TÁJÉKOZÓDÁS

Huoranszki Ferenc

DÖNTÉSELMÉLET ÉS ERKÖLCSI NORMÁK[*]

Quidquid agis, prudenter agas et rescipe finem.

A racionális döntések elméletének és a morális normakövetés cselekvésmagyarázatban betöltött szerepének viszonya problematikus. Szinte senki sem kételkedik abban, hogy cselekvéseinket nagymértékben meghatározzák azok az erkölcsi és más normák, amelyeket helyesnek fogadunk el, és amelyek ismerete elengedhetetlen mások viselkedésének megértéséhez. Mégis sokaknak úgy tűnik, hogy az erkölcsi normák szerinti cselekvés nem lehet része a racionális döntések elméletének.[1] A normakövető viselkedésről adott beszámolók jó része vagy azon alapszik, vagy elfogadásukból úgy látszik, az következik, hogy az erkölcsi normák szerinti döntés nem lehet része az önérdeken alapuló racionális kalkulációnak. Kétségtelen, hogy ha a racionális döntések elméletét oly módon értelmezzük, mint az önérdeken alapuló viselkedés modelljét, akkor bizonyos értelemben a morális megfontolásokon alapuló cselekvést máris a racionális döntések elméletének érvényességi körén kívül helyeztük el.

A racionális döntések elmélete azonban nem kötődik olyan szorosan az önérdek fogalmához, ahogyan azt gyakran feltételezik. A félreértés egyik oka talán az, hogy a társadalomtudományos magyarázatokról szóló elméletekben a racionális döntések elmélete a „módszertani individualizmus" tipikus reprezentánsaként szerepel, és sokan ebből arra következtetnek, hogy ez valamiféle „szubsztantív individualizmust" (tehát az egoizmus általános motivációs erejébe vetett meggyőződést) is maga után von. Ám ez természetesen nincs így. A másik, ehhez szorosan kapcsolódó magyarázat e félreértésre talán az lehet, hogy a racionális döntések elmélete az eredetileg a mikroökonómiában használt modelleket kiterjesztette a társadalomtudományok más területeire. A közgazdaságtan történetében pedig tagadhatatlanul jelentős szerepe van a *homo economicus*, az önérdekét követő egyén feltételezésének, hiszen a közgazdaságtan egyik alapvető kérdése Adam Smith óta arra vonatkozik, hogyan lehetséges, hogy az önérdeküket követő egyének cselekedetei egy társadalmilag optimálisan működő gazdasági rendszert hoznak létre.

[*] Köszönettel tartozom Orthmayr Imrének és Ambrus-Lakatos Lorándnak a tanulmány egy korábbi változatához fűzött kommentárjaikért.

[1] Tipikus példaként említhető az egyébként a normakövető viselkedés szerepét egyáltalán nem alábecsülő Elster, aki szerint „A racionális cselekvést, akár gazdaságilag, akár politikailag motivált, az eredmény érdekli." Ezzel szemben „A társadalmi normák vezérelte cselekvés nem az eredményekhez igazodik." (Elster 1995: 117).

Mint azt később szeretném megmutatni, az az elképzelés, miszerint a racionális döntések elméletének használhatósága az egoista motivációs feltevéshez kötődik, valóban összefügg a döntéselmélet közgazdaságtanbeli használatával, ám a kapcsolat sokkal mélyebben rejlik, s nem azon a meglehetősen felszínes (és talán hamis) feltételezésen nyugszik, miszerint a közgazdaságtani elméletek csak akkor működnek, ha kizárólag a szélsőségesen önző, egoista motivációkat tartják relevánsnak egy gazdasági rendszer működésének megértésekor. Ehhez azonban először is új megvilágításba kell helyeznünk a racionális döntésekre vonatkozó elmélet és a morális normakövetés viszonyát érintő problémát. A probléma szerintem a következőben áll: amennyiben a racionális döntések elméletére úgy tekintünk, mint a cselekvést megelőző döntés szerkezetének logikai-matematikai reprezentációjára, a morális normák követése melletti döntés, úgy tűnik, nem helyezhető el e megfontolások logikai terében. Az „úgy tűnik" tagmondat szerepe a fenti állításban nem csupán stiláris. Azt szeretném megmutatni, hogy létezik a racionális döntések elméletének egy olyan változata, amelyik képes a morális normák motiváló erejét beilleszteni a racionális megfontolások logikai terébe.

Az *1. fejezetben* azt szeretném pontosabban kifejteni, miben is áll racionalitás és moralitás viszonyának problémája. A *2. fejezetben* bemutatom és osztályozom azokat a főbb megoldási kísérleteket, amelyek a két állítólag egymással szembenálló motivációs feltevés viszonyának meghatározására vonatkoznak. A *3. fejezetben* vázlatosan ismertetem azt a döntéselméleti modellt, amelyet az előző fejezetben tárgyalt valamennyi megoldás feltételezni látszik. Végül a *4. fejezetben* azt szeretném megmutatni, hogy e modell módosítása segítségével hogyan juthatunk közelebb racionalitás és moralitás viszonyának megértéséhez.

1. A probléma

Mit értek tehát azon, hogy a morális normák alapján meghozott döntés nem helyezhető el a racionális döntést reprezentáló elmélet logikai terében? Nyilvánvaló, hogy a kijelentés, miszerint valami kívül áll valamilyen logikai téren, metaforikus. Ám ez nem jelenti azt, hogy ne tehetnénk egzakttá. A pontosítást talán érdemes annak vizsgálatával kezdeni, *mire nem* vonatkozik ez a kijelentés.

A morális normákkal kapcsolatos egyik (talán alapvető) filozófiai kérdés e normák megalapozására, illetve megalapozhatóságára vonatkozik. Bizonyos normákat egyszerűen helyesnek fogadunk el. A legjobb vagy legalábbis a leggyakrabban használt példák a tízparancsolatból származnak, amelyek közül van még néhány, amit ma is megfontolásra érdemesnek tartunk: ilyenek például a ne ölj, ne lopj, ne tanúskodj hamisan normái, vagy ilyen az ígéretek megtartására vonatkozó norma. Ezeket a normákat erkölcsi normáknak szokás tekinteni, mert azt feltételezzük róluk, hogy nem pusztán társadalmi konvenció termékei. Ezt a feltételezést persze korántsem osztja minden filozófus, de érvelésünk szempontjából ennek nincs is nagy jelentősége. Talán csak annyit érdemes megjegyezni, hogy e normákról feltételezzük, más típusú előírások, mint mondjuk azok, amelyek az étkezésre vagy az egymás köszöntésére vonatkoznak, vagy mások, mint a KRESZ szabályai.

Miután az erkölcsi normákról intuitíve feltételezzük, hogy elfogadásuknak a puszta megegyezésnél erősebb alapokon kell nyugodnia, e normákkal kapcsolatban természetesen módon merül fel megalapozásuk lehetőségének kérdése. E megalapozhatóság pedig nyilvánvaló módon összefügg a racionalitással, hiszen a megalapozás kérdése általában arra vonatkozik, miért nem a puszta önkény vagy véletlen következménye, hogy épp ezeket a normákat fogadjuk el. A kérdés tehát e normák elfogadásának ésszerű alapjaira vonatkozik. Látnunk kell azonban, hogy az elfogadás ésszerű alapja nem csak a racionális döntések elmélete értelmében vett racionalitás lehet. Létezik a racionalitásnak egy tágabb, nem a logikai konzisztencia és az optimalizáló viselkedés fogalmain nyugvó értelmezése is, amely fontos szerepet játszhat a normák megalapozásában, de amely kívül áll a racionális döntések elméletének hatókörén. Lehetséges – sőt szerintem bizonyos –, hogy erre a racionalitásfogalomra van szükségünk akkor, amikor meg akarjuk alapozni a morális normákat.

Egy példa talán segít annak megértésében, mire gondolok. Azt az embert, aki gyakran képtelen arra, hogy legjobbnak tartott döntéseinek megfelelően cselekedjék, akaratgyengének nevezzük. Általánosan elfogadott nézet, hogy az akaratgyengeség az irracionalitás egy formája, sőt adott esetben (vagy legalábbis egy külső megfigyelő értékelése szerint), a cselekvő számára végzetes következményekkel is járhat. Ám a racionális döntések elmélete az irracionalitásnak erről a formájáról nemigen tud mit mondani, hiszen az akaratgyengeség a meghozott döntés és a cselekvés viszonyára vonatkozik, nem pedig a döntés logikai struktúrájára. Ettől még helyes azt állítani, hogy az akaratgyenge ember irracionálisan viselkedik. De az a racionalitásfogalom, amit ebben az esetben használunk, nem lehet azonos azzal, amit a racionális döntések elmélete tételez. Ennek oka igen egyszerű. A döntéseknek megfelelő viselkedés melletti döntésről nincs értelme matematikai reprezentációt adni. Egyszerűen tudjuk, hogy aki nem legjobb belátása szerint cselekszik, az irracionális.[2]

Van tehát a racionalitásnak egy olyan fogalma, amely nem azonos azzal, amit a döntéselmélet vizsgál, és nem kizárt, hogy a morális normák megalapozhatóságának kérdése ennek segítségével könnyebben megközelíthető. Az a probléma azonban, amelyet itt vizsgálni szeretnék, nem a megalapozhatóságra vonatkozik, illetve csak közvetve vonatkozik arra. Annyit persze fel kell tételeznünk, hogy aki morális normák alapján cselekszik, az nem eleve irracionális lény. Fel kell tennünk, hogy a morális normák szerinti cselekvés nem az akaratgyengeség egy furcsa formája. Talán nem mindenki ért ezzel egyet. Azok viszont, akik nem értenek ezzel egyet, sokkal inkább egy problémára hívják fel a figyelmünket, mintsem megoldási javaslattal állnak elő.

A morális normák tartalmának problémájához szorosan kötődik egy másik tradicionális filozófiai kérdés, az tudniillik, hogy *honnan származik e normák motivációs ereje*. Úgy tűnik, hogy ez utóbbi kérdés vizsgálata közelebb visz a döntéselmélet és a morális normák kapcsolatának megértéséhez. A normák tartalmára vonatkozó probléma talán kívül marad a racionális döntések elméletének hatókörén, hiszen a tartalom meghatározása esetleg nincs is közvetlen kapcsolatban a cselekvéssel. El-

[2] Ez nem jelenti azt, hogy az irracionális viselkedés minden típusa kívül állna a racionális döntések vizsgálatának körén. Az irracionális preferenciák kialakulásáról például adható döntéselméleti modell (vö. Elster 1983, valamint Csontos 1985 vonatkozó részeit).

képzelhető, hogy nem lehet megindokolni, miért gondoljuk azt, hogy morálisan helyesnek tarjuk az ígéretek betartását, és helytelennek a hazudozást. Lehetséges, hogy csak a kulturális környezet kondicionálta esetlegességről van szó, amely nem alapozható racionális megfontolásra vagy döntésre. Ám ha tartalmuk meghatározásának nem is, e normák motivációs erejének valamilyen módon szerepet kell kapnia a döntéselméletben.

Úgy vélem, ez a feltételezés igaz is, meg nem is. Kétségtelen, hogy a motivációs erő egy bizonyos, rendkívül sovány értelemben szerepet játszik a normák és a döntéselmélet viszonyának tisztázása során. E sovány, ámde fontos értelmezés szerint *a morális normáknak megfelelő cselekvésnek a cselekvő értéket tulajdonít*. Látni fogjuk, milyen nagy szerepe lesz további érveink során e meglehetősen banális állításnak. Persze amikor a filozófusok az erkölcsi motiváció problémájára kérdeznek rá, nem azt a tényt firtatják, hogy a cselekvő értékeli-e a morális normák szerinti cselekvést. Ha ugyanis csak annyit mondunk, hogy a morális cselekvést értékeljük, és ezért motivál, akkor úgy tűnik, a moralitást egy bizonyos szempontból besoroltuk a vaníliafagylalt-nyalás, a technometál zene hallgatás vagy a ringlispílen ülés kategóriájába – hiszen ezeket is értékeljük, éppen ez motivációs erejük alapja. De ha nincs különbség a morális viselkedés értékelése és az utóbb felsoroltak értékelése között, akkor a morális cselekvést besoroltuk a *de gustibus non est disputandum* kategóriájába. Nem abszurd ez?

