



TÁJÉKOZÓDÁS

A MÁSODRENDŰ POTYAUTAS PROBLÉMA ÉS MEGOLDÁSAI*

KOVÁCS Benedek

Magyar Máltai Szeretetszolgálat
H-1033 Budapest, Miklós u. 32.; e-mail: kovacs.benedek@maltai.hu

Abstract: For the recent twenty years numerous various attempts have been developed to explain the co-operation problems on the area of social sciences based on the discipline of rational decision theory and similar intentions of evolutionary biology have joined these initiatives. Despite the outstandingly rich literature on suggested solutions, the problem has not been answered yet. Number of attempts on solving the Second-order Free-riding Problem suggest to involve an external factor in order to explain the co-operative behaviour (e.g. altruistic punishment), others attempt to modify the original problem, respectively insert rigorous assumptions (evolutionary explanations). It seems, that the altruistic punishment and the strong reciprocity will get outstanding roles in explaining the emerging and sustaining of norms. On the other hand this should not mean, that the rational calculation approach would not play any role in the future, as the various social control mechanisms expressed by the social networks can represent such cost reduction factors which can strengthen the motives based on fairness.

Összefoglaló: A kooperáció problémájának magyarázatára az utóbbi húsz évben rengeteg különböző megoldás született a racionális döntések elméletén nyugvó társadalomtudomány területén belül, ráadásul az evolúciós biológia hasonló irányú törekvései is bekapcsolódtak a kérdés tisztázásába. A javasolt megoldások irodalma rendkívül gazdag, ennek ellenére a kérdés tisztázása még korántsem zárult le. A másodrendű potyautas problémára tett megoldási kísérletek egy része valamilyen külső faktor beiktatásával kívánja megmagyarázni a kooperatív viselkedést (pl. altruista büntetés), más része viszont az eredeti probléma módosításával, illetve szigorú feltevések beiktatásával kísérletezik (evolúciós magyarázatok). Úgy tűnik, az altruista büntetés és az ezen alapuló erős reciprocitás kitüntetett szerepet fog kapni a normák keletkezése és fenntartása magyarázatában, ez azonban nem jelenti azt, hogy a racionális kalkuláción nyugvó megközelítés ne kapna szerepet. Ugyanis a társadalmi hálózatok által közvetített különféle társadalmi kontroll mechanizmusok olyan hatékony költségcsökkentő tényezők lehetnek, amelyek ráerősítenek a méltányossági elveken alapuló motívumokra.

* A tanulmány a Faludi Ferenc Akadémia valamint a *Társadalmi megoldások és társadalmi kudarcok* c. OTKA kutatás (T049432) támogatásával készült. Köszönettel tartozom az OTKA kutatás valamennyi résztvevőjének, különösen Takács Károlynak és Jankó Bélának.



Keywords: collective action, Second-order Free-riding Problem, social norm, sanctioning, altruistic punishments

Kulcsszavak: kollektív cselekvés, másodrendű potyautas probléma, társadalmi norma, szankcionálás, altruista büntetés

Olson (1997 [1965]) óta jól ismert, hogy a csoporttagok közös érdekére épülő kollektív cselekvési helyzetekben az egyéni cselekvések interakciója nem feltétlenül eredményezi a közös érdek megvalósulását. Sőt a szerző szerint inkább a sikeres kollektív cselekvés nevezhető kivételesnek és az csak meghatározott társadalmi feltételek között valósulhat meg. Ezzel szemben sokkal természetesebb kimenetel a társadalmi kudarc, hiszen a valós helyzetek nagy részében nem adtak az önérdék követő viselkedés számára a biztonságos együttműködés feltételei. A szerző későbbi bírálói közül többen rámutattak, hogy a sikeres kollektív cselekvés jóval elterjedtebb, mint azt Olson feltételezte (Marwell–Oliver 1993), és részletesebb analitikus elemzésnek vetették alá a kollektív cselekvési szituációt leíró feltevéseket (Frohlich et al. 1975; Marwell–Oliver 1993). Ettől függetlenül Olson alapproblémája valóságos társadalmi dilemmát jelenít meg, amelyet Hardin (1971) az „n-szereplős fogolydilemma” modellel ír le.

A kollektív cselekvés problémájának leküzdésére Olson bevezette a szelektív ösztönző fogalmát, amely olyan költségmódosító tényező, amely az adott egyént érdekelté teheti a közjósághoz való hozzájárulásban. A centrális szelektív ösztönzők mellett már Olson felhívta a figyelmet azokra a decentralizált mechanizmusokra, amelyeket a csoporttagok közötti kapcsolatok közvetítenek (társadalmi szelektív ösztönzők).

Attól függetlenül, hogy a szelektív ösztönző központi vagy decentralizált módon kerül biztosításra, mindenképpen felmerül az a kérdés, hogy ez vajon költséges vagy pedig nem az egyének, illetve az azt előállító ágens (pl. állam) számára. Ha azt feltételezzük, hogy a szelektív ösztönző költséges, akkor az felveti a másodrendű potyautas problémát (Oliver 1980; Elster 1989; Heckathorn 1989; Coleman 1990), amennyiben viszont költségmentesnek tekintjük, akkor a „másodrendű láthatatlan kéz” megoldja a szelektív ösztönző létrejöttének kérdését.

A MÁSODRENDŰ POTYAUTAS PROBLÉMA FELMERÜLÉSE

A közgazdaságtanban a tiszta közjavak meghatározása két tényező segítségével történik. Az első a kínálat egyetemlegessége, amely szerint egy újabb fogyasztó csatlakozása nem csökkenti a közjóság rendelkezésre álló mennyiségét. A második definíciós elem az, hogy nem lehetséges a fogyasztók kizárása a közjóság használatából, illetve az olyan költséges lenne, hogy gazdasági szempontból értelmetlenné válik (Samuelson 1954). A közjavak jelentős hányada azonban nem tiszta közjóság, ugyanis nem igaz rá a kínálat egyetemlegessége, éppen ezen a tényen alapul a csoportméret paradoxona (Oliver–Marwell 1988). Jóllehet a csoportméret paradoxona Olson számára még ismeretlen volt – hiszen egyik alaptétele szerint a csoportméret növekedése a kollektív cselekvés ellen hat – már ő sem tekintette a kínálat egyetemlegességét a közjóság szükségsszerű velejárójának. Így a (nem tiszta) közjóságot úgy definiálja,

Szociológiai Szemle 2008/2.



mint ami hasznosságot jelent egy csoport tagjai számára, ugyanakkor, ha egyszer rendelkezésre áll a csoport valamely tagja számára, akkor a többiek sem zárhatók ki belőle (Olson 1997 [1965]).

Más definíciók az előbbi definíciós elem (azaz, hogy a használatukból való kizárás nem lehetséges technikailag vagy nagyon költséges, és ezért gazdasági értelemben értelmetlen) mellett még egy kritériumot alkalmaznak, miszerint az előállításuk együttműködés által valósul meg, tehát a közjóság csak kollektív cselekvés által hozható létre (Marwell–Oliver 1993).

Vagyis a kollektív cselekvés a közjóság előállításának a módja. A kollektív cselekvést többféleképpen modellezték, de az egyik legáltalánosabban elfogadott az n szereplős fogoly dilemmája (FD) helyzet. Olson (1997 [1965]) óta tudjuk, hogy az egyének közös érdekei nem vezetnek automatikusan a csoportérdek megvalósulásához, ugyanis (Olson modellje szerint) a csoporttag haszna a létrejövő közjóság rá eső részéből és a hozzájárulás költségéből tevődik össze. Mivel a közjóság létrejötté a többi csoporttag cselekvésétől is függ, ezért – hacsak nem látszik eleve biztosítottnak a közjóság, vagyis az általánosnak tekinthető bizalomhiány feltevésével élve – a csoporttag számára a hozzájárulás a veszteség kockázatát rejti magában.¹ Csak abban az esetben racionális a hozzájárulás, ha az egyén hozzájárulása a többi csoporttag cselekvésétől függetlenül is akkora értéket képes létrehozni, hogy az meghaladja saját hozzájárulásának költségét. Olson ezt látta lehetségesnek a kicsoportok esetében.

A kollektív cselekvés FD alapú modellezését Hardin (1971) vezette be Olson (1997 [1965]) alapján. Az n-szereplős fogoly dilemma modellben feltételezzük, hogy a kollektív jóság mértéke az egyéni hozzájárulások számának függvényében növekszik.² A kétszereplős fogoly dilemmája helyzethez hasonlóan alapvetően minden egyes csoporttagnak érdeke, hogy ne járuljon hozzá a közjósághoz, és úgy részesedjen a többiek hozzájárulás következtében előálló hasznosságból. Ez a fajta potyautas jelenség a kétszereplős esetben nehezen képzelhető el, hiszen a helyzet könnyen áttekinthető és ezért kölcsönös dezertáláshoz vezet. Ugyanakkor sok szereplőt feltételezve, ha a közjóság növekedése folytonos, kevésbé nyilvánvaló egy-egy csoporttag dezertálásának hatása. Mivel azonban ez a felismerés könnyen általánossá válik, a kimenetel gyakran a közjóság meghiúsulása lesz.

Frohlich és szerzőtársainak (1975) elemzése kimutatta, hogy az, hogy a csoporttagok önként hozzájárulnak-e a közjósághoz vagy sem, nem kizárólag a kollektív javak meghatározásától függ, hanem a szituáció olyan elemeitől is, mint a hasznosságfüggvény és a termelési függvény alakja. Azonban a bináris választási lehetőség feltevése (csak kooperálni és dezertálni lehet) is egy olyan feltevés, amely szűkíti a lehetséges kimeneteket.

