, H., & Quirin, M. (2008). Gender differences in psychophysiological responses to disgust. *Journal of Psychophysiology*, 22(2), 65–75.
- Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic theory*, 25(1), 92–100.
- Rothko, M. (2006). *The artist's reality: Philosophies of art*. New Haven: Yale University Press.
- Rovó A. (2017). Annyit hazudtak régen és ma is, hogy az ember már nem hisz senkinek. Index 2017. február 28. Letöltve: 2018.10.20. [https://index.hu/gazdasag/2017/02/28/postabank\\_panik\\_20.\\_evfordulo\\_princz\\_gabor/](https://index.hu/gazdasag/2017/02/28/postabank_panik_20._evfordulo_princz_gabor/).
- Rozin, P., Haidt, J., & McCauley, C. R. Disgust. (2008). In Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.), *Handbook of emotions*. Guilford Press., pp. 757–776.
- Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science*, 8(2), 67–73.
- Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion. *Journal of economic theory*, 21(1), 1–9.
- Rubinstein, A. (1991). Comments on the interpretation of game theory. *Econometrica: Journal of the Econometric Society*, 909–924.
- Samuelson, P. A. (1947). *Foundations of economic analysis*. Harvard Economic Studies no. 80. Cambridge, MA.
- Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, 15(60), 243–253.

- Samuelson, Paul A. (1958). „An Exact Consumption-loan Model of Interest with or without the Social Contrivance of Money” *Journal of Political Economy* 66: 467–482.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755–1758.
- Scheff, T. J. (1967). Toward a sociological model of consensus. *American Sociological Review* 32(1): 32–46.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and social psychology bulletin*, 34(8), 1096–1109.
- Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. *Handbook of qualitative research*, 1, 118–137.
- Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of personality and social psychology*, 53(3), 550.
- Script (2012). Adapting 'a beautiful mind': Screenwriter Akiva Goldsman. *Script Magazine*. December 19. <https://www.scriptmag.com/features/interviews-features/adpating-a-beautiful-mind-screenwriter-akiva-goldsman>; leltve: 2018.09.10.
- Selten, R. (1978). The chain store paradox. *Theory and decision*, 9(2), 127–159.
- Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, 40(159), 241–259.
- Shilling, C., & Mellor, P. A. (2015). For a sociology of deceit: Doubled identities, interested actions and situational logics of opportunity. *Sociology*, 49(4), 607–623.
- Silverstein, M., Parrott, T. M., Angelelli, J. J. and Cook, F. L. (2000). „Solidarity and Tension between Age-groups in the United States: Challenge for an Aging America in the 21st Century.” *International Journal of Social Welfare* 9(4): 270–284.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359–1366.
- Simpson, B., & Willer, R. (2015). Beyond altruism: Sociological foundations of cooperation and prosocial behavior. *Annual Review of Sociology*, 41, 43–63.
- Sjoblom, Kriss (1985). „Voting for Social Security” *Public Choice* 45: 225–240.
- Slothuus, R. (2010). When Can Political Parties Lead Public Opinion? Evidence from a Natural Experiment. *Political Communication*, 27(2), 158–177.

