

# A POLITIKA MINT A FOGOLYDILEMMA MEGHALADÁSA

Tóth I. János  
(Szegedi Tudományegyetem)

## ÖSSZEFOGLALÓ

---

A nem kooperatív játékelmélet feltételei mellett a két és sokszemélyes fogolydilemmában szükségszerű a kölcsönös dezertálás, vagyis a Pareto inferior jellegű egyensúlyi állapot. A politikai filozófia terminológiáját használva a természeti állapotot a közrossz jellemzi, s ez a közösség szempontjából elfogadhatatlan. Joggal írja Elster (1976, 248.), hogy „[A] politikatudomány annak a vizsgálata, hogy milyen módon lehet meghaladni a fogolydilemmát.” Számos szerző (Olson, 1965; Ostrom, 1990; Boda, 2013) szintén azt hangsúlyozza, hogy a politika lényegében nem más, mint a kollektív cselekvés kudarcainak az elkerülése. A kooperatív játékelmélet segítségével rámutatott arra, hogy a fogolydilemmában sem spontán együttműködésre, sem társadalmi szerződésre nincs mód. Az evolúciós játékelmélet csak speciális esetekben ad lehetőséget a feltételes együttműködés spontán kialakulására. Emellett az evolúció lehetőségét ad a fogolydilemma „külső” meghaladására új típusú ágensek és a hozzájuk kapcsolódó új típusú intézmények kialakulása révén. A politikai filozófia terminológiáját használva a dolgozat azt a kérdést vizsgálja, hogy egy rossz természeti állapotból az evolúciós folyamatok eredményeképp hogyan jutunk el az egyre jobb társadalmi állapotba. Érvelésem szerint az új együttműködési megoldásoknak is súlyos fogyatékoságai vannak, amelyek újabb és újabb válaszokat igényelnek.

---

Kulcsszavak: játékelmélet ■ potyautas ■ erőszak ■ altruizmus ■ közőj, közrossz

---

## A KLASSZIKUS FOGOLYDILEMMA

A játékelmélet olyan interakciókkal foglalkozik, ahol a végeredményt (kimenetet) alapvetően az emberek (játékosok) döntései (stratégiái) határozzák meg. A játékelmélet implicit módon feltételezi a játékosok (i) formális egyenlőségét, sőt matematikai azonosságát, (ii) döntési szabadságát vagyis szuverenitását, (iii) individuális függetlenségét és elkülönültségét, illetve (iv) individuális racionalitását, azaz önérdekkövetését. Mindez azt jelenti, hogy az önérdekkövető és egymástól független játékosok interakciója mentes minden hatalmi struktúrától. A politikafilozófia terminológiája szerint ez megfeleltethető a természeti állapotnak.

A játékelmélet sokfajta interakciót ismer, amelyek sokféleképp osztályozhatók. Vannak olyan interakciók, ahol részben vagy teljesen érvényesül a láthatatlan kéz. Ekkor a játékelmélet terminológiája szerint részben vagy teljesen egybe esik az egyensúlyi és a Pareto-féle optimális állapot. A fentiekkel szemben a fogolydilemma az egyetlen olyan szimmetrikus interakció, ahol létezik egy egyetlen Pareto inferior jellegű egyensúlyi kimenetel. Ezekben a döntési helyzetekben az önérdék Platt (1973) szavaival élve láthatatlan ökölként lesújt a közjóra.

Mielőtt megvizsgálnánk a klasszikus fogolydilemmát, három módszertani kérdésre szeretnék reflektálni. (a) Valamit (esetünkben a kölcsönös dezertálást) meghaladni, azaz tagadni nagyon sokféleképpen lehet. A terjedelmi korlátok miatt ennek a sokféleségnek csak néhány alapvető elemét vizsgálom. (b) Az elemzés az együttműködés problémáját utilitarista szempontból vizsgálja és a felmerülő etikai és emberjogi szempontokat is csak ennek alárendelten elemzi. (c) A fogolydilemma egy játékelméleti *modell*, amelynek meghaladása is csak egy (meta-játékelméleti) modellt jelent, amelyet „kívülről” elemzek. Ezzel szemben a valóságos politikai folyamatokban résztvevő ágensek számára csak „belső” nézőpont létezik. Világos, hogy egy modell külső elemzése és megoldása sokkal egyszerűbb, mint a valóságos interakciók belső megfigyelése és értelmezése. Vegyünk például olyan fogalmakat, mint a közjó, a közrossz vagy a kooperálás stb. A fogolydilemma modelljének keretei között ezeknek a fogalmaknak a jelentése tiszta és világos, ezzel szemben a valóságos politikai folyamatokban jelentésükről és jelölésükről végtelen vita folyik (Tóth 2014/a). Sőt, már az is problémát jelent, hogy magát a fogolydilemma helyzeteket azonosítsuk a valóságban. Ridley (2011: 64.) szerint „Nagy általánosságban valószínűleg minden olyan helyzet fogolydilemma, amelyben csábítást érzünk valamit megtenni, ugyanakkor tudjuk, hogy komoly hiba lenne, ha mindenki ugyanezt tenné.”

A klasszikus fogolydilemma történetét Albert Tucker fogalmazta meg (McCain, 2004: 9).<sup>1</sup> Játékelméleti szempontból ez egy kétszemélyes, kétstratégias (C és D), egy lépéses, teljes információjú, nem kooperatív interakció, ahol a felek a döntéseiket egyidejűleg, de egymástól függetlenül hozzák meg. A játékosok szempontjából a fogolydilemma interakciónak csökkenő nyereségek sorrendjében a következő négy kimenetele lehetséges: egyoldalú dezertálás (DC = T); kölcsönös kooperálás (CC = R); kölcsönös dezertálás (DD = P); egyoldalú kooperálás (CD = S).<sup>2</sup> A legkisebb pozitív számok, amelyek a fogolydilemmát kielégítik a következők: T = 5; R = 3; P = 1; S = 0. Az előnyök és hátrányok megkülönböztetése érdekében a továbbiakban pozitív és negatív egész számokat használok, úgy mint: T = 3; R = 2; P = -1; S = -2. Felteszem, hogy a veszteséges egységek (egyedek, közösségek) viselkedési formái mint életképtelenek kiszelektálódnak.

A fogolydilemma definíciójából következik, hogy a kölcsönös együttműködés jelenti a közjót, mivel ekkor maximális a játékosok közössége által elérhető nyereség ( $2R > T + S$ ). A szokásos értékek mellett a kölcsönös dezertálás kimenetele jelenti a(z elsődleges) közrosszat, mivel ekkor a legkisebb a kollektív nyereség értéke ( $2R; T + S > 2P$ ). Ez nem minden fogolydilemmára igaz, de én a továbbiakban csak ezzel a lehetőséggel foglalkozom.<sup>3</sup>

Ebben az interakcióban a nem kooperatív játékelmélet szerint a játékosoknak dezertálnia (D) kell, ugyanis a dezertálás nemcsak az egyensúlyi, hanem a domináns stratégia is.<sup>4</sup> Tehát aki dezertál, az mindenképpen jobban jár, mint az, aki kooperál. A játékosoknak tehát egy domináns (önérdekkövető) és egy nem domináns (nem önérdekkövető) stratégia között kell választania. Tautologikus igazság, hogy egy önérdekkövető játékosok az önérdekkövető stratégiájukat választják.

A döntési helyzet szimmetrikus, azaz hasonló okok miatt a másik gyanúsított is dezertálni fog. A kölcsönös dezertálás (DD) az interakció egyensúlypontja, amely egyben egy Pareto-inferior kimenetelt ( $P = -1, P = -1$ ), vagyis a közrosszat jelenti. A fentiekből következik, hogy egy önérdekkövető játékos számára mind a dezertálás, mind a kooperálás mellett vannak érvek. Ennek ellenére, a nem kooperatív játékelmélet szerint egy játékos akkor viselkedik racionálisan a fogolydilemmában, ha dezertál. Ugyanis a „válassz domináns, illetve egyensúlyi stratégiát” elve rendkívül erős szempontot jelent, s mint ilyen az önérdekkövető játékosok esetében felülírja az egyéb megfontolásokat.

Pszichológiai szempontból a játékosoknak két okuk is van a másik elárulására: a maximális nyereség ( $T = +3$  egység) reménye; illetve a minimális veszteségtől ( $S = -2$  egység) való félelem. Mindkét motiváció a dezertálásra ösztönöz. A Hobbes által definiált természeti állapotban is kulcsszerepet játszik az önzés és a félelem motivációja. Gauthier (1969) elsőként kötötte össze a fogolydilemmát és a hobbesi természeti állapot fogalmát. Itt jegyzem meg, hogy a közösségi választások elméletében általában úgy értékeli hobbes koncepcióját, hogy a kölcsönös dezertálás (DDD ... D) természeti állapotából egy társadalmi szerződés vezet át a kölcsönös kooperálás (CCC ... C) társadalmi állapotába. Tanulmányomban amellet érvelek, hogy ez a folyamat nem egy egyszeri, absztrakt és univerzális jellegű társadalmi szerződés által történik meg, hanem sok, konkrét és parciális jellegű folyamat kapcsán.

## A SOKSZEMÉLYES FOGOLYDILEMMA FORMÁI

A kétstratégias és kétszemélyes fogolydilemmánál politikai szempontból sokkal fontosabbak a sokstratégias és sokszemélyes komplex fogolydilemma helyzetek. Például egy háromstratégias fogolydilemmát úgy lehet elképzelni, mint amelyet egy dezertáló (D), egy mérsékelt ( $C_1$ ) és egy radikális ( $C_2$ ) kooperáló

stratégia jellemez. A tényleges együttműködési lehetőségek sokfélesége ( $C_1$ ,  $C_1$  vagy  $C_2$ ,  $C_2$ ) pedig még az egyébként kooperáló személyek esetében is konfliktus forrása lehet (lásd később).