Lineáris termelési függvényt és hasznosságfüggvényt feltételezve, amely a legegyszerűbb eset, a kooperálók és a dezertálók haszna egymással párhuzamosan változik a kooperálók számának növekedésével, és ezért az egyének döntése arra vonatkozóan,

- 1 A defektálás két motivációs tényezője a potyautas magatartástól várható nyereség kísértése és – egy esetleges egyoldalú kooperálás esetében – a „balekká” válástól való félelem (Yamagishi–Sato 1986).
- 2 Két esetet különböztetünk meg, az elsőt az előállított közjóság folytonos mennyiség, a második esetben diszkrét (Gál 1988). Diszkrét esetben szükséges a hozzájárulások egy meghatározott küszöb értéke, ahhoz, hogy a közjóság létrejön. Folytonos esetben nincs ilyen küszöb érték, vagyis elvben a közjóság értéke csak akkor lesz nulla, ha egyetlen csoporttag sem járul hozzá.

Szociológiai Szemle 2008/2.



hogy kooperáljanak-e, független attól, hogy ténylegesen hányan járulnak hozzá a közjósághoz. Vagyis a játékosoknak van domináns stratégiájuk, amelyet az határoz meg, hogy mekkora költség árán lehet hozzájárulni az adott közjósághoz. Amennyiben a domináns stratégia a nem kooperálás, az egyének kollektív jósághoz való hozzájárulása csak szelektív ösztönzőkkel biztosítható.

Ezt az esetet írja le Hardin modellje, mely szerint a kooperálók haszna:

$$u_c = \frac{r}{n}(m+1) - 1$$

A nem kooperálók haszna pedig:

$$u_d = \frac{r}{n}m$$

A kettő különbsége pedig

$$\frac{r}{n} - 1$$

Ahol

n : a csoport létszáma

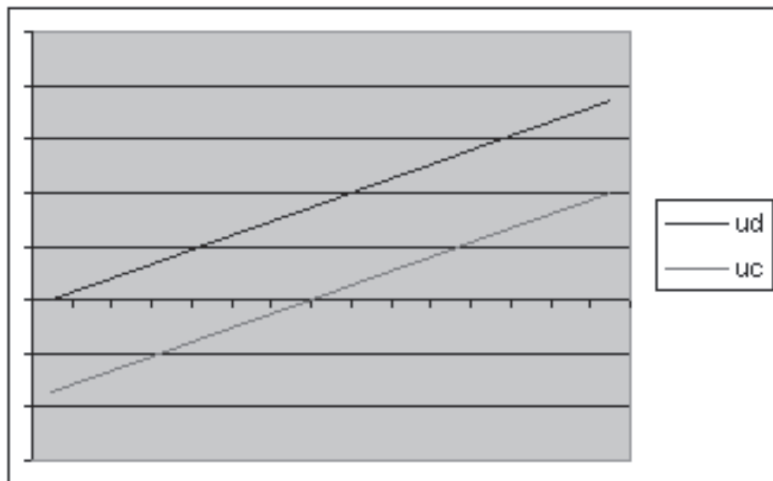
r : a hasznoknak a hozzájáruláshoz viszonyított aránya

u_c : a kooperálók haszna

m : a kooperáló játékos száma

u_d : a defektáló játékosok haszna

Ez esetet Schelling diagrammal ábrázolva jól látszik a kooperálók és a defektálók kifizetésfüggvénye közötti különbség.



1. ábra A kooperálók és a defektálók kifizetésfüggvénye



A valóságban sokkal gyakoribbak azonban azok az esetek, amikor a termelési függvény és/vagy a csoporttagok hasznosság függvénye nem lineáris. Ezt okozhatja tipikusan a hasznosságfüggvény esetében közjóság csökkenő határhaszna, a termelési függvény esetében pedig a növekvő határkölttség vagy épp magas kezdeti előállítási költség (Frohlich et al. 1975). Ekkor a kooperálók és a nem kooperálók kifizetési függvénye már nem feltétlenül párhuzamos, és keresztezi egymást, amely azt jelenti, hogy a csoporttagoknak már nincs domináns stratégiájuk, viselkedésük ezért attól függ, hogy hányan járulnak hozzá az adott közjósághoz. Ezek az esetek a kritikus tömeg modellekkel írhatók le (Schelling 1998 [1978]).

Az előzőekből látható, hogy nem minden esetben merül fel a csoporttagok kooperálástól való tartózkodásából adódó, a csoport egésze számára szuboptimális kimenetelt eredményező társadalmi csapdahelyzet.

A kollektív cselekvés problémájának megoldására több elméleti modell is létezik, de az egyik legrégebbi a szelektív ösztönzők (Olson 1997 [1965]) alkalmazása. Az egyik analitikus felosztás szerint léteznek centralizált és decentralizált megoldások (Elster 1989, 1995; Szántó 1999), de megadhatunk konkrétabb, objektum szerinti felsorolást is (mint morál, kormányzat, vallás, vagy a közjóság privatizációja).

A szelektív ösztönzők két formája a jutalom és a büntetés. Bár a defektálók szankcionálása vagy épp a kooperálók jutalmazása matematikai értelemben egyenértékű a formális modellekben, mégsem ugyanazzal a hatással bír. A két ösztönzési forma a kooperálók és a defektálók csoporton belüli arányától függ: ha a kooperálók vannak kevesebben, akkor a jutalmazás, ha a defektálók, akkor pedig a büntetés jelent kisebb költséget a szelektív ösztönzőt alkalmazók számára (Oliver 1980; Elster 1989; Gardner–West 2004). Éppen ezért olyan kollektív cselekvési helyzetekben, amikor nincs elfogadott csoportnorma, amely a kooperálás irányába hatna (vagyis amikor a viselkedési szabályszerűség a nem kooperálás, pl. civil mozgalmaknál), akkor a jutalmazás a hatékonyabb. Amikor viszont már kialakult egyfajta kooperációs norma (vagyis a viselkedési szabályszerűség inkább a kooperálás), akkor a normaszegők büntetése a hatékonyabb ösztönzési forma. Másrészt a büntetés a célszemélyt elidegenítheti, illetve provokálhatja, ami kockázatot rejt magában (Elster 1989; Horne 2004). Éppen ezért a büntetés könnyebben biztosítható, ha már van egy kooperációs minimum szint, amelyet valamely más mechanizmus hoz létre (Gardner–West 2004).

A hétköznapi tapasztalatok alapján ihletett, elterjedt megoldási javaslat a kollektív cselekvés problémájára a társadalmi normák alkalmazása. Azonban a társadalmi norma fogalmában nincs teljes konszenzus, egyes szerzők kizárólag tervezett, szándékolt normákról beszélnek (Coleman 1990) más szerzők nem szándékolt, evolúciós vagy tanulásalapú normafogalmat használnak (Axelrod 1986; Opp 2002).³

A másodrendű potyautas probléma Coleman (1990) elméletében – a szándékolt cselekvés eredményeként létrejövő – társadalmi norma keletkezésekor lép fel. A norma keletkezésének két feltétele Coleman szerint a norma iránti kereslet megjelenése és a normasértő viselkedés szankcionálása. Mivel a szankcionálás többnyire költséges a cselekvők számára (Hechter 1984), újabb problémaként merül fel az, hogy ki viselje a

3 Szintén nagy különbségek vannak az egyes szerzők között abban, hogy minek tekintik a társadalmi normát, pusztán szankciónak, szankcióval egybekötött viselkedési szabálynak, vagy morális kötelezettséget is társítanak hozzá. Erről bővebben Hechter–Opp 2001.



szankcionálás költségét, vagy hogyan osztozzanak a szankcionálás költségében? Ha szankcionálni költséges, akkor a szankcionáló egyének hátrányt szenvednek azokkal szemben, akik nem szankcionálnak, csupán élvezik a létrejövő (másodrendű) közjószág hasznát. Vagyis a szankcionálásban nem résztvevő cselekvők másodrendű potyautasoknak tekinthetők. A csoport érdeke hosszútávon hogy létrehozzák a normát, azonban az egyéni érdekek rövidtávon az diktálják, hogy ne vegyenek részt a szankcionálásban.

A probléma jól szemléltethető a „játékba ágyazott játék” módszerrel (Heckathorn 1991). A társadalmi kontroll játék két egymásba ágyazott játékból áll, egy alapjátékból, amely többféle társadalmi csapdahelyzet is lehet, és egy másodrendű „norma játékból” (*Compliant Control Game*), amelyben a játékosoknak lehetőségük áll a másik játékos cselekvési lehetőségeit beszűkíteni. Amennyiben a játékosoknak sikerül a második játékban kooperálni, az első játék kimenetele már nem kérdéses. Az n szereplős elsőrendű kollektív cselekvés probléma egy defektálót feltételezve egy $n-1$ szereplős másodrendű kollektív cselekvési problémát vet fel. Ez két szereplőnél nem merül fel, hiszen abban az esetben egyetlen játékosra hárul a szankcionálás költsége.