- Slothuus, R. – De Vreese, C. H. (2010). Political Parties, Motivated Reasoning, and Issue Framing Effects. *The Journal of Politics*, 72(3), 630–645.
- Smith, A. (1992 [1776]). *A nemzetek gazdagsága*. Budapest: KJK.
- Smith, A. (2010 [1759]). *The theory of moral sentiments*. New York: Penguin Books.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48(4), 813.
- Smith, K. B., Alford, J. R., Hibbing, J. R., Martin, N. G., & Hatemi, P. K. (2017). Intuitive ethics and political orientations: Testing moral foundations as a theory of political ideology. *American Journal of Political Science*, 61(2), 424–437.
- Smith, K. B., Oxley, D. R., Hibbing, M. V., Alford, J. R., & Hibbing, J. R. (2008). *The Ick Factor: Physiological Sensitivity to Disgust*. Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago.
- Smith, K. B., Oxley, D., Hibbing, M. V., Alford, J. R., & Hibbing, J. R. (2011). Disgust sensitivity and the neurophysiology of left-right political orientations. *PloS one*, 6(10), e25552.
- Smith, V. L. (2015). Adam Smith: Homo Socialis, Yes; Social Preferences, No; Reciprocity Was to Be Explained. *Review of Behavioral Economics* 2: 183–193.
- Sniderman, P. M., Crosby, G. C., & Howell, W. G. (2000). The politics of race. Sears, D. O., Sidanius, J., & Bobo, L. (Eds.). (2000). *Racialized politics: The debate about racism in America*. University of Chicago Press, pp. 236–279.
- Sokal, A., & Bricmont, J. (2008 [1997]). *Intellektuális impostorok*. Budapest: Typotex [*Impostures intellectuelles*. Paris: Éditions Odile Jacob].
- Sommerfeld, R. D., Krambeck, H. J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the national academy of sciences*, 104(44), 17435–17440.
- Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social Psychology Quarterly*, 224–237.
- Stets, J. E., & Carter, M. J. (2012). A theory of the self for the sociology of morality. *American Sociological Review*, 77(1), 120–140.
- Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *The american economic review*, 67(2), 76–90.
- Strand, M., & Lizardo, O. (2015). Beyond world images: Belief as embodied action in the world. *Sociological Theory*, 33(1), 44–70.
- Strand, M., & Lizardo, O. (2017). The hysteresis effect: theorizing mismatch in action. *Journal for the Theory of Social Behaviour*, 47(2), 164–194.

- Stryker, S. (1968). Identity salience and role performance: The relevance of symbolic interaction theory for family research. *Journal of Marriage and the Family*, 558–564.
- Stryker, S., & Serpe, R. T. (1982). Commitment, identity salience, and role behavior: Theory and research example. In *Personality, roles, and social behavior* (pp. 199–218). Springer, New York, NY.
- Sugden, R. (1984). Reciprocity: the supply of public goods through voluntary contributions. *The Economic Journal*, 94(376), 772–787.
- Syi (2008). Cselekvésemélet dióhéjban Játék, elmélet, módszer, tan – Weber: újratöltve. Budapest: Typotex.
- Syi (2009). Önzetlenségelméletek. Bereczkei Tamás: Az erény természete. *BUKSZ* 21(3): 226–233.
- Syi (2014). Syi.hu/cse. budapest: L'Harmattan. [http://syi.hu/pdf/action\\_semantics20140312.pdf](http://syi.hu/pdf/action_semantics20140312.pdf).
- Szántó Z. (1993). Racionalitás és társadalom: Peter Abell (ed.) „Rational Choice Theory” című könyvéről (Schools of Thought in Sociology, no. 8. An Elgar Reference Collection, 1991). *Szociológiai Szemle*, 3(4), 185–190.
- Szántó Z. (1999). A társadalmi cselekvés mechanizmusai. Budapest: Aula.
- Szántó Zoltán, Syi (2010). Fizikusok, bélyeggyűjtők, emberjárás jelentők. *BUKSZ* 22(3), 201–213.
- Szelenyi, I. (2015). The triple crisis of sociology. Contexts 15 April 2015 URL: <http://contexts.org/blog/the-triple-crisis-of-sociology> ; letöltve: 2018.10.15.
- Tabellini, Guido (1991). „The Politics of Intergenerational Redistribution” *Journal of Political Economy* 99(2): 335–357.
- Tabellini, Guido (1992). „A Positive Theory of Social Security” *Mimeo*, IGER, Milano.
- Tabellini, Guido (2000). „A Positive Theory of Social Security” *Scandinavian Journal of Economics* 102(3): 523–545.
- Takács, E. (2018). *A szolidaritás alakváltozásai*. Budapest: Napvilág.
- Takács, K., Janky, B., & Flache, A. (2008). Collective action and network change. *Social Networks*, 30(3), 177–189.
- Tavory, I. (2011). The question of moral action: A formalist position. *Sociological Theory*, 29(4), 272–293.
- Taylor, M. (1976). *Anarchy and cooperation*. John Wiley & Sons.
- Taylor, M. (1987). *The possibility of cooperation*. Cambridge University Press.
- Terrizzi, J. A., Shook, N. J., & McDaniel, M. A. (2013). The behavioral immune system and social conservatism: A meta-analysis. *Evolution and Human Behavior*, 34(2), 99–108.



- Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107(4), 657.
- Toff, B., & Suhay, E. (2018). Partisan Conformity, Social Identity, and the Formation of Policy Preferences. *International Journal of Public Opinion Research*. edy014.
- Trapa, P. E., & Nowak, M. A. (2000). Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, 41(2), 172–188.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, 46(1), 35–57.
- Tucker, A. W. (1950). A Two-Person Dilemma. Mimeo, Stanford University [Published: „On Jargon: The Prisoner’s Dilemma.” *UMAP Journal*, 1980, 1, p. 101. g]
- Turner, Bryan S. (1989). „Ageing, Politics and Sociological Theory.” *British Journal of Sociology* 40: 588–606.
- Turner, Bryan S. (1998). „Ageing and Generational Conflicts: A Reply to Sarah Irwin.” *British Journal of Sociology* 49: 299–304.
- Turner, J. C. – Oakes, P. J. – Haslam, S. A. – McGarty, C. (1994). Self and Collective: Cognition and Social Context. *Personality and Social Psychology Bulletin*, 20(5), 454–463.
- Turner, V., Turner, V. W., & Victor, T. (1967). *The forest of symbols: Aspects of Ndembu ritual* (Vol. 101). Ithaca, NY: Cornell University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Tybur, J. M. *et al.* & Consedine, N. S. (2016). Parasite stress and pathogen avoidance relate to distinct dimensions of political ideology across 30 nations. *Proceedings of the National Academy of Sciences*, 201607398.
- Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: individual differences in three functional domains of disgust. *Journal of personality and social psychology*, 97(1), 103.
- Tybur, J. M., Merriman, L. A., Hooper, A. E. C., McDonald, M. M., & Navarrete, C. D. (2010). Extending the behavioral immune system to political psychology: Are political conservatism and disgust sensitivity really related? *Evolutionary Psychology*, 8(4), 147470491000800406.
- Ulmann-Margalit, E. (1977). *The emergence of norms*. Oxford: Oxford University Press.