Amellett szeretnék érvelni, hogy bizonyos értelemben nem. Minden attól függ, milyen szerepet tulajdonítunk a *de gustibus*-elvnek. Ha azt értjük rajta, amit szó szerint jelent, tehát hogy az ízlésítéletekről nem lehet vitatkozni, akkor az elv nemcsak a morális értékelés esetében, hanem szinte bármely cselekvés vagy jószág esetében egyszerűen hamis. Talán legalapvetőbb biológiai adottságaink által meghatározott ízlésünkre áll, hogy nem lehet róla vitatkozni (mondjuk, bizonyos ízek iránt érzett vonzódás lehetne egy triviális példa, vagy valaki kedvenc színe), de szinte valamennyi más ízlésítéletünkről nemcsak hogy lehet vitatkozni, de ténylegesen vitatkozunk is róluk. Nem kell tehát a morál magaslataira emelkednünk ahhoz, hogy belássuk, ha szó szerint értelmezzük, az elv egyszerűen nem állja meg a helyét. Csakhogy a *de gustibus*-elv nem arra vonatkozik, hogy lehet-e vitatkozni az ízlésről, hanem arra, hogy mit kell a racionális döntés során adottnak vennünk. Tehát nem egy tényállítás, hanem egy posztulátum. Ebben az értelemben viszont nincs semmi abszurd vagy visszataszító abban, hogy a *de gustibus*-elvet alkalmazzuk a morális normák által előírt cselekvésre. Semmi mást nem jelent ez, minthogy posztuláltuk: a cselekvő értéket tulajdonít a normavezérelt cselekvésnek.

De akkor miért neveztem a motiváció ezen értelmezését soványnak? Azért, mert a filozófusok sokkal többet várnak a morális normák motivációs erejének elemzésétől, túl azon, hogy elfogadják, a cselekvő értéket tulajdonít nekik. Először is a normák tartalma és a motivációs erő, szemben az értékelés más típusaival, a morális normák esetében nem független egymástól. Sokan gondolhatják, hogy a vaníliafagylalt finom, de ebből még nem következik, hogy azt várnánk tőlük: igyekezzenek vaníliafagylalt közelébe jutni. Ellenben, ha valaki morálisan helyesnek tartja, hogy a parkban ne szemeteljen, akkor azt is várjuk, hogy valóban ne dobja el a kezében lévő csokoládépapírt.

Másodszor, a morális normákat univerzalizálhatónak tekintjük, tehát úgy gondoljuk, hogy adott helyzetben vagy körülmények között a helyesnek tartott norma alapján kellene cselekednie. Ezt azonban nem követeljük meg általában az ízléstől: vitatkozhatunk ugyan valakivel azon, hogy milyen badarságokat tart értékesnek, de azért a legmegátalkodottabb ízlésdiktátoroktól eltekintve, azt egyikünk sem kívánná, hogy mindenki ugyanazt szeresse, amit ő.

Végül a morális vitáknak van egy olyan jellegzetessége, amely nyilvánvalóan megkülönbözteti őket az ízlésvitáktól. A morális viták tárgya általában a morális konfliktushelyzet. A morális dilemma olyan helyzet, mikor két, általunk helyesnek tartott norma ellentétes cselekedetet ír elő. A puszta értékelésként értelmezett motiváció az ilyen helyzetről mindössze annyit tudna mondani, hogy azért tette valaki *A*-t *B* helyett, mert *A* normát többre értékelte. És ez bizony sovány állítás. Ennél jóval többet kell és lehet mondani arról, miért dönt valaki egy morális konfliktushelyzetben egyik vagy másik norma követése mellett.

De ha ennyire sovány a normák motivációs erejének a *de gustibus*-elv alapján történő értelmezése, akkor miért lehet ennek az értelmezésnek mégis jelentősége a racionális döntések elmélete és a morális normák viszonyának vizsgálatakor? Azért, mert mint fentebb láttuk, a döntéselmélet egyik posztulátuma, hogy az értékelést adottnak veszi, nem kutatja az értékelés melletti indokokat vagy az értékelés okait. Ezért akkor érthetjük meg igazán, mit jelent az, hogy az erkölcsi normákon alapuló viselkedés kívül marad a döntéselmélet logikai terén, ha pusztán abból a szempontból vizsgáljuk e normákat, amilyen szempontból legalábbis *elemei lehetnének* a döntéselméletnek. S ebből a szempontból a normák releváns tulajdonsága az, hogy a cselekvő (bármilyen alapon tegye is ezt) *értékeli* a normáknak megfelelő viselkedést.

Ennek feltételezése azonban nem elegendő ahhoz, hogy elhelyezhessük a morális normákat a racionális döntéseket reprezentáló elmélet logikai terében. A probléma abból ered, hogy amennyiben a morális viselkedés értékelését be akarjuk illeszteni a döntéselmélet logikájába, úgy a cselekvést megelőző megfontolások szerkezetéről egy olyan reprezentációt kell tudnunk adni, amely lehetővé teszi a morális és az egyéb értékelés alapján hozott döntések együttes vizsgálatát. Ám, ezt sem a klasszikus morál-, illetve társadalomfilozófiai elméletek, sem pedig a döntéselmélet általában elfogadott és leggyakrabban használt változata nem teszi lehetővé. Mint azt szeretném megmutatni, valóban kettőn áll a vásár: a racionális döntések elméletének leggyakrabban használt változata mintha arra kényszerítené a morális normák motiváló erejét megérteni kívánó filozófusokat, hogy a normákat követő viselkedést egyszerűen kívül helyezzék a döntéselmélet logikai terén. A döntéselméletészek pedig gyakorta lelkesen asszisztálnak ehhez az elegáns kivonuláshoz: megelégszenek azzal, hogy a morális normákat furcsa korlátoknak tekintik.

Én azonban nem vagyok elégedett ezzel a megoldással, amely sokkal több nehézséget okoz, mint amennyit megválaszol, és sokkal abszurdabb következményekhez vezet, mint egy olyan kísérlet, amely megpróbálja az erkölcsi normákat elhelyezni a cselekvést megelőző megfontolások döntéselmélet által vizsgált logikai terében. Azért gondolom ezt így, mert úgy hiszem, jelen pillanatban nincs jobb elméletünk e megfontolások struktúrájának vizsgálatára, mint a racionális döntések elmélete. Ugyanakkor arról is meg vagyok győződve, hogy a morális viselkedés melletti dön-

tés nem automatikus, reflexszerű és nem is irracionális. Ebből ugyanis az következne, hogy egy normális mentális élettel rendelkező ember valójában két személy: egy, amelyik képes a racionális megfontolásra és döntésre, és egy másik, amelyik időnként érthetetlen okokból furcsa normákat követ. Úgy tűnik, a jelenlegi elméletek többsége így tekint ránk, cselekvőkre. A következő fejezetben azt szeretném röviden megvizsgálni, melyek az erkölcsi normákon alapuló, illetve a racionális viselkedés kapcsolatára vonatkozó logikailag lehetségesek álláspontok.

2. Erkölcs és racionalitás

A következőkben tehát nem arra törekszem, hogy egyes filozófusok vagy társadalomtudósok racionalitás és moralitás kérdésében elfoglalt álláspontját elemezzem, inkább a logikailag lehetséges megoldásokat szeretném szemléltetni. Valamennyi álláspontnak van azonban jelentős képviselője.

Azok az elméletek, amelyek a döntéselmélet és az erkölcsi normák viszonyát tárgyalják, alapvetően két nagy csoportra oszthatók. Az egyik csoportba azok az elméletek tartoznak, amelyek szerint valójában csakis az önérdekkövető döntés tekinthető racionális megfontolás eredményének, ezért az erkölcsi normáknak vagy levezethetőknek kell lenniük az önérdekkövető viselkedést megelőző megfontolásokból, vagy ha ez nem lehetséges, akkor a morális normák szerinti cselekvést egyszerűen irracionálisnak kell tekintenünk. Az elméletek egy másik csoportja szerint a morális normák követése ugyan nem irracionális, de kívül marad az önérdekkövetés értelmében vett racionalitás fogalmán, amelyet a döntéselmélet reprezentálni hivatott, és egy attól teljesen különböző motivációs struktúra alapján magyarázandó. Ennek a megközelítésnek viszont szembe kell néznie azzal a már fentebb említett nehézséggel, hogy anélkül osztja ketté a cselekvő személyiségét, hogy aztán számot tudna adni az integráció elveiről. Másképp fogalmazva: nyilvánvaló módon az is döntés kérdése, hogy egy adott helyzetben a racionális érdekkalkuláció normáját követjük-e, avagy a morális normák alapján cselekszünk. Ám, ha a kétféle motiváció szerinti cselekvést logikailag más térbe helyezzük, akkor nincs értelme azt vizsgálni, hogy milyen elvek alapján dönthetünk köztük: döntésünk teljesen irracionális lesz. A morális döntés struktúrájának megértésébe tehát ekkor is óhatatlanul beszüremkedik az irracionalitás. Ebben az esetben talán nem tekinthető irracionálisnak a morális norma szerinti cselekvés általában, mégis minden egyes olyan döntés, amely adott helyzetben a morális motiváció relevanciája mellett vagy ellen szól, irracionálissá válik.[3]

Azok az elméletek tehát, amelyek szerint a morális döntések motiváló erejének levezethetőnek kell lennie az önérdekkövető döntés racionalitásából, további két csoportra oszthatók. Az egyik elmélet szerint nem létezik ilyen levezetés, ezért a morális motiváció szerinti cselekvés végső soron irracionális. A másik megközelítés szerint a moralitás csak azért nem irracionális, mert levezethető a racionális és önérdekkövető viselkedésből. Mindkét elmélet szemben áll a morális normákkal kapcsolatos

[3] Ez az alapja a Sartre és Camus által képviselt egzisztencialista morálfilozófiának: a morális és amorális viselkedés közti választás sohasem racionális megfontolásokon, hanem mindig irracionális döntésen nyugszik.

mindennapi meggyőződéseinkkel, de ez önmagában még nem elégséges ok arra, hogy elvessük őket. Bár arra talán elégséges ok, hogy némi gyanú merüljön fel velük kapcsolatban. Azt szeretném megmutatni, hogy a gyanú nem alaptalan.

Az egyik elképzelés szerint a morális normák alapján történő cselekvés irracionális. Ennek okát abban vélik felfedezni, hogy a cselekvő a norma alapján történő cselekvésnek *objektív értéket* tulajdonít. Valójában azonban az értékelés pusztán szubjektív (tehát a cselekvő kontingens mentális állapotától függő) attitűd. Miután az objektív értékekben való hit irracionális, az efféle meggyőződésen alapuló cselekvésnek is irracionálisnak kell lennie. Gilbert Harman szerint: „A morális hipotézisek nem segítenek abban, hogy meg tudjuk magyarázni, miért éppen azt figyeljük meg, amit megfigyelünk. Ezért az etika problematikus, és a nihilista álláspontot komolyan kell vennünk… A nihilizmus szélsőséges változata szerint a moralitás egyszerűen illúzió. E változat szerint el kell fordulnunk a moralitástól, éppúgy ahogyan az ateista elfordul a vallástól miután úgy döntött, hogy a vallási tények nem segítenek a megfigyelhető jelenségek magyarázatában" (Harman 1977: 11).

Ám Harman szerint ez a következmény nem szükségszerű, mivel annak a ténynek, hogy a morális normák nem segítenek a megfigyelhető viselkedés magyarázatában, létezik egy kevésbé szélsőséges értelmezése is. „... egy mérsékeltebb nihilizmus szerint a morális kijelentéseknek nem az a szerepük, hogy leírják a világot, hanem hogy kifejezzék morális érzéseinket, avagy hogy másoknak vagy magunknak címzett felszólítások legyenek" (Harman 1977: 12). Mi a probléma ezzel a megközelítéssel?

Azt gondolom a következő. A nihilista álláspont arra a kérdésre keresi a választ, hogy kell-e tételeznünk az objektív értékekre vonatkozó „erkölcsi tényeket" annak érdekében, hogy meg tudjuk magyarázni a megfigyelt viselkedést. Ez a kérdésfeltevés nem érdektelen, de sajnos elég homályos, és nem biztos, hogy a legszerencsésebb módja az erkölcsi motivációk mibenlétére vonatkozó kérdésnek. Lehetséges, hogy az „erkölcsi tények" posztulálása valóban nem segít a megfigyelhető cselekvés magyarázatában, különösképpen, ha előbbieket olyan metafizikai entitásoknak tekintjük, amelyek azonosítási kritériumait nem ismerhetjük.

Egy kérdés azonban, hogy milyen típusú tényeket tételezünk, és egy másik, hogy a morális normák cselekvő általi *elfogadása* szerepet játszik-e egy cselekvés magyarázatában. Gyakran magyarázunk egy cselekvést azzal, hogy azt azért tette valaki, mert úgy vélte, morálisan helyes. Harman szerint a cselekvő azon meggyőződése, hogy amit tesz, helyes, csupán egy „társadalmilag kondicionált érzékenységből" fakad. A mi kérdésünk azonban nem arra vonatkozik, honnan *ered* a morális normák motiváló ereje, hanem arra, hogy hogyan helyezhető el a cselekvést megelőző megfontolások logikai terében. Erről azonban a harmani megközelítés nem mond semmit. A normák szerinti cselekvést egyszerűen azonosítja az érzelmi alapú cselekvéssel, és ezzel száműzi is a racionális megfontolások területéről. Csakhogy, mint azt később szeretném megmutatni, ez a morális motiváción alapuló cselekvés szerkezetéről kialakított kép teljesen hamis, mégpedig éppen azért, mert nem helyes magyarázata a megfigyelhető cselekvésnek.

Persze nem azt akarom tagadni, hogy az érzések és érzelmek szerepet játszanak a cselekvést meghatározó döntéseinkben. Abból azonban, hogy részben érzelmeink

magyarázzák, miért értékelünk egy adott normát, még nem következik, hogy a norma szerinti döntést ki kellene vonnunk a racionális megfontolások köréből. Az érzelmek kiegészítői lehetnek egy, a normák motiváló erejével kapcsolatos elméletnek, de nem kell kizárniuk azokat a racionális megfontolás köréből.

Harman szerint elvileg létezik egy harmadik megközelítés is a moralitás mibenlétének megértésére, amely szerint az erkölcsi tényeket végső soron az érdek fogalmára kell visszavezetnünk. Bár ő maga nem ért egyet ezzel a megoldással, természetesen ez is egy logikailag lehetséges álláspont. Legjelesebb kortárs képviselője David Gauthier, aki szerint a morális normák megalapozhatóságával kapcsolatos probléma egyetlen lehetséges megoldása, ha sikerül megmutatnunk, hogy a morális normák elfogadásának racionalitása végső soron az önérdekkövető viselkedés racionalitásából származik. Hogyan lehetséges ez? Úgy, hogy bizonyos esetekben az önérdekkövető racionalitás által meghatározott cselekvés paradox eredményhez vezet. Ebben az esetben mintha „az ész önmagával kerülne ellentmondásba" (hogy kissé anakronisztikusan Kantot idézzük), s ilyen esetben a morális normák sietnek a zavarba jött döntéshozó segítségére.

E megközelítés lényegét egy példa segítségével világíthatjuk meg. A példa az olyan társadalmi szituációkra vonatkozik, amelyeket a fogolydilemma segítségével szoktunk jellemezni. A fogolydilemma típusú szituációban az önérdekkövető, racionális viselkedés olyan eredményhez vezet, amelyhez képest létezik egy minden résztvevő számára kedvezőbb alternatíva. A játékelmélet nyelvén kifejezve: a domináns stratégia alkalmazása Pareto-szuboptimális következményekkel jár. Minden egyes cselekvő önérdek-maximalizáló kalkulációja olyan kollektív eredményhez vezet, amelyikkel egyik szereplő sem elégedett, mivel mindannyiuk számára lehetséges lenne egy jobb eredmény. Ha azonban ez így van, úgy tűnik, bizonyos esetekben nem racionális a racionális haszonmaximalizáló stratégia alkalmazása. Ez bizony paradox következmény, és az elmélet szerint itt jöhet segítségünkre a morál. Gauthier (1991: 17–18) elmélete szerint tehát a morális norma az egyéni, haszonmaximalizáló kalkuláción a haszonmaximalizáció érdekében érvényesülő *korlát.*

Az az elképzelés, amely szerint a morális norma az önérdekkövető haszonkalkuláción értelmezett korlát, nagy múltra tekint vissza, és később még részletesebben is foglalkoznunk kell vele. Előbb azonban azt kell megvizsgálnunk, hogy amennyiben elfogadjuk a morális normák ezen értelmezését, vajon előbbre jutunk-e a normák motivációs erejének a racionális döntések logikai terében történő elhelyezésével. Látszólag nyilvánvalóan igen: hiszen mi hozhatja közelebb az erkölcsi normákat a racionális döntések logikájához, mint az a feltevés, hogy a normák valójában levezethetők az önérdekkövető és haszonmaximalizáló viselkedésből? E szoros kapcsolat azonban két okból is puszta látszat.

Először is mivel nem magától értetődő, hogy az erkölcsi normák motiváló ereje abból származik, hogy e normák levezethetők az egyéni haszonmaximalizációból, meg kell tudnunk mutatni, hogy valóban azok. David Gauthier egy részletesen kimunkált elmélet segítségével kívánja bizonyítani, hogy az erkölcsi normák levezethetők az önérdekkövető haszonmaximalizációból. Gauthier a már fentebb említett fogolydilemma-szerű szituációt használja példaként. Egy efféle szituációban a résztvevőknek nyilvánvalóan érdekükben áll, hogy megegyezzenek: azt a stratégiát kö-

vetik majd, amely a mindkettőjük számára optimális eredményre vezet. Egyik sem követelhet többet, mivel ezzel elijesztené a másikat a megegyezéstől, és ha nincs megegyezés, akkor mindegyikük kénytelen megelégedni a számára is előnytelen megoldással. Eddig rendben is volnánk, csakhogy ehhez még nincs is szükség semmiféle morális korlátra. A megegyezés mindenkinek *prima facie* érdekében áll, és valószínűleg létre is jön (Gauthier 1986: V. fejezet).

A probléma ott kezdődik – amint arra már Hobbes is rámutatott (1970: 117) –, amikor a megállapodás betartásáról van szó. A szituáció jellegéből fakadóan ugyanis minden egyes résztvevő számára az lenne a legelőnyösebb, ha a másik betartaná a megállapodást, ő maga viszont nem. Ha tehát pusztán az önérdekkövető hasznosságkalkuláció motiválja a résztvevőket, egyikük sem fogja betartani a megállapodást. Hobbesszal szólva, a „megállapodások puszta szavak", ott fejeztük be, ahol elkezdtük, senki sem fog amellett a viselkedés mellett dönteni, amely mindannyiuk számára kedvezőbb eredményt hozhatna. Történetünk e pontján lép színre a morális norma: a morál olyan belső kényszerítő feltétel, korlát, amely eltilt a „szerződésszegő" viselkedéstől, és ezáltal biztosítja a mindenki számára kedvező megoldás létrejöttét. A egyéni haszonmaximalizáció érdekében racionális korlátozni a haszonmaximalizáló viselkedést (Gauthier 1986 VI. fejezet).

Kétségtelen, hogy a fogolydilemma-szerű szituációk paradigmatikus esetei az olyan helyzeteknek, ahol a morális normáknak megfelelő cselekvés még a pusztán egoista résztvevők számára is előnyös lehet. Csakhogy van két súlyos probléma, amellyel szembe kell néznünk. Először is a morális normáknak *megfelelő* viselkedés nem feltétlenül azonos a morális normák *motiválta* viselkedéssel. Egyszerűen *de facto* nem igaz, hogy, amikor visszaadjuk a kölcsönkért pénzt, amikor átsegítjük a vakot az úttesten, vagy amikor visszaadjuk valaki elvesztett pénztárcáját, akkor egy olyan diszpozíció alapján cselekednénk, amelynek bármi köze van a kölcsönös optimum eléréséhez.

Másodszor abból, hogy egy bizonyos típusú cselekvés valamennyi résztvevő számára előnyös *lehet*, még nem következik, hogy *valóban* az is. Hogy valóban az legyen, ahhoz még számos más feltételnek is teljesülnie kell. A legfontosabb ilyen feltétel a kölcsönösség. Azonban sajnos könnyen belátható, hogy nem racionális a morális normák által korlátozott viselkedést választani egy olyan környezetben, ahol az interakciók nem ugyanazon, egymást ismerő személyek között ismétlődnek. Anonim közösségekben sokkal racionálisabb a „néha csalok – néha nem" stratégia alkalmazása, aminek ugyebár semmi köze nincs a morális értelemben vett normákhoz. Mégis vannak olyan társadalmak, ahol az emberek rendszeresen követik a morális normák által megkövetelt viselkedést.[4]

De nemcsak ez az oka annak, hogy a Gauthier-féle elmélet sem képes megmagyarázni, hogyan illeszkedik a morális normák követése a racionális döntések logikai terébe. A baj oka mélyebben keresendő. Még ha sikerülne is bizonyítani, hogy az önérdekkövető kalkuláción érvényesülő korlát valójában a normakonform módon viselkedő egyén érdekében áll, akkor sem sikerült a morális motivációkat elhelyeznünk a racionális megfontolások logikai terében. Ezen elmélet keretében kétféle

[4] Lásd erről Smith 1991: 229–253.

módon értelmezhetjük ugyanis a normakonform viselkedést. Az egyik értelmezés szerint a morális normák követését végső soron az önérdekalapú haszonmaximalizáció írja elő. Ebben az esetben nem azt magyaráztuk meg, hogy hogyan helyezkedik el a morális normák követése a racionális döntés logikájának terében, hanem eliminátuk a morális motivációt: kiderült, hogy az végső soron leplezett önérdek. Ha azonban azt mondjuk, hogy a normakövetés *hatása* előnyös csupán, és ezért jó dolog a morális normák szerinti viselkedés, akkor egyáltalában nem sikerült megmagyaráznunk, hogyan illeszkednek a morális normák a racionális megfontolások logikai terébe. Lehet, hogy adtunk egy elfogadható evolúciós magyarázatot a normakövető viselkedés fennmaradásáról (bár én ezt is kétlem), de eredeti problémánkra bizonyosan nem adtunk kielégítő választ.

Úgy tűnik tehát, nem vezethetjük vissza a normák motiváló erejét az önérdekkövető haszonmaximalizálásra, és ezért nem vagyunk képesek megmagyarázni, hogyan illeszthetők be a morális normák a racionális megfontolások logikai terébe. De talán nincs is erre szükség. Talán csak hasztalan filozófiai erőfeszítés, amikor megpróbáljuk összekötni a döntéselméletben használt racionalitásfogalmat és a morális cselekvés szerinti viselkedés megértését. A társadalomtudományban létezik egy olyan tradíció, amelyik éppen a két cselekvéstípus radikális megkülönböztetésén nyugszik, és amelynek számos különböző megfogalmazása létezik. Közülük a legismertebb talán a célracionális és az értékracionális cselekvés weberi megkülönböztetése. A hagyományos értelmezés szerint ugyanis az „értékracionalitás" valójában egy adott cselekvés magyarázatának kivonását jelenti a racionális döntések elméletének hatóköre alól.

A weberi hagyomány talán legjelentősebb modern folytatója Jon Elster. Szerinte a racionális és a normavezérelt cselekvések közti legfontosabb különbség, hogy míg a racionális cselekvés csak az eredményekre van tekintettel, egy cselekvést csak akkor tekinthetünk normavezéreltnek, ha az olyan motivációk alapján történik, amely figyelmen kívül hagyja az eredményeket. Elster szerint a társadalomtudomány egyik fontos feladata, hogy ne csupán azt vizsgálja, hogy melyek az individuális, racionális kalkuláción alapuló cselekedetek kollektív következményei, hanem azt is, hogy melyek bizonyos normák követésének kollektív hatásai. Ám azt is hangsúlyozza, hogy e normák létét még abban az esetben sem magyarázhatjuk e hatásokkal, ha e hatások pozitívak.

Mindezzel, úgy hiszem, egyetérthetünk. Csakhogy ezzel nem magyaráztuk meg, hogyan lehetséges, hogy bizonyos esetekben az erkölcsi normák alapján, más esetekben pedig a racionális érdekkalkuláció alapján döntünk. Vajon lehetséges-e élesen elkülöníteni ezeket az eseteket? Lehetséges olyan eseménytípusokat megkülönböztetni, amelyeket az egyik, és olyanokat, amelyeket a másik segítségével magyarázunk? Ez nem teljesen kizárt. Amikor arról kell döntenem, milyen típusú autót vásároljak, valószínűleg irrelevánsak a morális normák. Amikor az forog kockán, eljátszom-e a becsületem, valószínűleg nem. De számos olyan helyzet van, sőt valószínűleg a döntési helyzetek többsége ilyen, amikor a morális normák és a racionális érdekkalkuláció egyaránt szerepet játszik döntéseinkben. Talán nem érzékeljük ezt olyan erősen a tiltó normák esetében, amelyek motiváló ereje általában olyan nagy, hogy úgy tűnik, semmiféle eredményorientált érdekkalkulációval nem fér össze. De

az olyan esetekben, mint amilyen a „Segítsd a barátodat!" vagy a „Tartsd meg a haldoklónak tett ígéretedet!" igenis szerepet játszhat a norma betartásából az egyénre háramló következmény, s aki végül is nem a norma betartása mellett dönt, azt nem tekintjük feltétlenül erkölcstelennek, mint ahogy azt, aki a normáknak megfelelően dönt, sem tartanánk egyszerűen irracionálisnak.