	C_2	D_2												
C_2	<table border="1"> <tr> <td>C</td> <td>$5,5$</td> </tr> </table>	C	$5,5$	<table border="1"> <tr> <td>C</td> <td>$5,5$</td> </tr> <tr> <td>D</td> <td>$7,0$</td> </tr> </table>	C	$5,5$	D	$7,0$						
C	$5,5$													
C	$5,5$													
D	$7,0$													
D_2	<table border="1"> <tr> <td>C</td> <td>$5,5$</td> <td>D</td> <td>$0,7$</td> </tr> </table>	C	$5,5$	D	$0,7$	<table border="1"> <tr> <td>C</td> <td>$5,5$</td> <td>D</td> <td>$0,7$</td> </tr> <tr> <td>D</td> <td>$7,0$</td> <td>D</td> <td>$2,2$</td> </tr> </table>	C	$5,5$	D	$0,7$	D	$7,0$	D	$2,2$
C	$5,5$	D	$0,7$											
C	$5,5$	D	$0,7$											
D	$7,0$	D	$2,2$											

2. ábra Megoldások a kollektív cselekvés problémára

Mielőtt rátérnénk a másodrendű potyautas probléma lehetséges megoldásaira szükséges röviden kitérni azokra a lehetséges válaszokra, amelyek az elsőrendű potyautas problémára születtek. A kooperációt elősegítő tényezők között megkülönböztethetjük azokat a strukturális feltételeket, amelyek szükségesek a sikeres kollektív cselekvéshez és az effektív megoldási mechanizmusokat (Szántó 1999). Szántó (1999) a szükséges vagy épp kizáró feltételek között sorolja fel a kognitív feltételeket, a csoportnagyságot, a csoporttagok közötti kapcsolatokat, a csoport stabilitását, a csoport-





tagok kulturális heterogenitását, illetve a politikai vezetők jelenlétét vagy hiányát. A megoldási mechanizmusok között említtek elsősorban a szelektív ösztönzők, a normák, a politikai vállalkozók, vagy a kommunikáció (Szántó 1999). A szelektív ösztönzőknél szélesebb fogalom a társadalmi kontroll mechanizmusok (Baumgartner et al. 1975; Radelet 1979), amelybe Janky és Takács (1995) az informális szelektív ösztönzők (társadalmi szelektív ösztönzők, Olson 1997 [1965]) mellett más mechanizmusokat is beleértene, így a konformitást vagy viselkedés megerősítést. A társadalmi kontroll mechanizmusokat a társadalmi kapcsolatháló közvetíti szemben a centralizált szelektív ösztönzőkkel. A feltételes kooperálás (reciprocitás), a reputáción alapuló indirekt reciprocitás (Panchanathan–Boyd 2004; Nowak–Sigmund 2005), valamint a költséges jelzés elmélet (Smith–Bliege Bird 2000; Gintis et al. 2001) bizonyos feltételek mellett szintén elősegítik kooperációt.

A költséges jelzés elmélet magyarázatul szolgálhat a kollektív jószág csoport számára történő biztosítására (Smith–Bliege Bird 2000). A terepmunkán alapuló cikk szerint egy ausztrál törzs esetében a levesteknősök vadászatakor egy szűk alcsoport magára vállalva az előállítási költségét a teljes csoport számára biztosítja a kollektív jószágot, anélkül hogy ezt a számukra a többiek viszonozná. Vagyis ebben az esetben a reciprocitás kizárható. Ugyanakkor a magyarázó modell szerint a látszólagos altruizmus mögött az a motiváció húzódik meg, hogy a kollektív cselekvés biztosításának magas költsége hitelesíti a vadászok vezetői, fizikai stb. domináns képességeit. Vagyis a kollektív cselekvésben való részvétel egy verseny a magasabb státuszért és az azzal összefüggő magasabb haszonszerzési lehetőségekért. Részben ezekre a bizonyítékokra támaszkodva hozták létre sokszereplős játékelméleti modelljüket Gintis és szerzőtársai (2001). Az eredményeik szerint a közjószág biztosításán keresztül megmutatkozó a költséges jelzés lehet evolúciós értelemben stabil, és képes elterjedni egy olyan populációban, amelyben eredetileg ritkán fordul elő. Ám igaz az is, hogy nem csak a kollektív cselekvésben való részvétel képes költséges jelzéssé válni, hanem a magánjavakat (pl. hivalkodó fogyasztás) és a közrosszat (pl. büntetlen normasértés) előállító cselekvések éppúgy alkalmasak arra. Kérdés lehet, hogy a jelzés milyen feltételek mellett jelentkezik társadalmilag hasznos formában. Gintis és szerzőtársai (2001) szerint többek között a csoportszintű szelekció mechanizmusa kedvez a proszociális jellegű jelzések kialakulásának. De az is valószínűsíthető, hogy a proszociális jelzések nagyobb elismerést kapnak, mint a társadalmilag semleges vagy káros jelzések (Smith–Bliege Bird 2000).

A társadalmi csapdahelyzet ismétlődése bizonyos feltételek mellett magyarázattal szolgálhat a kooperációra racionális játékosok esetén is. Az ún. népi tétel (*folk theorem*) szerint végtelenül ismételt helyzetben sok egyensúlyi helyzet lehetséges (Fudenberg–Maskin 1986; Rasmusen 1992). Az egyensúlyi helyzetek szelekciójában segítséget nyújthat az evolúciós megközelítés (Binmore 1998), amelyet Axelrod (1984) is alkalmazott. Így a kétszereplős végtelenül ismételt helyzetben az ún. trigger stratégiák (pl. TIT-FOR-TAT – Axelrod 1984) – amelyek akár folyamatos defektálást is előírhatnak válaszlépésként az egyszeri defektálásra – alkalmazása lehetővé teszi a kooperáció létrejöttét racionális cselekvők feltételezése mellett. Ugyanis a válasz-defektálástól és az ennek következtében alacsonyabb kifizetéssel járó egyensúlyi pontban való megrekedéstől való félelem arra kényszeríti a feleket, hogy kooperáljanak. Ugyanez a megoldási mechanizmust (reciprocitás) azonban sokszereplős kollektív cselekvési helyzetben már nem érvényesül (Boyd–Richerson 1988; Fehr et al. 2002).

Szociológiai Szemle 2008/2.



Erős feltevésnek tűnik a végtelen ismétlődés, ugyanis a valóságban a cselekvők kapcsolatai nem végtelenek, de figyelembe kell azt is venni, hogy ha az ismétlődések száma viszonylag nagy és emellett az utolsó forduló bekövetkezésének időpontja bizonytalan – ami a valóságban sokszor teljesül –, akkor ez lényegét tekintve nem különbözik a végtelenül ismételt helyzettől. Ezzel szemben a (beláthatóan) véges számú ismételt fogoly dilemma játék a nem ismételt fogoly dilemma játékhoz válik hasonlóvá, ugyanis a sorozat utolsó aljátékában a defektálás a racionális cselekvés, amelyből visszafele következtetve pedig a folyamatos defektálás stratégiája következik. A kísérleti eredmények és hétköznapi tapasztalatok azt erősítik meg, hogy véges ismételt helyzetekben is gyakori a kooperáció kialakulása, aminek az a valószínű magyarázata, hogy az információhiány gátolhatja ezt a „visszafele következtetési mechanizmust” (Fudenberg–Maskin 1986).

Vagyis kétszereplős helyzetben, ha az interakció ismétlődésének a valószínűsége magas, akkor a defektálás rövidtávú hasznát ellensúlyozza annak hosszú távon jelentkező költsége, ami akár a teljes kooperációs egyensúlyhoz is vezethet racionális önérdékű cselekvőknél. Más a helyzet azonban, ha egyszerre sok szereplő van a játékban. Fischbacher és szerzőtársai (2001) kísérleti úton vizsgálták a feltételes kooperálás jelenségét. Egy nem ismételt kollektív cselekvési helyzetet hoztak létre, amelyben a játékosoknak egyrészt dönteniük kellett arról, hogy hozzájárulnak-e a közjószághoz, és milyen mértékben, másrészt arra is válaszolniuk kellett, hogy bizonyos előre meghatározott kooperációs átlagok mellett kooperálnának-e vagy sem. Eredményeik szerint annak ellenére, hogy a közjószág-játék nem ismétlődő szituáció volt, az átlagos kooperáció szintje nem a teljes defektálást mutatta. Ami kísérleti alanyok preferenciáit illeti, mintegy 30%-uk megrögzött potyautas preferenciákkal rendelkezett, vagyis a kooperáció bármely szintje mellett defektált volna, 50% pedig a feltételes kooperáció híve volt, azonban ők is hajlamosak lettek volna a kooperáció átlagos szintjénél valamivel kisebb mértékben hozzájárulni a közjószághoz. Feltételezve, hogy ismétlődő helyzetben nem változtatják játékosok a preferenciáikat, ezek az eredmények a kooperáció szintjének a csökkenését jelzik előre. A magas szintű és stabil kooperáció nem tartható fenn hosszú távon már akkor sem, ha a feltételes kooperálók, mindig valamivel az átlagos szint alatt járulnak hozzá a közjószághoz. A feltételes kooperáció kísérletben talált eredmények a szerzők szerint bizonyítékkal szolgálhatnak arra, hogy létezik egyfajta a feltételes kooperálási norma (Fehr–Fischbacher 2004b), amely akkor ír elő a kooperálást, ha mások is kooperálnak, míg a defektálás általánossá válása felmenti a csoporttagokat a kooperálás kötelezettsége alól. Azonban valódi megoldást csak a defektálók szankcionálása jelenthet.⁴

Minthogy az ismétlődés esetében a potyautas viselkedés büntetését a viszont-defektálás jelenti, azonnali költség nem merül fel (a kölcsönös defektálás hosszú távon jelentkező költségéről már esett szó). A sokszereplős kollektív cselekvési helyzetekben azonban az eddig eredmények szerint sokkal inkább az aktív, ún. *indirekt szankcionálás* jelenthet megoldást, amely viszont költséges az egyének számára, és ez magában rejti a másodrendű potyautas problémát. Mielőtt erre rátérnék, bemutatok egy megoldást, amelyben a sokszereplős helyzetre egy indirekt szankcionálási mo-

