5.2 Lecke Fogolydilemma (20 perc)

A fogolydilemmát a Princetoni Egyetem matematikus professzora, Albert Tucker találta ki, amikor 1950-ben matematika ismeretekkel nem rendelkező pszichológus kollégáknak bemutatott egy olyan fontos interakciót, amely játékelmélet segítségével jól elemezhető. Tucker választhatott volna jobb példát is. Nemcsak a negatív számok teszik nehézkessé a történetet, hanem az is, hogy bűnözők együttműködése nem kívánatos. Ennek ellenére ez a rövid történet jelenti a 20. századi társadalomtudományok legnagyobb hatású modelljét.

5.2.1 A klasszikus történet

Egy kirabolt bank előtt két pisztolyos egyént, Alt és Bobot letartóztatják a rendőrök és külön cellába zárják őket, hogy ne tudják egymással egyeztetni a vallomásaikat. A bíróságnak nincs elegendő bizonyítéka ahhoz, hogy elítélje őket bankrablásért, ezért fontos, hogy legalább az egyik gyanúsított beismerje, hogy ők követték el a rablást. Ennek érdekében az ügyész ugyanazt vádalkut ajánlja mindkettőnek:

(a) Ha mindketten továbbra is tagadjátok a rablást, akkor tiltott fegyverviselésért egy-egy év börtönbüntetést sózok a nyakatokba. (b) Ha egyikőtök bevallja, hogy ti követtétek el a rablást, akkor a vallomást tevőt szabadon engedem, de a másik tíz év börtönbüntetést kap. (c) Ha mindketten beismerő vallomást tesztek, akkor mindketten hat év szabadságvesztést kaptok.

Altnak és Bobnak két döntési lehetősége van a vallomástétel (confess) és a hallgatás (remain silent) vagyis a bűncselekmény tagadása. Ez négy különböző kimenetelhez vezethet. A következő ábrán az első szám Alt, míg a második szám Bob büntetését mutatja. Ebben a szituációban a hallgatást nevezik kooperatív (C), míg a vallomástételt dezertáló (D) viselkedésnek, illetve stratégiának. Kérdés, hogy miért? Intuitív módon a következőképp lehet érvelni: a bűncselekmény tagadása azért nevezhető kooperatív (C) viselkedésnek, mert aki tagad az kitart a másik mellett (betyárbecsület), míg a bűncselekmény beismerése, azért dezertáló viselkedés, mert a másik elárulását jelenti. Később egy formális kritérium segítségével is bemutatom, hogy miért volt helyes ez az elnevezés.

		B gyanúsított	
	(döntés)	Tagad (C)	Vall (D)
A gyanúsított	Tagad (C)	-1, -1	-10, 0
	Vall (D)	0, -10	-6, -6

Fogolydilemma (FD) eredeti története. Az első szám A, a második szám B gyanúsított börtönbüntetését mutatja.

Aki tagad (C), az 1 vagy 10 év börtönbüntetést, míg aki vall (D) az 0 vagy 6 év börtönbüntetést kockáztat a társa döntésétől függően. Tehát aki vall, az mindenképpen jobban jár, mint az, aki tagad; a szakemberek ezt úgy mondják,

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu

hogy a vall stratégia *szigorúan dominálja* a tagad stratégiát. A gyanúsítottaknak tehát egy domináns (racionális) és a nem-domináns (nem-racionális) stratégia között kell választania, s nem lehet kétséges, hogy ebben a helyzetben a szereplők, hacsak nincs valamilyen egyéb szempont, akkor a domináns és racionális viselkedést választják.

A döntési helyzet azonban szimmetrikus, azaz hasonló okok miatt a másik gyanúsított is beismerő vallomást tesz. Ezért mindketten 6–6 év börtönbüntetést kapnak. Ez az interakció egyensúlypontja. A játékelmélet szerint a racionális játékosoknak ebben a helyzetben vallomást kell tenniük, azaz dezertálniuk kell.

Tekinthetik-e a felek a kölcsönös vallomástételből fakadó 6–6 év börtönt maguk számára jó megoldásnak? Semmiképp sem, hiszen boldogan elcserélnék az 1–1 év börtönre. Tehát a szereplők, ha *kölcsönösen* eltérnek az egyensúlyi – sőt domináns és racionális – stratégiájuktól, akkor sokkal jobban járnak.