Sokat veszítünk azzal, ha megpróbáljuk kiemelni a normavezérelt cselekvést a racionális megfontolások logikai teréből. Igaz ez nemcsak a mindent megalapozni kívánó filozófusokra, hanem a társadalomtudósokra is. Vegyük példaként a már említett és heurisztikusan amúgy is mindig hasznos fogolydilemma-szerű szituációkat! Közismert tény, hogy a kollektív javak megvalósításának problémája, illetve azok egy része, reprezentálható fogolydilemma-szerű szituációként: a jószág megvalósulásából az egyénre háramló haszon meghaladná ugyan a reá jutó költségeket, de mivel a jószág megvalósulásában akkor is bízhat, ha ő maga nem járul hozzá, illetve akkor sem bízhat, ha hozzájárul, a közjószág iránti vágy, ha csak nincs olyan kényszer, mely ösztönöz a kooperatív viselkedésre, kielégítetlen marad. Ám azt is tudjuk, hogy ez nem minden esetben van így. Vannak közterek, amelyek tiszták, vannak köztéri rézszobrok, amelyet nem bontanak le a járókelők. És vannak választások, amelyekre az emberek többsége elmegy szavazni, még akkor is, ha tudja, hogy saját hozzájárulása a választások eredményéhez elhanyagolható. Hogyan lehetséges ez? Hogyan magyarázhatjuk például azt, hogy ki és miért megy el szavazni? Talán úgy, hogy egyesekről feltételezzük, hogy ők morálisak, de nem racionálisak, mások meg racionálisak, de képtelenek az erkölcsileg helyes viselkedés felismerésére? Ez nem az ígéretes megközelítés. Inkább arról van szó, hogy különböző személyek különböző módon súlyozzák a normákat és a normakövetés esetleges költségeit, és a racionális döntések elméletének képesnek kell lennie arra, hogy ezt reprezentálja.

A filozófus számára persze nem ez a fő nehézség, hanem az, amit már említettem e fejezet elején: ha így járunk el, önkényesen kettéosztjuk a cselekvőt, személyiségét nem vagyunk képesek egyetlen integráns egésznek tekinteni. Márpedig a legtöbben így szeretnénk magunkra tekinteni, nem pedig mint olyan lényekre, akiket a körülmények hol arra szorítanak, hogy cselekvésük következményeit latolgassák, hol pedig arra, hogy vakon kövessék a normákat. Önmagunk és mások személyiségét többek között éppen azon az alapon ítéljük meg, hogy milyen egyensúlyt tudunk kialakítani viselkedésünkben e kéttípusú motiváció között. Miután nem vagyunk egyformák, különböző személyek esetében e motivációk különböző erejűek lehetnek. Nem kell, hogy minden esetben egyetértsünk, egymás viselkedésének megértéséhez elegendő annyit feltennünk, hogy mindnyájan képesek vagyunk e mérlegelésre. Ha a másikat kritikával illetjük, talán azért tesszük, mert úgy véljük, helytelenül értékeli az egyes tényezők súlyát. Mint fentebb említettem, az a véleményem, hogy sok esetben *de gustibus est disputandum*.

A racionális döntések elméletének azonban nem az a feladata, hogy mások cselekedeteinek értékelésénél erkölcsi mércéül szolgáljon, hanem az, hogy logikai-matematikai eszközökkel reprezentálja a döntést megelőző megfontolásokat. Ha az erkölcsi normák által motivált cselekvés esetében erre nem képes, akkor bizony hatókörét nagyon nagy mértékben leszűkítettük. Azt jelentené ez, hogy egy olyan elméletet alkottunk a racionális megfontolásokról, amelyikben az, aki a morális nor-

mák alapján dönt, irracionális, de legalább is kívül helyezi magát a racionális megfontolások körén. Vagy talán azt – amennyiben elfogadjuk, hogy nem irracionális az erkölcsi normák alapján dönteni –, hogy a döntéselmélet valójában nem is a cselekvést megelőző racionális megfontolások reprezentációja. Ez szomorú következmény lenne. Szerencsére semmi okunk rá, hogy elfogadjuk.

3. Döntéselmélet: Savage

A weberi–elsteri megkülönböztetés megvilágítja, mit tekintünk a morális normák alapján történő cselekvés jellegzetességének, de egyben azt is jelzi, mi teszi e szerzők számára lehetetlenné, hogy a morális cselekvést integrálják a racionális megfontolások logikai terébe. A morális normák motiváló ereje abban rejlik, hogy azok bizonyos cselekedeteket a cselekedetek végrehajtójára háramló következmények számbavétele nélkül értékelnek. De miért kerülnek ezáltal kívül e cselekedetek a döntéselmélet által reprezentált megfontolások logikai terén? Vajon szükséges-e, hogy így legyen? A következőkben azt próbálom megmutatni, hogy miért tűnt sok filozófusnak és társadalomtudósnak úgy, hogy kívül kell kerülniük. Ezután viszont azt szeretném jelezni, miért nem hiszem, hogy szükségképp kívül kell kerülniük.

A probléma megértéséhez elengedhetetlen, hogy kicsit közelebbről szemügyre vegyük azt a logikai-matematikai struktúrát, amit a filozófusok és társadalomtudósok egy jelentős része a döntéselmélet alapjának tekint. E struktúra legkidolgozottabb klasszikus változata a bayesiánus döntéselmélet Leonard Savage által javasolt rendszere. A rendszer lényege a következőképpen foglalható össze.[5]

Egy döntéselméleti modell megadásához két dologra van szükségünk. Egyrészt meg kell határoznunk azokat az entitásokat, amelyeket a döntéselmélet posztulál, és amelyek között aztán különböző relációkat definiálhatunk, illetve amelyekhez különböző numerikus értékeket rendelhetünk. Másodszor meg kell adnunk azt az optimumkritériumot, amelynek alapján az elmélet szerint a döntésnek meg kell születnie. A Savage-féle elmélet a következő entitásokat posztulálja:

1. *természeti állapotok*, amelyek halmazait *eseményeknek* tekintjük S_i;

2. *eredmények* O_j;

3. *cselekvések* (függvények, amelyek a természeti állapotok halmazát az eredmények halmazára képezik le) $A_k (S_i) = O_j$;

4. *preferenciák* (amelyek a cselekvéseken értelmezett relációk).

Ezek jelentését legegyszerűbb egy példa segítségével szemléltetni.[6] Tegyük föl, hat tojásból akar valaki rántottát készíteni, amihez már összekevert öt friss tojást, s most azt kell eldöntenie, hogy a hatodikat, amelyik már nem friss, beleüsse-e a többi közé. Ebben az esetben a lehetséges esemény a következő *természeti állapotoknak* felelnek meg: *1.* ép tojás, *2.* záptojás. Az *eredmények* a következők lehetnek: *1.* hat

[5] A savage-i döntéselmélet kifejtésekor Ellery Eells kiváló munkájára támaszkodom (1982: 71–78).

[6] A példa Savage eredeti példájának módosított és egyszerűsített változata.

tojásból rántotta, *2.* oda a vacsora, *3.* rántotta öt tojásból, egy tojás veszteséggel, *4.* rántotta öt tojásból.

	S_1: ép tojás	S_2: záptojás
A_1: beleüti a hatodik tojást	O_{11}: hat tojásból rántotta	O_{12}: oda a vacsora
A_2: nem üti bele a hatodik tojást	O_{21}: rántotta öt tojásból, egy tojás veszteséggel	O_{22}: rántotta öt tojásból

A természeti állapotokhoz és az eredményekhez is numerikus értékeket rendelhetek: a természeti állapothoz rendelt értéket fogom (szubjektív) valószínűségnek nevezni. A valószínűségeket értelmezhetem úgy, mint a záptojásos lehetséges világok számának arányát azon világok számához képest, amelyekben a tojás ép. Miután lehetetlen, hogy a tojás ép is legyen, meg ne is, és szükséges, hogy vagy ép legyen, vagy ne, az utóbbi esemény bekövetkezésének valószínűsége 1.

Az eredményekhez, amelyek azt fejezik ki, mi történik a cselekvővel, szintén rendelhetek numerikus értékeket. Ezek az értékek az egyes eredmények a cselekvő számára nyújtott hasznosságát (az eredeti megfogalmazásban: kívánatosságát) reprezentálják. Egy cselekvés *várható hasznossága* a cselekvés valószínűségekkel súlyozott eredményértékeinek összege. Jelen esetben a szubjektív várható hasznosság *(subjective expected utility – SEU):*

$$SEU\ (A_1) = pu_1 + (1–p)\ u_4,$$
$$SEU\ (A_2) = pu_2 + (1–p)\ u_3$$

amennyiben,

	Prob(S_1: ép) = (p)	Prob(S_2: záptojás) = (1–p)
A_1: beleüti a hatodik tojást	$U(O_{11}) = u_1$	$U(O_{12}) = u_2$
A_2: nem üti bele a hatodik tojást	$U(O_{21}) = u_3$	$U(O_{22}) = u_4$

Természetesen a természeti állapotok (az eseményhalmaz részhalmazai) száma kettőnél – és így a lehetséges eredmények száma négynél – több is lehet. A cselekvés várható hasznossága Savage rendszerében tehát a következő lesz:

$$SEU\ (A_k) = \sum_i \Pr(S_i)\ U[A_k(S_i)],$$

vagy másképp (miután a cselekvés és a bekövetkező természeti állapot együttesen egyértelműen meghatározzák az eredményt):

$$SEU\ (A_k) = \sum_i \Pr(S_i)\ \mathrm{U}(\mathrm{O_{ki}}).$$

Ezek után már elég nyilvánvaló, hogy mit gondol Savage az optimumkritériumról. Az optimumkritérium azt mondja ki, hogy egy racionális cselekvő mindig a magasabb várhatóhasznosság-értékű cselekvést fogja preferálni.

$$A_k \mathrm{Pref}_x\ A_i \ \Leftrightarrow SEU(A_k) > SEU(A_i).$$

Ezt a döntéselméleti rendszert mármost többféleképpen lehet értelmezni. Savage eredeti célja az volt, hogy a statisztikai valószínűség korábban elfogadott, objektív gyakoriságként értelmezett fogalmát egy új, bayesiánus valószínűség-fogalommal helyettesítse. A bayesiánus felfogás szerint a valószínűség az egyének meggyőződéseinek erejét méri, ahol a „meggyőződés" alatt a cselekvést meghatározó mentális diszpozíciót értjük. Savage elképzelése (Bayes és Ramsey nyomán) az volt, hogy egy személy cselekedeteiből (pontosabban bizonyos helyzetekben történő választásaiból) következtetni tudunk arra, hogy meggyőződései milyen erejűek (tehát hogy milyen valószínűséget tulajdonít bizonyos események bekövetkeztének). A Savage által javasolt döntéselméleti rendszer azonban természetesen másképp is interpretálható. Értelmezhető oly módon is, hogy céljának nem a szubjektív valószínűség értékeinek meghatározását tekintjük, hanem azt, hogy meghatározza, adott helyzetben a racionális cselekvő a rendelkezésére álló cselekvési lehetőségekből melyiket választaná. Mindkét értelmezés feltételezi azonban, hogy a döntéselmélet képes matematikailag reprezentálni a cselekvést megelőző megfontolásokat.

Állításom mármost az, hogy a közgazdaságtanban – és a társadalomtudományokban általában – a döntéselméletnek ezt a Savage által megalkotott rendszerét használják és finomítják tovább. Teljesen nyilvánvaló például, hogy a Hirschleifer–Riley-féle döntéselméleti modell ezen a savage-i rendszeren alapszik. Savage rendszere valóban vonzó: egyrészt ökonomikus, másrészt úgy látszik, megfelel annak az intuitív felfogásnak, amit a racionális megfontolások szerkezetéről általában elfogadunk. Tulajdonképpen annak a már a 17. századi logikában ismert elvnek az egzakt matematikai változata, amit két híres karteziánus logikus, Arnauld és Nicole fogalmazott meg először az úgynevezett Port-Royal logikájában, s ami később a bayesianizmus alapja lett. „… mivel ahhoz, hogy megítélhessük, mit kell valamely jó eléréséhez vagy valamely rossz elkerüléséhez tennünk, nemcsak a jót, illetve a rosszat önmagában, hanem bekövetkezésük valószínűségét is meg kell vizsgálnunk, és mértanilag figyelembe kell vennünk azt az arányt, amelyet ezek együttesen kitesznek..." (Arnould–Nicole 1970: 428)

A döntéselmélet e változatának olyan következménye is van, amelyet érdemes külön megemlíteni, mivel egyrészt bizonyítja az elmélet ökonomikusságát és intuitív erejét, másrészt viszont, mint látni fogjuk, sajnos éppen e következmény mutatott rá az elmélet egyik gyengeségére is, amely azután arra ösztönzött egyes filozófusokat, hogy módosítsák Savage rendszerét. Savage rendszerének e nevezetes következménye az úgynevezett dominanciaelv. Az elv a következőt mondja ki:

Ha egy döntési szituációban létezik olyan cselekvés, amely minden egyes természeti állapot bekövetkezése esetén nagyobb hasznosságot ígér, mint a többi lehetséges cselekvés, akkor csakis ezt a cselekvést racionális választani. Vagyis, ha felidézzük döntési mátrixunkat:

	$Pr(S_1) = (p)$	$Pr(S_2) = (1 - p)$
A_1	$U(O_{11}) = u_1$	$U(O_{12}) = u_2$
A_2	$U(O_{21}) = u_3$	$U(O_{22}) = u_4$

de most azt feltételezzük, hogy $u_1 > u_3$ és $u_2 > u_4$, akkor bármilyen valószínűséggel következzék is be S_1 vagy S_2 állapot, a *racionális egyén mindenféleképpen* A_1 *cselekvést fogja választani*.

Ez az elv rendkívül meggyőzően hangzik, és az esetek többségében aligha merülhet fel kétely alkalmazásának racionalitását illetően. Ám mint látni fogjuk, vannak kivételek. Ahhoz ugyanis, hogy a bayesiánus döntéselmélet Savage-féle változata működjék, három nagyon fontos feltételnek is teljesülnie kell. Ezek pedig a következők:

α) *Az eredmények meghatározottsága*: minden egyes cselekvésről tudjuk, hogy adott természeti állapot bekövetkezése esetén milyen eredményhez fog vezetni. Ez a feltétel nagyon erős, és tényleges döntéseink egy jó részében valószínűleg nem is teljesül. Ám bennünket most nem az elmélet empirikus alkalmazhatósága érdekel, hanem az, hogy vajon intuitíve helyesen reprezentálja-e a cselekvést megelőző megfontolásokat. Önmagában az a tény, hogy használatához el kell fogadnunk bizonyos idealizációs feltételeket, még nem jár súlyos következményekkel.

β) *A cselekvéstől való függetlenség*: hogy milyen természeti állapot valósul meg, azt nem befolyásolhatja az, hogy milyen cselekvést hajtunk végre. Savage példáját alapul véve, az hogy valaki beleüti-e a hatodik tojást a rántottájába, vagy sem, nem fogja befolyásolni annak valószínűségét, hogy a tojás ép-e, vagy sem. Ez a feltevés, ha ezt a példát vesszük alapul, teljességgel elfogadhatónak tűnik. Ám nem minden esetben az. Mint látni fogjuk, ennek a feltételnek a feloldása motiválta elsősorban a Savage-féle döntéselméleti rendszer módosítását.

γ) *Az eredmények elérhetőségének neutralitása*: az eredmények értékelése független attól, hogy milyen cselekedet végrehajtása és milyen természeti állapot megvalósulása vezet el hozzájuk. A következőkben ez a feltevés lesz számunkra a legfontosabb. Amellett szeretnék érvelni, hogy ez az a feltevés, amely lehetetlenné teszi, hogy a morális motivációk alapján meghozott döntést el tudjuk helyezni a döntéselmélet logikai terében. Csakhogy nem lehetetlen e feltétel feloldása. Létezik a döntéselméletnek egy olyan változata, amely képes arra, hogy e feltételek bevezetése nélkül reprezentálja a cselekvést megelőző megfontolások logikáját.

Mielőtt azonban rátérnék a módosítás ismertetésére, szeretnék még néhány szót szólni e feltételekről, illetve a Savage-féle döntéselmélet alkalmazhatóságáról. Mindenekelőtt fontos leszögezni, hogy számos olyan szituáció (illetve olyan típusú szituáció) van, amelyben ezek a feltételek teljesülnek. Nyilvánvaló módon a közgazdaságtan által vizsgált helyzetek jelentős része (ugyan távolról sem mindegyike) ilyen. A második feltevés szerint a választott cselekvés nem befolyásolhatja a „természeti állapot" bekövetkezésének valószínűségét. S valóban, sok esetben abból indulunk ki, hogy bizonyos piaci interakciók során az egyes egyén hozzájárulása valamely eredményhez oly csekély, hogy a hozzájárulás maga aligha befolyásolhatja az eredmény bekövetkezésének valószínűségét. Ha például el kell döntenem, hogy milyen részvényt vegyek a prémiumomból, aligha számolhatok azzal, hogy döntésem befolyásolni fogja a vásárolt részvény várható hozamát. Amikor számolok a kockázattal, azzal is számolok, hogy nem leszek képes befolyásolni azt, milyen természeti állapot fog bekövetkezni. S az ökonómiai jellegűeken kívül számos más olyan döntési

helyzet van – például amikor szó szoros értelmében a természet dönt, vagy amikor, hogy a klasszikus példát említsük, szerencsejátékot játszunk –, amelyekben a Savage-féle elmélet kiválóan működik.

Ugyanez vonatkozik a harmadik feltevésre is. Amikor amellett döntünk, milyen bankba helyezzük el a pénzünk, vagy milyen részvényt, mekkora összegű biztosítást vásároljunk, magukat a cselekedeteket csak következményeik felől értékeljük. Hirschleifer és Riley például a következőképp érvel: „A hasznosság közvetlenül a következményekhez kapcsolódik és csak közvetetten a cselekvéseinkhez. Meglepő, mekkora intellektuális zavart okozott ezen egyszerű megkülönböztetés hiánya" (Hirschleifer–Riley 1998: 31). S valóban, ha döntéselméletünk alapja a Savage-féle rendszer lesz, érdemes ezt a megkülönböztetést szem előtt tartani: értékelni csak a következményeket vagy eredményeket értékeljük, a cselekvéseket önmagukban nem. Ámde ha ezt a feltevést nemcsak *bizonyos* döntési szituációkra tartjuk érvényesnek, hanem a racionalitás *egyetemes* normájának tekintjük, akkor bizony komoly nehézségeket fog okozni. Gondolom, már sejthető, hova szeretnék kilyukadni: ha elfogadjuk, hogy egy racionális egyén a cselekvést megelőző megfontolásai során csak cselekedetei következményeit, de magukat a cselekedeteket sohasem értékelheti, akkor eleve kizártuk annak lehetőségét, hogy a morális normák motivációs erejét beilleszthessük a döntéselmélet logikai terébe.

Úgy gondolom ez az oka annak, hogy a legtöbb filozófus és társadalomtudós számára nehézséget okoz a morális és a racionális viselkedés viszonyának értelmezése. Félreértés ne essék: eszem ágában sincs azt állítani, hogy a Savage-féle döntéselmélet fertőzte volna meg a filozófusokat és társadalomtudósokat, akik közül valószínűleg sokan nem is hallották Savage nevét. Épp ellenkezőleg, inkább azt szeretném mondani, hogy a Savage-féle elmélet azért lehetett olyan sikeres, mert *egy már létező konszenzusnak megfelelő matematikai reprezentációt adott a racionális megfontolásokról*.

Az elfogadott álláspont Hume racionalitás-felfogásából és Kant *hipotetikus imperatívusz* fogalmából ered. Mindkettő a cselekedeteknek *eszközértékjellegét* hangsúlyozza, és ezáltal mindkettő azt sugallja, hogy a racionális megfontolások körén kívül kell helyezni a moralitást. Ez a „sugallat" teljesen egyértelmű Hume esetében, aki szerint a moralitás alapja a morális érzület vagy „szenvedély", viszont „mivel csak az lehet ellentétes az igazsággal vagy az ésszel, ami arra vonatkozik, és kizárólag értelmünknek az ítéletei vonatkoznak rá, következésképpen a szenvedélyek csupán annyiban lehetnek ésszerűtlenek, amennyiben valamilyen ítélet vagy vélekedés kíséri őket..." Ésszerűségről és ésszerűtlenségről csak akkor beszélhetünk, „ha valamely szenvedélyünk cselekvésben nyilvánul meg, de *rosszul választjuk meg a célunk elérésére felhasznált eszközöket*, vagyis csalatkozunk az okoknak és okozatoknak a megítélésében". Ezért aztán „nincs semmi ésszerűtlen abban, ha inkább az egész világ pusztulását választom, mint azt, hogy egy karcolás essék a kisujjamon". Viszont „Nem ésszerűtlen akár saját romlásom árán is megmenteni egy indiánt vagy más, számomra ismeretlen embert a legkisebb kellemetlenségtől" (Hume 1976: 546. – kiemelés tőlem). Jól érzékelhető e sorokból, hogyan kapcsolódik össze a racionalitás azon felfogása, amely a cselekvéseket csak eszköznek tekinti a „szenvedélyek"

által meghatározott célok eléréséhez a moralitás egy olyan értelmezésével, amely kirekeszti azt a racionális megfontolások logikai teréből.

Kant ugyan – Hume-mal ellentétben – morális racionalista, aki az erkölcsi ítéleteket az észből eredezteti, ám az ész speciális, metafizikai felfogásából: a „fenomenális világban" a cselekvéseket valamely cél eléréséhez puszta eszköznek tekintő hipotetikus imperatívusz irányítja választásainkat. „A hipotetikus imperatívuszok egy olyan lehetséges cselekedet gyakorlati szükségességére utalnak, amely eszköz valaminek az elérésére, amit akarunk (vagy legalábbis akarhatunk)" (Kant 1991: 44). A morális cselekedetek esetében viszont olyan szabálynak kell meghatároznia akaratunkat, amely nem veheti figyelembe a döntésünkből „fenomenális Énünkre" háramló következményeket. „Kategorikus imperatívusz az, amelyik valamilyen cselekedetet önmagáért valóan, s nem egy másik célra vonatkoztatva állít objektíven szükségszerűnek" (Kant 1991: 44). A kategorikus imperatívusznak Kant szerint tehát éppen az a lényege, hogy kiemelje a morális döntéshozót a racionális érdekkalkuláció világából. Ám ezt csak akkor teheti meg, ha a morális motivációk *korlátozzák* azon cselekvések körét, amelyeket a hipotetikus imperatívuszok alapján hozott döntések során számba vehetünk. Az az elképzelés, hogy a morális döntés voltaképp az önérdekkövető döntésen értelmezett korlát, a kantiánus morálfilozófiai hagyomány öröksége. Ott, ahol nincs mit korlátozni, elvész a specifikusan morális motiváció lehetősége. Kant számára a világ nomenális világra (mely egyben a szabadság világa is) és fenomenális világra történő osztása lehetővé tette, hogy e korlátok alapján történő döntést is racionálisnak tekintse – a szónak persze nem döntéselméleti, hanem metafizikai értelmében. Azok számára azonban, akik a világok kettéosztását nem fogadták el, csak a motivációk szembeállítása maradt meg a kanti filozófiából. Hogy ez milyen szerencsétlen következményekkel járt, azt már láthattuk.