4 A feltételes kooperálásnál felmerülő büntetést (vagyis a „válasz”-defektálást) *indirekt szankcionálásnak* is nevezik, míg a szankcionálás olyan aktív formáit, mint a kizárás, fizikai büntetés, elutasítás, vagy helytelenítés *direkt szankcionálásnak* (Diekmann–Voss 2003).



delt dolgoztak ki. Panchanathan és Boyd (2004) a másodrendű potyautas probléma kikerülésével dolgozta ki evolúciós modelljét a kollektív cselekvés problémájára. Ebben a modellben a kollektív cselekvési helyzetet egy kölcsönös segítségnyújtási játékkal (közvetett reciprocitás) kapcsolták össze a szerzők. A játékosok információval rendelkeznek a többiek előző cselekvéseiről (reputáció), és a kollektív cselekvési játékban defektáló játékosok szankcionálása a társadalmi csere játékból történő kizárással történik, amely költségmentes, sőt monetáris értelemben hasznos a szankcionálók számára. A modellben a három lehetséges stratégia a kooperálóké (mindkét játékban kooperálnak), a defektálóké (mindkét játékban defektálnak) és a büntetőké (a kollektív cselekvésnél kooperálnak, a kölcsönös segítségnyújtásnál, csak a jóhírű játékosokat támogatják). A modell két stabil evolúciós egyensúlyi helyzetet eredményezett, az egyikben mindenki defektáló, a másikban mindenki büntető stratégiával rendelkezett. A két játék közötti kapcsolatot a büntetők jelentik, ezért ha a büntetők kezdeti aránya túl alacsony, akkor a defektálók fognak felülkerekedni. Vagyis az eredmények szerint az indirekt reciprocitás – a büntető stratégiák meghatározott kezdeti aránya mellett – elő tudja segíteni a kollektív cselekvés kialakulását. Ugyanakkor a modell kritikusa rámutat, hogy azzal, hogy a szankcionálók ki tudják zárni a potyautas játékosokat a második játékból, a szerzőpáros figyelmen kívül hagyja az esetleges másodrendű potyautasok létezését, másrészt az összes indirekt reciprocitáson alapuló rendszer nagyon érzékeny az információ (monitorozás) megbízhatóságára (Fowler 2005b). Panchanathan és Boyd (2005) is elismeri, hogy ez a modell csak abban az esetben jelenthet megoldást, ha feltételezzük, hogy van egy stabilan működő indirekt reciprocitáson alapuló rendszer.

Az eddigi eredmények az jelzik, hogy a kooperáció lehet racionális viselkedés is. Négy meghatározó mechanizmus (a rokonság alapú szelekció, a reciprocitás, az indirekt reciprocitás és a költséges jelzés) az evolúciós biológia az állatok viselkedésénél is kimutatott. Azonban a kísérleti eredmények szerint az emberi viselkedésben akkor is megfigyelhető a kooperáció, amikor a leírt feltételek nem teljesülnek, és a nem kooperálás a racionális viselkedés (Fehr–Gächter 2002).

TÁRSADALMI NORMÁK

Az ideológiák, az egyenlőségi és szolidaritási normák belső ösztönzők lehetnek a kooperálás szempontjából (Oliver 1980). A társadalomtudományokban jelenleg nincs széles körben elfogadott megegyezés a norma fogalmát illetően, és a különböző szerzők többféle dimenzióban ragadják meg a fogalmi eltéréseket (Gibbs 1965; Ulmann-Margalit 1977; Axelrod 1986; Elster 1989; Saam–Harrer 1999; Horne 2001b; Hechter–Opp 2001b).

Coleman (1990) elméletében a normák mindig valamilyen negatív externália kiküszöbölése vagy pozitív externália előállítása érdekében jönnek létre. S bár több szerző is említi a közrosszat előállító normákat (Elster 1989; Kuran 1997), Coleman álláspontja védhető annyiban, hogy mindig lehet találni egy haszonélvező csoportot, amelynek érdekében áll a norma előállítása. A pozitív externália előállítása vagy épp a negatív externália csökkentése - különösen, ha ez az externália egyszerre több embert is érint, márpedig többnyire így van -, többnyire közjószágot testesít meg, és eszerint

Szociológiai Szemle 2008/2.



tulajdonképpen minden norma kooperációs (vagy koordinációs⁵) normának tekinthető. A más szerzők által említett normák közül a teljesség igénye nélkül meg kell említenünk a morális normákat, a konvenciót, allokációs normákat és a privát normákat (Elster 1989; Gál 1989).

Elster (1989) szerint a normák általában nem eredményorientáltak. A kooperációs norma ebben kivételes (vagyis ebben az esetben Coleman elképzelését erősíti meg), de a kooperációs normákra rásegítenek olyan mélyebben gyökerező normák, amelyek nem eredményorientáltak (méltányossági normák). Így a racionális megfontolások mellett nem racionális tényezők is szerepet játszanak a kooperációs viselkedésben. Ezt, mint később kitérünk rá, kutatási eredmények is megerősítik (DeQuervain et al. 2004).

A kooperációs norma speciális helyzetű és ezért viszonylag jól definiált. Eszerint a normák felfoghatók úgy, mint eszközök egy csoport céljai elérése érdekében, vagyis a normák *másodrendű közjavak*, amelyek az *elsőrendű közjavak* létrehozásához szükségesek (Opp 2002). Ez magával hozza a norma viselkedésközpontú megközelítését, amely elsősorban a szankció jelenlétével ragadja meg annak létét (Axelrod 1986). A norma fogalmát három elemre szokták bontani, melyek a viselkedési szabályszerűség, a szankcionálás és a kötelezettség érzés (Gibbs 1965; Hechter–Opp 2002). Az informális szankcionálás területén jelentkező legfrissebb eredmények arra hívják fel a figyelmet, hogy a méltányossági illetve morális elvek szerepétől sem nem lehet eltekinteni, ugyanis egyrészt ezek állhatnak a büntetés mögött motivációs háttérként (Falk et al. 2005), másrészt ezek biztosítják a szankcionálás legitimitását (Fehr–Rockenbach 2003) is. A humán ökológia pedig megerősíti, hogy az emberi kooperáció sikerességében a nyelv is szerepet játszik (Smith 2003). Mindez arra enged következtetni, hogy a norma fogalma mégsem szűkíthető le a szankció elemére. Elster (1989) elképzelése szerint egy többszintű norma-rendszerrel állunk szemben, ahol vannak mélyebb gyökerező (érzelmi alapú), és magasabb szintű (társadalmi) normák, és ahol az előbbieket az utóbbiaknak, amely viszont már racionális elemeket is tartalmaz. Normákra épülnek a különböző méltányossági és altruista büntetés modellek (Fehr et al. 2004a). Az ultimátum játékokra alapozott kísérletek eredményei szerint a „negatív reciprocitás” típusú büntetés még igen magas költségek esetén is megfigyelhető, jóllehet nem racionális az egyén számára (Camerer 2003). A méltányossági normák megsértése esetén, legalább is kétszereplős helyzetben, az egyének komoly költségek mellett is érdemesnek látják megbüntetni a normasértőt.

A társadalmi normákra való hivatkozás ezért egyre inkább jelen van a racionális döntések elméletén nyugvó társadalomtudományokban is, ám a társadalmi normák keletkezése maga is megoldatlan probléma. Horne (2001b) kétféle magyarázó modellt különböztet meg, az első az evolúciós, amely lehet próba-hiba tanulás alapú vagy a racionális cselekvés mellékterméke, ide sorolhatjuk többek között Axelrod (1986), vagy Opp (2002) tanulmányait. A második az externália alapú, amelynél a norma szándékolt cselekvés és racionális kalkuláció eredménye (Coleman 1990). Opp (2002) és Orthmayr (2004) ettől a felosztástól némileg eltérően a spontán keletkezésű és a terve-

5 Ulmann-Margalit (1977) felosztásában szerepel a koordinációs és a kooperációs normák megkülönböztetése. A koordinációs normákhoz való igazodás a csoporttagok legjobb érdeke, a kooperációs normák követése viszont csak második legjobb érdeke.



zett normákat különbözteti meg. Ezek a felosztások különbözik abban, hogy nem tárgyalják azt az esetet, amikor a norma alulról szerveződve jön létre, ám szándékolt cselekvés eredménye (Coleman 1990), és a centralizált illetve decentralizált mechanizmusok szerint különböztetik meg a normák keletkezését.