Számos interakció működik a fogolydilemma logikájára, úgy mint a potyautas (Free-rider_problem), A_közlegelők_tragédiája.

Példa a fogolydilemma kifizetési mátrixára

	Kooperálás	Defektálás
Kooperálás	3, 3	0, 5
Defektálás	5, 0	1, 1

A "nyer-veszt" terminológiát használva a táblázat a következőképpen néz ki:

	Kooperálás	Defektálás
Kooperálás	nyer-nyer	többet veszt-többet nyer
Defektálás	többet nyer-többet veszt	veszt-veszt

Fogolydilemma

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu

SZÉCHENYI 2020
MAGYARORSZÁG KORMÁNYA
Európai Unió
Európai Szociális Alap
BEFEKTETÉS A JÖVŐBE

Forrás: https://hu.wikipedia.org/wiki/Fogolydilemma

5.2.2 Parfit értelmezése

A fogolydilemmának két felettébb zavaró tulajdonsága is van.

Egyrészt ha mindketten racionálisan viselkednek (dezertálnak), akkor mindketten rosszabb eredményt (-6, -6 év börtön) érnek el, mintha mindketten nem-racionálisan viselkednek (kooperálnak), amikor az eredmény -1, -1 év börtön.

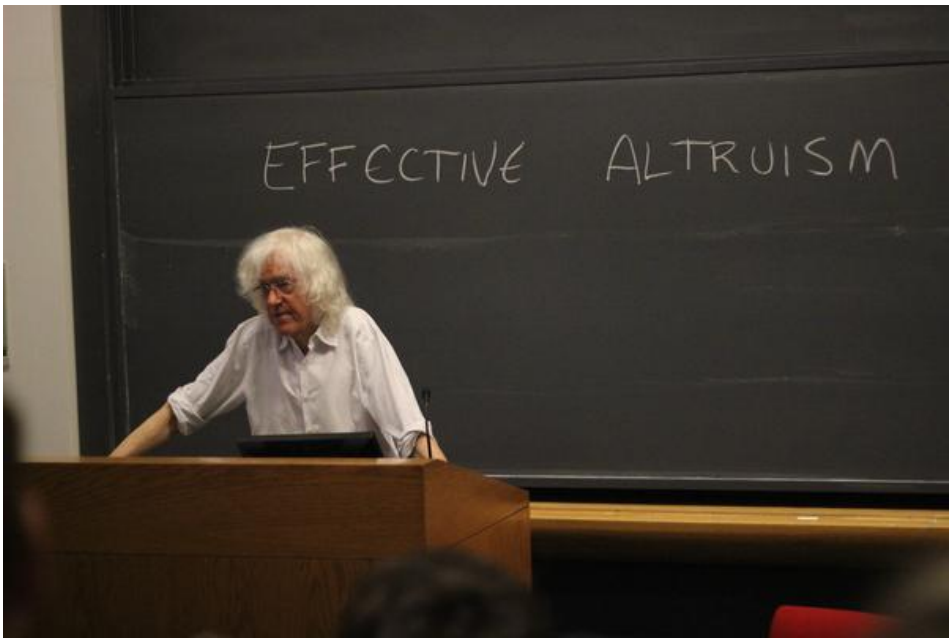
Másrészt a kölcsönösen racionális viselkedés az érintettek számára a legrosszabb kimenetelhez (-12 év) vezet. Tehát a racionalitás – Platt szavaival éve – *láthatatlan ökölként* súlyt le a közjóra.

kollektív viselkedés	individuális nyereség	kollektív nyereség	Megjegyzés
CC	-1 év, -1év	-2 év	nem-racionális viselkedés & csoportmaximum
DC	0 év, -10 év	-10 év	
CD	-10 év, 0 év	-10 év	
DD	-6 év, -6 év	-12 év	racionális viselkedés & csoportminimum

Individuális és kollektív nyereség a fogolydilemmában.