Hadd idézzek fel két példát! A kanti hagyományt veszi aztán át és fejleszti tovább Weber, amikor megkülönbözteti a „célracionális" és az „értékracionális" cselekvést (Weber 1987: 53). Maga az értékracionális kifejezés Kantra (és persze a weberi tudományfilozófiának alapul szolgáló Rickertre) utal. A *célracionalitás* pedig arra, hogy ebben az esetben a cselekvés önmagában nem hordoz értéket, neutrális, a cselekvés magyarázata során csak azt kell azonosítanunk, hogyan értékeli a saját céljait a döntéshozó egyén. Az értékek szerinti cselekvést a neokantiánus hagyománytól távolabb álló szerzők ugyan már nem illették a megtisztelő „racionális" jelzővel, de a morális normák szerinti cselekvést, mint azt Elster esetében láthattuk, teljesen kivonták a racionális döntések elméletének hatókörébe tartozó megfontolások közül. Hogy miért, azt Elster teljesen világossá tette: a racionális döntést hozó egyén csak cselekvéseinek eredményeit veheti figyelembe, míg a morális normák alapján döntést hozó egyénnek nem szabad a döntéséből reá háramló következményeket figyelembe venni.[7]

S a filozófusokkal sem áll máshogy a helyzet. Rawls, aki filozófiáját több helyütt is kantiánusnak nevezi, szembeállítja egymással a morális személy két alapvető ké-

[7] „A racionalitás lényegileg feltételes és jövőbe irányuló. Imperatívuszai hipotetikusak, azaz azoktól a jövőbeli céloktól függenek, amelyeket meg akar valaki valósítani. A társadalmi normák által kifejezett impertívuszok viszont vagy feltétlenek, vagy ha nem azok, akkor sem a jövőre irányulnak" (Elster 1989: 98). Világos, hogy e sorokban is Kant köszön vissza.

pességét. Az egyiket „racionalitásnak” nevezi, ami megfelelne az általában „közgazdaságtaninak” tekintetett racionalitásfogalomnak, annak tehát, mely szerint a racionális megfontolás során a cselekvő a kívánt eredmények eléréséhez keresi a megfelelő eszközt. Ezzel áll szemben az „ésszerűség”, amely tulajdonképpen a racionális érdekkalkuláció igazságosság elvei által történő korlátozását jelenti.[8] És bármilyen elszánt kritikusa legyen is Gauthier Rawls állítólagos kantianizmusának, ő maga is e tradíciót követi, amennyiben a morális normákat az önérdekkövető cselekvéseken érvényesülő racionális korlátoknak tekinti, s amennyiben úgy kezeli a racionális döntéseket, mint amelyek során a cselekvő csakis cselekedetei következményeit értékeli. Miután Gauthier elméletéről formális modellt is ad, az ő esetében végképp bizton állíthatjuk, hogy a Savage-féle modellt alkalmazza.[9]

Azt állítom tehát, hogy a Savage-féle döntéselmélet csak azt a már meglévő konszenzust vette alapul, amely a társadalomtudósok és a filozófusok többségének felfogását uralta, sőt uralja még ma is. Hogy ez milyen problémákhoz vezet, azt már láttuk. Lássuk most, vajon lehet-e módosítani a döntéselmélet rendszerét úgy, hogy képes legyen a morális motivációkat integrálni a cselekvést megelőző megfontolások matematikai reprezentálása során.

4. Döntéselmélet: Jeffrey

Azok, akik a Savage-féle döntéselmélet módosítására törekedtek, természetesen nem azért tették ezt, mert elkeserítette őket, hogy az elmélet ezen értelmezése nem teszi lehetővé a morális normák racionális döntések logikai terében történő elhelyezését. Először tehát azt szeretném röviden jelezni, mi késztette a filozófusokat arra, hogy újragondolják a döntéselmélet alapjait, és igyekezzenek feloldani azokat a feltételeket, amelyek érvényesülése nélkül a Savage-féle elmélet nem működik.

Tekintsünk egy önérdekkövető és racionális egyént, akinek a következő problémával kell szembenéznie! Leszokjon-e a dohányzásról vagy sem? Tegyük fel, hogy emberünk meg van győződve arról, hogy a dohányzás káros az egészségére, és ha sokat dohányzik, nagy valószínűséggel 65 éves kora előtt elhuny. De persze tisztában van azzal a ténnyel is, hogy ez akkor is előfordulhat, ha nem dohányzik. Nézzük meg, hogyan alkalmazható a fentebb ismertetett modell e probléma reprezentálására! A legszemléletesebb az lesz, ha a problémát mátrix segítségével ábrázoljuk:

	p valószínűséggel több mint 65 évet él	1–*p* valószínűséggel 65 éves kora előtt elhuny
Leszokik	leszokik és sokáig él *(LÉ)*	leszokik és korán elhuny *(LH)*
Dohányzik	dohányzik és sokáig él *(DÉ)*	dohányzik és korán elhuny *(DH)*

[8] Lásd különösképpen Rawls (1980).

[9] Igaz, nem Savagera hivatkozik, hanem R. D. Luce és H. Raiffa híres könyvére, akik viszont részben Savage alapján, részben pedig az ebből a szempontból Savagéval rokon Neumann–Morgenstern-féle hasznosságértelmezés alapján tárgyalják e témát (Luce–Raiffa 1958).

Feltételezhetjük, hogy emberünk a következőképpen értékeli a lehetséges következményeket: *DE* eredményt szeretné a leginkább, *LH*-t a legkevésbé. *LÉ* viszont jobban vonzza, mint a korai halál *DH*. Ha már most egy pillanatra elfelejtjük a fenti matematikai reprezentációt, intuitíve azt mondanánk, hogy minél valószínűbbnek tartja valaki, hogy a dohányzás megrövidíti az életet, annál racionálisabb lesz számára, hogy leszokjon a dohányzásról. Ám ha ránézünk a táblázatra, és felidézzük a Savage-féle elmélet fontos következményét, a dominanciaelvet, akkor arra a következtetésre kell jussunk, hogy az illető számára teljesen mindegy, mekkora valószínűséggel rövidíti meg életét a dohányzás, számára az egyetlen racionális megoldás a dohányzás folytatása lehet. Miért? Azért, mert bármely esemény következzék is be, *DÉ* értéke magasabb, mint *LÉ*-é, és *DH* értéke magasabb, mint *LH*-é. De valóban úgy gondoljuk, hogy mindenki, aki leszokott a dohányzásról, irracionális? Talán nem. És ha nem, akkor a döntéselméletnek illenék olyan modellt ajánlania, amely képes számot adni a dohányzásról leszokók racionális megfontolásairól is.

Vegyük észre, hogy ebben a döntési szituációban két fentebb tárgyalt kikötés (β és γ) is sérül! Először is azt feltételezzük, hogy az, meddig él valaki, nem független attól, leszokik-e a dohányzásról, vagy sem *(β)*. Másodszor a következmények értékelésében az is szerepet fog játszani, hogy le kell-e mondjon a dohányzás okozta élvezetről, vagy sem *(γ)*. Egy efféle helyzet döntéselméleti elemzésére tehát nem alkalmas Savage eredeti modellje. Olyan modellt kell találnunk, amelyik képes kifejezni a cselekedetek hatását az eredmények valószínűségére, és amelyik nem függetleníti az eredmények értékelését a hozzájuk vezető cselekvés értékelésétől, vagyis magukat a cselekvéseket is értékeli.

Richard Jeffrey (1983) olyan döntéselméleti modellt dolgozott ki, amely nem feltételezi a valószínűségek és az eredmények értékének cselekedetektől való függetlenségét. Jeffrey modelljében csupán egyetlen entitást kell feltételeznünk: propozíciókat, vagyis absztrakt igazságérték-hordozókat. Viszont ebben a modellben is különbséget tehetünk az eredményeket kifejező propozíciók, a természeti állapotokat kifejező propozíciók és a cselekvéseket kifejező propozíciók között. Mi haszna van akkor annak, hogy propozíciókról beszélünk? Az, hogy ugyanazon típusú entitásoknak tulajdoníthatunk valószínűségeket és értékeket, s egyben azt is lehetővé tettük, hogy a cselekvéseket kifejező propozíciók is ezen entitások között szerepeljenek. E megoldásnak számos más előnye van, amelyekre azonban most nem szükséges kitérnünk.

Számunkra most az a fontos, hogy a cselekedet várható hasznosságát meghatározó formulában képesek legyünk kifejezni, miként függ az eredmények értéke és bekövetkezésük valószínűsége a választott cselekvéstől. Ezt pedig a Savage-féle elv igen egyszerű módosításával érhetjük el: elemi valószínűségek helyett feltételes valószínűségekkel kell számolnunk, hasznossági függvényünkbe pedig be kell építeni a cselekvés értékét is. A cselekvés várható értékét meghatározó – feltételes várható érték – formulánk (*Conditional Expected Value – CEV*) így a következőképpen módosul:

$$\mathrm{CEV}\ (A_j) = \sum_i \Pr(O_{ji}|A_j)V(O_{ji}\ \&\ A_j)$$

Szavakban kifejezve ez a következőt jelenti: egy cselekvés várható értéke azonos a cselekvés és következményeinek a feltételes valószínűségekkel súlyozott értékével. Egyes esetekben nincs is szükségünk arra, hogy az eredményeket és a természeti állapotokat megkülönböztessük: az eredményeket egyszerűen egy adott cselekvés végrehajtása esetén bekövetkező természeti állapottal azonosíthatjuk. Ilyen eset a példaként idézet dohányzásról való lemondással kapcsolatos probléma. Döntési mátrixunkat a következő szavakkal is kifejezhetjük:

	Több mint 65 évet él	65 éves kora előtt elhuny
Leszokik	több mint 65 évig él, feltéve hogy leszokik a dohányzásról	65 éves kora előtt elhuny, feltéve hogy leszokik a dohányzásról
Dohányzik	több mint 65 évet él, feltéve hogy dohányzik	65 éves kora előtt elhuny, feltéve hogy dohányzik

Ezért a cselekvés várható értéke a következő módon is kifejezhető:

$$\mathrm{CEV}\ (A_j) = \sum_i \Pr(S_{ji}|A_j)V(S_{ji}\ \&\ A_j).$$

A formula még tovább csiszolható oly módon, hogy feltételezzük: a lehetséges eredmények különböznek ugyan a lehetséges természeti állapotoktól, viszont nem függetlenek a cselekvések és a természeti állapotok együttes végrehajtásától, illetve bekövetkezésétől. Problémánk szempontjából azonban ezek a további módosítások, illetve finomítások már nem érdekesek. Csak azért említettem őket, mivel jelezni szeretném, hogy milyen tág a Jeffrey-féle döntéselmélet alkalmazásának köre. Mielőtt rátérnék a számunkra érdekes alkalmazásokra, még két technikai részletet kell megemlítenünk. Az egyik arra vonatkozik, hogyan milyen a döntési mátrix a Jeffrey-féle döntéselméletben, a másik pedig arra, hogyan viszonyul Jeffrey elmélete a Savage-féle elmélethez. E technikai problémák tisztázása után rátérhetünk annak megmutatására, miért gondolom úgy, hogy a Jeffrey-féle döntéselmélet lehetővé teszi a morális motiváción alapuló megfontolások elhelyezését a döntéselmélet logikai terében.

Optimalizációs szabályunk ugyanaz lesz, mint Savage esetében: a racionális cselekvő mindig a legnagyobb várható értékű cselekvést választja:

$$A_k \mathrm{Pref}_{\underline{x}}\ A_i \ \Leftrightarrow \mathrm{CEV}(A_k) > \mathrm{CEV}(A_i).$$

Ám a Jeffrey-féle döntéselméletben feltételes valószínűségekkel kell súlyoznunk a várható következményeket, ezért hasznos, ha az eredmények *értékére* és a *bekövetkezésük valószínűségére* vonatkozó táblázatot külön írjuk fel:

Az értékeket kifejező mátrix:

	S_1	S_2
A_1	$V(S_1\ \&\ A_1) = v_1$	$V(S_2\ \&\ A_1) = v_2$
A_2	$V(S_1\ \&\ A_2) = v_3$	$V(S_2\ \&\ A_2) = v_4$

A valószínűséget kifejező mátrix:

	S_1	S_2
A_1	$\Pr(S_1 \mid A_1) = p_1$	$\Pr(S_2 \mid A_1) = p_2$
A_2	$\Pr(S_1 \mid A_2) = p_3$	$\Pr(S_2 \mid A_2) = p_4$

Miután azonban sok esetben nem áll az, hogy $\Pr(S_1 \mid A_1) = \Pr(S_1 \mid A_2)$, illetve, hogy $\Pr(S_2 \mid A_1) = \Pr(S_2 \mid A_2)$, az azonos oszlopban található valószínűségek különbözőek lehetnek. Ismét csak fenti példánkat idézve: nem áll, hogy annak valószínűsége, hogy 65 évnél tovább él valaki, feltéve, hogy leszokik a dohányzásról, azonos annak valószínűségével, hogy 65 évnél tovább él, feltéve, hogy dohányzik. Ha ez nem lenne így, akkor irracionális lenne minden olyan cselekvés, ami bizonyos áldozatok árán az egészség megőrzésére irányul. Márpedig ezeket a cselekedeteket nem tartjuk feltétlenül irracionálisnak.