Nemcsak a stratégiák számának, hanem a személyek számának a növekedése is sok különböző esetet eredményez. Például egy háromszemélyes játék alapvetően kétféleképp képzelhető el: három páros (1&2, 1&3, 2&3) illetve egyetlen háromszemélyes (1&2&3) interakcióként. Nevezzük az előbbit *páros*, míg az utóbbit *homogén* interakciónak. Az ötszemélyes játékok esetében is különbséget tehetünk a 10 páros (1&2, 1&3, 1&4 ...) illetve az egyetlen ötszemélyes homogén interakció (1&2&3&4&5) között. Sőt, e két véglet között is lehetnek „köztes” kombinációk, úgy mint egy háromszemélyes és egy kétszemélyes interakció ((1&2&3)&(4&5)). Az utóbbi játékot *strukturált* interakciónak nevezem. A páros, homogén és strukturált interakciók sajátosságai jelentős mértékben különböznek egymástól. A továbbiakban korlátozzuk a figyelmünket a páros és a homogén sokszemélyes fogolydilemma interakciókra. Az előbbinek első elemzője Axelrod volt, míg az utóbbi helyzethez olyan jól ismert interakciók tartoznak, mint a közlegelő tragédiája, potyautas vagy a közjó játék bizonyos formái.

Egy *homogén háromszereplős fogolydilemma* esetében mindhárom játékos egyetlen interakcióban van, mindenki csak egyetlen döntést hoz, ez a döntés azonban befolyásolja a másik két játékos kimenetelét is, ahogy a másik két játékos döntése is visszahat az első játékos eredményére. Egy ilyen játék úgy is interpretálható, mintha (virtuális értelemben) minden játékos külön játszana egy páros interakciót a másik két játékosal.<sup>5</sup> Ebben az interakcióban összesen  $2^3 = 8$  különböző kimenetel lehetséges. Ha eltekintünk az individuális különbségektől, vagyis pusztán csak a kimenetek kollektív vagy makro mintázatát nézzük, akkor az elvileg lehetséges 8 kimenetel tartalmilag 4 különböző kimenetelre egyszerűsíthető. Úgy mint senki sem dezertál (CCC), csak egyetlen játékos dezertál (DCC, CDC és CCD), két játékos dezertál (DDC, DCD, CDD), illetve mindhárom játékos dezertál (DDD); ezen kimenetekhez tartozó kollektív nyereségek rendre pedig 6, 4, 0 és -6 pont.<sup>6</sup>

A fentieket általánosítva elmondható, hogy egy  $N$ -személyes homogén fogolydilemmában kombinatorikai szempontból  $2^N$  kimenetel lehet, ebből kollektív szempontból  $N + 1$  lehetőség a fontos (0, 1, 2 ...  $N$  játékos dezertál). Azon belül pedig a négy legfontosabb kimenetel a következő:

- $N$  játékos dezertál és 0 játékos kooperál; ebben az esetben a dezertálók nyeresége alacsony [ $E_D=(N-1) \times P$ ] és a kollektív nyereség minimális, vagyis a társaság elszenvedti az elsődleges közrosszat, [ $E=E_D \times N$ ]. Ez játékelméleti szempontból a valószínű és stabil, míg társadalmi szempontból elfogadhatatlanul rossz kimenetelt jelent.
- $N-1$  játékos dezertál és 1 játékos kooperál; ebben az esetben a dezertáló játékosos nyeresége alacsony [ $E_D=1 \times T + (N-2) \times P$ ], az egyetlen kooperá-

ló játékos, a balek nyeresége pedig minimális [ $E_C=(N-1)\times S$ ], a kollektív nyereség pedig majdnem minimális [ $E = E_D\times 1 + E_C\times(N-1)$ ]. Ez játékelméleti szempontból valószínűtlen és instabil, illetve társadalmi szempontból is rossz és kerülendő kimenetelt jelent.

- 1 játékos dezertál és  $N-1$  játékos kooperál; ebben az esetben a dezertáló játékos nyeresége maximális [ $E_D=(N-1)\times T$ ], míg egy kooperáló játékos nyeresége is relatíve magas [ $E_C = 1\times S + (N-2)\times R$ ], a kollektív nyereség pedig majdnem maximális [ $E = E_D\times 1 + E_C\times(N-1)$ ]. Játékelméleti szempontból ez egy instabil kimenetel, a közös nyereség viszont majdnem maximális. Ezt a helyzetet a később bevezetendő úr-rabszolga viszony stabilizálja, ami a politikai hatalom első formája.
- 0 játékos dezertál és  $N$  játékos kooperál; ebben az esetben a kooperáló játékosok nyeresége nagy [ $E_C = (N-1)\times R$ ], és a kollektív nyereség maximális, vagyis a kollektíva eléri a közjót [ $E = E_C\times N$ ]. Ez egy játékelméleti, sőt politikai szempontból is instabil, de kívánatos kimenetelt jelent.

A többszemélyes homogén fogolydilemma interakciók logikai szempontból megegyeznek a potyautas interakciójával. A potyautas problémájának három fontos tulajdonsága van (Schelling, 1973 vagy Mclean, 1987: 145.). (a) A dezertáló mindig jobban jár, mint a kooperáló játékos, bármennyien is kooperálnak. Tehát a dezertáló stratégia szigorúan dominálja a kooperatív stratégiát. (b) Ha mindenki kooperál, akkor mindenki jobban jár, mintha mindenki dezertálna. (c) Minél többen kooperálnak, azaz minél nagyobb az  $N_C$  értéke, annál nagyobb a kooperálók ( $E_C$ ) és a dezertálók nyeresége ( $E_D$ ), sőt a kollektív nyereség ( $E$ ) értéke is. Tehát  $E$  értéke akkor a legkisebb (elsődleges közrossz), ha senki sem kooperál és akkor a legnagyobb (közjó), ha mindenki kooperál.

Ezek a sajátosságok is jól mutatják, hogy a potyautas esetében (vagyis a sokszemélyes homogén fogolydilemmában) ellentétes motivációk vannak jelen: az (a) pont a dezertálásra, míg a (b) és (c) pont a kooperálásra ösztönzi a játékosokat. Mindamellet a nem kooperatív játékelmélet előfeltevései mellett a *tipikus* önérdékkövető ember viselkedését az (a) pont, vagyis a dominancia-elv határozza meg. *Tehát a nem kooperatív játékelmélet keretei között a tipikus játékos dezertálni fog, s ezért szükségszerű az elsődleges (dezertálásból származó) közrossz.* Ugyanakkor már ilyen feltételek között is előfordulhat, hogy egy *bölcs* (de önérdékkövető) aktor nemcsak felismeri az együttműködés előnyét, hanem gyakorolja is azt.

#### A FOGOLYDILEMMA „BELSŐ” MEGHALADÁSA

A nem kooperatív játékelmélet mellett létezik a kooperatív és az evolúciós játékelméletet is. A következő lépésben vizsgáljuk meg, hogy a standard játékelméletnek ez a két másik megközelítési módja lehetőséget ad-e a közjó elérésére vagy legalább a közrossz elkerülésére.

(*Kooperatív játékelmélet*) Mi történik akkor, ha a játékosok kommunikálhatnak egymással? Megváltozik-e akkor a stratégiai viselkedésük? Meg tudnak-e ekkor egyezni a kölcsönösen előnyös (CCC...) kimenetelben?

A kooperatív játékelmélet feltételezi, hogy a játékosok a kommunikáció eredményeképpen képesek kötelező erejű megállapodásokat (szerződéseket) kötni (Zagare, 2006: 100.). A játékosok csak akkor kötnek meg egy szerződést, ha az érdekükben áll. Ez az érdek egyaránt vonatkozik a szerződésben foglalt kimenetelre, a szerződés betartására és a többiek viselkedéséből származó kockázatokra. A kooperatív játékelmélet terminológiája szerint ezeknek az interakciónak van magjuk (core). A mag a nem dominált kimenetek halmaza (Zagare, 2006: 107.), vagyis az olyan kimenetek összessége, amelynek a realizálásában minden játékos érdekelt. Egy játékos csak akkor azonosul a kollektív érdekekkel, ha ezzel szembefordulva nem nyerhet többet (Aumann, 1961; Scarf, 1967).<sup>7</sup>

Sajnos a fogolydilemma szituációban a szabad, független és önző játékosok nem fognak tudni egyezséget kötni. Hiába van konszenzus abban, hogy a kölcsönös együttműködés ( $CC = R$ ) mindkettőjük számára előnyös, ha a játékosok érdekeltek a megállapodás megszegésében ( $DC = T$ ) és mindegyik retteg a másik szerződésszegésétől ( $CD = S$ ). A kölcsönös együttműködés egy olyan instabil Pareto-optimális kimenetelt jelent, amelynek üres a magja. Ezért a kooperatív játékelmélet által tételezett játékosok *nem fognak megállapodni*, s így elkerülhetetlen a dezertálásból származó közrossz.

Tehát a kooperatív játékelmélet szerint a fogolydilemma keretei között még a „spontán” együttműködés sem lehetséges, és persze a *társadalmi szerződés* sem realizálható. A társadalmi szerződés egy olyan átfogó (meta) megegyezés, ahol mindenki egyetértésével létrehoznak egy hatalmi struktúrát. Mivel ennek a hatalomnak – többek között – joga lenne súlyos büntetéseket kiszabni, ezért az önérdékkövető felek sohasem fognak ehhez önként hozzájárulni. Viszont, ha már van egy központi hatalom, akkor a felek képesek megállapodni egymással, mivel a hatalom csökkenti a szerződésszegés kockázatát, ugyanis a várható büntetés elrettent a szerződésszegéstől. Természetesen a nem önző aktorok megjelenése is csökkenti a szerződésszegő viselkedés valószínűségét, és így növeli az együttműködés, beleértve a szerződéses együttműködés lehetőségét is (lásd később).

(*Evolúciós játékelmélet*) Az evolúciós játékelmélet nem tételezi fel a játékosok individuális racionalitását, viszont az interakciókat mindig egy hosszabb időperspektívában, vagyis mint többlépéses vagy többmenetes interakciókat vizsgálja, ahol a szelekciós nyomás miatt csak a legjobb és legeredményesebb viselkedési formák maradhatnak fenn. Itt a szelekció helyettesíti a rációt. Így az

evolúciós játékelmélet alapvetően olyan helyzetekkel foglalkozik, amikor az interakció sokszor kialakul egymás után. A diszkontálás segítségével kiszámítható a jövőben várható nyereségek, illetve veszteségek jelenlegi értéke. Általában elmondható, hogy minél nagyobb a következő interakció valószínűsége, annál nagyobb a diszkontált összeg értéke.<sup>8</sup> Elvileg a 2x2-es, többlépéses fogolydilemmában a négy tiszta kimenetel lehetséges, ugyanakkor kényszerítő feltételek hiányában teljesen valószínűtlen a folyamatos és egyoldalú kooperálás (S, S, S ...) és dezertálás (T, T, T ...).

A sokszemélyes, páros és többlépéses fogolydilemmát vizsgálva Axelrod (1984) mutatta ki elsőként az ún. tit for tat (TFT) szabály eredményességét. Ez a metastratégia azt jelenti, hogy egy játékos az első lépésben megkockáztatja a kooperálást, majd a továbbiakban azt lépi, amit a másik fél korábban. Ezt a viselkedést joggal tekintik barátságosnak (elsőként nem dezertál), kölcsönösnek (megismétli a másik lépését) és megbocsátónak (egy dezertálást csak egyszer büntet). A TFT egy *feltételes együttműködés*, amelyben fontos szerepet játszik a kölcsönösség, a talio-elv, fogat fogért.

Később kiderült, hogy ez a szabály csak egy „zajtalan”, egyszerű és determinált világban biztosítja a kölcsönös kooperálást. Ha a kommunikáció egy zajos térben történik, illetve ahol egy viselkedés több szempontból is értelmezhető, ott a TFT kudarcot vall. Ugyanis egy zajos és komplex világban könnyen előfordulhat, hogy A játékos kooperatív viselkedését a másik dezertálásként értelmezi, s majd a talio-elv jegyében arra dezertálással válaszol, amire pedig A szintén dezertálással válaszol. Így a felek beleszorulnak a kölcsönös dezertálásba, s ebből a TFT merevsége miatt nem tudnak kiszabadulni.

Tehát egy zajos és komplex világban nagyobb együttműködési hajlandóság szükséges. Például a TF2T-stratégiát követő A játékos csak B második dezertáló lépése után fog dezertálni, s így csökken az esélye annak, hogy a felek véletlenül belesodródjanak a kölcsönös dezertálásba. („Ha megdobnak kővel, dobd vissza kenyérrel.”) Ugyanakkor az együttműködési hajlandóság növekedésével párhuzamosan újra megnő a potyázók és csalók aránya is, akik éppen ezt az engedékenységet kihasználva bátrabban próbálkoznak az egyoldalú dezertálással. (Miközben a csalók ugyanerre a „szemtelenségre” a TFT mellett nem mernének vetemedni.)

A fentiek alapján elmondható, hogy az evolúciós játékelmélet keretei sem biztosítják a játékosok együttműködését. Ugyanakkor speciális körülmények (pl. páros és többlépéses interakció, baráti viszony) esetében a hosszú távú szempontokat is figyelembe vevő, azaz az értelmes önzés talaján álló egyedek képesek spontán módon együttműködni. A legfontosabb előfeltétel, hogy a következő interakció valószínűsége magas legyen, vagyis a személyes kapcsolat hosszú távra vonatkozzon; továbbá a korábbi együttműködések pozitív tapasztalatait is meghatározóak.

## A FOGOLYDILEMMA „KÜLSŐ” MEGHALADÁSA

Tehát a játékelmélet keretein belül a sokszemélyes, homogén (potyautas jellegű) fogolydilemma interakciókban spontán módon nem alakulhat ki az együttműködés. Ezért célszerű kiterjeszteni a vizsgálódást az önző játékos mellett új szereplőkre is, úgy mint *erőszakos*, *altruista*, *morális és közösségi* (közjóra törekvő) ágensekre. A modellnek az ilyen jellegű kibővítése összhangban van az evolúciós játékelmélet szemléletével is, hiszen az nem csak az önző, hanem az új típusú ágensek véletlenszerű (mutáció révén történő) kialakulására is lehetőséget ad. Ugyanakkor felteszem, hogy egy organikus lény alapvető sajátossága az önzés, amelytől csak jól meghatározott feltételek esetében képes eltérni. Az evolúció logikájából következik, hogy a szereplők logikai értelemben nem azonosak, hanem az egyenlőségben belül különbözőek. Az ilyen játékosokat tartalmazó fogolydilemmát megkülönböztetésül *kiterjesztett fogolydilemmának* nevezem (Tóth, 2014/b).

## Az erőszak

Először vizsgáljuk meg egy olyan populációt, ahol az ágensek sajátosságai a következők: (i) formális egyenlőség és *egyedi különbözőség*, (ii) döntési szabadság, (iii) individuális függetlenség és (iv) *erőszakos nyereségmaximalizálás*. Tehát a standard játékelmélet előfeltevéseit a (i) és a (iv) pontban módosítottuk. Mit jelent az erőszakos nyereségmaximalizálás fogalma?

A fogolydilemma interakciójában az egyoldalú dezertálás (DCCC ...) biztosítja a maximális nyereséget (T) egy játékosnak. Tehát egy aktor csak akkor tudja megszerezni a nyereségmaximumot, ha a többi játékos viselkedését is képes meghatározni, azaz ha képes őket rávenni a kooperálásra. Hosszú távon egy játékos csak erőszakkal tudja kikényszeríteni a riválisok egyoldalú kooperálását. Általában *erőszakosnak* nevezünk egy ágenszt, ha céljai érdekében képes és hajlandó erőszakot alkalmazni. Az, hogy pontosan mikor válik a potenciálisan erőszakos személy ténylegesen is erőszakossá, illetve agresszívvé, az a felek erőviszonyaitól függ.<sup>9</sup> A továbbiakban elsősorban a támadó jellegű erőszak, vagyis az agresszió kérdésével foglalkozom.

A fogolydilemma sajátosságaiból következik, hogy a maximális nyereség megszerzése érdekében rá kell kényszeríteni a többi felet az egyoldalú kooperálásra. Azaz az agresszornak (pl. A-nak) fel kell számolnia a többiek egyenlőségét, döntési szabadságát és függetlenségét. Ennek érdekében A-nak olyan helyzetbe kell hozni B-t, hogy az lemondjon a fenti jogairól, röviden a szabadságáról. B játékos csak a legvégső esetben mond le a szabadságáról, azaz csak akkor, ha ellenkező esetben az erősebb fél nagyon nagy rosszat képes okozni neki, mondjuk képes megsemmisíteni őt. Nevezzük ezt a kimenetelt a *legfőbb*



*individuális rossznak* (summum malum), ami rosszabb kimenetelt jelent, mint az egyoldalú hátrány.

Gyakran már az erőszakkal való fenyegetés is képes biztosítani a fenyegető számára a nyereségmaximumot. Ekkor az erősebb fél (A) *megfenyegeti* a másikat (B-t), hogy a legfőbb rosszat fogja okozni neki, ha B nem hajlandó az engedelmességre és azon belül az egyoldalú kooperálásra. Egy fenyegetés csak akkor hatékony, ha a fenyegető sokkal erősebb, mint ellenfele. A sikertelen fenyegetést követően csak a következő lehetőségek maradnak a fenyegető számára: (i) erőszakot alkalmaz, vagyis beváltja a fenyegetését; (ii) lemond a nyereségmaximumról és megpróbál tárgyalni a kölcsönösen előnyös kimenetelről (ennek sikerét a korábbi fenyegetés nyilvánvalóan rontja); vagy (iii) lemond az erőszakról és egyszerűen dezertál.

Az erőszak kölcsönös alkalmazása elvezet a *hatalmi harchoz*, illetve kollektív egységek esetében a *háborúhoz*, ami a politikai filozófia fontos kategóriája. „A gazdagsáért, megbecsülésért, katonai parancsnokságért vagy más hatalomért folyó versengés viszályhoz, ellenségeskedéshez és háborúhoz vezet, mert minden versengő, *hogy vágyát elérje*, igyekszik versenytársát megölni, legigázni, félreállítani vagy visszaszorítani.” (Hobbes, 1999: 146.). „Harcnak nevezünk egy társadalmi kapcsolatot, amennyiben a cselekvőt az a szándék vezeti, *hogy a saját akaratát a másik vagy a többi féllel szemben keresztülvigye*.” (Weber, 1987: 64.). „A háború tág körben dúló párviadal. [...] csupán két küzdő fél áll előttünk, akiknek mindegyike arra törekszik, *hogy a másikat saját akaratának teljesítésére kényszerítse*, ellenfelét leverje s ezáltal további ellenállásra képtelenné tegye. A háború tehát az erőszak ténye, amellyel az ellenséget saját akaratunk teljesítésére kényszeríteni igyekszünk.” (Clausewitz, 1999: 2/13.) (Kiemelések tőlem – T. I. J.)

A harc végső célja a győzelem vagy legalábbis a vereség elkerülése. A harc kimenetelét elsődlegesen a felek harci, illetve katonai ereje határozza meg, bár a végeredményt a véletlen is befolyásolja. A porosz szerző is az akarat érvényesítésében látja a harc, illetve a háború célját. „Hogy az ellenség a mi akaratunkat teljesítse, oly helyzetbe kell őt hoznunk, mely reá nézve károsabb, mint az áldozat, melyet tőle követelünk. [...] A legrosszabb helyzet, melybe valamely hadviselő fél juthat: a teljes védképtelenség. [...] Ebből következik, hogy a hadi működés céljául mindenkor az ellenség lefegyverzését vagy – ha úgy tetszik – tönkretételét kell kitűznünk.” (Clausewitz, 1999: 4/16.)

Clausewitz számára közömbös a harc tartalmi célja (a nyereségmaximum vagy éppen a közjó), hiszen annak formális célja mindig ugyanaz: az ellenfél akaratának a felszámolása és viselkedésének a meghatározása (Földi, 2006: 86.). Tehát a harc, illetve a háború egy olyan mechanizmus, amely képes a szabad, egyenrangú és független felek viszonyát egy alá-fölé rendeltségi viszonyná, vagyis hatalmi viszonyná alakítani.

Minden háború háromféleképpen végződhet: győzelem, döntetlen, vereség. Az egyik fél győzelme szükségszerűen együtt jár a másik fél vereségével, de lehetséges a kölcsönös döntetlen is. Az a fél a *győztes*, aki képes a másikat elpusztítani, miközben a másik nem képes őt elpusztítani. Ekkor a vesztes csak aközött választhat, hogy engedelmeskedik a győztesnek vagy elpusztul. Döntetlentről beszélünk, ha egyik sem tudja elpusztítani a másikat, vagy ha csak kölcsönösen tudják elpusztítani egymást. Ez utóbbi helyzet jellemezte a hidegháború éveit (1947–1991).

Hobbes szerint a polgárháború nem jelent megoldást a természeti állapot(ban létező fogolydilemma) problémájára, mivel az egyenlő erőt képviselő emberek esetében nincs győztes, csak végtelen erőszak. Ezért Hobbes szerint a polgárháború a lehető legrosszabb állapot. Álláspontom szerint a polgárháború egy lehetséges módszer arra, hogy a játékosok meghaladják a kölcsönös dezertálást, ugyanakkor nyilvánvalóan ez a lehető legrosszabb módszer, vagyis joggal nevezhetjük a polgárháborút a *másodlagos közrossznak*.

A harcban vesztes fél lényegében egy „hadifogoly”, aki csak úgy tudja megőrizni a létét, ha elfogadja a győztes fél felsőbbbségét, uralmát és akaratát. A győztes bármit megtehet a vesztesrel. Így például a győztes-vesztes viszonyt átalakíthatja úr-rabszolga viszonyrá, amelyben szabadon transzferálhatja a nyereségeket, sőt ezt a hatalmi struktúrát örökítheti is, amennyiben az úr utódaiból urak, míg a rabszolgák utódaiból rabszolgák lesznek. Az úr-rabszolga viszony egy olyan hatalmi struktúra, amelynek lényege a *kikényszerített együttműködés*. Ebben a relációban az úr, megőrizve a győztes jogosítványait, bármit megtehet a rabszolgával. A magasabb státuszú úr tehát önkényes és zsarnoki hatalmat gyakorol az alacsonyabb státuszú és neki alárendelt szolga felett. Vizsgáljuk meg ezt az úr-rabszolga viszonyt, azaz a zsarnoki hatalmat különböző számú ágensek esetében.

A *kétszemélyes úr-rabszolga* viszonyban az úr tartósan és egyoldalúan dezertálhat, míg a rabszolgának tartósan és egyoldalúan kooperálnia kell. Az első interakcióban az úr  $DC = T = +3$  egységet nyer, míg a rabszolga  $CS = S = -2$  egységet veszít, míg a közös nyereség  $+1$ . Mivel veszteséges rabszolga elpusztul, így, vagy megszűnik az úr-szolga viszony vagy az úrnak kompenzálnia kell a rabszolga veszteségeit. Ezért az úrnak a nyereségéből, minimálisan  $2,1$  nyereséget át kell adni a szolgának, s így az úr végső nyeresége:  $+0,9$ , míg a szolga nyeresége:  $+0,1$ . Ez mindkét félnek rosszabb, mintha együttműködve megszereznék  $+1$ ,  $+1$  kimenetelt. Ennek ellenére mégis van „evolúciós” értelme az úr-rabszolga viszonynak már egy kétszemélyes kapcsolatban is. Ugyanis a hatalmi struktúra esetében biztosan elérhető a  $+0,9$ ;  $+0,1$  kimenetel, míg a természeti állapotban a felek ki vannak téve a kölcsönös dezertálásnak és az abból származó  $-1$ ;  $-1$  veszteségnek. A kétszemélyes úr-szolga viszonyt tehát a formális egyenlőtlenség, a viszonylagos materiális egyenlőség, a stabilitás és az alacsony hatékonyság jellemzi.

A *három személyes* úr-rabszolga viszony egy úrból és annak két rabszolgájából áll. Itt az úr elsődleges nyeresége +6 egység, míg a rabszolgáké -1; -1. A minimálisan szükséges nyereségtranszfer után a végső nyereség rendre: +3,8; 0,1; 0,1. Az ötszemélyes úr-szolga viszonyban már nincs szükség nyereségtranszferre, hiszen minden szereplő nyeresége pozitív: +12 és +1 egység. A *tíz személyes* fogolydilemma esetében az egyetlen úr és a kilenc rabszolga „természetes” nyeresége +27; +6. Általában elmondható, hogy minél több rabszolga felett uralkodik az úr, annál nagyobb a felek természetes nyeresége, sőt a kollektív nyereség is annál közelebb van a közjóhoz. Ugyanis az úr – saját érdekét követve – a rabszolgák együttműködését nemcsak vele szemben, hanem egymással szemben is megköveteli. Ebből következik, hogy minél nagyobb létszámú a hatalmi struktúra, a kollektíva nyereségét annál inkább a rabszolgák együttműködéséből származó kölcsönös előny határozza meg. Természetesen az úrnak nem célja a közboldogság, ő a definíció szerint csak a saját nyereségét akarja maximalizálni, de ennek egy járulékos következménye a közrossz két alapformájának az elkerülése.

Az úr-rabszolga viszony alapvető fogyatékoságát mégsem az úr dezertálása jelenti, hanem a rabszolgák természetes nyereségének az elvonása. Például a már említett 10 személyes úr-rabszolga viszony esetében az úr fejenként 5 egységet fog elvonni a rabszolgáktól és így felépít egy olyan rendszert, ahol az ő nyeresége 72 egység, míg a rabszolgák nyeresége csak 1 egység. Egy ilyen elosztási rendszerben a rabszolgák csak nagyon alacsony érdekelttséggel rendelkeznek, s ezért vagy igyekeznek kimaradni a különböző fogolydilemma-helyzetekből, vagy ha megtehetik, akkor dezertálnak, hiszen a teljesítményüktől függetlenül csak a létminimumot kapják. Joggal nevezhetjük tehát ezt a közrossz egy újabb (harmadik) formájának. A zsarnoki hatalom másik problémája, hogy kifelé viszonylag gyenge katonai erővel rendelkezik, ugyanis a lakosság többsége rabszolga, akiknek a kezébe az urak nem adhatnak fegyvert.

Érdemes rámutatni arra, hogy az a makromintázat (társadalmi struktúra), amelyben két egyenrangú úr szerepel, két szempontból is rosszabb, mint az társadalmi struktúra, amelyet egyetlen úr dominál. Egyrészt kevesebb rabszolga jut egy úrra, s ezért a szolgák és az urak nyeresége is kisebb; másrészt az egyenrangú urak és csapatuk között újra kialakulhat egy fogolydilemma-helyzet, illetve egy háború, amely pedig a háborús közrossz állapotával fejeget.

Sokszemélyes interakcióban az úr közvetlenül (személyesen) már nem tudja irányítani a rabszolgákat, ezért kénytelen intézményesíteni a hatalmát. Ennek fontos eszköze a köztisztviselő, illetve a törvény. Ugyanakkor ezek az intézmények nemcsak megkönnyítik, hanem korlátozzák is az úr önkényes hatalomgyakorlását. Így az önkényes úr-rabszolga viszony átalakul egy szabályozott úr-rabszolga viszonyná.

Tehát az úr-rabszolga viszony egy működőképes hatalmi struktúra, hiszen képes kikényszeríteni az alárendeltek (rabszolgák) engedelmisségét és együttműködését, és így képes elkerülni a dezertálásból és a háborúból fakadó közrosszt. Másrészt a zsarnoki hatalom elszenvedi az érdektelenségből és az alacsony hatékonyságból fakadó *harmadlagos közrosszat*. A kikényszerített együttműködés alacsony hatékonyságának két különböző oka is van. Egyrészt a zsarnoki kényszer eleve csökkenti az alattvalók együttműködési hajlandóságát. Másrészt a nyereségmaximalizáló (és az önző) ágensek csak kényszer hatására működnek együtt. Mindamellet az erőszakos aktorok számára a zsarnoki hatalom és a kikényszerített együttműködés jelenti az *adekvát* hatalmi struktúrát.

### Az altruizmus

Most vizsgáljunk meg egy olyan populációt, ahol minden ágenst a már említett (i-iii) előfeltevés és a (iv) partikuláris altruizmus jellemez. Egy ilyen ágens általában önérdekkövető, kivéve néhány meghatározott személyt (szerelmi partnert, rokont, barátot), akikkel – a fogolydilemmától független okok miatt – eleve altruista. Tehát a partikuláris altruista éles különbséget fog tenni az emberek két csoportja között, vagyis a szeretett (számára szent) és a nem szeretett (számára profán) emberek között. Az előbbi csoport tagjaival szemben önzetlenül cselekszik, vagyis csakis az ő jólétükre koncentrálnak, míg az utóbbi csoport tagjaival szemben önérdekkövető módon viselkedik.

Az altruista viselkedése egy páros relációban teljesen egyértelműen a kooperálás, mivel ez szolgálja a másik érdekét.<sup>10</sup> Azonban az altruizmus nemcsak páros, hanem homogén kiscsoportokban is képes realizálni a spontán együttműködést és a közjót. Legyen egy háromszemélyes homogén fogolydilemma (például egy család), ahol az egyik fél altruizmusának az előnye a másik két ágenst is érinti és fordítva. A fogolydilemma egyik paradox sajátossága, hogy benne a kölcsönösen altruisták egyénileg is sokkal jobban járnak, mint az egoisták.