A MÁSODRENDŰ POTYAUTAS PROBLÉMA MEGOLDÁSAI

A metanorma feltételezésén nyugvó megoldás

A norma keletkezésének és stabilitásának lehetséges magyarázatára dolgozta ki szimulációs modelljét Axelrod (1986), amely számításba veszi a másodrendű potyautas problémát is. A szerző az evolúciós (tanuláselméleti) játékelméleti megközelítést alkalmazta, a korlátozott racionalitás feltevése mellett. A szimulációs modell a következőképp épül fel. Minden játékosnak két döntést kellett hoznia: kooperál vagy defektál, illetve megbünteti a defektálókat, vagy futni hagyja őket. A kifizetések úgy lettek kialakítva, hogy a defektálás kis mértékben csökkentette valamennyi játékos kifizetését, a defektáló játékos megbüntetése költséges volt (ennek költsége a defektálás okozta kárnál magasabb volt), a büntetés maga pedig mindkét értéknél jóval nagyobb kárt okozott a defektálóknak. A játékosok stratégiáját két tényező határozza meg, a defektálási hajlandóság (boldness) és a büntetési hajlandóság (vengefulness). A játékosok akkor defektáltak, ha a defektálási hajlandóságuk meghaladta a monitorozás hatékonyságát (a lebukásuk valószínűségét), a lebukott defektálók megbüntetése pedig a büntetési hajlandóságnak megfelelően történt. A sikeres stratégia kiválasztódása az utódok száma alapján történt: az egy-egy több döntéssorozatból álló kör után történt a stratégiák értékelése és a sikeresebb stratégiákat képviselő játékosok több utódot szerezhetek, amelyek növelték az adott stratégia részarányát a populációban. Az eredmények szerint az öt 100 körből álló teljes sorozatból mindössze egyben sikerült létrehozni normát, a többi négyben a stabil kimenetel a defektálás előterbe kerülése és a büntetés elgyengülése volt. Az evolúciós folyamat a következő logika szerint zajlott. A kezdő állapot az átlagos defektálás hajlam és az átlagos büntetési hajlam volt. A viszonylag hatékony büntetés következtében erősen lecsökkent a defektálási hajlam a populációban, amit az egyébként is hátrányos tulajdonságnak számító büntetési hajlandóság eltűnése követett. Végül a büntetés megszűnése újra teret adott az erősebb defektálási hajlam elterjedésének.

Ezután Axelrod (1986) módosított a játékon és bevezette a büntetést elszalasztók megbüntetését (metanorma). Az új feltételek mellett az öt teljes sorozat kimenetele már a normák létrehozását jelezték: az erős büntetési hajlandósággal rendelkező játékosok száma gyorsan növekedett, míg a defektálási hajlam összességében csökkent. Ugyanakkor a játék csak relatíve erős kezdő büntetési hajlandóság mellett vezetett pozitív eredményre, enélkül a norma ismét elbukott. Vagyis az eredmények szerint a defektálók büntetésének mechanizmusa önmagában nem áll meg, szükséges egy azt támogató metanorma.

Mivel a modellben a büntetés költséges és nem igazán térül meg (ha nagyon kevés defektáló van jelen a populációban, akkor is csak hosszú távon térül meg és akkor is magasabb kifizetéssel zárnak azok, akik nem szankcionálnak), törvényszerű ennek a

Szociológiai Szemle 2008/2.



tulajdonságnak a leépülése az evolúciós logika szerint. Ha azonban a büntetés elmulasztást is büntetjük, akkor már elvesztik az előnyüket a másodrendű defektálók a büntetőkkel szemben.

Felmerül azonban a kérdés, hogy miért nem jelenik meg a harmadrendű potyautas probléma Axelrod modelljében? Erre az a kapcsolat a magyarázat, ami a két különböző szinten megjelenő büntetési hajlandóság és monitorozási hatékonyság között van: a másodrendű büntetési hajlandóság ugyanolyan erős, mint az elsőrendű, valamint a defektálók lebukási valószínűsége is azonos a játék két szintjén. A két szinten megjelenő büntetési hajlandóság összekapcsolása nélkül, a játékosoknak lehetősége lenne a másodrendű büntetési hajlandóság csökkentésére, ami egy idő után maga után vonná az első rendű büntetési hajlandóság csökkenését is, ami a norma erózióját eredményezné. A modell ezen tulajdonságát vitatja Yamagishi és Takahashi (1994). A szerzők kidolgoztak egy alternatív modellt Axelrod (1986) alapján, amelyben nincs szükség metanormára. Ebben egy Axelrodéhoz (1986) hasonló összekapcsolást végeztek az elsőrendű dilemma és a másodrendű dilemma között. Azaz, aki az első döntésnél valamilyen valószínűséggel kooperált, az ugyanolyan valószínűséggel szankcionált a másodrendű dilemma esetében. Ezzel az n szereplős fogoly dilemma játékot átalakították egy n szereplős biztosítási játékra (*assurance game*). Míg az eredeti dilemma esetében a nem büntetés volt a domináns stratégia, addig a kooperálás és a büntetés összekapcsolása után megváltozott a helyzet, és a döntés már attól vált függővé, hogy a többség milyen döntés hoz. Vagyis volt egy kritikus kooperálási szint, amely fölött a büntetés sikeres stratégiának bizonyult. Az eredmény Axelrodéhoz (1986) hasonló volt.

Yamagishi és szerzőtársa (1994) arra a következtetésre jut, hogy racionális önérdékű cselekvőket feltételezve nem lehet megoldani a másodrendű potyautas problémát csak bizonyos módosításokkal. Járható út az olyan külső tényezők beiktatása, mint a csoport kollektív büntetése (Heckathorn 1988), de ekkor valójában egy másodrendű centrális megoldást alkalmazunk. Vagy a játék átalakítható egy n -szereplős biztosítási játékká (Yamagishi–Takahashi 1994).

Úgy tűnik több szerző is arra a következtetésre jutott, hogy a racionális önző indíték kevésnek bizonyul a normák megmagyarázásához. Elster (1995) úgy véli, hogy bár a racionális cselekvés logikája elsőbbséget élvez, a sikeres kollektív cselekvést nem lehet egyetlen mechanizmussal magyarázni, hanem többféle indíték együttes eredménye, amelybe beletartozik az önérdékű cselekvés mellett a társadalmi norma vagy altruizmus. Falk és szerzőtársai (2005) egy kísérletben vizsgálták az informális szankcionálás mögött meghúzódó motívumokat, és azt találták, hogy az önző stratégiai indíték nem játszik szerepet a büntetésben, ezzel szemben két nem stratégiai motívum igen (Falk et al. 2005). Ez azon álláspontot erősíti, hogy a társadalmi (vagyis decentralizált keletkezésű) norma alapvetően nem eredményorientált (Elster 1989) illetve nem szándékolt (Opp 2002), sokkal inkább mellékterméknek tekinthető.



Az altruista büntetés elmélete

Az altruista büntetés és az ezen alapuló erős reciprocitás⁶ elmélet az előzőek alapján abból indul ki, hogy a racionális cselekvésen alapuló modellek és az önző cselekvésre épülő evolúciós modellek önmagukban nem képesek megmagyarázni a kooperációt (Fehr–Gächter 2002). Szemben az ökonomia szempontból hagyományosnak tekinthető megközelítésekkel, amely a kooperációt mindig valamilyen közvetlen (Axelrod 1984), vagy közvetett, a jövőben teljesülő anyagi haszon megszerzésével magyarázza (reputációs hatás, Raub–Weesie 1990), az altruista büntetés elmélete a másokra irányuló cselekvés kulcsfontosságára építi modelljeit.

Az elmélet képviselői az altruista büntetés jelenségét tartják a kooperáció alapvető előmozdítójának olyan helyzetekben, amikor a nem kooperálás a racionális viselkedés. Szankcionálás hiányában – még a feltételes kooperálás híveinek jelenlétét feltételezve is – a kooperáció szintjének csökkenésével kell számolni a potyautas jelenség miatt. A büntetés, mint szelektív ösztönző megoldást jelent a kollektív cselekvés problémájára, csak hogy a büntetés – ebben az elméleti keretben feltételezve – költséges az egyének számára. A másodrendű potyautas probléma megoldása azonban lehetséges úgy, ha feltételezzük, hogy vannak olyan csoporttagok, akik önként felvállalják a büntetés költségeit. A másokat altruista módon fegyelmező játékosok számára a büntetés monetáris értelemben nem megtérülő viselkedésforma. Ennek ellenére Fehr és szerzőtársai kísérleti úton igazolták az altruista büntetést jelenlétét (Fehr et al. 2004b).

Fehr-ék egy olyan kollektív cselekvés kísérletet terveztek meg, amelyben büntetés hiányában és büntetés mellett is vizsgálhatták a csoporttagok kooperációs hajlandóságát. A kísérleti eredményei azt igazolták, hogy minél inkább eltértek a defektálók a kooperáció csoportos szintjétől, annál keményebb büntetést szenvedtek el (Fehr–Gächter 2000). Továbbá míg büntetés hiányában, ismétlődő helyzetben a kooperáció átlagos szintje fokozatosan romlott, sőt a teljes defektálás felé közelített, addig a büntetés mellett a kooperáció szintje növekvő tendenciát mutatott, és a teljes kooperáció közelébe került (Fehr–Gächter 2000, 2002).

Mindebből következik, hogy ha az altruista büntetést alkalmazók száma meghalad egy meghatározott küszöbértéket az adott csoportban, akkor a potyautazás már nem kifizetődő a racionális csoporttagok számára, és a játék kimenetele a teljes kooperáció felé mozdul el.

Fehr és Fischbacher (2004a) a „büntető harmadik fél” kísérletében azt igazolta, hogy még a kívül állók is motiválva vannak a nem kooperáló játékos megfegyelmezésében. A kísérlet egy egyszeri, nem ismételt fogoly dilemma szituációban zajlott, melyet megfigyelt egy kívül álló harmadik játékos, akinek lehetőségében állt megbüntetni a másik két játékos. Ez a kísérlet különösen releváns a nem racionális büntetés szempontjából, hiszen a kívül álló harmadik félnek semmilyen materiális haszna nem származott a kétszereplős fogoly-dilemma játékban dezertáló fél megfegyelmezéséből. Az eredmények azt mutatták, hogy a kívül álló (harmadik) fél pozíciójában lévő kísérleti alanyok közel fele meg kívánta büntetni a defektáló játékos, míg a

6 Az erős reciprocitás (*strong reciprocity*) egy olyan preferenciát vagy diszpozíciót jelent, amely a feltétlen kooperáláson és a defektálók megbüntetésén (vagy kooperálók jutalmazásán) alapul (Fehr et al. 2002).



kooperáló játékosok büntetése elhanyagolható mértékű volt. A dezertáló játékos büntetésének mértéke ráadásul még magasabb volt akkor, ha a másik játékos kooperált. A “büntető harmadik” kísérletek azt támasztják alá, hogy az erős reciprocitás lehet a társadalmi normák kikényszeríthetősége magyarázatának egyik kulcsa (Fehr et al. 2002).

Az erős reciprocitás mindenhol kimutatható, de kultúra függő is, amelyet egy több társadalmat összehasonlító ultimátum játék kísérlettel bizonyítottak: azon natív közösségek tagjai, ahol a kooperáció erősebb a mindennapi szituációkban, az ultimátum játék során többet ajánlottak fel és szigorúbb büntetést is alkalmaztak (Gintis et al. 2005).

A büntetés nem minden esetben vezet a kooperáció szintjének növekedéséhez. Ha ugyanis a büntetés a szankcionáló fél pusztán önértékét szolgálja, az a hagyományos közgazdaságtudományi megközelítés várakozásaival ellentétben másik fél dezertálásához vezethet. Ekkor ugyanis a büntetést elszenvedő játékos nem tartja „jogosnak” a szankcionálást. A kísérleti eredmények szerint a büntetésnek meg kell felelnie bizonyos méltányossági vagy morális követelményeknek (Fehr–Rockenbach 2003). Ez a jogfilozófia területén már korábban tárgyalt elképzelés (a jog belső szemlélete, Hart 1995 [1961]) a társadalmi normák területén is érvényesülni látszik.

Az altruista büntetés és az ezt a jelenséget magába foglaló erős reciprocitás magyarázattal szolgálhat az egymással rokonsági kapcsolatban nem álló emberek kooperálására. A viselkedésforma létezését a kísérleti eredmények (Fehr–Gächter 2000, 2002) mellett a hétköznapi megfigyelések is megerősítik. Ugyanakkor a közgazdaságtan és a biológia hagyományos felfogásába sorolható nem-altruista magyarázatok, mint a rokonság alapú szelekció (*kin selection*), a kölcsönös altruizmus (*reciprocal altruism*), az indirekt reciprocitás (*indirect reciprocity*) és a költséges jelzés elmélet (*costly signaling theory*)⁷ nem tudják megmagyarázni az erős reciprocitás jelenségét (Fehr et al. 2002; Burnham–Johnson 2005).

Éppen ezért többféle újabb evolúciós modellt is kidolgoztak az erős reciprocitás kialakulásának magyarázatára (Gintis 2000; Bowles–Gintis 2004; Sethi–Somanathan 2001), azonban az erős reciprocitást, mint adaptív viselkedésformát nem sikerült maradéktalanul levezetni, inkább csak a sikeres fennmaradása igazolható bizonyos feltételek bevezetése mellett.⁸ Így Gintis (2000) evolúciós modelljében a csoport kihálásának fenyegetettsége olyan sajátos körülményt teremt, amely mellett a csoportok közötti szelekciós mechanizmus az erős reciprocitás sikerességéhez vezet, ám ez csak a tisztán önértékű cselekvéssel együtt érvényesül. A probléma az, hogy mivel szankcionálni költséges, ezért az altruista büntetők átlagos kifizetése mindig a kooperáló és a racionális önértékű cselekvők kifizetése alatt marad. Ez pedig nyitva hagyja azt a kérdést, hogy hogyan magyarázható meg evolúciósan az altruista büntető viselkedés kialakulása (Sethi–Somanathan 2001).

Egy lehetséges megoldást nyújt Fowler (2005a). Modelljében a kooperálók, defektálók mellett megkülönböztet nem résztvevőket is, akik stabil kifizetéssel rendel-

7 Ezek a magyarázatok az elnevezések ellenére racionális önértékű cselekvőket feltételeznek, hiszen a befektetett költségek valamilyen módon megtérülnek, ezért beleillenek a biológia és a közgazdaságtan hagyományos felfogásába és a gyenge reciprocitás (*weak reciprocity*) gyűjtőfogalom alá soroljuk (Burnham–Johnson 2005).

8 Legalább is az újabb evolúciós modellek sikeressége erősen vitatott (Burnham–Johnson 2005).



keznek, amely az általános defektálás állapotában meghaladja a kollektív cselekvésben részt vevők kifizetését. Ezért a modell nem engedi meg az általános defektálás állapotának kialakulását, mint pl. Panchanathan és Boyd (2003) modellje, amely egyszerre teszi lehetővé a teljes defektálás és a teljes kooperálás egyensúlyi helyzetének kialakulását. Itt a magasabb kifizetéssel járó nem-resztvevőkké válás csábítóbb. Fowler (2005a) modelljében ugyan nem jön létre stabil egyensúlyi helyzet, a stratégiai ciklikusan változnak, azonban van egy kritikus pont, ahol csaknem büntetőkből áll az egész populáció. Az egyetlen stabil kimenetel a kizárólag büntetőkből álló populáció lehetne. De fontos eredmény, hogy egyetlen büntető mutáns is képes elterjedni a populációban. Csakhogy az is igaz, hogy mind Fowler (2005a), mind Gintis (2000) modellje nem pontosan az olsoni helyzetre ad választ, azt némileg módosították: Fowler modelljében nem mindenki részesül a közjószágból, Gintis-nél pedig lehetőség van a defektálók közjószágból való kizárására.

Az altruista büntetés mögött meghúzódó mechanizmusokat vizsgálták DeQuervain és szerzőtársai (2004). Kísérleti eredményeik szerint az altruista büntetés stimulálja az agy jutalomközpontját, tehát pozitív érzelmeket vált ki, amelyek megerősítik ezt a viselkedésformát. Ez egyben vita forrása is: Fehr és Gächter (2002) az altruista büntetés mögött érzelmi illetve egalitáriánus motívumokat ugyanakkor a mechanizmusnak a két aspektusának tekintették. Ugyanakkor ezt nem mindenki fogadja el. Fowler és szerzőtársai (2005) újra elemezve Fehr és Gächter (2002) kísérleti eredményeit, arra hívták fel a figyelmet, hogy a kifizetésekben lévő különbségek sokkal inkább motiválják az altruista büntetést, mint a másik játékos defektálásának a ténye, vagyis a méltányossági elvek megsértése szerintük fontosabb motívum a büntetésben, mint az érzelmek. Ezzel szemben Falk és szerzőtársai (2005) újabb kísérleti eredményeik alapján úgy foglalnak állást az informális szankcionálás mögött a defektáló fél megbüntetésének szándéka húzódik meg, és nem a kifizetésekben jelentkező különbségek korrekciója. Ebben a kísérletben háromféle méltányossági modellt⁹ hasonlítottak össze és ezek közül a „megtorlás” modell került ki győztesen.

Az eddigi eredmények szerint az altruista büntetés a normák keletkezése és fenntartása szempontjából kitüntetett szerepet kaphat. Diekmann és Voss (2003) Fehr-ék mechanizmusát használják fel 2 fázisú norma-játékukba, melyben az első fázis egy FD szituáció, a második fázisban lehetőséget kapnak a játékosok a másik megbüntetésére, ahol a másodrendű potyautas problémát a „másokra irányuló preferenciák” alkalmazása oldja fel. Úgy érvelnek, hogy racionális és önérdékű cselekvőket feltételezve a szankcionálás fenyegetése nem lenne hiteles a defektáló játékosok számára, ugyanis tudják, hogy a szankcionálás költséges. Azonban, ha altruista motívációkkal rendelkező játékosokkal kerülnének szembe a racionális önérdékű játékosok, akkor a szankcionálás fenyegetése már hiteles lenne a számukra. A méltányossági modellekre épülő büntetés fogalmak a normák fenntartására megfelelő magyarázatot nyújtanak nem ismétlődő társadalmi dilemmák esetében is (Diekmann–Voss 2003).

DeQuervain és szerzőtársai (2004) eredményei szerint a szankcionálók figyelembe veszik a szankcionálás költségét is az érzelmi megtérülés hasznához képest. Vagyis,

9 A három említett méltányossági modell: a játékosok összehasonlítják kifizetéseiket a csoport átlagos kifizetésével; a játékosok az egyenlőtlenséget akarják minimalizálni; és végül a játékosok egyszerűen meg akarják büntetni a defektáló felet (Falk et al. 2005).



ha el is fogadjuk az altruista büntetés elméletet, abból nem következik az, hogy a szankcionálás mindenhol jelen lesz tekintet nélkül annak társadalmi feltételeire. A defektálók megbüntetésének érzelmi indítéka ellensúlyozhatja a szankcionálás költségét, ugyanakkor lehetnek további költségcsökkentő tényezők is. Olyanok, amelyeket a társadalmi kapcsolatháló közvetít. Ezek a társadalmi kontroll mechanizmusok. Vagyis az altruista büntetés elmélet nem zárja ki a játékosok racionális mérlegelésének lehetőségét, csupán azt állítja, hogy a büntetés indítéka nem lehet racionális önző indíték.

Kapcsolathálón alapuló megoldások

Racionális önérdékű cselekvőket feltételezve a büntetés - tekintve, hogy felveti a másodrendű potyautas problémát - csak akkor jelenthet akár részleges megoldást a kooperációs problémára, ha vagy olcsóbb büntetni, mint kooperálni (Gardner–West 2004), vagy pedig a büntetést támogatja egy másik olyan mechanizmus, amelyik magát a kooperációt nem. Itt kap szerepet a társadalmi kapcsolatháló. A kollektív cselekvés sikeressége és a társadalmi kapcsolatháló közötti összefüggés Olson (1997 [1965]) privilegizált csoport fogalmába is belevetíthető. Olson szerint a kiscsoportok hatékonysága a közjósághoz való hozzájárulás egy csoporttagra jutó sokkal nagyobb részaránya mellett abban áll, hogy az informális szelektív ösztönzők („szelektív társadalmi ösztönzők”) csak ennél a csoportméretnél működőképesek, mivel csak ezek elég kicsik ahhoz, hogy a tagjaik közvetlen kapcsolatban álljanak egymással. Több empirikus vizsgálat is megerősíti, hogy a társadalmi kapcsolatháló elősegítheti a normák kialakulását (Hechter–Opp 2001b). Smith (1976) a közösség és a társadalom Tönnies elméletén alapuló hagyományos jellemvonásait vizsgálta és kimutatta, hogy bár a közösség/társadalom dichotómia összetettebb jelenséget takar annál, minthogy egyértelmű összefüggést mutasson a kooperáció sikerességével, azt találta, hogy a közjósághoz való hozzájárulás csökken, ha csökken a csoport sűrűsége vagy növekszik a csoporton belüli cserélődés. Természetesen nem maga a kapcsolatháló, hanem az általa hordozott ösztönzők azok, amelyek a norma kialakulásához vezetnek.

Coleman-nél (1990) a társadalmi tőke, a csoporttagok közötti társadalmi kapcsolatok egy speciális sajátossága segít érvényre juttatni a norma kialakulását és fenntartását. A norma kialakulásának első fázisa valamely negatív vagy pozitív externália következtében megjelenő normakereslet. A másodrendű potyautas jelenséget a szerző elméletében a játékosok közötti együttműködéssel oldható fel. A kapcsolatháló (a kapcsolatok zártsága) biztosítja azt a csoporttagok közötti kommunikációs csatornát, amely az együttes cselekvés összehangolását lehetővé teszi. Ez a kommunikáción kívül egyben társadalmi kontroll mechanizmust is közvetít: a csoporttagok kifejezhetik egymás felé az elvárásaikat, elköteleződéseiket, és pszichológiai értelemben jutalmazhatják vagy büntethetik is egymást. A kapcsolatok által közvetített szociális jutalmak vagy épp büntetések végül racionális önérdékű cselekvőt a kooperáció irányába mozgathatják el. Eszerint a kapcsolatháló zártsága (sűrűsége) elősegíti a kooperáció szintjének a növekedését.

A sűrűség-kooperáció hipotézis újabb elméletét képviseli Horne (2001a, 2004, 2007). A modell a csoport kohézió és a szankcionálás összefüggéséről tárgyalja, amelyben a potenciális szankcionálók és a kooperálók közötti kötések a meghatározó-



ak. E kötések erőssége határozza meg annak a támogatásnak az erősségét, amit a csoporttagok a szankcionálók számára nyújtanak, így a struktúra ezen keresztül, vagyis indirekt módon hat a szankcionálásra. Eszerint a szankcionálás gyakorisága magasabb a koherensebb csoportokban. Horne kísérleti eredményei alátámasztották a hipotézis, miszerint a kohézív csoportok egy jutalmazási mechanizmuson keresztül támogatják a szankcionálást (Horne 2001). Flache (1996) korábban a sűrűség és a kooperáció egyértelmű összefüggését kétségbevonó eredményekre jutott. Modelljében a csoporton belüli szociális cserekapcsolatok bizonyos feltételek mellett a kollektív jószággal szemben mutatott közömbösségre vezetnek. Éppen ezért hangsúlyozza Horne egy későbbi cikkében (2004), hogy a sűrűség csak abban az esetben segíti elő a norma kikényszerítését, ha a szankcionálás a defektálók büntetésén túl elősegíti a csoporttagok szociális cserében való nagyobb fokú részvételét. Vagyis a szankcionálás a potenciális szankcionálóra jutó közvetlen haszon (defektáló hozzájárulása a közjószághoz, ill. az ebből való részesedés) túl egy olyan közvetett hasznot is eredményez, amely sikeresen átváltható a szociális cserefolyamatokban, tehát ha a szociális csere és a szankcionálás haszna összekapcsolódik. A kérdés az, hogy milyen feltételek mellett állít elő a szankcionálás szociális cserébe bocsátható indirekt hasznot? Horne szerint ez függhet a norma tartalmától vagy attól, hogy az adott társadalmi szituáció tartalmaz-e egyéb kooperációt előmozdító mechanizmust. Ha például jelen van egy centralizált szelektív ösztönző, az csökkentheti a decentralizált szankcionálás indirekt hasznát (Horne 2004). Ehhez hasonló eredmény hozott egy szimulációs modell, amely szerint bár a nagyobb csoportosűrűség elvben kedvez a csoport teljesítményének, ha van kommunikáció a csoporttagok között, a modellt kiegészítve egy újabb segítő mechanizmussal, a vezető jelenlétével a két segítő mechanizmus csökkenti egymás hatékonyságát. Azaz a csoporttagok közötti kommunikáció és a vezető együttes jelenléte összességében kedvezőtlenebb hatást eredményez a csoport teljesítményére, mintha csak az egyik mechanizmus működne (Gigliotta et al. 2007).

A társadalmi tőke szempontjából vitatott, hogy a kapcsolatok zártsága vagy épp a strukturális lyukak játszanak inkább szerepet (Burt 2001), ugyanakkor úgy tűnik, hogy a normák keletkezése és fenntartása szempontjából a zártság kedvezőbb. Janky és Takács (2005) eredményei szerint a társadalmi kapcsolatháló töredezettsége aláássa a kooperációt. Ostrom (1990) a közjavak előállítására irányuló decentralizált szankcionálás (és monitorozás) működése szempontjából a csoport stabilitásának és a gyakori kommunikációnak a feltételeit hangsúlyozza. Elképzelhető, hogy a strukturális lyukak akkor töltenének be fontosabb szerepet, ha a kollektív cselekvésben nem résztvevő csoporttagokat néhány evolúciós modellhez hasonlóan (Bowles–Gintis 2004; Gintis 2000; Fowler 2005a) ki lehetne zárni a közjószágból. Ekkor valószínűleg néhány kisebb zárt alcsoportokban (ahol valamilyen okból a nagyobb kooperációs hajlammal rendelkező csoporttagok aránya magasabb) kialakulhatna a norma, amely, mint sikeres „minta”, a hídkötéseken keresztül tudna elterjedni.

Mindez azt valószínűsíti, hogy a társadalmi normák keletkezése és fenntartása elsősorban a kisközösségekben és a zártabb csoportokban valószínűsíthető, és a tagolt hálózatokkal rendelkező urbanizált világban az alulról építkező normák kialakulása és fenntartása külső beavatkozás nélkül problematikus.



IRODALOM

- Axelrod, R. (1984): *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R. (1986): An Evolutionary Approach to Norms. *American Political Science Review*, 80(4): 1095–1111.
- Baumgartner, T.–Buckley, W.–Burns, T.R. (1975): Meta-power and Relational Control is Social Life. *Social Science Information*, 14: 49–78.
- Binmore, K. (1998): *Just Playing: Game Theory and the Social Contract*. II., Cambridge (M.A.): MIT Press.
- Boyd, R.–Richerson, P.J. (1988): The Evolution of Reciprocity in Sizable Groups. *Journal of Theoretical Biology*, 132(3)Jun: 337–356.
- Boyd, R.–Richerson, P. J. (2005): Solving the Puzzle of Human Cooperation. In Levinson, S.C.–Jaisson, P. eds.: *Evolution and Culture*. MIT Press, 105–132.
- Bowles, S.–Gintis, H. (2004): The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Theoretical Population Biology*, 65: 17–28.
- Burnham, T. C. – Johnson, D. D. P. (2005): The Biological and Evolutionary Logic of Human Cooperation. *Analyse & Kritik*, 27(1): 113–135.
- Burt, R.S. (2001): Structural Holes Versus Network Closure as Social Capital. In Lin, N.–Cook, K.S.–Burt, R.S. eds.: *Social Capital: Theory and Research*. Aldine Transaction.
- Camerer, C.F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Coleman, J.S. (1990): *The Foundations of Social Theory*. Cambridge (Mass.): The Belknap Press.
- DeQuervain, D.–Fischbacher, U.–Treyer, V.–Schellhammer, M.–Schnyder, U.–Buck, A.–Fehr, E. (2004): The Neural Basis of Altruistic Punishment. *Science*, 305: 1254–1258.
- Diekmann, A.–Voss, Th. (2003): Social Norms and Reciprocity. Working Paper, .
- Egas, M.–Riedl, A. (2005): The Economics of Altruistic Punishment and the Demise of Cooperation. Tinbergen Institute Discussion Papers, [_www1.fee.uva.nl/creed/people/riedl/econpunish.pdf](http://www1.fee.uva.nl/creed/people/riedl/econpunish.pdf).
- Elster, J. (1989): *The cement of society*. Cambridge University Press.
- Elster, J. (1995): *A társadalom fogaskerekei*. Budapest: Osiris-Századvég.
- Falk, A.–Fehr, E.–Fischbacher, U. (2003): On the Nature of Fair Behavior. *Economic Inquiry*, 41(1): 20–26.
- Falk, A.–Fehr, E.–Fischbacher, U. (2005): Driving Forces behind Informal Sanctions. *Econometrica*, 73(6): 2017–2030.
- Falk, A.–Fehr, E.–Fischbacher, U. (2005): Driving Forces behind Informal Sanction. *Econometrica*, 73(6): 2017–2030.
- Fehr, E.–Fischbacher, U. (2004a): Third-party Punishment and Social Norms. *Evolution and Social Behavior*, 25, 63–87.
- Fehr, E.–Fischbacher, U. (2004b): Social Norms and Human Cooperation. *TRENDS in Cognitive Sciences*, 8(4)April: 185–190.
- Fehr, E.–Fischbacher, U.–Gächter, S. (2002): Strong Reciprocity, Human Cooperation and the Enforcement of Social Norm. *Human Nature*, 13: 1–25.
- Fehr, E.–Gächter, S. (2000): Altruistic Punishment in Public Goods Experiments. *American Economic Review*, 90: 980–994.
- Fehr, E.–Gächter, S. (2002): Altruistic Punishment in Humans. *Nature*, 415(Jan): 137–140.
- Fehr, E.–Rockenbach, B. (2003): Detrimental Effects of Sanctions on Human Altruism. *Nature*, 422(March): 137–140.
- Fehr, E.–Schmidt, K. (2001): Theories of Fairness and Reciprocity – Evidence and Economic Applications. Institute for Empirical Research in Economics – IEW Working Papers, iewwp075, .

Szociológiai Szemle 2008/2.



- Fischbacher, U.–Gächter, S.–Fehr, E. (2001): Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economics Letters*, 71: 397–404.
- Flache, A. (1996): *The Double Edge of Networks*. Amsterdam: Thesis Publishers.
- Flentge, F.–Polani, D.–Uthmann, Th. (2001): Modelling the Emergence of possession Norms Using Memes. *Journal of Artificial Societies and Social Simulation*, 4(4), .
- Fowler, J.H. (2005a): Altruistic Punishment and the Origin of Cooperation. *Proceedings of the National Academy of Sciences*, 102(19)May: 7047–7049.
- Fowler, J.H. (2005b): Human Cooperation: Second-order Free-riding Problem Solved? *Nature*, 437(September): E8.
- Fowler, J.H.–Johnson, T.–Smirnov, O. (2005): Egalitarian Motive and Altruistic Punishment. *Nature*, 433(January): E1–E2.
- Frohlich, N.–Hunt, Th.–Oppenheimer, J.–Wagner, R.H. (1975): Individual Contribution for Collective Goods. *Journal of Conflict Revolution*, 19(2)June: 310–330.
- Fudenberg, D.–Maskin, E. (1986): The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3)May: 533–554.
- Gál R.I. (1989): A hetedik parancsolat: Tanulmány a társadalmi normák erőzójáról. In Gál R.I. szerk.: *Racionális választások – irracionális következmények*. Budapest: Marx Károly Közgazdaságtudományi Egyetem Szociológiai Tanszék.
- Gardner, A.–West, S.A. (2004): Cooperation and Punishment, Especially in Humans. *American Naturalist*, 164(6)December.
- Gibbs, J.P. (1965): Norms: The Problem of Definition and Classification. *American Journal of Sociology*, 70: 586–94.
- Gigliotta, O.–Miglino, O.–Parisi, D. (2007): Groups of Agents with Leader. *Journal of Artificial Societies and Social Simulation*, 10(4), .
- Gintis, H. (2000): Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206: 169–179.
- Gintis, H.–Smith, E.A.–Bowles, S. (2001): Costly Signal and Cooperation. *Journal of Theoretical Biology*, 213: 103–119.
- Hardin, R. (1971): Collective Action as an Agreeable N-prisoners' Dilemma. *Behavioral Science*, 16(5): 472–481.
- Hardin, R. (1980): The Emergence of norms, *Ethics*, 90(Jul): 575–587.
- Hart, H.L.A. (1995 [1961]): *A jog fogalma*. Budapest: Osiris Kiadó.
- Hechter, M. (1984): When Actors Comply: Monitoring Costs and the Production of Social Order. *Acta Sociologica*, 27: 161–183.
- Hechter, M.–Opp, K.D. (2001a): Introduction. In Opp, K.D.–Hechter, M. eds.: *Social Norms*. New York, Russell Sage Foundation.
- Hechter, M.–Opp, K.D. (2001b): What Have We Learned about the Emergence of Social Norms? In Opp, K.D.–Hechter, M. eds.: *Social Norms*. New York, Russell Sage Foundation.
- Heckathorn, D.D. (1989): Collective Action and the Second-order Free-rider Problem. *Rationality and Society*, 1: 78–100.
- Heckathorn, D.D. (1988): Collective Sanctions and the Emergence of Prisoner's Dilemma Norms. *American Journal of Sociology*, 94: 535–562.
- Horne, C. (2001a): The Enforcement of Norms: Group Cohesion and Meta-norms. *Social-Psychology Quarterly*, 64(3): 253–266.
- Horne, Ch. (2001b): Sex and Sanctioning: Evaluating Two Theories of Norm Emergence. In Opp, K.D.–Hechter, M. eds.: *Social Norms*. New York: Russell Sage Foundation.
- Horne, C. (2004): Collective Benefits, Exchange Interests, and Norm Enforcement. *Social Forces*, 82(3): 1037–1062.
- Horne, C. (2007): Explaining norm Enforcement. *Rationality and Society*, 19(2): 139–170.

Szociológiai Szemle 2008/2.



- Janky B. (2005): *Szolidaritás és jóléti preferenciák*. Budapest: Akadémiai Kiadó, Philosophiae Doctores.
- Janky, B.–Takács, K. (2005): Social Control, Participation in Collective Action and Network Stability. In Bodó B. és társai: *Társadalmi térben*. [In Social Space.] Budapest: BMGE Szociológia és Kommunikáció Tsz, 157–188; műhelytanulmányként HUNNET Online Working Paper, 2003, .
- Kuran 1997
- Marwell, G.–Oliver, P. (1993): *The Critical Mass in the Collective Action*. Cambridge: Cambridge University Press.
- Nowak, M.A.–Sigmund, K. (2005): Evolution of Indirect Reciprocity. *Nature*, 437(27)Oktober: 1291–1298.
- Oliver, P. (1980): Rewards and Punishment as Selective Incentives for Collective Action: Theoretical Investigation. *American Journal of Sociology*, 85(6): 1356–1375.
- Oliver, P.–Marwell, G. (1988): The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II., *American Sociological Review*, 53(February): 1–8.
- Olson, M.J. (1997 [1965]): *A kollektív cselekvés logikája*. Budapest: Osiris.
- Opp, K.D. (2001): Social Networks and the Emergence of Protest Norms. In Opp, K.D.–Hechter, M. eds.: *Social Norms*. New York: Russell Sage Foundation.
- Opp, K.D. 2002. When do Norms Emerge by Human Design and When by the Unintended Consequence of Human Action? The Example of the No-smoking Norm. *Rationality and Society*, 14(2):131–158.
- Orthmayr, I. (2004): A társadalmi normák döntésméleti és evolúciós magyarázata. *Szociológiai Szemle*, 3.
- Ostrom, E. (1990): *Governing the Commons. The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Panchanathan, K.–Boyd, R. (2003): A Tale of Two Defectors: the Importance of Standing for Evolution of Indirect Reciprocity. *Journal of Theoretical Biology*, 224: 115–126.
- Panchanathan, K.–Boyd, R. (2004): Indirect Reciprocity Can Stabilize Cooperation without the Second-Order Free Ride Problem. *Nature*, 432: 25.
- Panchanathan, K.–Boyd, R. (2005): Human Cooperation: Second-order Free-riding Problem Solved? *Nature*, 437: September, E8–9.
- Radelet, M.L. (1979): Introduction: The Ethnographic Study of Social Control. *Journal of Contemporary Ethnography*, 8: 267–273.
- Rasmusen, E. (1992): Folk Theorems for the Observable Implications of Repeated Games. *Theory and Decision*, 32: 147–164.
- Raub, W.–Weesie, J. (1990): Reputation and Efficiency in Social Interaction: An Example of Network Effects. *American Journal of Sociology*, 96(November): 626–54.
- Saam, N.J.–Harrer, A. (1999): Simulating Norms, Social Inequality, and Functional Change in Artificial Societies. *Journal of Artificial Societies and Social Simulation*, 2(1): .
- Samuelson, P. (1954): The Pure Theory of Public Expenditure. *Review of Economics and Statistics*, 36(November): 387–389.
- Schelling, Th. C.(1998 [1978]): Termosztátok, tragacsok és más modelleszaladók. In Csontos L. vál.: *A racionális döntések elmélete*. Budapest: Osiris Kiadó–Láthatatlan Kollégium.
- Sethi, R.–Somanathan, E. (2001): Norm Compliance and Strong Reciprocity. Discussion Paper, Columbia University: Department of Economics.
- Smith, E.A. (2003): Human Cooperation: Perspectives from Behavioral Ecology. In Hammerstein, P. ed.: *Genetic and Cultural Evolution Of Cooperation*. MIT Press.
- Smith, E.A.–Bliege Bird, R.L. (2000): Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signal. *Evolution and Human Behavior*, 21: 245–261.



- Smith, J. (1976): Communities, Association and the Supply of Collective Goods. *American Journal of Sociology*, 82(2): 291–308.
- Szántó Z. (1999): *A társadalmi cselekvés mechanizmusai*. Budapest: AULA Kiadó.
- Ulmann-Margalit, E. (1977): *The Emergence of Norms*. Oxford: Clarendon Press.
- Yamagishi, T.–Sato, K. (1986): Motivational Bases of the Public Goods Problem. *Journal of Personality and Social Psychology*, 50: 67–73.
- Yamagishi, T.–Takahashi, N. (1994): Evolution of Norms without Metanorms. In Schulz, U.–Albers, W.–Mueller, U. eds.: *Social Dilemmas and Cooperation*. Berlin: Springer-Verlag, 311–326.