- Vaisey, S., & Valentino, L. (2018). Culture and choice: Toward integrating cultural sociology with the judgment and decision-making sciences. *Poetics*, 68, 131–143.
- van Leeuwen, F., Dukes, A., Tybur, J. M., & Park, J. H. (2017). Disgust sensitivity relates to moral foundations independent of political ideology. *Evolutionary Behavioral Sciences*, 11(1), 92.
- van Oorschot, W., & Roosma, F. (eds.) (2017). *The Social Legitimacy of Targeted Welfare: Attitudes to Welfare Deservingness*. Cheltenham: Edward Elgar Publishing.
- Varnum, M. E., Grossmann, I., Kitayama, S., & Nisbett, R. E. (2010). The origin of cultural differences in cognition: The social orientation hypothesis. *Current directions in psychological science*, 19(1), 9–13.
- Verbon, Harrie A. A. (1987). „The Rise and Evolution of Public Pension Systems” *Public Choice* 52: 75–100.
- Verbon, Harrie A. A. (1988). *The Evolution of Public Pension Schemes*. Berlin: Springer Verlag.
- Verbon, Harrie A. A. and Verhoeven, Marijn M. J. (1992). „Decision Making on Pension Schemes under Rational Expectations” *Journal of Economics* 56: 71–97.
- Verhoeven, Marijn M. J. and Verbon, Harrie A. A. (1991). „Expectations on Pension Schemes under Non-stationary Conditions” *Economics Letters* 36: 99–103.
- Von Wright, G. H. (1987 [1971]). Magyarázat és megértés. In: Bertalan L. (szerk.) Magyarázat, megértés, társadalomtudomány. Budapest: Tömegkommunikációs Kutatóközpont [*Explanation and understanding*. Ithaca: Cornell University Press.]
- Wagemans, F., Brandt, M. J., & Zeelenberg, M. (2018). Disgust Sensitivity Is Primarily Associated With Purity-Based Moral Judgments. *Emotion forthcoming*.
- Walque, Gregory de (2005). „Voting on Pensions: A Survey” *Journal of Economic Surveys* 19(2): 181–209.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129–133.
- Watts, D. J. (2014). Common sense and sociological explanations. *American Journal of Sociology*, 120(2), 313–351.
- Weber, M. (1987 [1922]). *Gazdaság és társadalom: A megértő szociológia alapvonalai 1*. Budapest: KJK.
- Weeks, J. (1995). *Invented moralities: Sexual values in an age of uncertainty*. Columbia University Press.

- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological science*, 16(10), 780–784.
- Wilson, W. J. (1987). *Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. London: The University of Chicago Press.
- Wilson, W. J. (2009). *More than just race: Being black and poor in the inner city (issues of our time)*. New York: WW Norton & Company.
- Winter, F., Rauhut, H., & Helbing, D. (2012). How norms can generate conflict: An experiment on the failure of cooperative micro-motives on the macro-level. *Social Forces*, 90(3), 919–946.
- Yamagishi, T. et al. & Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, 109(50), 20364–20368.
- Young, H. P. (2007). *Social Norms*. University of Oxford Department of Economics Discussion Paper Series No 307.
- Zaller, J. (1996). The Myth of Massive Media Impact Revived: New Support for a Discredited Idea. In Mutz, D. C. – Sniderman, P. M. – Brody, R. A. (szerk.). *Political Persuasion and Attitude Change* University of Michigan Press, Ann Arbor, 17–78.



## Névmutató

- Aarøe, L. 143, 161, 169, 172–176,  
179, 212, 269, 284  
Aaron, H. J. 113, 269  
Abend, G. 32, 269  
Adams, T. G. 170, 269  
Ahn, W. Y. 169, 269  
Akerlof, G. A. 120, 124, 139, 269  
Albrekt Larsen, C. 141, 213, 269  
Alchian, A. A. 33, 52, 55, 192, 270  
Alesina, A. 112, 140, 141, 270  
Alford, R. R. 59, 270, 295  
Almås, I. 143, 270  
Allais, M. 268, 270  
Alvard, M. S. 151, 270  
Andersen, S. 134, 270  
Anderson, C. 26, 270, 275, 276  
Andreoni, J. 118, 129, 270  
Arceneaux, K. 269, 270  
Aron, R. 59, 270  
Arrow, K. 55, 117, 271  
Auerbach, A. J. 112, 271, 291  
Aumann, R. J. 97, 99, 112, 147, 148,  
194, 195, 199, 203–205, 271  
Axelrod, R. 93, 102, 103, 109, 117,  
119–122, 128, 130, 132, 138, 152,  
271  
Baldwin, P. 86, 271  
Banks, A. J. 28, 164, 171, 173, 271  
Bardsley, N. 131, 271  
Bargheer, S. 32, 271  
Barrett, L. F. 259, 271, 293  
Becker, G. 53–55, 59–61, 70, 84, 85,  
91, 92, 115, 117, 118, 127, 138,  
139, 234, 271, 272, 295  
Bereczkei, T. 20, 244, 272, 296  
Berelson, B. R. 58, 59, 272  
Berger, P. L. 31, 232, 272  
Bernau, P. 125, 272  
Bernhard, H. 128, 272, 279  
Bicchieri, C. 42, 108, 109, 149, 152,  
153, 185, 203, 205, 209, 210, 226,  
230, 231, 240, 241, 250, 256, 261,  
263, 272  
Binmore, K. 33, 56, 62, 64, 74, 79, 97,  
99, 101, 103, 120, 123, 132–133,  
145–147, 149, 152, 185, 191–195,  
203, 205, 207, 208, 210–212, 222,  
226, 228, 229, 231, 232, 234, 237,  
242, 245, 256, 272, 273  
Bisgaard, M. 217, 273  
Blau, P. M. 59, 273  
Bloch, M. 208, 273  
Boadway, R. W. 114, 273

