

# AZ „ARANYSZABÁLYOK” EVOLÚCIÓJA

Marosán György

filozófiai tudományok kandidátusa, főiskolai tanár  
Budapesti Gazdasági Főiskola Külkereskedelmi Kar  
marosan.gyorgy@mail.datanet.hu

Az „aranyszabályok” – a közösségek elvárt viselkedési normáinak alapelemei – fontos szerepet töltenek be életünkben. Tudatosan vagy öntudatlanul, ezek a szabályok vezérlik többnyire cserére vonatkozó döntéseinket. Márpedig az életben szinte mindenhez csere útján jutunk. Adjuk, ami a mienk – pénzt, árut, segítséget, kedvességet, tiszteletet vagy éppen szerelmet – azért, hogy hozzájussunk ahhoz, aminek hiányát érezzük. A cserét szabályozó elvek jelentőségét az adja, hogy alapvetően meghatározzák a megszereshető „örömjavak” mennyiségét. Vannak szabályok, amelyek erősítik a bizalmat, és ezért ösztönzik az együttműködést. Ám vannak olyanok is, amelyek bizalmatlanságot szülnek, így lerombolják a kapcsolatot. A kutatások egyértelműen igazolták: a társadalmak sikeressége „intézményeiktől” – a közösséget összefűző stabil, folyamatosan újratemelő és együttműködést szolgáló viszonyrendszerektől – függ. (Marosán, 2005a). A magunk formálta intézmények – és ezek alapjaként az „aranyszabályok” – segítik elő egyes társadalmak fejlődését, vagy ellenkezőleg, tartanak szegénységben másokat. (Hibbs et al., 2004)

Az „aranyszabályoknak” többféle változata létezik. Legtöbbször a Hegyi beszédben elhangzó krisztusi tanácsot: „Csak azt tedd felebarátoddal, amit kívánnál, hogy ő veled tegyen” nevezik ennek. Néha így hivatkoznak a Krisztusnál egy generációval fiatalabb neves zsidó rabbi, Hillél intésére: „Amit nem

akarsz, hogy veled cselekedjenek, azt ne tedd felebarátoddal sem”, és gyakran ugyanígy utalnak az ötszáz évvel korábban élt kínai bölcs, Konfuciusz elvére is: „Tégy úgy másokkal, amint ők tesznek veled”. De vajon honnan származnak ezek a szabályok? Az egyik lehetséges válasz: Isten – mindenkinek saját hite szerinti Istene – alkotta meg, és írta elő számunkra. S valóban, a legtöbb vallás hirdeti követésre méltóként javallt „parancsolatokat”, amelyek évezredek óta szolgálják az emberek együttélését és a közösségek fejlődését. A szabályok isteni eredete erősíti „legitimításukat”, s ezzel hatékonyságukat. Ám van egy másik ok, ami miatt elkerülhetetlennek tűnt Istenre hivatkozni. Nehezen elképzelhető, hogy az önérdek rángatta és gyakran végig nem gondolt cselekedetek káoszából megszülethessen a közösséget összetartó és az érdekek folyamatos összehangolódását segítő viselkedési modell. De vajon tényleg nem szülhet-e a káosz – intelligens „tervező” nélkül – szabadságot?

A társadalmi együttélés – szabályok vezérelte – apróbb-nagyobb cseréi, közvetlen szükségletkielégítő hatásuk mellett, egyben e szabályok versenyét is jelentik. Miközben az emberek a lehetséges szabályból kiválasztják azt, ami alapján berendezik életüket, akarva-akaratlanul a szabályok evolúcióját valósítják meg. A mindennapok során kiválasztódnak azok a maximák, amelyek támogatják hirdetőjük rátermetségét, míg kisselektálódnak és eltűnnek az embe-

rek életéből – azok, amelyek nem segítik hirdetőjük túlélőképességét. Ám van itt egy különös ellentmondás: a közhiedelem úgy tartja, hogy a becsületes, a bizalomteljes, és jószándékú viselkedés eleve hátrányban van az agresz-szív, önző, a másokat kihasználó viselkedéssel szemben. A kérdés tehát az, vajon lehet-e kísérletileg ellenőrizni, milyen a valós élet körülményei között a legsikeresebb viselkedési stratégia?

### *A foglyok dilemmája*

A *foglydilemma* néven elhíresült szituáció a játékelmélet közkezdvelt és napjainkra közismertté vált modellje. Két eltérő érdekű partner együttműködésre vagy ellenkezőleg, a másik kijátszására irányuló döntéseinek lehetséges kimeneteit elemzi, és keresi a legelőnyösebb megoldását. (Mérő, 1996. 3. és 4. fejezet) Az együttműködő vagy önző lépések kölcsönös következményeit az alábbi kifizetési mátrix mutatja (1. táblázat).