Amint az már sejthető, a Jeffrey-féle elméletben módosul a dominanciaelv értelmezése is. Mint láthattuk, a dominanciaelv elfogadásának racionalitása abban állt, hogy bizonyos esetekben nem érdemes az eredmények bekövetkezésének valószínűségét figyelembe vennünk, amikor arról döntünk, hogyan cselekedjük. Bármely természeti állapot következzék is be, ha az egyik cselekvés végrehajtásával nagyobb haszonra tehetünk szert, mintha a másikat választanánk, felesleges a bekövetkezés valószínűségét is számításba vennünk. Ez az elv önmagában természetesen aligha kérdőjelezhető meg. Ám megkérdőjelezhető az a mód, ahogyan Savage alkalmazni próbálta. A savage-i alkalmazás ugyanis azon a feltevésen nyugszik, hogy az eredmények bekövetkezése független a cselekvés végrehajtásától. Így a dominanciaelv egy nagyon egyszerű alkalmazási formájához jutunk. Csak össze kell hasonlítanunk az eredménymátrix oszlopaiban található hasznosságértékeket, és ha találunk olyan sort, ahol az érték minden esetben magasabb lesz, mint a hozzá tartozó oszlopban található többi érték, az adott sorhoz tartozó cselekvést kell választanunk.

Természetesen ez az az eljárás, amit a Jeffrey-féle elméletben nem alkalmazhatunk. Mint láttuk, ebben az esetben a feltételes valószínűségek értéke nem lesz azonos egy adott oszlopban. Ezért a dominanciaelv nem alkalmazható oly módon, ahogy az a Savage-féle elmélet esetében történt. Ez azonban hasznos következmény. Ezért tudjuk megmagyarázni, miért nem irracionális az egészség érdekében áldozatokat hozni. Mindez nem jelenti azt, hogy a dominanciaelvet, amelynek racionalitásáról igen erős intuíciónk van, fel kellene adni. Inkább arra mutat rá, hogy a Savage-féle döntéselmélet a Jeffrey-féle elmélet egy határesete: abban az esetben alkalmazható, amikor a természeti állapotok bekövetkezése független a választott cselekvéstől. Ekkor ugyanis:

	S_1	S_2
A_1	$\Pr(S_1 \mid A_1) = \Pr(S_1)$	$\Pr(S_2 \mid A_1) = \Pr(S_2)$
A_2	$\Pr(S_1 \mid A_2) = \Pr(S_1)$	$\Pr(S_2 \mid A_2) = \Pr(S_2)$

Tehát az azonos oszlophoz tartozó eredmények bekövetkezésének valószínűsége azonos lesz. Ilyenkor természetesen érvényes a dominanciaelv. Ám ettől még nem tekinthető a racionális döntés univerzális szabályának.

De a dominanciaelv univerzalitásának feladása és a feltételes valószínűségek figyelembevétele más áldásos következményekkel is jár, elsősorban azokban az esetekben, amikor interaktív döntési szituációkat vizsgálunk. Az interaktív döntési helyzetek vizsgálatának terepe a játékelmélet. A játékelmélet alapproblémája nem azonos a döntéselmélet által vizsgált kérdéssel, hiszen míg utóbbi számára a várható hasznosság maximalizációjaként értelmezett cselekvést megelőző megfontolások matematikai modelljének megalkotása az alapvető feladat, a játékelmélet az interaktív szituációkban kialakuló *egyensúly*, illetve a *megoldás* fogalmait vizsgálja. Ám a két elmélet nem teljesen független egymástól. Nyilvánvaló, hogy a racionalitás fogalma és a racionális megfontolások megértésének igénye hidat kell verjen köztük. A Jeffrey-féle döntéselmélet alkalmazása talán segíthet e kapcsolat megértésében. E kapcsolat vizsgálata egyben visszavezet eredeti kérdésfeltevésünkhöz, racionalitás és moralitás viszonyának problematikájához is.

A legcélravezetőbb, ha felidézzük a fogolydilemma-szerű szituációk jól ismert példáját. Mint láttuk, a domináns stratégia alkalmazása (vagyis annak a stratégiának az alkalmazása, amely, bármit is tegyen a másik, az adott cselekvő számára nagyobb nyereményt ígér) kollektív módon szuboptimális eredményhez vezet. Létezik olyan stratégiakombináció, amelynek alkalmazása valamennyi résztvevő számára előnyösebb megoldást kínálna. Nevezzük ezt a stratégiát kooperatív stratégiának! Hogyan reprezentálhatjuk mármost egy, a fogolydilemma-szerű helyzetben lévő személy megfontolásait a döntéselmélet segítségével? Először érdemes megnézni, milyen lenne a döntési helyzet ábrázolása a Savage-féle modellben.

	Pr(a másik kooperál) = p	Pr(a másik nem kooperál) = 1–p
A: kooperálok	$U(O_{11})$	$U(O_{12})$
Ā: nem kooperálok	$U(O_{21})$	$U(O_{22})$

Miután feltevésünk szerint a legjobb eredményt az hozza számomra, ha csak én nem kooperálok, tehát $U(O_{21}) > U(O_{11})$, a legrosszabbul pedig az egyoldalú kooperálással járhatok, tehát $U(O_{22}) > U(O_{12})$, a Savage-féle döntéselméletet alapul véve, az egyszer játszott fogolydilemma-szerű szituációban a nem kooperatív stratégia alkalmazása lesz a racionális, hiszen bármit tegyen is a másik, számomra mindig jobb eredményt ígér az, ha nem kooperálok: a nem kooperálás a domináns stratégia. Lássuk, hogyan ábrázolhatunk egy hasonló szituációt a Jeffrey-féle elmélet segítségével:

Értékek:

	A másik kooperál	A másik nem kooperál
A: kooperálok	$V(O_{11}$ & kooperálok) = v_1	$V(O_{12}$ & kooperálok) = v_2
Ā: nem kooperálok	$V(O_{21}$ & nem kooperálok) = v_3	$V(O_{22}$ & nem kooperálok) = v_4

Valószínűségek:

	A másik kooperál	A másik nem kooperál
A: kooperálok	Pr(a másik kooperál | kooperálok) = p_1	Pr(a másik nem kooperál | kooperálok) = p_2
Ā: nem kooperálok	Pr(a másik kooperál | nem kooperálok) = p_3	Pr(a másik nem kooperál | nem kooperálok) = p_4

A cselekvések várható értéke mármost a következő lesz:

$$CEV\ (kooperálok) = Pr(a\ másik\ kooperál \mid kooperálok)\ V(O_{11}\ \&\ kooperálok) + Pr(a\ másik\ nem\ kooperál \mid kooperálok)\ V(O_{12}\ \&\ kooperálok)$$

$$CEV\ (nem\ kooperálok) = Pr(a\ másik\ kooperál \mid nem\ kooperálok)\ V(O_{21}\ \&\ nem\ kooperálok) + Pr(a\ másik\ nem\ kooperál \mid nem\ kooperálok)\ V(O_{22}\ \&\ nem\ kooperálok)$$

Vajon módosíthatja-e a fogolydilemma-szerű szituációkban a „nem kívánatos" (hiszen szuboptimális) eredményt az, ha a Jeffrey-féle elméletet vesszük alapul? Azt gondolom, igen. Persze nem abban az értelemben, hogy a Jeffrey-féle elméletre alapozott döntés minden esetben racionálissá tenné a kooperációt. Egy ilyen javaslat triviálisan hamis lenne. Nem vitathatjuk ugyanis, hogy vannak olyan helyzetek, amikor irracionális a kooperáció. De a Jeffrey-féle elmélet segítségével meg tudjuk magyarázni, mi teszi némely esetben tökéletesen irracionálissá a döntést hozó személy számára a kooperatív magatartást, máskor pedig nem.

Mint korábban láttuk, ahhoz, hogy a Savage-féle elmélet működjön, teljesülnie kell két fontos feltételnek, amelyeket a cselekvéstől való függetlenségnek *(β)*, illetve az eredmények elérhetősége neutralitásának *(γ)* neveztünk. Ha ezek a feltételek érvényesülnek, racionális a domináns stratégia alkalmazása. Mit jelent ez a fogolydilemma esetében? Azt, hogy azokban az esetekben egészen bizonyosan irracionális a kooperatív stratégia alkalmazása, amelyekben biztos lehetek abban, hogy a másik cselekvése független lesz az én döntésemtől, és ahol az a tény, hogy egy adott cselekedetet végrehajtok-e vagy sem, nem befolyásolja azt, hogy hogyan értékelem az eredményt. Természetesen sok olyan fogolydilemma-szerű szituáció van, amelynek esetében e feltételek teljesülnek. Vegyük példaként a fogolydilemma eredeti kerettörténetét. Ha azt feltételezzük, hogy a két fogoly csupán a bankrablás erejéig kötött alkalmi szövetséget, amúgy utálja egymást, és mindkettő biztos benne, hogy többet látni se fogja a másikat, vitathatatlanul az árulás stratégiájának (tehát a nem kooperatív stratégiának) a választása racionális. Egyikük sem feltételezheti, hogy a másik döntését befolyásolhatja az ő döntése, és egyikük sem tulajdonít az árulásnak negatív értéket.

Módosítsuk azonban most az eredeti helyzetet. Tegyük fel, hogy a két gyanúsított bankrabló nem először áll a vizsgálóbíró előtt, és azt is sejtik, nem utoljára. Már régóta együtt dolgoznak, és a továbbiakban is szeretnék fenntartani a már jól bevált együttműködést. Ebben az esetben döntésük egyben *jelzés* is lesz a másik számára, hogy az mennyiben számíthat rájuk. Saját döntéseivel ezért mindegyikük befolyásolni tudja a jövőbeli lehetséges eredményeket. Ha tehát a fogoly dilemmáját több-

ször játszott játéknak tekintjük (ahol a szereplők csaknem zérus valószínűséget tulajdonítanak annak, hogy a következő alkalom lesz az utolsó, amelyben hasonló szituációban találkoznak), a Jeffrey-féle elmélet segíthet annak megértésében, hogyan alakulhat ki kooperáció.

Ám a tanulmányunkban vizsgált probléma szempontjából nem ez az igazán fontos eset. Módosítsuk a kerettörténetet inkább a következőképpen! Nem két rablót, hanem két forradalmárt tartanak fogva. Számukra nemcsak az együttműködés eredményének van értéke, hanem magának az együttműködés fenntartásának is. Nekik tehát nemcsak az számít, hogy hány évet kell börtönben tölteniük, hanem hogy ezt az árulás vagy a kooperatív stratégia alkalmazásával érik-e el. Ebben az esetben a kooperatív stratégia értéke olyan mértékben megnő, és annak valószínűsége, hogy a másik nem ezt a stratégiát alkalmazza, olyan mértékben csökken, hogy egyértelműen racionális lesz a kooperatív stratégia alkalmazása.