Egy kiscsoportos interakcióban legfeljebb a fogolydilemma helyzet felismerése okozhat problémát az altruisták számára. Elvileg előfordulhat, hogy az egyik fél (például a férj) nem ismeri fel, hogy fogolydilemmában van és önérdekkövető viselkedése kárt okoz a többi félnek (a családtagoknak), vagyis dezertál velük szemben. Ezt a hibát azonban segít elkerülni az intenzív kommunikáció, amely megmutatja a többi fél szempontjait is.

Összességében tehát azt mondhatjuk, hogy a kölcsönös altruizmus jelenti a fogolydilemma tökéletes meghaladását és megoldását. Sajnos az altruizmus egy szűkkeblű érzés, így ez a megoldás csak a legszorosabb személyes (páros és kis létszámú) kapcsolatokban működik.

## A moralitás

Egy morális ágensnek belső késztetése van arra, hogy bizonyos helyzetekben ne érdek, hanem szabály, illetve érték alapon cselekedjen. A morális ágens megvizsgálja, hogy viselkedésének van-e erkölcsi tartalma, ha a válasz igen, akkor követi saját morális szabályait, ha a válasz nem, akkor pedig az önérdekét követi. A morális ágens mindig a saját erkölcsi szabályait követi, ugyanakkor ezeket a szabályokat számos tényező határozza meg, többek között a környezetében levő emberek moralitása, a kapcsolódó absztrakt etikai, ideológiai, vallási rendszer(ek) stb. Egy etikai, ideológiai rendszeren belül interszubjektív konszenzus van az alapértékekben. Ezért egy etikai rendszeren belül a véleménykülönbségek áthidalhatók és a kommunikáció segítségével megteremthető a konszenzus, azaz megfogalmazhatók az aktuális morális törvények és szabályok. Ezek a morális törvények elvileg korlátlan számú morális ágens együttműködéséhez vezethetnek.

Vizsgáljuk meg, hogy a különböző formális etikai szabályok milyen viselkedést tartanak helyesnek és követendőnek a fogolydilemma esetében. Az *arany szabály* csak a páros fogolydilemma interakcióban alkalmazható, de ott nyilvánvalóan az együttműködést írja elő. A *klasszikus utilitarista maxima* szintén az együttműködést parancsolja, hiszen csak így érhető el a közjó. A *katégorikus imperatívusz* nem közvetlenül a viselkedést, hanem annak indítékát (maximáját) szabályozza. A fogolydilemmában a dezertálás mögött önző, míg a kooperálás mögött nem önző indíték húzódik meg. A katégorikus imperatívuszt alkalmazva pedig kimondhatjuk, hogy csak a nem önző maxima „szolgálhat általános törvényhozás elvéül”. Tehát a három legfontosabb formális etikai szabály a fogolydilemma esetében C-típusú viselkedést parancsol. Ráadásul maga a 'kooperál' és 'dezertál' igék is erős morális tartalommal rendelkeznek, s így ez a fogalomhasználat közvetlenül befolyásol minden morális ágens. Tehát a közösség a fogalomhasználat révén azt közvetíti, hogy a C-típusú viselkedés a helyes és a D-típusú viselkedés a helytelen (Tóth, 2003).

A '*ne ölj!*' parancs minden emberi közösséget jellemez. Ha egy ágens elfogadja és betartja a *ne ölj* morális parancsot, akkor nemcsak az emberölést, hanem az emberöléssel való fenyegetést is kizárja az életéből. Ez pedig a fogolydilemmában arra kényszeríti őt, hogy lemondjon a nyereségmaximum erőszakos megszerzéséről. Szintén fontos morális parancs az emberi méltóság elismerése, amely arra kényszerít egy morális ágens, hogy ne pusztán eszközként tekintsen a többi emberre. Ez a fogolydilemma keretei között azt jelenti, hogy el kell utasítania a rabszolgaság intézményét. Sokan nem gondolnak arra, hogy az önző ember (*Homo oeconomicus*) viselkedése is számos morális előfeltevést tartalmaz, úgymint az erőszakos viselkedés tilalma, a magántulajdon tisztelete, mások döntési szabadságának az elismerése és a megkötött szerző-

dések betartásának a kötelezettsége stb. Ezek az univerzális jellegű szabályok, nemcsak a piacon, hanem számos emberi interakcióban fontosak.

Több stratégias fogolydilemma esetében formálisan többfajta együttműködés lehetséges. Egy 3 stratégias fogolydilemma esetében a felek között vita lehet arról, hogy a mérsékelt ( $C_1$ ) vagy a radikális kooperálás ( $C_2$ ) tekinthető morálisan helyesnek. Például a keresztény felfogás a terhességmegszakítás minden formáját kárhóztatja, mert azt gyilkosságként értékeli, míg a modern szekuláris etika számos ága a terhességmegszakítást nem tekinti gyilkosságnak és így azt erkölcsileg nem ítéli el. Nincs az a párbeszéd, amely ezt a véleménykülönbséget képes lenne felszámolni. A világban sok különböző ideológiai, etikai rendszer (vallásos, világi, materiális, formális stb.) létezik, amelyek között fundamentális különbségek vannak, ezért az őket követő morális ágensok között is fundamentális véleménykülönbségek vannak, amelyeket párbeszéd útján sem lehet felszámolni. Azaz a különböző ideológiai, etikai rendszerek követői között nincs mód a konszenzusra.

A fentiek is mutatják a moralitás és az altruizmus közötti különbséget. Az altruizmus mindig konkrét és partikuláris, nemcsak abban az értelemben, hogy meghatározott személyekre vonatkozik, hanem abban az értelemben is, hogy a kedvezményezettekhez is egyedileg és személyesen viszonyul. Ezzel szemben az erkölcs mindig fundamentális és univerzális, egyrészt azért, mert az *összes morális lényre* vonatkozik, másrészt azért, mert *alapvető univerzális elvet* követ. Hangsúlyozandó, hogy a morális lények halmaza is folyamatosan bővül. Régen csak a *törzs* tagjaira terjedt ki a morális figyelembevétel; a világvallások számára a morális figyelem határát az *azonos vallás* jelenti, a modernitás szerint *minden ember* morális lény; a posztmodern etika (pl. fenntarthatósági etika) morális státuszt tulajdonít a *jövő generációk tagjainak* is. Másrészt nem minden univerzális szabály válik morális törvénnyé, hanem csak azok, amelyekben morális konszenzus van. Így például az ölés, lopás, házasságtörés, stb. tilalma morálisan egyértelmű, ezzel szemben a jövő generációk életfeltételeinek védelmében ma még egyáltalán nincs konszenzus, s így a moralitás erre a kérdésre nem terjed ki.

Tehát a moralitás is csak egy korlátozott választ ad a fogolydilemmára. Egyrészt a moralitás csak „in foro interno [belső téren, lélekben] köteleznek minket”. (Hobbes, 1999: 195.) Immorális (például önző vagy erőszakos) ágens viselkedését a morál eleve nem befolyásolja. Másrészt a moralitás nem minden fogolydilemma, hanem csak a rossznak felismert fogolydilemma helyzeteket szabályozza. Harmadrészt a moralitásból fakadó együttműködés csak azonos ideológiai, etikai rendszeren belül érvényesülhet, mivel csak itt lehetséges konszenzus. Ráadásul morális véleménykülönbség esetében egy morális ágenst sokkal nehezebb kényszeríteni, mint egy önző ágenst, mert nemcsak az érdekeit, hanem az értékeit is meg kell törni.

## A közjóra való törekvés

Most vizsgáljunk meg egy olyan populációt, ahol az ágensek sajátosságai a következők: (i) formális egyenlőség és *egyedi különbözőség*, (ii) döntési szabadság, (iii) *közösséghez való tartozás* és (iv) *közjóra való* törekvés. Ezek az előfeltevések a standard játékelmélethez képest három előfeltevésben (i, iii és iv) is különbséget jelentenek.

Először is mit jelent a közösséghez való tartozás fogalma? A *közösség* olyan aktorok egysége, akiket erős kötelék fűz egybe, akik rendelkeznek közösségi tudattal, és akik együtt próbálnak érvényesülni. Egy közösség Tönnies (2004) kifejezésével élve lehet kicsi és helyi jellegű (Gemeinde), illetve nagyobb és átfogóbb, például nemzeti jellegű (Gemeinschaft). A fentiekkel szemben definiálja Tönnies a társulást (Gesellschaft), ahol a felekre nem jellemző a szolidaritás, illetve a közösségi tudat. Sokfajta közösség (családi, baráti, szakmai stb.) lehetséges, de a továbbiakban csak a politikai közösségek kérdésével foglalkozom. Politikai közösség az, amely a közösség egészére vonatkozó kérdésekben hoz döntéseket.

A fogolydilemma esetében a független és a közösségben élő ember közötti különbséget jól megvilágítják Parfit (1998: 145.) szavai: „Mindenkinek egyenként kell-e a maga számára a lehető legjobb eredményre törekednie; vagy nekünk együtt kell a legjobb eredményt biztosító viselkedést követnünk?” Nyilvánvalóan éles különbség van az *egyenként* és az *együtt* logikája között. Mindkét esetben értelmezhető a racionalitás fogalma, a csak társulásban (Gesellschaft) élő független ember számára ez az individuális racionalitást, vagyis a dezertálást, addig a közösségi ember számára ez a kollektív racionalitást, vagyis az együttműködést jelenti (Málik, 2013).

A közjő az adott közösség maximális nyereségét jelenti, amely a fogolydilemmában megegyezik a kölcsönös kooperálással (Tóth, 2014/a). A közjóra törekvő ember számára maga a közjó az a szent érték, amelyet meg kell valószínűsíteni. Egy közjóra törekvő ágens viselkedését a következő sajátosságok jellemzik. Éles különbséget tesz a közjó szempontjából irreleváns és releváns interakciók között. Az előbbieken mindig az önérdékét követi, míg az utóbbiban egy bonyolult (c–e pontban leírt) algoritmust követ. Az aktor szintén éles különbséget tesz a közösséghez tartozó és nem tartozó ágensek között. A közösséghez nem tartozó ágensekkel – definíció szerint – dezertál. A közösséghez tartozó ágensekkel szemben pedig a fogolydilemma esetében a következő algoritmust követi. (i) Ha bízik abban, hogy a közösség többi tagja is a közjóra törekszik, és osztja a közjóról vallott speciális nézetét, akkor a  $C_1$  stratégiát választja, ellenkező esetben (ii) először megpróbálja meggyőzni őket az együttműködés általa képviselt módjáról, és ha ez nem sikerül (iii), akkor megpróbálja rákényszeríteni a többieket az együttműködés általa képviselt módjára. (iv) Legjobb esetben a közösség tagjai kidolgoznak és elfogadnak egy békés

eljárást (alkotmányt), amit a véleménykülönbség esetében alkalmaznak. Vizsgáljuk meg ezeket a lehetőségeket.

(*Konszenzus*) Legegyszerűbb esetben a felek konszenzusra jutnak a közjó megteremtésének a módjában, majd ezt a megállapodást realizálják. Közjóra törekvő ágensek esetében a konszenzus gyakran megteremthető. Problémát „csak” az jelenthet, ha a felek között fundamentális véleménykülönbségek vannak az együttműködés módjában. Ez a helyzet általában akkor alakul ki, ha a közösség tagjai különböző ideológiai, etikai, vallási rendszerekben hisznek. Lehet, hogy a szemben álló felek (véleményközösségek) ezt úgy élik meg, hogy a másik fél tagjai valójában nem a közjóra törekszenek, hanem pusztán önzők. Ez a bizalmatlanság a feleket egy alacsonyabb szintű együttműködési formába és hatalmi struktúrába zárja. Azonban még ha nem is következik be a másik oldal démonizálása, a fundamentális véleménykülönbségek miatt akkor sincs konszenzus, és így nincs együttműködés, hanem csak tehetetlenség, döntésképtelenség, amely a *közrossz* egy további (*negyedleges*) formáját jelenti.

(*Erőszak*) Ebből a válságból a legegyszerűbb kivezető utat az erőszak jelenti. Első lépésben a fundamentális kérdésekben azonos véleményű emberek koalíciót alkotnak egymással. A koalíció tagjai formálisan és materiálisan is egyenlők, de nem azonosak; ugyanakkor a várható katonai összecsapás miatt a koalícióból katonai erőt képviselő hadsereget kell létrehozni, amelyet egyértelmű vezetés és hierarchia jellemez. Csak annak a koalíciónak van esélye a katonai győzelemre, amely képes erre az átalakulásra; vagyis ahol a tagok konszenzussal megegyeznek, egy *társadalmi szerződést* kötnek egy *méltányos hatalmi struktúráról*. Ez a méltányos hatalmi struktúra feltehetően a következő pontokat tartalmazza: a véleményközösség alapvető értékeit, az emberek jogát arra, hogy spontán kövessék ezeket az értékeket, a hatalom érték alapú kontrollját. Majd a második lépésben a szemben álló felek megütköznek egymással és várhatóan az erősebb hadsereg győz.

Ha minden egyéb feltétel azonos, akkor az a hadsereg, az a véleményközösség, az az ideológia fog győzni, amelyikhez többen csatlakoztak. A győztes oldal határozza meg az együttműködés módját. Nyilvánvaló, hogy a győztes vezetőnek erős kontrollt kell gyakorolnia mindazok felett, akiknek más véleményük van a közjó megvalósításáról. A legyőzött oldal tagjait kényszerítenie kell, míg a saját oldal tagjaival szemben a hatalomnak csak korlátozott jogosítványai vannak. Mivel egyetlen közösségről van szó, ezért egyidejűleg kell kényszerít alkalmazni és bizalmat adni mindenkinek az együttműködés meghatározott módjára vonatkozólag. Ebből fakad, hogy ez a hatalmi struktúra alapvetően különbözik az úr-rabszolga viszonytól, amely mindig csak a magánérdekre irányul.



– A közjóra törekvő vezetőnek a definíció szerint részt kell vennie az együttműködésben, azaz alá kell rendelnie magát az együttműködés törvényének; vagyis itt *'a törvény a király'* elve érvényesül. Ezért a közösségi vezető és az alattvalók közötti formális különbség „eredendően” nem eredményez tényleges (materiális) különbséget, hiszen a fogolydilemmában minden együttműködő azonos nyereségre tesz szert.

– A közösségi vezető nem teheti meg, hogy a javakat a saját hasznára csoportosítja át, ugyanis ekkor az érdektelenségből és az alacsony hatékonyságból fakadó harmadlagos közrossz következne be. Ennek tilalmából szintén a politikai vezető és az alattvalók materiális egyenlősége következik. A fogolydilemma világában az együttműködők nyeresége egyenlő, legfeljebb abból fakadhat különbség, hogy ki milyen sok interakcióban vesz részt.

Mivel a politikai vezető tudja, hogy az alattvalók is a közjóra törekednek, ezért a hatalmi kontroll kényes egyensúlyát kell kialakítania. Az uralkodónak elégséges hatalmat kell gyakorolni az alattvalók felett, hogy érvényesítse a közjóról vallott felfogását, másrészt elég szabadságot kell adnia az alattvalóknak ahhoz, hogy azok önként és szabadon együttműködhessenek a közjő meghatározott rendszerében.

(*Alkotmány*) A fentebb említett döntésképtelenségből és a hozzá tartozó negyedeleges közrosszból elvileg a másik kivezető utat a különböző tartalmú alkotmányok jelentik. Ezek közös sajátossága, hogy a közösség tagjai konszenzussal megegyeznek egy békés meta-döntéshozatali eljárásban. Azaz a közösség tagjai és vezetői ígéretet tesznek arra, hogy ezt a békés meta-döntéshozatali eljárást szentként tisztelik, s saját véleményüket alárendelik az így született döntésnek. A békés döntéshozatal legfőbb előnye, hogy a fundamentálisan különböző vélemények közötti döntés nem erőszakosan születik meg.

A meta-döntéshozatali eljárás sokféle lehet. A legegyszerűbb forma, hogy kijelölnek egy *politikai vezetőt*, akit feljogosítanak arra, hogy konszenzus hiányába önkényesen döntsön a politika alapvető kérdéseiben. Hangsúlyozandó, hogy a politikai vezető döntése nem magán döntés, hanem közösségi döntés, továbbá, hogy ezt a döntést a közösség tagjai önként követik. Persze a politikai vezető kiválasztása is vitát generálhat. Ezt elkerülendő, dönthet úgy a közösség, hogy a politikai vezetői tisztség öröklődjön, de dönthet úgyis, hogy a politikai vezetőt *többségi elv* segítségével válasszák meg. Mind a választók, mind a választhatók sokassága a közösség különböző részére kiterjedhet.

A közösség dönthet úgy, hogy szétválasztja a különböző hatalmi ágakat, vagyis különböző személyeket ruház fel szabályalkotó (törvényhozó), a szabályokat az adott helyzetben a közösség érdekében alkalmazó (végrehajtó), illetve a tagok vitás ügyeit rendező (bírói) hatalommal. Lehetséges persze az is, hogy az új alkotmány egy forradalom (vagy polgárháború) eredményeként

alakul ki. Ebben az esetben a győztes csak a saját politikai elképzelésein belül engedélyezi az aktuális politikai döntéseket.

Érdeemes rögzíteni, hogy a közösségi ágens miben különbözik más ágensektől. A közjóra törekvő ágensnek is partikuláris jellegű célja van (a saját közösség jóléte), hasonlóan az altruistához. Ugyanakkor a közösségi ember célja sokkal absztraktabb, mint az altruistáé, aki csak a személyes ismerőseinek a jólétét szolgálja. A közjóra törekvő ember abban is hasonlít az altruistához, hogy éles különbséget tesz az emberek („mi-ők”) között. A közösségi ember, amikor a közjóra törekszik, akkor a saját jólétére is törekszik, hasonlóan az egoistához. Ugyanakkor a közösségi ember a magánérdekét a közérdekkel összhangban próbálja realizálni, szemben az egoistával, aki közömbös a közjával kapcsolatban. A közösségi aktor annyiban hasonló a morális aktorhoz, hogy egy viszonylag absztrakt elvet követ, amely sokféleképp realizálható, másrészt annyiban különbözik tőle, hogy érdeklődése csak saját közösségének a tagjaira (a „mi”-re), illetve a közjő szempontjából releváns interakciókra (alapvetően a fogolydilemma interakciókra) terjed ki.

Sajnos, a közjóra törekvés, s azon belül az alapértékeket és a konfliktusok feloldási módját szabályozó alkotmányok sem oldanak meg minden problémát a fogolydilemma világában. Egyrészt a többségi vélemény, illetve az alkotmány nem feltétlenül helyes. Másrészt a közjóra törekvő ember számára alapvető kérdést jelent annak meghatározása, hogy ki tagja („mi”) és ki nem tagja („ők”) a közösségnek. Az ősi közösségek vérségi és rokonsági alapon szerveződtek, ezzel szemben a modern politikai közösségeket a közös kultúra, nyelv, terület stb. határozza meg. Az átmenet időszaka nyilvánvalóan még inkább felerősíti azt a problémát, hogy ki tekinthető a közösség tagjának.

Harmadrészt a közösség időbeli határai sem egyértelműek. Kezdetben nyilvánvalóan az a felfogás érvényesül, hogy a közösség azonos az őt alkotó emberekkel, sőt, csak a választójoggal rendelkező nagykorú ágensekkel. Ez az értelmezés a döntéshozatal, illetve a közösség fogalmából kizárja a gyerekeket, illetve a még meg sem született generációkat. A közösség aktuális és potenciális tagjai között azonban létezik egy speciális fogolydilemma interakció, ahol csak az előbbi csoport van hatással az utóbbi csoport jólétére. Ha a közösségi döntéseket mindig csak az aktuális tagok többségi döntése határozza meg, akkor az éppen aktuális közösség kizsákmányolja a „jövőbeli” közösséget, pontosabban magát a közösséget. Jó példa erre a közösség eladásítása (vagyis a jelenlegi célok költségeinek a jövő generációira való terhelése), a közösség környezeti feltételeinek a felélése, vagy éppen a közösség demográfiai fennmaradásának a veszélyeztetése, történjen az túlnépesedés vagy éppen „alulnépesedés” formájában. Persze az aktuális közösség ezeket a károkat nem gonoszságból okozza, hanem pusztán azért, mert saját jólétét próbálja maximalizálni, amelynek szükségszerű következménye a jövő közösségének minimális nyeresége.

Végül további problémát jelent, hogy egy politikai közösség kollektív nyereségének maximalizálása (vagyis a közjó realizálása) nyilvánvalóan árt más politikai közösségek érdekeinek. Számos esetben a különböző politikai közösségek közötti interakció a fogolydilemma logikáját követi, s ekkor az egyik politikai közösség nyereségmaximalizálása a többi politikai közösség minimális nyereségéhez vezet. (Ahogy már említettem a fogolydilemmában nincs olyan kimenetel, hogy minden entitás maximális nyereségre tegyen szert.) Azaz „magasabb szinten” a politikai közösségek mint ágensek szintjén is kialakulnak a fogolydilemma típusú interakciók. Tehát az együttműködés problematikája nemcsak az individuális ágensek, hanem a politikai közösségek (pl. államok) esetében is felmerül. Általános szinten a politikai közösségek között kialakuló fogolydilemma helyzetekre ugyanazok a szabályok és megfontolások érvényesek, mint az individuális aktorok esetében. Ugyanakkor a nemzetközi aktorok számos tekintetben (kisebb létszám, nagyobb individuális racionalitás, altruizmus hiánya) különböznek a természetes személyektől. Ezért nem lehet mechanikusan alkalmazni a természetes személyek esetében kialakult megoldási formákat a nemzetközi aktorok közötti fogolydilemma helyzetekre.

#### KITEKINTÉS ÉS ÖSSZEFOGLALÁS

Gondolatkísérletemben megpróbáltam összekapcsolni a játékelméletet a politológiával. Elemzésem központi kérdése, hogy a sokszemélyes és sokstratégias homogén fogolydilemma (potyautas jellegű) interakciókban hogyan lehetséges a szereplők együttműködése és a közjó megteremtése. A tanulmány szándékosan nem érintette a gazdasági jellegű speciális fogolydilemma helyzeteket (például a közlegelő tragédiája, árháború) és azok speciális megoldási formáit (magántulajdon, elvándorlás, árverseny stb.). Az evolúció logikájával összhangban feltettem, hogy bármilyen individuális vagy kollektív viselkedési forma megjelenhet, de csak a nyereséges és egyre hatékonyabb struktúrák maradhatnak fenn és terjedhetnek el.

Konklúzióm szerint ennek az interakciónak nincs egyetlen jó és univerzális megoldása, hanem csak különböző korlátozott hatókörű és hatékonyságú megoldásai vannak, amelyek maguk is további problémákat (közrosszakat) okoznak. Így az együttműködés megteremtése korlátlan evolúciós folyamatként értelmezendő. Ebben a folyamatban fontos tényezőt jelent a népesség eddig tapasztalt folyamatos növekedése. Ez nemcsak mennyiségi bővülést, hanem új típusú interakciók megjelenését is jelenti, a különböző rokonsághoz, véleményhez, generációhoz, közösséghez, stb. tartozó emberek között. A kezdeti dezertálást és az abból fakadó elsődleges közrosszt az egyre átfogóbb együttműködési struktúrák evolúciós „kialakulása” követi. Régen a kisméretű és zárt közösségek esetében az evolúció az együttműködést még *természetes*

*képességekkel és erényekkel* is segítette, úgymint kommunikáció, altruizmus, közösségi érzések. Később a nagyobb közösségek külső és belső együttműködését már olyan *mesterséges intézmények és erények* koordinálják, mint az uralom, a többségi elv, az alkotmány.

A fogolydilemma külső meghaladásának az elemzésekor a különböző, logikailag „tisztá” tulajdonságú aktorok homogén, vagyis tiszta populációjának a kialakulásával számoltam. Nyilvánvalóan létezhetnek olyan további „tisztá” aktorok (bölc, csaló, hazug), amelyeket a tanulmány nem tárgyal. A modell továbbfejlesztésének egyik módja a különböző „tisztá” aktorok sokaságának, másrészt a „kevert” tulajdonságú aktorok sokaságának a vizsgálata. Terjedelmi korlátok miatt azonban ezeket a lehetőséget egyáltalán nem vizsgáltam.

Tanulmány legfontosabb tézisei a következőkben foglalhatók össze. (i) A klasszikus (nem kooperatív) fogolydilemma előfeltételei mellett (azonos, szabad, független és önző játékosok), vagyis a természeti állapotban szükségszerű a kölcsönös dezertálásból fakadó elsődleges közrossz bekövetkezése.

(ii) A természeti állapotban a kommunikáció és a kooperáció elvi lehetősége *sem* biztosítja a spontán együttműködést, illetve a társadalmi szerződés megkötését.

(iii) Az evolúciós játékelmélet csak a „páros” sokszemélyes fogolydilemma esetében képes előidézni a játékosok együttműködését. Ugyanakkor az evolúciós folyamatok új ágenseket hoznak létre, amelyek képesek az együttműködés új formáira.

(iv) A nyereségmaximumra törekvő erőszakos ágensek között mindenki háborúja folyik mindenki ellen, amit joggal tekinthetünk egy másodlagos közrossznak. A harc azonban általában a legerősebb fél győzelmével végződik. A győztes létrehozza az úr-rabszolga viszonyt, vagyis a kikényszerített együttműködés rendszerét. Egy stabil zsarnoki hatalom elkerüli a kölcsönös dezertálásból és a háborúból származó közrosszt, viszont nem tudja elkerülni a rabszolgák érdektelenségéből származó harmadlagos közrosszt. Ennek ellenére az önző és erőszakos ágensek számára csak ez az együttműködési forma lehetséges.

(vi) Az altruizmus csak szoros személyes kapcsolatokban működik, ott viszont tökéletes együttműködést biztosít.

(vii) A moralitás csak az alapvető fogolydilemma helyzeteket képes koordinálni egy meghatározott etikai, ideológiai, vallási rendszeren belül. Az etikai rendszerek különbözősége elvezet a véleménykülönbségből és döntésképtelenségből fakadó negyedleges közrosszhoz.

(viii) Ha egy közösségen belül mindenki a közjóra törekszik, akkor a közjó elérhető, kivéve, ha a felek között fundamentális jellegű véleménykülönbség van az együttműködés módjáról. A fundamentális véleménykülönbséget a közösség vagy erőszakkal (polgárháborúval) vagy békés módon hidalja át. Nyilvánvaló, hogy a közjó szempontjából az utóbbi megoldás tekinthető jobbnak,

amely mindig valamilyen meta-döntéshozatali (politikai vezető, többségi elv, hatalom megosztása stb.) eljárásra épül. Sajnos ezek az intézmények is súlyos problémákat rejtenek magukba: mi legyen a kisebbségi véleménnyel, ki tartozik a közösséghez, szabad-e a jövő, illetve a szomszédos közösségek terhére maximalizálni a közösség kollektív nyereséget? A közösségiség fejlődésének egy paradox következménye a szupraindividuális jellegű aktorok (klán, törzs, állam) kialakulása, amelyek viszonyát, többek között a fogolydilemma típusú interakciók is jellemzik. Ezekre a kihívásokra az individuális és szupraindividuális aktorok elvileg találhatnak különböző, korlátozott hatékonyságú válaszokat, amelyek újabb problémákat vetnek fel és így tovább a végtelenségig.

## IRODALOM

- Aumann, Robert J. (1961): The core of a cooperative game without side payments. *Trans. Amer. Math. Soc.* 98, 539–552. <http://www.ma.huji.ac.il/raumann/pdf/0039.pdf>.
- Axelrod, Robert M. (1984): *The Evolution of Cooperation*. Yale University Press, New Haven.
- Boda Zsolt (2013): *Legitimitás, bizalom, együttműködés. Kollektív cselekvés a politikában*. Argumentum Kiadó, Budapest.
- Clausewitz, Carl von (1999): *A háborúról*. Göttinger Kiadó, Veszprém, A könyv az 1917-es kiadás reprintje. Fordította Báró Hazai Samu.
- Elster, Jon (1976): Some Conceptual problems in Political Theory. In: Barry, Brian (ed.): *Power and Political Theory: Some European Perspectives*. Wiley, London, 245–270.
- Földi Ferenc (2006): Gondolatok a fegyver szerepéről a harcban. *Hadmérnök*, 1. szám, 83–101.
- Gauthier, David (1969): *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes*. Oxford, Clarendon Press.
- Hobbes, Thomas (1999): *Leviatán (vagy az egyházi és világi állam formája és hatalma)*. (Ford. Vámosi Pál–Ludassy Mária) Kossuth, Budapest.
- Olson, Mancur: (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups* (Revised ed.). Harvard University Press.
- Ostrom, Elinor (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- Málik József Zoltán (2013): Dilemmák a lojalitással. *Jogelméleti Szemle*, 3. szám. 84–92. <http://jesz.ajk.elte.hu/malik56.pdf>
- McCain, Roger A. (2004): *Game Theory. A Non-Technical Introduction to the Analysis of Strategy*. Thomson & South-Western.
- McLean, Iain (1987): *Public Choice. An Introduction*. Basil Blackwell, Oxford–New York.
- Parfit, Derek (1998): Körültekintés, erkölcsiség és a fogolydilemma. In: Csontos László (szerk.): *A racionális döntések elmélete*. Osiris–Láthatatlan Kollégium, Budapest.
- Platt, John (1973): Social Trape. *American Psychologist*, 128 (8): 641–651.
- Ridley, Matt (2011): *Az erény eredete*. Akadémia Kiadó Zrt, Budapest. 64.
- Scarf, Herbert E. (1967): The core of an n-person game. *Econometrica*, 35: 50–69.

- Schelling, Thomas.C (1973): Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *Journal of Conflict Resolution*, 17: 381–428.
- Tóth I. János (2003): Játékelmélettől az etikáig. *Világosság*, 44 (5–6): 167–174.
- Tóth I. János (2014/a): A közjó meta-játékelméleti elemzése. *South East Europe International Relation Quarterly*, 5 (17): 1–9. [http://www.southeast-europe.org/pdf/17/dke\\_17\\_m\\_b\\_Toht-I-Janos\\_A-kozjo-meta-jatekelmeleti-elemzese.pdf](http://www.southeast-europe.org/pdf/17/dke_17_m_b_Toht-I-Janos_A-kozjo-meta-jatekelmeleti-elemzese.pdf)
- Tóth I. János (2014/b): A foglydilemma kiterjesztése. *Magyar Tudomány*, 175 (1): 90–98.
- Tönnies, Ferdinand (2004): *Közösség és társadalom*. Társadalomtudományi Könyvtár, Budapest.
- Weber, Max (1987): *Gazdaság és Társadalom*. Közgazdasági és Jogi könyvkiadó, Budapest.
- Zagare, Frank C. (2006): *Játékelmélet. Fogalmak és alkalmazások*. Helikon Kiadó, Budapest.

## JEGYZETEK

<sup>1</sup> Egy kirabolt bank előtt két pisztolyos egyént, Altot és Bobot letartóztatják a rendőrök. Az ügyésznek nincs elegendő bizonyítéka ahhoz, hogy bankrablásért vádat emeljen a gyanúsítottak ellen. Ezért az ügyész külön cellába zárja a foglyokat, hogy ne tudják egymással egyeztetni a vallomásaikat, majd ugyanazt a vádalkut ajánlja mindkettőnek:

		B gyanúsított	
	(döntés)	tagad (C)	vall (D)
A gyanúsított	tagad (C)	R = -1 év, R = -1 év kölcsonösen előnyös PO, (=közjó)	S=-10, T=0 <sup>PO</sup>
	vall (D)	T = 0, S = -10 <sup>PO</sup>	P = -6, P = -6 <sup>PI</sup> (=közrossz)

1. ábra. Foglydilemma (FD) eredeti története

Az első szám A, a második szám B gyanúsított börtönbüntetését mutatja években. PO: Pareto-optimális, PI: Pareto-inferior kimenetel.

\* egyensúlypont. Általában foglydilemmáról, akkor beszélünk, ha  $T > R > P > S$  és  $2R > T + S$ . Az adott értékek mellett igaz az, hogy  $2R$  (közjó)  $> T + S > 2P$  (közrossz).

Rögzítsük tehát a következőket: a situációt az ügyész szabályai strukturálják (szabályalkotás); a szabályozott situációban az egyenlő, független és racionális szabad szereplők hozzák meg a végső döntést; az interakció a *nem kooperatív játékelmélet* törvényszerűségeit követi. Altnak és Bobnak két döntési lehetősége van, a 'vallomástétel' és a 'bűncselekmény 'tagadása'. A szakirodalomban az előbbit dezertálásnak (defect), míg az utóbbit kooperálásnak (cooperate) nevezik. A valóságban a 'kooperációt' általában megelőzi valamilyen egyeztetési folyamat, amelyre a klasszikus foglydilemmában nincs lehetőség, ezért célszerű különbséget tenni a procedurális és szubsztantív kooperáció között (Tóth, 2014/b: 91.).

<sup>2</sup> A rövidítések a következő kifejezésekből származnak: Sucker's payoff, Punishment for mutual defection, Reward for mutual cooperation, Temptation to defect. Fontos, hogy a foglydilemmában a kölcsönös nyereségmaximum [T; T] kimenetele nem létezik.

- <sup>3</sup> Lehetnek azonban olyan fogolydilemma interakciók, ahol a közrossz állapotát a nyereségmaximum és minimum összege jelenti. Például ha  $T = 5 > R = 4 > P = 3 > S = 0$ , akkor egyrészt igaz az, hogy  $2R > T + S$ , de az is igaz, hogy  $2P > T + S$ . A továbbiakban ezzel az esettel nem foglalkozom.
- <sup>4</sup> A dezertáló játékos +3 vagy -1 egység nyereséget, míg az együttműködő +1 vagy -2 egység nyereséget kockáztat a társa döntésétől függően.
- <sup>5</sup> Például, ha mindhárom játékos kooperál (CCC), akkor virtuális értelemben mindhárom játékosnak van két kölcsönös kooperálása ( $CC = R = 1 + CC = R = 1$ ) és így fejenként 2 egység nyereségre tesznek szert. A CDC kimenetel esetében az első játékos „virtuális értelemben” egyoldalúan kooperál a másodikkal, nyeresége:  $CD = S = -2$  egység és kölcsönösen kooperál a harmadikkal, nyeresége:  $CC = R = 1$ , tehát összesen -1 egységet nyer. Ugyanez igaz a harmadik játékosra is. A második játékosnak pedig van két egyoldalú dezertálása ( $DC = T = 3$ ), vagyis ő 6 egységet nyert. Ezt a szituációt még éppen lehet mátrixformában is ábrázolni:

				3. játékos	
				C	D
1. játékos	C	2. játékos	C	CCC (2, 2, 2→6) kölcsönös előny, közjó	CCD (-1, -1, 6→4)
			D	CDC (-1, 6, -1→4)	CDD (-4, 2, 2→0) nyereségminimum az 1. játékosnak
	D	2. játékos	C	DCC (6, -1, -1→4) nyereségmaximum az 1. játékosnak	DCD (2, -4, 2→0)
			D	DDC (2, 2, -4→0)	DDD (-2, -2, -2→-6) kölcsönös hátrány, közrossz

2. ábra. Háromszereplős homogén fogolydilemma.

A zárójelben levő számok rendre: az 1., 2., 3. játékos individuális nyeresége és a három játékos együttes nyeresége. (Ezeket az értékeket a kétszemélyes fogolydilemma önkényes számértékeiből:  $T = 3; R = 2; P = -1; S = -2$  számoltam ki.)

Egy homogén 5 személyes fogoly dilemma esetében elvileg (kombinatorikai szempontból) összesen  $2^5 = 32$  kimeneteli lehetőség van. Kollektív mintázat szempontjából ebből csak a következő 6 jelent tartalmi különbséget, úgy mint: 5C (—; 4; 20); 1D4C (12; 1; 16); 2D3C (8; -2; 10); 3D2C (4; -5; 2) 4D1C (0, -8; -8) 5D (-4; —; -20). A zárójelben megadott értékek sorrendben a dezertáló nyereségét, a kooperáló játékos nyereségét, illetve a kollektív nyereséget mutatja. (Tehát ha mind az 5 játékos dezertál, akkor egy dezertáló játékos nyeresége -4 egység, kooperáló játékos nincs és így az ő nyeresége nem értelmezhető erre utal a „—” jel, míg az 5 játékos nyereségének összege -20 egység, stb.) Egy homogén 10 személyes fogolydilemma interakcióban elvileg összesen  $2^{10} = 1024$  kimeneteli lehetőség van. Makromintázat szempontjából ebből csak 11 kimenetel különbözik tartalmilag. S ebből a 11-ből pedig a következő 4 az, amelyik igazán fontos: 10C (—; 9; 90); 1D9C (27; 6; 81); 9D1C (-5; -18; -58); 10D (-9; —; -90). A zárójelben megadott számok jelentése ugyanaz mint előbb.

- <sup>6</sup> Vegyük például a szarvasvadászat (stug hunt) nevű koordinációs játékot, amelynek két egyensúlypontja is van: a kölcsönös együttműködés (CC: pl.  $R = 5, R = 5$ ) illetve a kölcsönös dezertá-

lás (DD: pl.  $P = 3$ ,  $P = 3$ ). A dilemma abból fakad, hogy az együttműködés eredménye egy nagyobb, de bizonytalan nyereség, míg a dezertálás eredménye egy biztos, de kisebb nyereség. Mivel CC kimenetel mindkét játékosnak egyénileg is a legjobb kimenetelt jelenti, ezért a kooperatív feltételek között kialakulhat a megegyezés, mivel senkinek sem áll érdekében konszenzusosra épülő megállapodástól eltérni.

<sup>7</sup> Például legyen a következő interakció valószínűsége  $w$ , ahol  $w$  értéke 0 és 1 között változhat, akkor az egyoldalú dezertálás jelenlegi értékét a következő összeg adja meg:

$$T + wT + w^2T + \dots + w^nT + w^nT \dots = T/(1-w).$$

Ha  $w$  értéke tart az 1-hez (vagyis szinte biztos a következő interakció), akkor  $T$  értékek sorozatának az összege tart a végtelenhez. Hasonló módon kiszámítható a másik három kimenetelhez tartozó nyereség jelenbeli értéke is úgy mint:  $R/(1-w) > P/(1-w) > S/(1-w)$ .

<sup>8</sup> A standard játékelmélet keretei között az egyik fél nem tudja kényszeríteni a másikat, mert a játékosok minden szempontból azonosak. Az evolúció logikájából viszont az következik, hogy még az azonos típusú aktorok is különböznek egymástól. Így a felek általában katonai erőben is különböznek egymástól.

<sup>9</sup> Ugyanis, ha A (pl. az anya) kooperál, akkor B (pl. a csecsemő)  $C_A C_B = R_B = +1$  vagy  $C_A D_B = T_B = +3$  egységet nyer, míg ha A dezertál, akkor B  $D_A C_B = S_B = -2$  vagy  $D_A D_B = P_B = -1$  egységet veszít. Ezért az altruista feltétel nélkül C-t fogja választani, és sohasem választhatja a D-t. Ha mindkét személy altruista (pl. egy szerelmespár), akkor mindkettő feltétel nélkül együttműködik egymással.