A valóságos élet szemszögéből nagyobb jelentőségre tett szert az ún. ismétlődéses foglydilemma-játék (IFDJ). Ez nem egyetlen döntés következményeit, hanem meghatározott döntési stratégiát követő „partnerek” több meneten keresztül folytatott játékának kimeneteit és felhalmozódó pontokban kifejeződő eredményességét elemzi. A közvélemény figyelmét egy negyedszázada lezajlott különös verseny irányította a modellre. Egy fiatal politológus – Robert Axelrod – arra kérte ismerőseit, hogy a foglydilemma több meneten keresztül történő játszására vezérfonalául szolgáló stratégiát fogalmazzanak meg. (Axelrod, 1990) A verseny résztvevőinek le kellett írniuk, milyen „maximák” szerint játszanák ismétlődően ezt a

játékot. A beküldött és többnyire szavakban megfogalmazott stratégiákat – stratégián a továbbiakban a döntéseket vezérlő szabályokat értjük – Axelrod számítógépes programokká írta át, majd e programokat 200 meneten keresztül játszatta egymással. A játék minden menetében a programok – éppen úgy, mint az emberek – „döntéseket hoztak”, majd a saját és a „játstótárs” lépéseitől függően a kifizetési mátrix alapján pontokat kaptak. A pontok halmozódtak, így a játék végén kirajzolódott, melyik stratégia bizonyult a legsikeresebbnek. Axelrod később azzal tette még izgalmasabbá a játékot, hogy folyamatosan kihagyta a legkevesebb pontot elért programokat. A játékban így egyre „versenyképesebb” programok vetélkedtek, és feltételezhető volt, hogy amelyik az „evolúciós versenyt” végül megnyeri, az a legrátermettebb.

A „nyertes” végül a kanadai politológus, Anatol Rapoport által beküldött, és Tit-for-tat-nak (TFT) – „szemet, szemért” – nevezett program lett. Ez mindössze két sorból állt: (1) induláskor előlegezd meg a bizalmat, (2) majd tedd mindig azt, amit partnered az előző lépésben („tükrözd” a viselkedését). A TFT-ről kiderült, hogy nem hagyományos értelemben mindenképpen győzelemre törekvő versenyző. Inkább köszönhetne annak sikerét, hogy partnereit minél több meneten keresztül való együttműködésre készítette.

Az IFDJ az elmúlt csaknem fél évszázadban a társadalomtudományok rendkívül népszerű modelljévé vált. Sokféle elrendezésben, különböző körülmények között, és meghökkenően eltérő „aktorok” – emberek (Mérő, 1996), állatok (Dugatkin, 1997), és számítógépek (Szabó – Tőke, 1998) – rész-

*A partnerek viselkedése*

Önző

Együttműködő

Önző

(-2, -2)

(-4, +4)

Együttműködő

(+4, -4)

(+2, +2)

1. táblázat

vételével játszották. Én magam is már több mint tíz éve rendszeresen játszom – *Piros-Kék játék* címmel – különböző összetételű csoportokban, és főként főiskolai hallgatókkal. Ennek során tanúja lehettem annak, ahogyan a résztvevők – a próba-szerencse módszerrel – keresik, és gyakran rátalálnak a sikeres viselkedés stratégiájára. A játék szokásos körülményei azonban nem tették lehetővé az eredmények kiértékelését. Ezért pontosan megtervezett kísérletsorozatot végeztem, hogy ellenőrizhető körülmények közt választ kapjak a következő kérdésekre:

- Rendelkeznek-e a résztvevők azonosítható stratégiával?
- Tükröződik-e ez a stratégia a játék során meghozott döntéseikben?
- Változtatnak-e a résztvevők az induló stratégián a játék közben szerzett tapasztalataik függvényében?
- Melyik lesz a nyerő (a legsikeresebb) stratégia?
- Levonják-e a résztvevők akár a siker, akár a kudarc tanulságait?

#### *A kísérlet leírása*

A kísérletben a BGF Külkereskedelmi Főiskola, az Általános Vállalkozási Főiskola és a Zsigmond Király Főiskola összesen 68 hallgatója – 19-22 éves fiatalok, 39 lány és 29 fiú – vett részt, 12-16 fős csoportokban (Marosán, 2005b). A párokba osztott játékosoknak – írásban megadott szabályok szerint – 30 meneten keresztül kellett döntéseket hozni. Minden menetben – életfelfogásuk, a játékban szerzett tapasztalataik alapján, valamint az *1. táblázatban* megadott kifizetési mátrix figyelembe vételével – P-t vagy K-t írnak a döntési lapra. E döntéstől függően pontokat kaptak, amely a játék során felhalmozódott. A játék célja – ezt a leírásban kiemeltük – „*minél több pontot szerezni*”.

A játék során – hacsak ezt kifejezetten nem tettük lehetővé – a párok nem kommunikálhattak egymással. A játék első szakaszá-

ban – az 1-10. menet között – a 4., a 6., és a 9. menet után a partnerek megbeszélhették, kölcsönösen milyen döntéseket hozzanak. Jeleztük azonban: „tárgyalni lehet, de nem kötelező. Meg lehet állapodni bármiben, de a megállapodást nem kötelező betartani”. A második szakaszban – a 10-20. menet között – a 12., a 15. és a 18. menetet követően lehetőség nyílt a „partnercserére”. Aki kívánta, a „válás” helyett tárgyalhatott. A harmadik szakaszban – a 20-30. menet között, – a 22., a 24., a 26. és a 28. meneteket követően, az addig a legkevesebb pontot elért két játékos „kiesett” a játékból. A partner nélkül maradt játékosokat újra összesorsoltuk, és a játék egészen a 30. menetig folytatódott. A játék közben a kifizetési mátrix pontjainak értéke növekedett: az első rész 2 (vagy 4) pontja a második részben 3-at (6-ot), a harmadik, befejező részben pedig 4 (8) pontot „ért”.

A kísérlet közben egy táblán látható volt a résztvevők minden döntése, az ezért kapott és az összesen elért pont. Ez az elrendezés a többiekkel párhuzamosan, de tőlük – elvileg – függetlenül játszó és az eligazítás szerint csak egymással törődő párokból versengő játékteret teremtett. A résztvevők tanúi lehettek a bizalom megelőlegezése, az együttműködés megszakítása, a megegyezés felrúgása, a tárgyalás visszautasítása és a mindenáron való győzelem diktálta csalások hatásának. Ez az elrendezés alapvetően eltért a számítógépes programokkal lefolytatott kísérletektől. Ez utóbbi előnye, hogy a „résztvevők” (a programok) stratégiája egyértelműen beállítható, a játszma minden lépése pontosan nyomon követhető, a körülmények tökéletesen „kézben tarthatók”, a menetszám és a résztvevők száma jelentősen növelhető. Hátránya viszont, hogy feltételezései – a „játékosok” pontosan meghatározott stratégiát követnek, az nem is változik, és az érzelmeknek nincs szerepük, – az életben nem állják meg a helyüket.

A „hús-vér” résztvevőre építő játék elő-

nye, hogy a körülmények életszerűek. A játékosok kevert összetételű – és a valóságos társadalmat többé-kevésbé tükröző – stratégiát követő populációt alkotnak. Az egyes résztvevők viselkedése is „homályos” stratégiával írható le. A környezet véletlen hatásai – például a félreértések – befolyásolják a résztvevők döntéseit. Az, hogy mindenki eredménye látható, teret nyit az érzelmeknek, de lehetővé teszi a tanulást. Ugyanakkor az élő személyekkel folytatott játéknak komoly hátrányai is vannak. A résztvevők száma és a végigjátszható menetek száma szükségképpen korlátozott. A játékosok megváltoztathatják induló stratégiájukat, és döntéseikben szerepet kapnak az érzelmek. (Amikor például a megállapodást a partner nem tartotta be, kifejezetten dühödtt reagálásokat váltott ki.) Mindez megnehezítheti az eredmények pontos számszerűsítését. Ám ezekért a hátrányokért – úgy véljük – kárpótol, hogy a kísérlet elrendezése nyomon követhetővé tette a résztvevők által követett stratégiák eloszlásának tendenciaszerű elmozdulását és az adott körülmények között legsikeresebb stratégia azonosítását.

### *A résztvevők beállítódása*

A résztvevők – a játék során követett stratégiáját meghatározó – beállítódását a kísérlet megkezdése előtt kérdőívvel azonosítottam. A játék befejezését követően a változást hasonló szerkezetű kérdőívvel követtem nyomon. A viszonylag egyszerű kérdőív a résztvevők döntéseit leginkább befolyásoló nyolc értékre – a bizalom, a versenyszellem,

az együttműködési készség, a sikerorientáció, hajlandóság mások kihasználására, mások érdekeinek tekintetbe vétele, az igazságosság, a kölcsönösség – vonatkozó véleményét tárta fel.<sup>1</sup> Ennek alapján felrajzolható volt a résztvevők által preferált viselkedés eloszlásának spektruma, illetve ennek változása a játék folyamán. (lásd *1. táblázat*)

A hétköznapi gondolkodás – néha a kutató is – általában két alapvetően ellentétes beállítódást ismer: a bizalomra építő *együttműködőt* és a bizalmatlan, csaló, *önzőt* (Boyd et al., 2003). Az élet tipikus helyzetei azonban összetettebb és változatosabb viselkedésre utalnak. Az emberekkel és a számítógépes programokkal végrehajtott játékokban több, részben eltérő stratégiát azonosítottak. Ilyenek a „pavlovi” stratégia – „ha nyertél, folytasd, ha veszítettél, válts” –, az „elnéző” TFT – „a csalásra ne válaszolj csalással azonnal, de egyébként kövesd a TFT-t”, a „bizalmatlan csaló” – „használd ki a másikat, amíg lehet, majd állj tovább” stb. Egy legutóbbi, emberi résztvevőkkel végrehajtott kísérlet háromféle, egymástól lényegileg eltérő beállítódást azonosított: együttműködőt, óvatossággal együttműködőt és a versengő nem együttműködőt. (Kurzban, et al., 2005).

A 68 kérdőív adatait elemezve úgy találtam, hogy a résztvevők négy eltérő beállítódással jellemezhető csoportba oszthatók:

1. Elnéző Tit-for-Tat (ETFT – 108–130 pontig). Életelve: „*Jótt helyébe, jót várj*” Bizalmat megelőlegező, feltétlen együttműködő, a partner csalását botlásnak tekintő, azt több meneten keresztül elnéző, a másik

<sup>1</sup> A kérdőív a 8 érték esetén 3 lehetséges választ kínál az adott értékre vonatkozóan: a „teljes bizalom”, a „feltételes együttműködés” és „versengő csaló” megközelítést, amelyek között kellett a válaszadónak 10 pontot szétosztania. A kiértékelés során az egyes válaszlehetőségek pontjait még megszoroztam 1-gyel, 2-vel vagy 3-mal, a kérdés „jólelkűségének”, „bizalmatlanságának” és „önzőségének” mértékétől függően. A válaszok alapján a résztvevők egy konkrét pontszámot kaptak. Ez a pontszám kijelölte helyüket a feltétlen bizalom

– önző versengés skálán. A skála minimuma – *Teréz anyja* – 80 pontjától az önző és csaló *Gordon Gekko* 240 pontos elméleti maximumáig terjedt. A játék résztvevői esetén a minimum 108 pont volt, míg a maximum 218. Az általam használt skála 100–220-ig terjedt; ezt osztottam fel hatpontos egységekben egy 20 osztású skálára. Így felrajzolható volt induláskor, majd a játék végén a kérdőív alapján „mért” beállítódás eloszlása a bizalom–versengés skálán, és elemezhetővé vált a skálán való elmozdulás. Lásd *1. ábra*.

javát kereső, a közösségbe befektető, és azt értékek tekintő.

2. Tit-for-Tat (TFT – 131–160 pontig) Életelve: „*Kölcsönkenyér visszajár*”

Saját érdek érvényesítésére törekvő, de a kölcsönösséget elfogadó, feltételekhez kötöten együttműködő, bizalmat („remegő kézzel”) megelőlegező, ám azt csalódás esetén azonnal visszavonó, csalást megbüntetni kész, a közösséghez hozzájáruló.

3. Bizalmatlan Tit-for-Tat (BTFT – 161–190 pontig). Életelve: „*Amilyen az adjonisten, olyan a fogadjisten*”

Bizalmatlan és induláskor kockázatminimalizáló (gyakran P-val induló!), a partner várható viselkedésében bizonytalan, feltételeken együttműködő, kölcsönösséget megkövetelő, a saját érdeke rövid távú érvényesítése által vezérelt, gyenge szolidaritású, közösségi hozzájárulását hamar visszavonó.

4. Önző versengő (ÖV – 191–220 pontig) Életelve: „*Szemesnek áll a világ*”

Az életet versenynek tekintő, nyeresre játszó, a győzelem érdekében a csalást (hítségést) is vállaló, bizalmatlan, ám ha érdekében áll, együttműködő, de a szabályokat saját érdekeinek megfelelően értelmező, a bizalmat tanúsító partnert és a közösséget könyörtelenül kihasználó, de a vele szembeni csalást megtorló.

Feltételeztem, hogy a játék résztvevői – elmentében a számítógépes programokkal – a pillanatnyi helyzettől, múltbeli tapasztalataiktól és a partnerek viselkedésétől függően átlagos beállítódásuk körüli „stratégia-mix”-ből választanak. Döntéseiket alapvetően befolyásolják a környezeti hatások. Mivel láthaták egymást, tárgyalásokat is folytathattak és tanúi lehettek a megegyezések felrúgásának, a résztvevők viselkedése gyakran eltért a kérdőívre adott válaszaiktól. A játék hevében gyakran kiestek választott szerepükből.

### *A stratégiaverseny eredménye*

A kérdőívek elemzése alapján a 2. táblázatban látható kép alakult ki a résztvevők viselkedési stratégiáját illetően a játék kezdete előtt és azt követően:

A legfontosabb változás: a kísérlet végére csökkent az ÖV-t és az ETFT-t választók száma. A vélemények a TFT felé „gravitáltak”. Ezt tanúsítja a választott viselkedési stratégiák eloszlásának az induláshoz képesti módosulása is. (1. ábra) Az induló kérdőív átlaga 12,5, szórása 4,2, míg a záró átlaga 11,06, szórása pedig 3,98.<sup>2</sup> Csökkent a „szélsőséges” véleményűek száma, miközben a nézetek szórása is csökkent. A szórás csökkenése annak a jele, hogy a szélsőséges – feltétel nélküli bizalom vagy a versengő önző – viselkedési stratégiák visszaszorultak.

Hagyományosan a stratégiaverseny sikerkritériumának a pontversenybeli helyezést tekintik. A versenyt a két ETFT nyerte, őket két TFT követte. Valójában azonban többféle, az eredményeket más és más oldalról bemutató sikerkritériumot célszerű használni. Jól jellemzi a sikerességet a „nyertesek” száma.<sup>3</sup> A stratégiaverseny végeredményét a 3. táblázat foglalja össze. Ez a viselkedési stratégiák – induló és a záró kérdőívekben feltáruló – helyezési számait, illetve annak változását mutatják az első tízbe, az első húszba és végül az „állva maradt” 32-be kerültek esetén. A számok közvetlenül érzékeltetik az eltérő viselkedési stratégiák „evolúciós sikerességét”.

A táblázat arról tanúskodik, hogy a TFT típusú stratégiák kerültek ki „győztesen” a versenyből. A játék végére, és eredményeként a többség felismerte és el is fogadta a TFT típusú viselkedési stratégia sikerét.

<sup>2</sup> Az átlag és a szórás az 1. táblázat versengés mértékét jelző 1–20-as skálájú vízszintes tengelyére, s a függőleges tengelyen mért gyakoriságeloszlásra vonatkozik.

<sup>3</sup> A nyertesek száma = 1–10-be kerültek száma az induló és a záró kérdőív szerint.

A választott viselkedési stratégia	Követők száma az induló kérdőív szerint	Követők száma a záró kérdőív szerint
Elnéző Tit-For-Tat (ETFT)	5 (7,4%)	3 (4,4%)
Tit-For-Tat (TFT)	15 (22,1%)	26 (38,2%)
Bizalmatlan Tit-For-Tat (BTFT)	27 (39,7%)	28 (41,2%)
Önző versengő (ÖV)	21 (30,8%)	11 (16,1%)

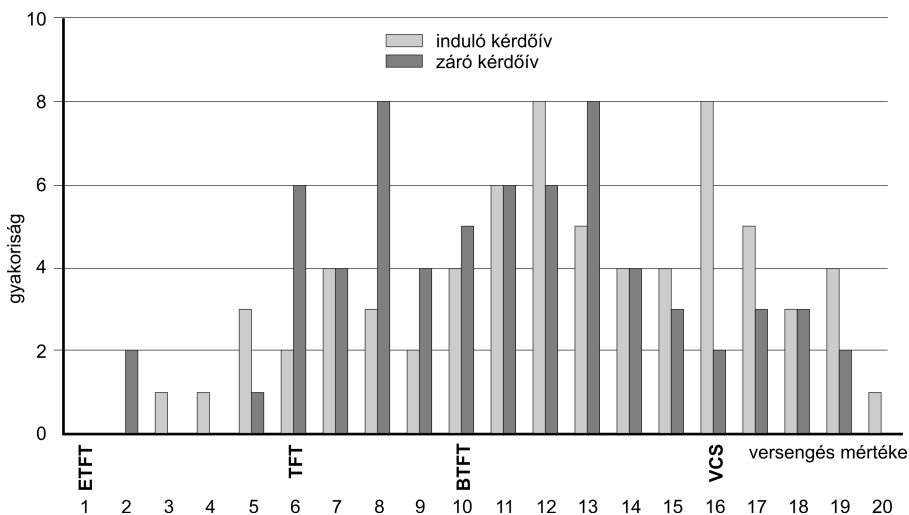
2. táblázat

	Első10-ben		Első 20-ban		Állva maradt 32-ben	
	Induló kérdőív	Záró kérdőív	Induló kérdőív	Záró kérdőív	Induló kérdőív	Záró kérdőív
ETFT	2	2	2	2	3	2
TFT	3	5	7	8	11	14
BTFT	4	3	8	8	10	13
ÖV	1	0	3	2	9	5

3. táblázat

A sikeresség mutatói	ETFT	TFT	BTFT	ÖV
Nyertesek (első 10-be kerültek) száma	2 (i), 2 (z)	3 (i) 5 (z)	4 (i) 3 (z)	1 (i) 0 (z)
Átlagos helyezési szám	0,4	0,9	0,5	0,24
Elért átlagpontszám	84	92	80	76

4. táblázat



1. ábra

Induló kérdőív	Záró kérdőív ETFT (3) (kap)	TFT (26) (kap)	BTFT (28) (kap)	ÖV (11) (kap)
ETFT (5) (ad)	<b>2</b>	2	1	0
TFT (15) (ad)	1	<b>9</b>	5	0
BTFT (27) (ad)	0	11	<b>14</b>	2
ÖV (21) (ad)	0	4	8	<b>9</b>

5. táblázat

A sikeresség további fontos szempontja –miként az olimpiákon számolt pontversenyek – a versenyben részt vevő stratégiák által elért átlagos helyezési szám.<sup>4</sup> További lényeges információt szolgáltat az egyes stratégiák sikerességéről az azt követők által elért átlagpontoszám.<sup>5</sup> Ez jól tükrözi egy adott stratégia „rátermettségénél” függően – hisz a verseny kinyilvánított célja minél több pontot szerezni. A verseny eredménye e sikemutatók alapján a 4. táblázatban látható:

Az adatok értelmezésénél tekintetbe kell venni, hogy a játékból a 21–30. menet során a játékosok nagyjából fele, az addig követett stratégiája sikerességétől függően – kiszekelődött. Ez értelemszerűen csökkentette az ETFT-k, az ÖV-k és részben a BTFT-k átlagpontoszámaikat és helyezési számaikat. Ám mivel a szelekció éppen a stratégiák átlagos „rátermettségétől” függő mértékben „ritkította” meg a „populációt”, ez az adatokból levonható következtetést inkább megerősíti. Valamennyi sikerkritérium a TFT sikerét mutatja.

Az elért eredmények részletes elemzése – a játék első szakaszának nyomonkövetése, a „nyertesek” és „lemaradtak” életpályájának összevetése – jellegzetes képet fest az „életfelfogásbeli” különbségeikről. Indulásakor a 68 résztvevő közül 44-en léptek P-t és 24-en K-t. Az első húsz helyezett közül

azonban csak heten indultak P-vel és 13-an K-val. Az első tárgyalás után – ahol szinte mindig K-ban állapodtak meg – 51 résztvevő írt K-t, ám a következő lépésben már csak 35- tartott ki a K mellett. Az első húszba bekerült résztvevők a tárgyalás után 18-an írtak K-t, és a következő három lépésben sem csökkent ez 16 alá.