A fogolydilemma-szerű szituációk olyan interaktív döntési szituációt reprezentálnak, amely – mint az közismert – igen hasznos lehet számos mindennapi döntési probléma elemzése során. Mint láthattuk, a morális motivációk szerinti cselekvésnek az az egyik jellegzetessége, hogy ilyenkor nemcsak a cselekvés eredményét, hanem magát a cselekvést is értékeljük. Azonban ez nem megkülönböztető jegye a moralitásnak: mint a fenti, dohányzásra vonatkozó példa mutatja, más esetek is vannak, ahol a cselekvő nemcsak a következményeket, hanem magát a cselekvést is értékeli. A morális normák szerinti viselkedésnek tehát nem *elégséges* feltétele, hogy olyan cselekvésnek tekintsük, ahol magát a cselekvést is értékeljük. De ez vizsgált problémánk szempontjából nem is fontos kérdés. Ami fontos az az, hogy *csakis* olyan cselekvést tekintünk morális indíttatásúnak, amely magát a cselekvést is értékeli. A morális cselekvés *szükséges* feltétele tehát, hogy a cselekvésének az egyén önértéket tulajdonítson. Röviden: az erényes cselekvés önmaga jutalma. Nem abban a – triviálisan hamis – értelemben, hogy az egyén számára mindig hasznos következményekkel jár, hanem abban az értelemben, hogy a cselekedet erényes (illetve erkölcstelen) volta is hozzájárul ahhoz, ahogy egy döntési szituációt értelmezünk. Formálisan a következőről van tehát szó. A cselekedetek értékelését az általuk meghatározott szituációk segítségével a következő módon osztályozhatjuk:

1. erkölcsileg semleges szituáció:

$$\forall(A_j)\ \forall(O_{ji})\ \{\mathrm{V}(O_{ji}\ \&\ A_j) = V(O_{ji})\},$$

2. erkölcsileg nem semleges szituáció:

$$\exists(A_j)\ \forall(O_{ji})\ \{\mathrm{V}(O_{ji}\ \&\ A_j)\ \#\ V(O_{j})\}.$$

Az első esetben a racionális cselekvőnek csak cselekedeteinek következményeit kell értékelnie, míg a második esetben az értékelésnek az is része, hogy végrehajt-e, illetve nem hajt-e végre egy bizonyos típusú cselekedetet. A cselekvésnek tulajdonított érték ekkor nem pusztán az értékelő függvénytől és az eredmények bekövetkezésének valószínűségétől függ: maga a választott cselekvés típusa is befolyásolhatja a valószínűségekkel súlyozandó értékeket. Ezért amikor ilyen helyzeteket vizsgá-

lunk (szemben például a *3. fejezetben* példaként említett „közgazdaságtani" döntésekkel, ahol maguknak a cselekedeteknek nem tulajdonítunk értéket), az értékfüggvényt úgy kell meghatároznunk, hogy igaz legyen rá a következő:

$$\exists(A_j) \forall (O_{ji}) \{[V(O_{ji} \& A_j) > V(O_{ji})] v [V(O_{ji} \& A_j) > V(O_{ji})]\}.$$

A fogolydilemmát segítségével a következőképp értelmezhetjük:

	Pr(a másik kooperál) = p	Pr(a másik nem kooperál) = 1–p
A: kooperálok	$V(O_{11}$ & kooperálok) = v_1	$V(O_{12}$& kooperálok) = v_2
Ā: nem kooperálok	$V(O_{21}$ & nem kooperálok) = v_3	$V(O_{22}$ & nem kooperálok) = v_4

O_{11}: kölcsönösen kooperatív viselkedés következménye,
O_{12}: „balekviselkedés" következménye,
O_{21}: „potyautas-viselkedés" következménye,
O_{22}: kölcsönös árulás következménye.

Mint fentebb láthattuk, ha értékelő függvényünk pusztán a következményeket foglalja magában, egy ilyen helyzetben a nem kooperatív viselkedés választása a racionális, hiszen ez a domináns stratégia. Ám ha feltesszük, hogy $V(A) > V(\bar{A})$, és hogy a cselekvésnek tulajdonított érték befolyásolhatja értékelő függvényünket, akkor lehetséges, hogy miután feltételezésünk szerint $\Pr(O_{11}) = \Pr(O_{21})$, viszont $V(O_{11}$ & A) $> V(O_{21}$ & Ā), az árulás már nem lesz domináns stratégia. Ebben az esetben minden attól függ, mekkora értéket tulajdonítunk egy cselekvésnek, és hogy milyen kockázattal és veszteségekkel jár az egyoldalú kooperáció.

Összefoglalva tehát egy racionális személy döntését egy ilyen szituációban az határozza meg, hogy

$$\Pr(O_{11} \mid \text{A})\ V(O_{11} \& A) + \Pr(O_{12} \mid A)\ V(O_{12} \& A) \text{ vagy}$$
$$\Pr(O_{21} \mid \bar{A})\ V(O_{21} \& \bar{A}) + \Pr(O_{22} \mid \bar{A})\ V(O_{22} \& \bar{A})$$

képviseli-e számára a nagyobb értéket.

Miután azonban a fenti példában $\Pr(O_{ji} \mid A_j) = \Pr(O_{ji})$, a választott cselekvés nem befolyásolja valamely eredmény bekövetkezésének valószínűségét. Ezért ekkor:

$$\text{CEV}(A) = \Pr(O_{11})\ \text{V}(O_{11} \& A) + \Pr(O_{12})\ V(O_{12} \& A)$$
$$\text{CEV}(\bar{A}) = \Pr(O_{21})\ V(O_{21} \& \bar{A}) + \Pr(O_{22})\ V(O_{22} \& \bar{A}).$$

Ebben az esetben jól látható tehát, mekkora jelentősége van annak, hogy miként értékelünk egy cselekedetet, de az is látható, hogy megfontolásaink nem lesznek függetlenek a választásunkból ránk háramló következményektől. Lehetséges például, hogy egy fogolydilemma-szerű szituációban egy fogoly nagyra értékeli a kooperatív viselkedést, de ha rendkívül alacsony annak valószínűsége, hogy a másik nem fog vallani, akkor nem racionális számára a kooperatív stratégia választása, még akkor sem, ha amúgy utálja az árulkodást. Hasonló módon: minél nagyobb a beígért büntetés mértéke az egyoldalú kooperáció esetén, tehát minél nagyobb $\text{V}(O_{12}$ & A) abszolút értéke, annál kevésbé racionális megkockáztatni az együttműködést. Ilyenkor valóban „hősies" vagy „aszkétikus" jellemre van szükség: a hős esetében V(A)

nagyon magas, az aszkétánál pedig $V(O_{12})$ abszolút értéke alacsony (utóbbi tehát csekély negatív értéket tulajdonít a veszteségeknek).

Igen sok olyan eset van azonban, amikor a kooperatív stratégia alkalmazása nem fenyeget hatalmas veszteségekkel. Ilyen esetben az erkölcsi normák értékelésének nagy hatása lehet a cselekedetekre. Lehet, hogy mások össze fogják rondítani a parkot, de nekem megéri azt a kis kényelmetlenséget, hogy a szemetet elcipelem az első ládáig. Nem olyan nagyok a veszteségeim, ha a többiek szemetelnek, hogy megérné számomra feladni azt a morális normakövetésből származó értéket, hogy *én* nem teszek ilyesmit. Fordított esetre is van példa. A legtöbben értékeljük az ígéretek betartását. Ám az ígéret megtétele és a betartása által megkövetelt cselekvés között bizonyos idő telik el, és előfordulhat, hogy a cselekvőnek az ígéret megtételekor várt költségek helyett jóval nagyobb veszteségekkel kell számolnia akkor, amikor az ígéret betartására sort kell kerítenie. Ebben az esetben lehetséges, hogy valaki nem fogja betartani adott szavát, még abban az esetben sem, ha úgy véli, az ígéretek megtartása morális érték.

A következmények és a cselekedet morális érteke közti kapcsolatot még világosabb a pozitív normák esetében. Még ha el is fogadnánk, hogy a negatív (tiltó) normák mind feltétlenek, és ezáltal kívül kerülnek a racionális megfontolások terén, a pozitív normák tekintetében senki sem állítana ilyesmit. Senki sem állítaná, hogy valaki, akiből nem lesz Teréz anya, ne értékelhetné a „Segítsd felebarátodat!" normáját. Ám hogy a segítség milyen mértékére vagyunk képesek és alkalmasak, azt az dönti el, hogy hogyan súlyozzuk a segítségnyújtás értékét és az abból ránk háramló költségeket. Ez utóbbit pedig jellemünk és személyiségünk határozza meg.

Racionalitás és moralitás tehát nem feltételezik a személyiség kettészakítását pusztán a cselekvés eredményeit számba vevő egoista Énre és a morális normáknak valamilyen irracionális (vagy „transzracionális") módon engedelmeskedő Énre. Cselekedeteinket számos norma befolyásolja, és ezek ereje nem egyforma. A „Ne ölj!" parancsa olyan norma, ami legtöbbünk számára gyakorlatilag feltétlenként kezelendő: nincs olyan következmény, amelyet pozitívan értékelnénk, ha egy másik ember életének kioltása vezet hozzá. De a „Segítsd a szegényeket!" vagy a „Támogasd a kulturális értékeket!" normái nem követelik meg, hogy a ránk háramló következmények figyelembevétele nélkül mindenünket felajánljuk a szembejövő koldusnak vagy a Nemzeti Színház építésére létrehozott alapítványnak.

Van még egy fontos következménye annak, ahogyan a Jeffrey-féle döntéselmélet segítségével megpróbáltam elhelyezni a normakövető viselkedés magyarázatát a racionális megfontolások logikai terében. Jeffrey döntéselmeléte ugyanis, csakúgy, mint a Savage-féle elmélet, bayesiánus. Ez azt jelenti, hogy az ezen eleméletekben használt valószínűségi és értékelő (hasznossági) függvények idioszinkratikusak. Éppen ezért hangsúlyoztam a *2. fejezetben*, hogy az erkölcsi motivációk racionális döntések logikai terében történő elhelyezésének kérdése nem a cselekvés külső megítélésére vonatkozó probléma. És ezért érveltem amellett, hogy a *de gustibus*-elvet mint posztulátumot ez esetben is alkalmaznunk kell. Cselekvésének magyarázata során adottnak kell tehát vennünk, hogy valaki miként értékeli a normakövető viselkedést. De ez nem jelenti azt, hogy cselekvését morális értelemben igazoltuk volna. A morális értelemben vett igazoláshoz, vagyis egy cselekvés harmadik sze-

mélyű értékeléséhez, mint azt már korábban is jeleztem, bizonyosan túl kell lépnünk a racionális döntések elméletének fogalmi keretein. Annak megértéséhez azonban, hogy a cselekvő által értékelt norma hogyan lehet része a cselekvést megelőző megfontolásoknak, nem kell kilépnünk a döntéselmélet fogalmi keretéből, csak a probléma reprezentálására alkalmas modellt kell választanunk.

5. Konklúzió

Fejtegetéseimet azzal a kijelentéssel kezdtem, hogy a racionális döntések elméletének és a morális normakövetés cselekvésmagyarázatban betöltött szerepének viszonya problémákat rejt magában. Ennek okát abban látom, hogy a társadalomtudósok és a filozófusok jó része a racionális döntések elméletének olyan modelljét használja, amelyben a cselekvésnek csak származtatott értéke lehet. Megpróbáltam megmutatni, hogy ez a felfogás egy igen régi filozófiai tradíció része (nem része azonban, s ezt talán fontos megemlíteni, az antik morálfilozófiai tradíciónak). A döntéselmélet e változata azonban csak bizonyos feltételek teljesülése esetén ad adekvát képet a döntést megelőző megfontolások logikai struktúrájáról. Ám létezik a döntéselméletnek olyan változata, amely képes figyelembe venni azt a tényt, hogy a racionális cselekvő nemcsak a cselekedeteiből származó következményeket, hanem magát a cselekedetet is értékeli. A döntéselmélet e változata lehetővé teszi, hogy megértsük, hogyan helyezkednek el a morális normák mint motiváló erők a cselekvést megelőző megfontolások logikai terében.

Hivatkozások

Arnould, A.–Nicole, P. 1970. *La logique de Port-Royal ou l'art de penser*. Párizs: Flammarion (első megjelenés: 1660)

Csontos László 1985. *Szituációs-logikai modellek a társadalomtudományokban.* Kézirat.

Eells, E. 1982. *Rational Decision and Causality*. Cambridge: Cambridge University Press

Elster J. 1983. *Sour Grapes. Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press

– 1995. *A társadalom fogaskerekei.* Budapest: Osiris–Századvég

– 1989. *The Cement of Society. A Study of Social Order*. Cambridge: Cambridge University Press

Gauthier, D. 1986. *Morals by Agreement.* Oxford: Clarendon

– 1991. Why Contactarianism? In: P. Vallentyne (ed.) *Contractarianism and Rational Choice*. Cambridge: Cambridge University Press

Harman, G. 1977. *The Nature of Morality. An Inroduction to Ethics*. New York: Oxford University Press

Hirschleifer, I. J.–Riley, J. G. 1998. A bizonytalanságban hozott döntések elemei. In: Csontos László (szerk.): *A racionális döntések elmélete.* Budapest: Osiris

Hobbes, T. 1970. *Leviatán*. Budapest: Magyar Helikon

Hume, D. 1976. *Értekezés az emberi természetről.* Budapest: Gondolat

Jeffrey, R. 1983. *The Logic of Decision*. New York: McGraw-Hill

Kant, I. 1991. *Az erkölcsök metafizikájának alapvetése.* Budapest: Gondolat

Luce R. D.–Raiffa, H. 1958. *Games and Decisions.* New York: John Wiley & Sons

Rawls, J. 1980. Kantian Constructivism in Moral Theory. *The Journal of Philosophy*, 77.

Smith, H. 1991. Deriving Morality from Rationality. In: Peter Vallentyne (ed.) *Contractarianism and Rational Choice*. Cambridge: Cambridge University Press

Weber, M. 1987. Gazdaság és társadalom. *A megértő szociológia alapvonalai 1.* Budapest: Közgazdasági és Jogi Könyvkiadó