- Boeri, T. 113, 273  
Boldrin, M. 273  
Bolsen, T. 217, 273  
Bolton, G. E. 126, 273  
Borg, J. S. 159, 273  
Bosker, B. 26, 274  
Boudreau, C. 217, 219, 274  
Bourdieu, P. 208, 274  
Bower, G. H. 274  
Bowles, S. 140, 270, 274, 279, 281, 283  
Boyd, R. 134, 135, 270, 274, 281, 283  
Brams, S. J. 101, 274  
Breyer, F. 274  
Bridges, B. 274  
Browning, E. K. 113, 115, 274  
Buchanan, J. 51, 78, 274  
Bullock, J. G. 217, 218, 274  
Burke, P. J. 31, 244, 295  
Burton-Chellew, M. N. 153, 274, 275  
Busby, E. 275  
Butler, M. 275
- Camerer, C. F. 128, 270, 275, 283  
Cameron, L. A. 134, 179, 183, 275  
Cameron, C. D. 275  
Camic, C. 61, 275  
Cappelen, A. W. 270, 275  
Capraro, V. 235, 275, 292  
Carsey, T. 217, 275  
Cass, D. 201, 275  
Chamley, C. 220, 275  
Chan, C. 275  
Chapman, H. A. 159, 178, 181, 258, 275, 276  
Charness, G. 20, 129, 276  
Chwe, M. S. 205, 206, 276  
Ciuk, D. J. 183, 188, 217–220, 256, 276
- Clifford, S. 160, 162, 172, 176, 177, 276  
Coleman, J. 105, 276, 287  
Conde-Ruiz, J. 276  
Conesa, J. C. 276  
Cookson, R. 129, 276  
Cooley, T. F. 114, 276  
Cox, D. 276  
Cukierman, A. 276
- Csontos, L. 78, 79, 81, 206, 276
- Dana, J. 130, 131, 151, 277  
De Quervain, D. 137, 277  
DeSteno, D. 277  
Deutsch, M. 122, 277  
Diamond, D. W. 224, 277  
DiMaggio, P. 277  
Downs, A. 51–53, 217, 277  
Druckman, J. 217, 218, 273, 275, 277  
Duckitt, J. 277  
Dufwenberg, M. 138, 277  
Duncan, L. A. 277  
Durkheim, É. 20, 23, 29, 31, 47–49, 206, 222, 244, 251, 276, 277
- Earp, B. D. 25, 277  
Eckel, C. 134, 149, 154, 194, 278  
El Mouden, C. 274, 275  
Ellingsen, T. 278  
Ellison, G. 106, 278  
Ellsberg, D. 51, 129, 278  
Elster, J. 32, 278  
Embrey, M. 100, 148, 278  
Emler, N. 157, 278  
Engelmann, D. 129, 278  
England, P. 53, 57, 278  
Evans, G. 178, 241, 278  
Exadaktylos, F. 133, 278

- Falk, A. 136, 138, 278  
Federico, C. M. 165, 220, 278, 282  
Fehr, E. 54, 55, 124–129, 131, 132,  
134–137, 145, 150, 270, 272, 275,  
277–279, 281, 283, 286  
Fischbacher, U. 128, 138, 272, 277,  
278  
Feinberg, M. 279  
Ferge, Z. 59, 279  
Ferri, G. 202, 279  
Fields, J. M. 204, 279  
Flache, A. 110, 111, 236, 237, 242,  
248, 279, 296  
Flood, M. M. 87, 279  
Foner, A. 279  
Fong, C. M. 140, 212, 279  
Fontaine, P. 13, 34, 53, 85, 90, 118,  
139, 280  
Forges, F. 201, 280  
Franklin, M. N. 86, 280  
Franklin, R. 266, 280  
Friedell, M. F. 122, 123, 133, 203,  
204, 206, 280  
Friedman, M. 33, 52, 191, 192, 280  
Fudenberg, D. 101, 280  
  
Gabaix, X. 268, 280  
Gächter, S. G. 128, 137, 231, 279,  
280, 283  
Gál, R. I. 280  
Galasso, V. 114, 276, 280, 281  
Gale, D. 15, 281  
Gelissen, J. 116, 281  
Gintis, H. 42, 49, 125, 134, 139, 149,  
154, 185, 194, 202, 203, 206–209,  
231–233, 240, 245, 256, 258, 270,  
274, 278, 279, 281, 283  
Gneezy, U. 109, 270, 281  
Golder, S. A. 26, 281  
  
Graham, J. 165, 182, 281  
Granovetter, M. 76, 77, 80, 105, 110,  
281  
Gray, K. 275  
Green, D. 53, 79, 281  
Greif, A. 107, 210, 249, 282  
Guala, F. 195, 202, 207, 232, 245,  
261, 282, 284  
Gulyás, A. 121, 126, 282  
Güth, W. 123, 133, 148, 282  
  
Haidt, J. 14, 21, 55, 155–159, 162,  
164–166, 174, 177, 180, 182, 281,  
282, 284, 293, 294, 299  
Haley, K. J. 151, 282  
Hamil-Luker, J. 113, 116, 282  
Hamilton, W. D. 60, 117, 120, 271,  
282  
Han, J. 275, 282, 283  
Hansson, I. 283  
Hardin, G. 89, 283  
Harsanyi, J. C. 50, 283  
Hatemi, P. 164, 283, 295  
Hédoin, C. 189, 203, 283  
Heijden, E. 283  
Henrich, J. 125, 134, 135, 144, 150,  
151, 265, 270, 283  
Herrmann, B. 283  
Herzog, H. A. 177, 283  
Highton, B. 217, 283  
Hindriks, F. 261, 284  
Hirshleifer, J. 104, 105, 258, 259,  
263, 284, 292  
Hitlin, S. 32, 284  
Hobbes, T. 47, 284  
Hochman, H. M. 117, 139, 284  
Hogg, M. A. 218, 284  
Hu, S-C. 284  
Huddy, L. 219, 284

- Imai, K. 25, 27, 28, 284  
Inbar, Y. 157, 165–167, 169, 174, 284  
Inglehart, R. 58, 85, 284  
Ioannidis, J. P. 27, 284  
Irwin, S. 285, 297  
Iyengar, S. 218, 219, 285
- Jackendoff, R. 216, 285  
Janky, B. 79, 110, 112, 121, 126, 143,  
151, 236, 237, 285, 296  
Jasso, G. 122, 285  
Jensen, C. 142, 285  
Jerit, J. 176, 276  
Johnson, D. J. 101, 181, 285  
Jones, A. 158, 285, 293
- Kahneman, D. 33, 123, 129, 285  
Kam, C. D. 283, 286  
Kandori, M. 105–108, 110, 111, 114,  
127, 151, 232, 233, 237, 249, 286  
Kaplan, M. A. 89, 286  
Kertész I. 32, 255, 286  
Kimbrough, E. O. 231, 286  
Kinder, D. R. 160, 173, 286, 290  
Klarreich, E. 202, 286  
Knoch, D. 128, 286  
Kolm, S. C. 124, 279, 286  
Konow, J. 129, 132, 140, 143, 286  
Kotlikoff, L. J. 112, 114, 271, 286  
Köthenbürger, M. 286  
Kranz, S. 231, 286  
Kreps, D. M. 98, 232, 286  
Krupka, E. L. 235, 286  
Kurz, M. 102, 103, 286
- Lakatos, I. 9, 11, 244, 286  
Lamba, S. 154, 286  
Lambertini, L. 287  
Lave, L. B. 89, 94, 99, 287
- Lazarus, R. S. 34, 257, 258, 287  
Leeper, T. J. 217, 218, 287  
Lelkes Y. 217, 218, 285, 288  
Lewis, D. 203, 204, 208, 287, 293  
Lindbeck, A. 90, 287  
Lindenberg, S. 25, 56, 61, 78, 251,  
287  
Lindquist, K. A. 183, 259, 275, 287  
List, J. A. 270, 287  
López-Pérez, R. 231, 233, 287  
Luce, R. D. 87, 287  
Luckmann, T. 31, 232, 272, 287  
Luttmer, E. F. 140, 288
- Mackie, D. M. 257, 288  
Macy, M. W. 24–26, 110, 248, 279,  
281, 288  
Malka, A. 217, 218, 288  
Manza, J. 86, 288  
Marosán, G. 139, 288  
Mauss, M. 208, 288  
Megyeri D. 230, 288  
Meier, S. 129, 134, 288  
Meltzer, A. H. 112, 276, 288  
Mendez, M. F. 162, 289  
Minturn, L. 214, 289  
Morehous, L. G. 94, 289  
Moretti, L. 181, 289  
Morsky, B. 193, 289  
Mulligan, C. B. 272, 289  
Mühlenbernd, R. 208, 289  
Myerson, R. B. 202, 223, 289
- Nash, J. F. 6, 7, 63–75, 88–91, 97,  
98, 101, 106, 145, 147, 148–150,  
152, 154, 185, 187–195, 197, 198,  
200–202, 206, 207, 221–224, 226,  
228, 229, 250, 258, 284, 287, 289,  
292, 297



- Navarrete, C. D. 289, 297  
Nee, V. 249, 289  
Nelson, T. E. 160, 290, 294  
Neumann, J. V. 46, 48, 50, 60, 63, 80, 189, 195, 268, 290  
Nosek, B. A. 281  
Nowak, M. A. 120, 121, 153, 193, 208, 290, 297
- Oaten, M. 290  
Ockenfels, A. 126, 273  
Olatunji, B. O. 163, 167, 290  
Oliver, P. 104, 290  
Olson, M. 84, 89–92, 108, 290  
Oosterbeek, H. 133, 290  
Oosterhoff, B. 163, 290  
Orthmayr, I. 23, 29, 35, 49, 290  
Ostrom, E. 20, 108, 163, 290
- Pampel, F. C. 290  
Pareto, V. 51, 87, 97, 284, 290  
Parsons, T. 20, 108, 214, 215, 290  
Pascal, B. 101, 290  
Patterson, O. 215, 291  
Pearl, J. 28, 291  
Persson, T. 286, 291  
Petersen, M. B. 142, 143, 172–176, 212, 217, 219, 269, 277, 285, 291  
Petrescu, D. C. 170, 177–179, 292  
Phelps, E. S. 53, 118, 292  
Piazza, J. 210, 292  
Popper, K. 35, 79, 83, 292  
Portes, A. 249, 292  
Preston, S. H. 112, 292  
Profeta, P. 276, 281, 286, 292
- Rand, D. G. 275, 292  
Rangel, A. 292  
Rapoport, A. 94, 103, 119, 123, 292
- Rasmusen, E. B. 104, 105, 292  
Raub, W. 93, 292  
Rawls, J. 122, 292  
Reali, F. 216, 293  
Rege, M. 151, 293  
Richerson, P. J. 135, 274  
Roell, S. 267, 293  
Rohrmann, S. 293  
Rosenthal, R. W. 97, 293  
Rothko, M. 266, 293  
Rozin, P. 159, 177, 282, 293  
Rubinstein, A. 70, 74, 101, 146, 242, 293
- Samuelson, P. A. 50, 51, 58, 60, 122, 148, 193, 293, 294  
Sanfey, A. G. 180, 294  
Scheff, T. J. 203, 206, 294  
Schelling, T. C. 50, 89, 150, 192, 294  
Schmidt, K. M. 126, 127, 129, 132, 279  
Schnall, S. 294  
Schwandt, T. A. 35, 294  
Schwartz, S. H. 58, 294  
Selten, R. 96–98, 123, 148, 283, 294  
Sen, A. 294  
Shilling, C. 294  
Silverstein, M. 113, 116, 294  
Simmons, J. P. 27, 294  
Simpson, B. 20, 23, 32, 294  
Sjoblom, K. 114, 294  
Slothuus, R. 217–219, 273, 277, 287, 294  
Smith, A. 20, 47, 66, 90, 92, 125, 295  
Smith, C. A. 295  
Smith, K. B. 156, 159, 161, 163, 164, 169, 182, 183, 295

- Smith, V. L. 126, 138, 261, 295  
Sniderman, P. M. 295, 299  
Sokal, A. 26, 295  
Sommerfeld, R. D. 210, 295  
Stets, J. E. 20, 31, 244, 295  
Stigler, G. J. 53, 61, 234, 295  
Strand, M. 33, 295  
Stryker, S. 31, 60, 244, 296  
Sugden, R. 120, 296  
Syi (Szakadát István) 19, 20, 24, 27, 34, 77, 79, 81, 139, 244, 256, 296  
Szántó Z. 10, 296  
Szelenyi, I. 57, 267, 296  
Tabellini, G. 115, 273, 291, 296  
Takács, E. 9, 86, 296  
Takács, K. 11, 110, 238, 296  
Tavory, I. 20, 21, 32, 296  
Taylor, M. 102, 103, 108, 109, 120, 226, 296  
Terrizzi, J. A. 161, 164, 168, 171–173, 296  
Thomas, K. A. 206, 297  
Toff, B. 220, 297  
Trapa, P. E. 208, 297  
Trivers, R. L. 117, 297  
Tucker, A. W. 87, 297  
Turner, B. S. 297  
Turner, J. 218, 297  
Turner, V. 208, 297  
Tversky, A. 129, 297  
Tybur, J. 161, 163, 166–169, 172, 174–176, 178, 182, 297  
Ulmann-Margalit, E. 226, 297  
Vaisey, S. 32, 267, 284, 298  
van Leeuwen, F. 160, 298  
van Oorschot, W. 142, 298  
Varnum, M. E. 20, 298  
Verbon, H. A. 114, 283, 298  
Verhoeven, M. M. J. 298  
Von Wright, G. 79, 298  
Wagemans, F. 159–161, 180, 298  
Walque, G. de 298  
Wasserstein, R. L. 28, 298  
Watts, D. J. 26, 27, 298  
Weber, M. 23, 29, 33, 47, 48, 49, 51, 54, 57, 58, 79, 82, 242, 244, 253, 298  
Weeks, J. 226, 298  
Weiner, B. 142, 298  
West, S. A. 153, 274, 275  
Wheatley, T. 164, 180, 299  
Willer, R. 20, 23, 32, 279, 294  
Wilson, W. J. 32, 85, 86, 139, 299  
Winter, F. 299  
Yamagishi, T. 133, 299  
Yost, B. A. 217–218, 276  
Young, H. P. 261, 299  
Zaller, J. 218, 299