Derek_Parfit a kooperálást egyénileg önpusztítónak nevezi, hiszen ez a döntés azt jelenti, hogy valaki egy nem domináns stratégiát választ egy domináns stratégiával szemben. Ezzel szemben a dezertálást kollektívan önpusztítónak tartja, hiszen ha ez a tipikus döntés, akkor a két fő közös (kollektív) nyeresége minimális lesz. Parfit szerint az ellentmondás a következő: „*Mindenkinek egyenként* kell-e a maga számára a lehető legjobb eredményre törekednie; vagy *nekünk együtt* kell a legjobb eredményt biztosító viselkedést követnünk? Ezeket az eseteket „egyenként-együtt" dilemmáknak (Each-We Dilemmas) Parfit (1998, 145) szerinte a kanti morál lényege éppen az, hogy a szereplők elmozduljanak *az egyenként* logikájától *az együtt* logikája felé.

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu



Derek Parfit (1942-2017)

Forrás: https://en.wikipedia.org/wiki/Derek_Parfit#/media/File:Derek_Parfit_at_Harvard-April_21,_2015-Effective_Altruism.jpg

5.2.3 Általános akarat

A fogolydilemma és a potyautas helyzetére jól alkalmazható Rousseau megjegyzése az akarat és az érdek különböző formáiról. *„A közakarat gyakran eltér az általános akarattól; ez utóbbi csak a közérdeket nézi, míg az előbbi a magánérdekeket, s nem több a különös akaratok összegénél."* A magánérdek (volonté particuliére) arra ösztönzi a feleket, hogy dezertáljanak, ezzel párhuzamosan a magánérdekek mechanikus összege a mindenki akarata (volonté de tous) is a dezertáláshoz vezet, szemben az *általános akarattal* (Volonté _générale), amely együttműködésre ösztönöz.

A fenti idézet így folytatódik: *„De ha a különös akaratokból elvesszük azt, amiben az egyik több vagy kevesebb a másiknál, márpedig ezek a különbségek kölcsönösen megsemmisítik egymást, úgy a kivonás eredményeként az általános akarat marad fenn."* (Társadalmi szerződés, II. könyv. 3. fejezet).

Úgy gondolom, hogy Rousseau azért írta ezt, mert a saját korában még így volt. Abban az időben a racionalitásban még nagyon erősen jelen volt a kollektív racionalitás és az értékracionalitás elemei. S így volt remény arra, hogy az emberek egy ilyen helyzetben az erényt és az általános akaratot kövessék. "Az erény nem más, mint a személyes *volonté*

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu

particuliére hozzáigazítása a nyilvános *volonté généralé*-hoz"- mondja Rousseau.

Napjainkban – összhangban az elmúlt évszázadok individualizációjával – azonban teljesen más a helyzet. A racionalitás fogalma alatt ma már a többség az individuális racionalitást érti. Tehát egy ilyen helyzetben,'a *kivonás eredményeként mindenki dezertálása marad fenn.'*

A fogolydilemma a népszuverenitás és demokratikus döntéshozatal problémájára is felhívja a figyelmet. Ha mindenki akarata egy népszavazás esetében nem vezet el az általános akarathoz, sőt a közösség számára legrosszabb kimenetelt, a kölcsönös dezertálást jelenti, akkor hogyan bízhatunk a népszuverenitásban és a többségi akaratban, mint legfelső fórumban.

Ezeknek a tényeknek az ismeretében térjünk vissza a nomenklatúra kérdésére. Itt két kérdés is felmerül. Miért nevezzük a tagadást kooperálásnak és az árulást dezertálásnak? Továbbá miért tekintjük a kooperálást nem-racionális, míg a dezertálást racionális viselkedésnek?

Azért nevezzük a tagadást kooperálásnak, mert a tagadó csökkenti a másik rabló, illetve a „rablóbanda" büntetését. Tehát a rablók szempontjából helyes a hallgatót kooperálónak nevezni, mert ha a kooperálás a tipikus viselkedés, akkor maximális a rablók közös nyeresége. Ezzel szemben az áruló egyaránt növeli a társa, illetve a banda közös büntetésének a mértékét. Tehát a rablók szempontjából helyes az árulót dezertálónak nevezni, mert ha a dezertálás a tipikus viselkedés, akkor minimális a rablók közös nyeresége.

A második kérdés sokkal problematikusabb. Vegyük számba a lehetőségeket.