A játék eredményeként végbemenő átrendeződések részleteit a transzformációs mátrix (5. táblázat) mutatja. Ez az induló kérdőív szerinti választásokat és a játék hatására bekövetkező „átlépések” részleteit, valamint a záró kérdőív szerint eredményeket mutatja.

Jól látható, hogy a játék során jelentősen nőtt a TFT-t követők száma, és a növekedés döntő forrása a BTFT-t és ÖV-t követők voltak. Így a játék bizonyította, hogy – ellentétben a hétköznapi logikával – a játékosok többsége (örömmel vagy kénytelen-kelletlen) felismerte a TFT (típusú) stratégia hatékonyságát. Ez azt mutatja, hogy a valóságos életet tükröző körülmények között – bizonytalan helyzet, ismeretlen partner, a becsapás kockázata, a döntést befolyásoló zaj – a TFT típusú viselkedés a legsikeresebb.

Csökkent az ETFT-t követők száma: ötről kettőre. Az a három játékos, akiknek tapasztalatai alapvetően negatívak voltak, megtanulták: amíg potenciális játszótársaik versengők és önzők, csak „kipróbált és tesztelt” partnerrel célszerű a feltétel nélküli bizalom stratégiáját követni. Az ETFT, a csalást megbocsátható botlásnak tekintő és azt több menetben keresztül elnéző – „ha

<sup>4</sup> Átlagos helyezési szám = a versenyben állva maradt 32-ben elért helyezések száma / az ezt a stratégiát követők száma

<sup>5</sup> Az átlagpontoszám = az egyes stratégiát követők által elért összes pontoszám / az ezt követők száma. Ez a szám az adott stratégia átlagos „rátermettségére” utal.

megdobnak kövel, dobd vissza kenyérral” elveként érvényesülő – életmodellje a valós élet körülményei között ön- és közveszélyes. Az önző csalót, a partner sorozatos K-ja – a kísérlet tanulsága szerint – nem „téríti meg”. A megbocsátás – a P-kre adott feltétel nélküli K – inkább megerősíti a „potyautast” stratégiájának helyességében, semmint stratégiájának megváltoztatására ösztönöznék.

A fentieknek nem mond ellent, hogy az „egyéni” pontversenyt két ETFT nyerte. Ez a két személy, akiknek „bejött” a stratégiájuk, megerősödtek véleményükben. „Aki rátalált a megfelelő partnerre – fogalmazott egyikük – az ne kísérletezzen csalással és partnercserével, hanem legyen boldog partnerével, és újra és újra erősítse meg ragaszkodását”. Mindez arra utal, hogy a valóságos életben nincs minden körülmények között hatékony stratégia. Viselkedésünket a menedzsment „kontingencia”-elméletének megfelelően a körülményekhez, s főként a partnerek stratégiájához célszerű illeszteni.

A BTFT-k száma lényegében nem változott. Egy részük belátva, hogy a tartós és jó kapcsolat érdekében induláskor nagyobb bizalmat célszerű tanúsítani, inkább együttműködőként kell viselkedni, és a megállapodását kevésbé szabad felrúgni, csatlakozott a TFT-hoz. A „jól megválasztott” P – mint az a bizonyos makorenkői pofon – azt üzeni: „térj észre, neked is érdeked az együttműködés”. A BTFT-től a TFT-hez „átigazoltakat” a ÖV-k közül ide „átpártoltak” helyettesítették. Ők a játék során belátták, hogy hosszú távon a szerzett pontszám csak a kölcsönösség tudomásul vételével és az együttműködést megalapozó szabályok betartásával növelhető.

A ÖV-k száma jelentősen csökkent. Sokan közülük – éppen a saját és a többiek eredményét látva – a játék menetében maguk változtattak. Meghatározó szerepe volt tanulásukban – a partner „büntető” P-jei mellett – a játék közbeni tárgyalásnak. Ezeket a partnereket világosan elmondták egymásnak, mit

kívánnak a másiktól, értésére adva, mit nyerhet ezzel, és mit veszíthet, ha visszautasítja. A ÖV átlagos viselkedési modellje az adott feltételek között, hosszú távon kudarcra vezetőnek bizonyult. A ÖV-k több mint fele ezért váltott: „átállt” leginkább a BTFT-k, kisebb részben a TFT-k közé. Többen azonban mindvégig azt keresték, hogy – szó szerint idézve egyikük megfogalmazását – „vajon hol követtem el a hibát, milyen trükkkel lehetett volna mégis nyerni?”

### *Az arany szabály születése*

A kísérlet eredményei kirajzolják, hogy a szabályok szelektációjával miként erősödik meg a közösség együttélését leginkább segítő norma (Orthmayr, 2004). A norma kialakulására utal a populációviselkedés „változatosságának” csökkenése, ami a közösség által követett norma egyneműbbé válását mutatja. Az, hogy a többség a TFT típusú viselkedés mellett döntött, úgy értelmezhető, hogy a játék eredményeképpen a bizalmatlan, önző, csalásra hajlamos emberek közöségéből kiválasztódik és fokozatosan meghatározóvá válik a TFT típusú viselkedési stratégia. Ez az eltolódás alapvetően fontos a közösségi „rátermettség” szempontjából, hiszen a TFT generálja a legnagyobb egyéni és „csapat”-pontokat. Ez úgy is értelmezhető, hogy az ilyen „viselkedésű” egyedekből álló közösségben az egyéni és a közösségi „vagyon” gyorsabban nő, mint a ÖV-kből álló közösségé. Némi túlzással, a Tőzsdecápák Gordon GeKKo-ja kénytelen-kelletlen „Teréz anyja” jellegű viselkedést vesz fel, és fordítva, „Teréz anyja” is ráébred az önérdek fontosságára. Mindenfajta „felsőbb” beavatkozás nélkül – isteni eredetű (és legitimitású) etikai elvek hiányában is – fokozatosan kiformalódhat tehát az együttműködő és bizalmat megelőlegező viselkedési stratégia (Bendor et al., 2001, 1495).

Az induló és a záró kérdőív elemzése arra utal, hogy az „életstratégiákkal” kapcsolatos



vélemények érzékelhetően megváltoztak. Egyrészt az 1. táblázat tanulsága szerint az ETFT–ÖV skálán a vélemények „középre húztak”. Ez azt jelenti, hogy a „szélsőségesen bizalomteli” („Teréz anya”) és a „szélsőségesen önző-versengő” („Gordon Gekko”) viselkedés a „némi bizalmatlansággal vegyes” TFT-hez közeledett. A két „szélsőséges” viselkedés volt a játék igazi „vesztése”. Kiderült, sem a feltétel nélküli bizalom, sem a következményekkel nem számoló csalás nem hatékony viselkedési stratégia napjaink körülményei között, átlagos partnerekkel szemben. A résztvevők véleménye és viselkedése a játék során, s annak hatására, közelebb kerül a magas szintű bizalmat és együttműködési készséget kifejező origóhoz.

A kiértékelések során vetődött fel egy további izgalmas kérdés: vajon melyik stratégia „tesz” a legtöbbet a „közösség boldogulásáért”? Többféle tényező készítheti a résztvevőket viselkedésük módosítására. Először, a táblán látható eredményeik demonstrációs hatást követve hány pontot érhetünk el. Másodsor, a partnerek tárgyalásaik során érvelve vagy nyomást gyakorolva – esetleg a válással fenyegetve – hathatnak egymás viselkedésére. Harmadsor, az adott stratégiát követő játékos a játék során folyamatosan jutalmazza vagy bünteti partnerét, s így minden lépésben az addig követett stratégia újragondolására készíti őt.

A „közösség hőse” megtisztelő cím adományozásánál – nem tévesztve szem elől, hogy az egyéni siker és a közösségi „vagyon-

generálási képesség” környezet- és partnerfüggő – két szempontra célszerű figyelni. Egyrészt, számításba vehetjük a párosok versenyében elért sikereket.<sup>6</sup> Ezt az egyetlen ETFT páros nyerte, ám a párosok versenyének élmezőnyében a TFT-k kerültek többségbe. E mellett vizsgálni lehet, *ki teszi a legtöbbet* a közösséget károsító ÖV típusú viselkedés visszaszorításáért. Itt a demonstrációs hatás és a tárgyalások során kifejtett nyomás mellett a legfontosabb tényező a konkrét döntések – a partner P-jére vagy K-jára adott és jutalmazásként vagy büntetésként értelmezhető válaszok – gyakorlati nevelő hatása (Bienenstock et al., 2001). A viszonylag kis gyakoriság miatt óvatosan kell bánni az eredménnyel, mégis úgy tűnik, hogy a TFT típusú stratégiák „nevelnek” a leghatékonyabban. Mindez azt tanúsítja, hogy BTFT–ÖV környezetben – a kísérlet, és általánosabban, a mai Magyarország feltételei között – a TFT nemcsak egyénileg a legsikeresebb, hanem ez a stratégia tesz egyben a legtöbbet a közösség „boldogságáért”. A BTFT–ÖV típusú társadalom inkább „nevelhető” a realista TFT figyelmeztető P-jeivel (és jutalmazó K-jaival), mint az ETFT jószándékú és elnéző („ha megdobtak követ, dobd vissza kenyérrrel”) típusú reakcióival, de akár a ÖV megtorló P sorozatával.<sup>7</sup> Úgy tűnik tehát: nemcsak az egyéni, hanem a közösségi boldogság nélkülözhetetlen tartozéka a „büntetni”, sőt – mint egy másik kísérletben feltárult – a másokat (a közösséget) ért inzultust saját káron is megtorló, ún. altruista büntetést végrehajtani kész aktor. (Fowler, 2005, 7048.)

<sup>6</sup> A kísérlet eredetileg egyéni versenyre épült. Értelmezhető azonban a párok egyesített pontszáma is. A TFT típusú párok mindvégigke viszonylag magas pontot ér el, így „csapatpontszámuk” ennek megfelelően magasak. Ezzel szemben az ÖV-ETFT párok esetén az egyik magas, a másik alacsony egyéni pontszámot gyűjtött, így a „csapatpontszámuk” kicsi lett.

<sup>7</sup> A TFT által végrehajtott „nevelési” folyamat meghökkenően emlékeztet a „zsetongzdaság” (token economy) néven ismert módszerre. A zsetongzdaság

meghatározása szerint olyan viselkedésmódosító eljárás, amely az egyén viselkedését követően azonnal jutalmazással vagy büntetéssel „igazolja vissza” a viselkedés kívánatos vagy elkerülendő voltát. A viselkedéshez kapcsolt azonnali (a tettet és következményét „rövidre záró”) jutalmazó, illetve büntető ingerek viselkedésmódosító hatására utal Hans és Michael Eysenck *Elmevadászat* című könyvében leírt zsetongzdaság típusú, börtönbéli nevelési kísérletben. (Eysenck – Eysenck, 2002, 364–365.)

*Befejezés helyett*

A kísérlet alapfeltételei megítélésünk szerint jól modellezték életünk hétköznapi viszonyait. A résztvevők között voltak együttműködők és önzők, eltérő volt a megelőlegezett bizalom szintje és a csalásra való hajlamosság, a körülmények és pillanatnyi élmények hol erre, hol arra billentették a döntéshozót. A populáció egyedei tehát kevert stratégiát követtek. A játék folyamán kapott pontjaik alapján közvetlenül érzékelték követett stratégiájuk sikerességét és kudarcát. A kísérlet során nyert tapasztalataik hatására a többség változtatott eredeti stratégiáján. Hajlandók voltak tanulni. Ilyen általános feltételek esetén a legsikeresebb – legtöbb pontot gyűjtő – egyéni stratégiának a TFT bizonyult. A TFT azonban nem csupán az egyéni, hanem a közösségi siker leghatékonyabb „kovácsa”. Ez a típusú viselkedés rákényszeríti a bizalmatlan és nem kooperatív résztvevőket, hogy bizalomteljes és együttműködő stratégiára térjenek át. Emiatt a TFT fokozatosan

elterjedt a közösségben. Mindez arra utal, hogy a hétköznapi emberek saját élettapasztalataik alapján ráébrednek Kant kategorikus imperatívuszára – „cselekedj úgy, hogy cselekedeted maximája egyúttal egyetemes törvényhozás elvéül szolgálhasson”. Ennek alapján – minden „fentről” jövő sugalmazás nélkül – arany szabállyá emelik a TFT-t. Ezzel a Hobbes-i „mindenki harca mindenki ellen” kaotikus állapotát az egyéni tapasztalás, majd a társadalmi tanulás, végül „arany szabályok” megalkotása és kulturális átörökítése révén kiszámítható és gyorsan halmozódó emberi tőkét társadalmakká formálja. (Vogel, 2006). A történelemtudomány és az antropológia pedig végül már csak konstatálja: a TFT maximája által vezérelt egyénekből álló közösségek fejlődnek, míg akik nem ezt követik, a fejlődés perifériájára szorulnak.

Kulcsszavak: *evolúció, arany szabály, foglyadilemma, tit-for-tat, versengés, bizalom, önzés, szelekció, kategorikus imperatívusz, zsetongazdaság*

**IRODALOMJEGYZÉK**

- Axelrod, Robert (1990): *The Evolution of Cooperation*. Penguin Books, London
- Bienstock, Elisa Jayne – McBridge, Michael (2004): Explication of the Cultural Transmission Model. *American Sociological Review*. 69, 138–143.
- Bendor, Jonathan – Swistak, Piotr (2001): The Evolution of Noms. *American Journal of Sociology*. 106, 6, 1493–1545.
- Boyd, Robert – Gintis, H. – Bowles, S. – Richerson, P. J. (2003): The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Sciences of the USA*. 100, 3532.
- Dugatkin, Lee Alan (1997): *Cooperation among Animals: An Evolutionary Perspective*. Oxford University Press, Princeton
- Eysenck, Hans – Eysenck, Michael (2002): *Elmevadászat*. Kairosz, Budapest
- Fowler, James H. (2005): Altruistic Punishment and the Origin of Cooperation. *Proceedings of the National Academy of Sciences of the USA*. 102, 7047–7049.
- Hendrich, Joseph (2006): Cooperation, Punishment, and the Evolution of Human Institutions. *Science*. 312, 5770, 60.
- Hibbs, Douglas A. – Olsson, Olov (2004): Geography, Biogeography, and Why Some Countries Are Rich, and Others Are Poor. *Proceedings of the National Academy of Sciences of the USA*. 101, 3715–3720.
- Kurzban, Robert – Houser, Daniel (2005): Experiments Investigating Cooperative Types in Humans. *Proceedings of the National Academy of Sciences of the USA*. 102, 1803–1807.
- Marosán György. (2005a): *Hogyan készül a történelem?* Money-Plan, Budapest
- Marosán György. (2005b): *A bizalom és az együttműködés evolúciója*. Tudományos Évkönyv. Budapesti Gazdasági Főiskola, Budapest, 114–129.
- Mérő László. (1996): *Mindenki másként egyforma. (A játékelmélet és a racionalitás pszichológiája)* Tertium, Budapest
- Orthmayr Imre (2004): A társadalmi normák döntéseméleti és evolúciós magyarázata. *Szociológiai Szemle*. 3.
- Szabó György – Töke Csaba (1998): Evolutionary Prisoners's Dilemma Game on a Square Lattice. *Physical Review E*. 58, 1, 69.
- Vogel, Gretchen (2004): The Evolution of the Golden Rule. *Science*. 303, 1129.