(i) Nevezhetjük a *dezertálást* racionális viselkedésnek, ragaszkodva a nem-kooperatív játékelmélet eredeti terminológiájához. Ennek a megoldásnak az az előnye, hogy a 'racionalitás' értelmezése összhangban marad a köznapi és a közgazdasági értelmezéssel. A hátránya viszont az, hogy esetenként a racionalitás – Platt szavaival éve – *láthatatlan ökölként* súlyt le a közjóra.

(ii) Nevezhetjük a *kooperálást* a racionális viselkedésnek. Elsősorban etikai megfontolások szólnak emellett. Ha alkalmazzuk az aranyszabályt, a kategorikus imperatívuszt vagy az utilitarista princípiumot a fogolydilemmára, akkor azt kapjuk, hogy a helyes válasz a kooperálás. Sajnos az emberek nem így gondolkodnak. Ezt a problémát vizsgálta Hofstadter (1983) egy félig empirikus tanulmányában. Ő is azt várta, hogyha mindenki világosan látja, hogy mások is fogolydilemma helyzetben vannak, akkor a szereplőket a racionálisan felfogott önérdek elvezeti a kooperáláshoz. Nem így történt; akik kooperáltak azok mind azt hangoztatták, hogy morális és nem racionális okok miatt kooperáltak. Ez az eredmény Hofstadtert meglepte, de nem titkolta el.

(iii) Nevezhetjük *mindkét viselkedés* racionálisnak. A játékelmélet a dezertálást individuális, míg a kooperálást kollektív racionális viselkedésnek tekinti. Ennek a megoldásnak az a nehézsége, hogy a racionális önérdek fogalmát értelmezés és kontextus függvényévé teszi. A racionális önérdek túl fontos fogalom a társadalomban ahhoz, hogy a jelentése értelmezési vitáktól függjön.

Véleményem szerint a társadalom számára csak az (i) járható. Tehát a dezertálás jelenti a racionális viselkedést. Elsősorban azért mert ha arra kérünk

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu

valakit, hogy a racionális önérdeke alapján döntsön, akkor ő mindig a dezertálást, vagyis a szigorúan domináns stratégiát fog választani. Nincs az a bölcsész okoskodás, amelynek elhinné, hogy a kooperálás, vagyis egy szigorúan dominált opció választása jelenti az önérdekét.

Ez azonban nem a „világvége", hiszen az átlagember jól tudja, hogy a ráció csak lehetséges viselkedési mód sok más mellett. Számára nem probléma elfogadni, hogy a racionális önérdek esetenként rossz eredményre vezet, ezért esetenként nem racionálisan (erkölcsi vagy éppen érzelmi alapon) kell döntenünk. Bár ez azt jelenti, hogy a ráció felett kell lenni egy felsőbb hatalomnak (Istennek, tradíciónak, morálnak, szokásrendszernek, evolúciónak), aki vagy ami megmondja, hogy mikor kell és mikor nem szabad a rációt követni. A racionalisták számára azonban nem fogadható el, hogy nem a ráció a legfelső autoritás.

Kérdések

1. Ismertesse a klasszikus fogolydilemmát!
2. Mit jelent a „nyer-nyer" (win-win) kifejezés?
3. Miért nevezi Parfit a fogolydilemmát egyenként-együtt dilemmának?
4. Az általános akarat fogalma hogyan értelmezhető a fogolydilemmára?
5. Mit jelent a láthatatlan ököl fogalma a fogolydilemmában?

Irodalom

Hofstadter D.R (1983): Computer tournaments of the Prisoner's Dilemma suggest how cooperation envolves. *Scientific American* 248 May pp 14-20 o.

Parfit, Derek (1998). Körültekintés, erkölcsiség és a fogolydilemma. In Csontos László (szerk.): *A racionális döntések elmélete*. Budapest: Osiris–Láthatatlan Kollégium.

Platt, J. (1973): Social Trape. *American Psychologist*, 128, 8, pp. 641-651.

Rousseau, Jean-Jacques (1997): *A társadalmi szerződésről*. Budapest: PannonKlett.

Tóth I. János (2014/b): A fogolydilemma kiterjesztése. *Magyar Tudomány*175:(1) pp. 90-98.

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu