

Moral na temelju dogovora? Jedna kritika Davida Gauthiera

Sažetak:

*Nakon što uvodno daje osnovne kontekstualne informacije o teoriji racionalnog izbora, autor se u radu bavi Gauthierovom inačicom te teorije. Razmatrajući problem poštivanja moralnih pravila u društvu od strane njegovih članova i definirajući ga kao problem poštivanja ugovora/dogovora, američki teoretičar racionalnog izbora David Gauthier ponudio je u svojoj knjizi *Morals by Agreement* alternativu Smithovoj "nevidljivoj ruci" tržišta i Hobbesovoj "vidljivoj nozi" Levijatana. Umjesto sila tržišta i prisile vladara, Gauthier predlaže internalizaciju sklonosti poštivanja dogovora i to na temelju racionalnog slijeđenja vlastitog interesa. Osnova njegove ideje je svojevrсни "račun koristi", kojim se može pokazati da je, uz određene uvjete koje Gauthier smatra relativno skromnima, isplativije odabrati sklonost ka djelovanju kao da smo (moralnim pravilima) ograničeni, nego kao da smo izravni maksimizatori vlastite koristi. Nažalost, Gauthier sam odstupa od pozicije moralne neutralnosti, utjelovljene u konceptu interesno motiviranog poštivanja moralnih pravila, i u svoj račun koristi ne uvrštava sve potencijalne koristi i gubitke. Time se gubi smisao samog računa. U ovom tekstu autor pokušava ispraviti tu Gauthierovu pogrešku te unosi značajne izmjene u njegov izvod. Tako revidi-*

ran račun koristi mogao bi poslužiti kao temelj za zasnivanje "morala na temelju ugovora".

Ključne riječi:

racionalni izbor, David Gauthier, poštivanje ugovora, zatvorenikova dilema, moral na temelju ugovora, izravni maksimizator, ograničeni maksimizator.

Uvod: teorija racionalnog izbora

Već je engleski ekonomist Adam Smith u svojem djelu *Bogatstvo naroda*, objavljenom 1776. godine, pretpostavio da se društvo sastoji od mnoštva "sebičnih" (*self-interested*) pojedinaca koji, sljedeći vlastiti interes i stupajući u međusobne ekonomske interakcije, kroz sustav potpune konkurencije dovode društvo do maksimalnog blagostanja. Smith je time napao prethodne fiziokratske i merkantilističke ekonomske teorije koje su podrazumijevale snažno uplitanje države u regulaciju ekonomije. Takav *laissez-faire* koncept pružio je teorijsku podlogu interesima rastuće građanske klase koja je htjela što više ograničiti djelovanje kralja i države, između ostalog i zbog tadašnjeg običaja dijeljenja trgovinskih i poreznih povlastica plemstvu. Utjecaj Smithovog djela bio je izniman ne samo u političkoj ekonomiji, već u cjelokupnoj anglosaksonskoj filozofiji te u kasnije nastalim društvenim znanostima.

Teorija racionalnog izbora utemeljena je upravo na metodološkim postavkama što ih je nagovijestio još Smith, a koje su svojstvene neoklasičnoj školi ekonomije. Te postavke su postavka metodološkog individualizma i postavka "ekonomskog čovjeka". Metodološki individualizam usmjerava fokus na pojedinca i njegovo pojedinačno djelovanje, a društvene grupe i strukture ne uzima kao zasebne entitete i aktere, već na njih gleda kao na agregate pojedinačnog djelovanja. Postavka "ekonomskog čovjeka" znači da se pojedinac shvaća kao *homo oeconomicus* koji racionalno djeluje s ciljem maksimizacije vlastite koristi.¹ Ipak, treba napomenuti da neki autori odbacuju takvo vezivanje racionalnog izbora uz metodološki individualizam (Ward, 2005.). Također, treba naglasiti da se "sebični" racionalni pojedinci mogu ponašati i "altruistično", uz pretpostavku da im tuđa dobrobit čini zadovoljstvo (tj. daje im korist). No, za razliku od univerzalističkih koncepcija racionalnosti, poput kantovske, kod "sebične" racionalnosti

¹ Vidi npr.: Ostrom i Ostrom, 1971.

(raz)um ne nalaže pojedincu da djeluje u korist drugoga iz pukog principa (kao što to čini u slučaju Kantovog kategoričkog imperativa).

Povijesno gledano, teorija racionalnog izbora javila se 1950-ih i 1960-ih, djelomice pod utjecajem biheviorističke psihologije koja je tvrdila da se ljudska bića ponašaju racionalno i da se njihovo ponašanje tj. reagiranje temelji na sustavu očekivanih nagrada i kazni. Važni za razvoj racionalnog izbora bili su i radovi Johna von Neumanna i Oscara Morgensterna kojima je utemeljena tzv. teorija igara (*game theory*). Ona služi za razmatranje situacija u kojima izbor strategije drugih igrača utječe na naš najbolji izbor i obratno (Ward, 2005: 62).

Anthony Downs je u svojem djelu *An Economic Theory of Democracy* iz 1957. prvi primijenio koncept racionalnog izbora u političkoj znanosti, predlažući model koji je objašnjavao ponašanje birača i stranaka. Birači glasaju za stranke od kojih očekuju najveće “koristi”². Stranke, pak, biraju one platforme koje će privući najveći broj glasača jer im je primarni cilj i smisao postojanja osvajanje vlasti, moći i položaja. Tim radom udareni su temelji teorije javnog izbora.

No, ubrzo se pokazalo da se koncept racionalnog izbora može upotrijebiti u političkoj teoriji na posve drugačiji način. U svojem djelu *Teorija pravednosti* iz 1971., progresivni³ američki filozof John Rawls pokušao je ponuditi alternativu smithovskom duhovnom nasljeđu *laissez-fairea* koje je sprječavalo širu primjenu koncepta države blagostanja (*welfare-state*) u SAD-u. Prividno paradoksalno, Rawls se pritom poslužio upravo teorijom racionalnog izbora i njenim inherentno konzervativnim metodološkim pretpostavkama. No, i unutar takve prividno “neprijateljske” paradigme uspio je konstruirati koncept društvenog ugovora u kojem bi racionalni sebični pojedinci, pod određenim skromnim ograničenjima, odabrali puno “pravedniji” sustav od nemilosrdnog socijalnog darvinizma.

Dijelom potaknut Rawlsovim primjerom, američki teoretičar društvenog ugovora David Gauthier na sličan je način primijenio racionalni izbor u djelu *Morals by Agreement (Moral temeljen na dogovoru)* iz 1986. godine. Njegov cilj je bio pokazati da je (uz određena ograničenja) moguće utemeljiti moralne norme isključivo na racionalnom interesu sebičnih pojedinaca. U ovome radu prvo ću prikazati Gauthierov izvod računa koristi, a zatim ću iznijeti svoju kritiku tog izvoda i ponuditi vlastite modifikacije.

² Koje ne moraju biti nužno samo materijalne, nego mogu biti i duhovne ili ideološke.

³ Riječ “progresivan” (*progressive*) u američkom političkom diskursu označava osobu koja ima “lijeva” stajališta u ekonomskim pitanjima (podržava ideju države blagostanja), te liberalna stajališta u socijalnim pitanjima (tolerancija prema manjinama - vjerskim, nacionalnim, rasnim, seksualnim; podrška pravu na pobačaj, itd.).

1. Gauthierov “moral na temelju dogovora”

Gauthier je pokušao pokazati kako se moralna pravila mogu temeljiti na ugovoru te kako se potonje poštivanje tih pravila može rastumačiti svojevrsnim računom koristi. Moralna pravila nekog društva postavljaju određena dogovorena ograničenja na ponašanje članova. Kako je djelovanje pojedinaca motivirano prvenstveno brigom za maksimiziranje vlastite koristi tj. svodi se na zadovoljavanje vlastitih želja, poštivanje takvih pravila predstavlja za svakog člana odricanje određenog dijela potencijalnih koristi, uslijed žrtvovanja apsolutne slobode djelovanja. Ipak, s obzirom da se pravila pridržavaju i drugi članovi društva, svaki član uživa koristi koje proizlaze iz postojanja urednog društva. Te koristi su veće nego one kojih se pojedinac odrekao pridržavajući se ugovora.

No, kod poštivanja takvog tipa ugovora javlja se problem jer on ima oblik igre “zatvorenikove dileme”: bolje je da ugovori postoje i da se poštuju, nego da ih uopće nema; no, najbolje je ako mi kršimo dogovore, a drugi ih se pridržavaju; najgore je ako je obratno.⁴ Dakle, iako svatko ostvaruje neto-koristi od postojanja ugovora, svi pojedinci ujedno imaju i poticaj da jednostrano prekrše dogovor i steknu dodatne koristi na račun ostalih. Sličnu primjedbpu upućuje i Budala (*Foole*) u Levijatenu, kritizirajući Hobbesov treći prirodni zakon (“osobe se pridržavaju svojih zavjeta”)⁵ tvrdeći da se on ne može obraniti principom vlastitog interesa jer je ponekad korisnije (a time i racionalnije) prekršiti dogovor.

Hobbes odgovara da takva osoba, koja je sklona kršiti dogovore, neće biti primljena u društvo (tj. u kooperativne aranžmane) te da ne može očekivati da će uopće biti u prilici požnjati ni ekstra-koristi od suradnje ni eventualne ekstra-koristi od kršenja dogovora (Gauthier, 1986.: 160-161). Dakle, čak i iako pojedinačna kršenja dogovora mogu biti u njenu korist, takva sklonost joj dugoročno šteti. S druge strane, sklonost poštivanju ugovora čini osobu poželjnim partnerom u suradnji na obostranu korist pa je sama po sebi korisna! Korisno je, stoga, poštivati pojedinačne dogovore jer time održavamo reputaciju poštenja.⁶

Regulaciju tako može vršiti “nevidljiva ruka” tržišta (Smith), u ovom

⁴ Kant je problem “zatvorenikove dileme” riješio svojim kategoričkim imperativom, koji određuje da se treba činiti o opći zakon i tako ograničava ostale, a da ga mi osobno imamo slobodu kršiti.

⁵ Usp. Hobbes, 1991.: 153.

⁶ Takvo razmišljanje su već ponudili Machiavelli i Guiccardini u 16. stoljeću, razmatrajući problem poštivanja dogovora među vladarima. Vidi: Berridge, Keens-Soper i Otte, 2005.: 17-21, 49.

slučaju kroz vrijednost dobre reputacije. Hobbes je, naravno, prethodno ponudio svoje rješenje u vidu “vidljive noge” vladara (Levijatana) koja putem prijetnje silom omogućuje željeni ishod. Uzeti zajedno, ovi koncepti čine moralnost, shvaćenu u smislu stavljanja određenih (internaliziranih) ograničenja na djelovanje “sebičnog” pojedinca koji maksimizira svoju korist, prividno suvišnom. No, svaki od ta dva sustava ima i svoju slabost. Smithovsko savršeno tržište zakazuje u slučaju da postoje eksternalije. Institucija “vladara”, s druge strane, nosi i određene operativne troškove, stoga je konačni rezultat suboptimalan (čak i ako pretpostavimo da je vladar potpuno pošten). Stoga bi svima bilo bolje kada bi se svatko dobrovoljno pridržavao dogovora, odnosno kada bi u društvu postojala općeprihvaćena moralna pravila.

Čak i Budala priznaje da postoje prilike za suradnju oko kojih bi se svaka osoba složila kada bi mogla očekivati da će dogovori biti poštivani, ali koje ostaju nerealizirane jer svatko očekuje da se drugi neće držati dogovora. Budala smatra da je za racionalno kršenje dogovora dovoljno da koristi kršenja budu veće od koristi poštivanja, a da bi pristanak na suradnju bio racionalan, osoba mora smatrati da je i za druge racionalno pristati. U slučaju potpunih informacija i potpune racionalnosti ostvaruju se samo oni ugovori koje se nikome ne isplati ne poštivati (*igra suradnje* – npr. vožnja desnom stranom), a svi oni koji bi mogli donijeti koristi svakome, no kod kojih postoji poticaj da se “vara”, neće se uopće ni sklapati (*zatvorenikova dilema*). (Gauthier, 1986.: 161-164).

Već Hobbesov rad pokazuje da postoje dvije perspektive s obzirom na korisnost pridržavanja dogovora. Njima odgovaraju dvije vrste igrača u Gauthierovoj igri poštivanja moralnih pravila. Gauthier definira **izravnog** (*straightforward*) **maksimizatora (IM)** kao osobu koja teži maksimizirati svoju korist s obzirom na strategije drugih igrača (IM je poput Hobbesove Budale). **Ograničeni** (*constrained*) **maksimizator (OM)**, pak, osoba je koja u nekim situacijama teži maksimizirati svoju korist ne s obzirom na tuđe strategije, već s obzirom na koristi drugih igrača u zajedničkoj strategiji (*joint strategy*). Kod zajedničke strategije rezultat je unaprijed poznat, pod uvjetom da je se svi sudionici pridržavaju, a u pitanju je samo raspodjela neto-koristi (tj. razlike između koristi koja proizlazi iz dogovora i žrtve koju pojedinac mora podnijeti). Ograničeni maksimizator ima uvjetnu sklonost temeljiti svoja djelovanja na zajedničkoj strategiji, bez da razmatra da li bi mu neka individualna strategija mogla donijeti veće koristi (Gauthier, 1986.: 166-168). Svaki oblik dogovora u društvu, uključujući zakone i moralna pravila, oblik je takve zajedničke strategije. Ograničeni maksimizator je **spreman** na suradnju ako će ishodi biti korisni

(*beneficial*) i poštenu (*not unfair*), a *doista* će i *surađivati* ako očekuje da će djelovanje biti korisno (ovdje mora uzeti u obzir da će neke osobe varati ili odbiti surađivati). **OM**-ovi tako mogu iskoristiti koristi suradnje koje su nedostupne **IM**-ovima (neovisno o njihovoj mudrosti), a **IM**-ovi mogu u nekim situacijama iskorištavati **OM**-ove. Oboje su racionalni, no tko je u pravu? (Gauthier, 1986.: 168)

Račun koristi za **OM**-ove i **IM**-ove

Gauthier pokušava ponuditi dokaz da je racionalno izabrati **sklonost ograničenom** maksimiziranju, a ne *izravnom*. Ne radi se o konkretnom pojedinačnom djelovanju u određenom slučaju, već o racionalno-interesno utemeljenoj internalizaciji sklonosti ka poštivanju dogovora, odnosno moralnom djelovanju. Tako preformulirano pitanje sada glasi: ako uzmemo u obzir tuđe sklonosti kao fiksne, je li očekivana korist veća ako maksimiziramo vlastite koristi s obzirom na strategije ostalih (**IM**) ili ako smo skloni surađivati u dogovorima koji donose približno poštene i optimalne ishode (**OM**)?

Naravno, Gauthier ističe da ćemo razmatrati samo one slučajeve u kojima se djelovanje po tim strategijama razlikuje. Takvi se slučajevi javljaju uz dva uvjeta: (1.) odnosi moraju sadržavati priliku za obostrano korisnu i poštnu suradnju (poticaj **OM**) i (2.) odnosi moraju sadržavati priliku za ostvarenje dodatne koristi kršenjem dogovora (poticaj **IM**).

Označimo s **x** korist od *individualne strategije*, s **y** korist od *zajedničke strategije* (dogovora), a sa **z** korist od *varanja*. Kako bismo zadovoljili gornje uvjete, odnos spomenutih koristi mora biti $x < y < z$.⁷ Neka **p** bude procijenjena vjerojatnost da će drugi surađivati, $0 < p < 1$. Osoba koja treba odabrati između sklonosti djelovanju kao **OM** ili **IM** može razmišljati na dva načina:

1. Ako sam **IM** onda će mi očekivana korist biti $p*z + (1-p)*x$ → (ja varam, ostali surađuju) + (nema suradnje), a ako sam **OM** očekivana korist će mi biti $p*y + (1-p)*x$ → (svi surađujemo) + (nema suradnje), s obzirom da je **z** veće od **y**, bolje je biti **IM**!

2. Ako sam **IM** očekujem da su to i ostali pa mi je korist **x** (nikad nema suradnje), a ako sam **OM** korist mi je $p*y + (1-p)*x$, s obzirom da je **y** veće od **x** za bilo koji **p** veći od **0** (a jednako za

⁷ Te vrijednosti možemo definirati i kao relativne u odnosu na maksimalnu dobit z, tj. kao $0 < x_{rel} < y_{rel} < 1 (= z)$.

0), bolje je biti **OM**!

Naravno, Gauthier ističe da jedan od ova dva argumenta mora biti pogrešan! Prvi bi vrijedio samo kada bi vjerojatnost da drugi surađuju bila neovisna o našoj sklonosti ka suradnji. No, kako to nije slučaj, **IM** nema prilika za stjecanje koristi koje se otvaraju za **OM**-a, jer druge osobe neće surađivati s njime osim ako nisu sigurne da se radi o igri koordinacije, a ne o zatvorenikovoj dilemi.

No, Gauthier priznaje da drugi argument ima slabost jer ovisi o ispravnoj identifikaciji osobe kao **IM**-a ili **OM**-a. Stoga bi logični dio **IM** strategije bio da se osoba pretvara da je **OM** i tako biva pripuštena u zajedničke aranžmane iz kojih može izvući **z**. Naravno, to možemo spriječiti uvođenjem uvjeta **transparentnosti** (potpune prozirnosti) koji onemogućava varku. Realističnija bi pretpostavka bila da su osobe **transluentne** (poluprozirne) tj. da možemo s određenom vjerojatnošću znati njihovu pravu orijentaciju. U takvom svijetu **OM**-ovi ponekad neće prepoznati jedni druge i neće stupiti u korisnu interakciju. **OM**-ovi će također ponekad kooperativno djelovati prema **IM**-ovima, a ovi će ih u tom slučaju (ako ih također prepoznaju!) moći iskoristiti, tvrdi Gauthier. Tako će **OM** proći gore nego kod transparentnosti, a **IM** bolje, ovisno o stupnju transluentnosti.

Gauthier potom razmatra situacije u kojima postoji razlika u ishodima između **OM**-ova i **IM**-ova, tj. kada se **OM**-ovi prepoznaju i stupe u interakciju, te kada **OM**-ovi ne prepoznaju **IM**-ove, a ovi njih prepoznaju pa ih iskoriste. Postoje četiri moguća ishoda interakcije s obzirom na korist koje može ostvariti svaka od strana u dogovoru: **0** (prevareni **OM**) < **x** (nesuradnja) < **y** (suradnja) < **z** (varalica **IM**).⁸ Postoje i tri vjerojatnosti:

1. $p_G \rightarrow$ vjerojatnost da će se **OM**-ovi prepoznati i surađivati

⁸ Treba istaknuti da *0 de facto* predstavlja ekstra-gubitak za sudionika dogovora, jer on gubi i onu korist koju bi ostvario da dogovora nema (*x*). Npr. ako se dva seljaka dogovore da neće jedan drugome krasti stoku, moći će računati na povećanu korist *y*, (*y*>*x*). Neto-korist (*y*-*x*), koja će se ostvariti dogovorom, proizlazi iz toga što neće morati paziti jedan na drugoga ili kupovati alarmne uređaje, nego će se moći potpunije posvetiti uzgoju stoke i povećati svoju produktivnost. Korist *x* je bila ona koju je svaki seljak ostvarivao u situaciji bez dogovora, a činila ju je korist od uzgoja stoke umanjena za iznos stoke koja je mu je ukradena, a uvećana za iznos stoke koju je ukrao. Važno je primijetiti da čak i ako seljak ukrade točno toliko stoke koliko njemu ukradu te time poništi efekt krađe, njegova produktivnost i dalje nije optimalna jer je dio svojeg vremena utrošio na krađu umjesto na uzgoj. No, ako se samo jedan od njih pridržava dogovora, on će izgubiti svu svoju stoku (jer neće više paziti na nju, vjerujući da je problem krađe riješen) ostvarujući tako korist 0, tako da će za njega konačni ishod u odnosu na stanje prije dogovora biti gubitak od 0-*x* tj. -*x*.

($\mathbf{OM}=\mathbf{y}$);

2. $\mathbf{q}_G \rightarrow$ vjerojatnost da \mathbf{OM} -ovi neće prepoznati \mathbf{IM} -ove, a da ovi njih hoće te će ih iskoristiti ($\mathbf{OM}=\mathbf{0}$, $\mathbf{IM}=\mathbf{z}$);

3. $\mathbf{r}_G \rightarrow$ vjerojatnost da je neki član populacije \mathbf{OM} ;

Očekivane koristi za \mathbf{OM} će biti jednake $\{\mathbf{x} + [\mathbf{r}_G \mathbf{p}_G (\mathbf{y} - \mathbf{x})] - (\mathbf{1} - \mathbf{r}_G) \mathbf{q}_G \mathbf{x}\}$, a za $\mathbf{IM} = \{\mathbf{x} + [\mathbf{r}_G \mathbf{q}_G (\mathbf{1} - \mathbf{x})]\}$. Razmatrajući pitanje je li razboritije odabrati sklonost izravnom ili ograničenom maksimiziranju, Gauthier donosi sljedeći zaključak: sklonost djelovanju kao \mathbf{OM} bit će racionalna samo ako je korist od toga veća nego ako smo \mathbf{IM} , a to vrijedi jedino ako je $\mathbf{p}_G/\mathbf{q}_G$ veće od odnosa između koristi od varanja ($\mathbf{1}$) i koristi od suradnje (\mathbf{y}). Kako vrijednost \mathbf{r} raste, tako se odnos $\mathbf{p}_G/\mathbf{q}_G$ potreban da bi bilo racionalno biti \mathbf{OM} smanjuje. Na primjer, ako je raspodjela koristi $\mathbf{0} < \mathbf{1}/3 < \mathbf{2}/3 < \mathbf{1}$, onda $\mathbf{p}_G/\mathbf{q}_G$ mora biti veće od $(\mathbf{r}_G + \mathbf{1})/\mathbf{r}_G$. Tada, ako je $\mathbf{r}_G = \mathbf{1}/2$, onda je dovoljno da se \mathbf{OM} -ovi prepoznaju u $\mathbf{2}/3$ slučajeva, i izbjegnu biti eksploatirani od strane \mathbf{IM} -ova u $\mathbf{4}/5$ slučajeva, da bi sklonost \mathbf{OM} bila racionalna.

2. Ispravan račun koristi za \mathbf{OM} i \mathbf{IM}

Gauthier pokušava utemeljiti odabir osobe između sklonosti djelovanju kao \mathbf{OM} ili kao \mathbf{IM} na etički indiferentnom principu koristi. Naravno, smatra da je uz određene uvjete (koje ne drži pretjerano zahtjevnima) svakoj osobi racionalno odabrati sklonost djelovanju kao \mathbf{OM} . Na tom računu Gauthier temelji postojanje i pridržavanje moralnih načela u društvu.

Nažalost, usprkos silnom trudu oko nepristranosti analize (Gauthier čak dopušta da je u nekim uvjetima razboritije biti \mathbf{IM}), Gauthier je upravo u trenutku kada je pokušavao izgraditi most između ekonomističke logike i racionalističke etike, propao u duboko blato morala. U želji da dokaže svoju tezu da je (uglavnom) ipak bolje surađivati, počinio je nesmotrenu pogrešku te je svoj dokaz nastavio razrađivati s implicitnom (etičkom) pretpostavkom da su \mathbf{OM} -ovi “dobri”, a \mathbf{IM} -ovi “loši momci”. Stoga eksploataciju \mathbf{OM} -ova od strane \mathbf{IM} -ova, koja se dešava kada dvije takve osobe sklope ugovor, interpretira kao *namjernu*. To ga dovodi do daljnje pogreške jer uvodi varijablu \mathbf{q} kao “vjerojatnost da će \mathbf{OM} pogrešno identificirati \mathbf{IM} , a da će sami biti ispravno identificirani, što će dovesti do nesuradnje i eksploatacije” (Gauthier, 1986.: 172-173). No, zapravo nije važno da li \mathbf{IM} -ovi uopće identificiraju svoje partnere kao \mathbf{IM} -ove ili kao \mathbf{OM} -ove, jer je njihov odabir uvijek isti: nesuradnja. Dakle, kod

IM-ova uopće ne postoji pitanje (a dakle ni vjerojatnost) prepoznavanja suprotne strane. **IM**-ovi će uvijek, neovisno o tome je li druga osoba **OM** ili **IM**, odabrati nesuradnju jer ona predstavlja dominantnu strategiju u igri “zatvorenikove dileme”.

Potrebno je stoga unijeti značajne izmjene i u definiranje varijabli i u Gauthierovu formulu korisnosti. One su, po mom uvidu, sljedeće:

p → vjerojatnost da će **OM** prepoznati orijentaciju svoga potencijalnog partnera (ujedno i prava mjera translucenosti);
 $1-p$ → vjerojatnost pogrešnog prepoznavanja;
 r → vjerojatnost da je neka slučajno odabrana osoba s kojom stupamo u interakciju **OM**, tj. udio **OM**-ova u ukupnoj populaciji;
 $= r_G$;

Gauthier definira *translucenost* kao osobinu koja “drugim ljudima omogućuje da utvrde sklonost suradnji neke osobe s određenom sigurnošću, koja nije stopostotna, ali je više od pukog nagađanja” (Gauthier, 1986.: 172). S obzirom na to da postoje samo dvije vrste osoba u populaciji (**OM** i **IM**), puko nagađanje daje nam 50 postotnu (0.5) šansu ispravnog utvrđivanja sklonosti suradnji neke osobe. Translucenost *per definitionem* mora biti bolji indikator. Stoga moguće vrijednosti varijable p (mjere translucenosti) možemo ograničiti na: $0.5 < p < 1$. Varijabla r može poprimiti bilo koju vrijednost između 0 i 1, uključujući i granične vrijednosti: $0 \leq r \leq 1$.

Kada razmatra moguće koristi i štete koje racionalna osoba može očekivati od odabira da bude sklona (**OM**) ili nesklona suradnji (**IM**), Gauthier ponovo odstupa od čisto logičke analize jer u jednadžbu očekivanih koristi (i gubitaka) **OM**-ova namjerno ne uvrštava situacije u kojem, zbog pogrešnog prepoznavanja, **OM** osoba eksploatira drugu **OM** osobu ili sama biva od nje “eksploatirana” (Gauthier, 1986.: 172). Čini se da je razlog tomu što ne želi da u kalkulaciju koristi, od odabira “sklonosti ka suradnji”, ulaze i koristi stečene na “nepošten” način tj. na način na koji ih inače ostvaruju **IM**-ovi kada eksploatiraju **OM**-ove. Obrnutu situaciju, kada naš **OM** biva krivo identificiran te potom “eksploatiran” od strane drugog **OM**-a, Gauthier uopće ni ne spominje. No, sebična i racionalna osoba koja tek treba odabrati hoće li se odrediti kao **OM** ili **IM**, svoj odabir sigurno će htjeti temeljiti na *potpunoj* analizi mogućih dobiti i gubitaka.

Razmotrimo sada račun dobiti i gubitka za **OM** i **IM** u svakoj pojedinoj situaciji u kojoj se “djelovanje po tim strategijama [OM i IM]

razlikuje⁹:

Za OM	Međusobno prepoznavanje (p^2) s drugom OM osobom (r) koje rezultira suradnjom i ostvarenjem ekstra-dobiti ($y-x$)	$rp^2(y-x)$
	Pogrešno prepoznajemo ($1-p$) IM osobu ($1-r$) radi čega bivamo eksploatirani tj. ostvarujemo ekstra-gubitak ($0-x$)	$(1-r)(1-p)(0-x)$
	Pogrešno prepoznajemo ($1-p$) OM osobu (r), koja nas ispravno prepoznaje (p), te ju stoga nehotice eksploatiramo i ostvarujemo ekstra-	$rp(1-p)(1-x)$
	Ispravno prepoznajemo (p) OM osobu (r), ali ona nas pogrešno prepoznaje ($1-p$), te nas stoga nehotice eksploatira i mi ostvarujemo	$rp(1-p)(0-x)$
Za IM	Bivamo pogrešno prepoznati ($1-p$) od strane OM osobe (r) te ostvarujemo ekstra-dobit ($1-x$)	$r(1-p)(1-x)$

Stoga, izmjenom nekih varijabli (u odnosu na Gauthierov koncept) i unošenjem tih dodatnih mogućih situacija jednadžbe troškova i koristi za **OM** i **IM** izgledaju ovako:¹⁰

$$OM = x + rp^2(y-x) - (1-r)(1-p)x + rp(1-p)(1-x) - rp(1-p)x$$

$$IM = x + r(1-p)(1-x)$$

gdje je x dobiven od obostrane nesuradnje, y od obostrane suradnje, a u situaciji eksploatacije **OM** dobiva 0 (tj. gubi x), dok **IM** (ili **OM**) dobiva 1 (tj. ima ekstra dobit jednaku $1-x$). Vrijedi da je $0 < x < y < 1$ tj. z .

Ako sada u Gauthierov primjer ($x=1/3$, $y=2/3$) uvrstimo neke ogledne vrijednosti za r i p dobit ćemo vrlo zanimljive podatke:

⁹ Treba spomenuti da smo izostavili tri dodatne situacije u kojima ili ne dolazi do sklapanja ugovora (OM ispravno prepoznaje IM-a, dva OM-a se međusobno pogrešno prepoznaju) ili dolazi do njegovog sklapanja između dva IM-a koji ga se nikada nisu ni mislili pridržavati. U svim tim situacijama konačna korist svakog sudionika je ista i jednaka je x , odnosno koristi koju sudionici ostvaruju kada dogovor ne postoji.

¹⁰ Naravno, $(0-x)$ smo sada napisali kao x , ali smo promijenili predznak ispred cijelog izraza: $(1-r)(1-p)(0-x) = - (1-r)(1-p)x$;

vrijednosti p (vjerojatnost prepoznavanja)	minimalne vrijednosti r (učestalost OM u populaciji)	vrijednosti r (učestalost OM u populaciji)	minimalne vrijednosti p (vjerojatnost repoznavanja)
0,6	2 (nemoguća vrijednost)	0,2	0,857
0,7	0,75	0,5	0,75
0,8	0,333°	0,9	0,679

Za vrijednost $p = 0,6$ i zadane parametre ($x=1/3$, $y=2/3$), r bi morao poprimiti vrijednost veću od 2 da bi OM-ovi bili u prednosti u odnosu na IM-ove. S obzirom na uvjet da r mora biti manje ili jednako od 1 ($r \leq 1$), minimalnu vrijednost p , pri kojoj bi uopće moglo biti isplativije odabrati ograničenu maksimizaciju, možemo izračunati kroz jednadžbu $(1-p)/(2p-1) \leq 1$, i ona iznosi $p_{\min} = 2/3$.

Odstupanja ovako modificirane jednadžbe od Gauthierovog modela najbolje se vide kada je $r=1$ tj. kada su sve osobe OM. Uz uvjet da su parametri $x=1/3$, $y=2/3$, tada je potrebna vrijednost $p > 2/3$ da bi se isplatilo biti OM, a ne IM. Jer čak i kada su sve osobe OM (uz pretpostavku da ta činjenica nije poznata samim članovima društva¹¹), mi kao OM ostvarujemo gubitke uzrokovane situacijama u kojima mi ispravno prepoznamo drugu osobu kao OM, ali ona nas pogrešno identificira kao IM, te bivamo eksploatirani. Također, mi ostvarujemo i koristi od međusobne suradnje i od naše nehotične eksploatacije drugih OM osoba koje smo zamijenili za IM-ove. Naravno, u svijetu u kojem su svi OM-ovi, ne može nam se dogoditi da nas eksploatira krivo prepoznati IM.

Ako bi, kao Gauthier, zanemarili situacije “nehotične eksploatacije”¹², onda bi nam u gore navedenom primjeru minimalni p bio jednak $p = \sqrt{3} - 1$ tj. $p \approx 0,732$. To predstavlja značajno odstupanje od stvarno potrebnog minimalnog $p = 2/3 = 0,666^\circ$.

No, nema nikakvog razloga za pretpostavku da su koristi od suradnje i nesuradnje tako ravnomjerno postavljene kao što je to u Gauthierovom primjeru: 0 (“žrtva”) $< 1/3$ (nesuradnja) $< 2/3$ (suradnja) < 1 (“eksploatator”). Možemo razmotriti što se zbiva kako se mijenjaju

¹¹ Jer tada više translucentnost i prepoznavanje ne bi bili potrebni, s obzirom da bi svatko znao da su svi ostali OM, te bi svi surađivali.

¹² Tada bi jednadžba glasila: $rp^2(y-x) > r(1-p)(1-x) \Rightarrow 1/3p^2 > 2/3 - 2/3p \Rightarrow p^2 > 2 - 2p \Rightarrow p^2 + 2p - 2 > 0 \Rightarrow (p+1)^2 - 3 > 0 \Rightarrow (p+1)^2 > 3 \Rightarrow p+1 > \sqrt{3} \Rightarrow p > \sqrt{3} - 1$ (naravno, uz ograničenje da p mora biti pozitivna vrijednost);

odnosi između koristi od suradnje i nesuradnje te njihove vrijednosti s obzirom na dobit od eksploatacije. Drugačije rečeno, do kakvih promjena dolazi kada vrijednosti y i x rastu ili padaju.

Za pojednostavljenu jednadžbu uzet ćemo da se vrijednosti x i y izražavaju kao udjeli ili postoci od z , tj. najveće moguće dobiti u igri (situacija kada mi eksploatiramo drugog), koju možemo označiti sa 1 . Kada smo mi sami eksploatirani, korist koju ostvarujemo je 0 . Naravno, povećanjem vrijednosti y raste korist od suradnje u odnosu na nesuradnju. Povećanje x dovodi do nekoliko promjena: ekstra-korist od suradnje ($y-x$) pada, ali i pada i ekstra-korist od eksploatacije ($1-x$). S druge strane, ekstra-gubitak ($0-x$), u slučaju da smo sami eksploatirani, raste.

Kako bismo točno odredili efekt promjene vrijednosti x ili y na promjenu vrijednosti jednadžba korisnosti OM i IM izračunat ćemo njihove parcijalne derivacije po x i y .

$$\begin{aligned}\frac{\partial OM}{\partial y} &= rp^2 \\ \frac{\partial IM}{\partial y} &= 0 \\ \frac{\partial OM}{\partial x} &= -rp^2 - (1-r)(1-p) - rp(1-p) - rp(1-p) = -rp^2 - 3rp + r + p - 1 \\ \frac{\partial IM}{\partial x} &= -r(1-p) = rp - r\end{aligned}$$

Parcijalne derivacije funkcija OM i IM po y pokazuju da će za neku promjenu vrijednosti Δy , vrijednost funkcije koristi za OM porasti za $rp^2\Delta y$, a funkcija koristi za IM se neće mijenjati. Parcijalni diferencijali funkcija OM i IM po x pokazuju da će se za neku promjenu vrijednosti Δx , vrijednost funkcije koristi za OM promijeniti $(rp^2 - 3rp + r + p - 1)\Delta x$, a vrijednost IM za $(rp - r)\Delta x$.

Na iznos promjene vrijednosti funkcija OM i IM s obzirom na promjene y i x , utjecat će vrijednosti r i p . Vidimo da će promjene OM -a, izazvane porastom Δy , uvijek biti pozitivne te će biti veće što su veći r i p , s time da je efekt vrijednosti p snažniji (jer se javlja kao p^2). Ono što nas zapravo zanima je saldo ukupnog utjecaja na isplativost odabira OM pred IM , no kako porast y nema nikakvog utjecaja na promjenu IM , zaključak ostaje da će rastom y rasti i primamljivost odabira OM pred IM .

Utjecaj porasta x na primamljivost odabira OM pred IM je nešto složeniji. S jedne strane, rastom x , OM osoba gubi u svim situacijama: smanjuje joj se korist od suradnje ($y-x$), povećava joj se gubitak ako biva eksploatirana ($0-x$), a čak dobiva i manje ekstra-profit kada sama "nenamjerno" eksploatira drugog OM -a. No, pada i ekstra-profit ($1-x$) koji

IM ostvaruje u situacijama eksploatacije. Pitanje je gdje je negativni efekt veći? Odgovor možemo lako vidjeti ako parcijalnu derivaciju **OM** po **x** napišemo nešto drugačije (tj. zbrojimo efekte treće i četvrte situacije):

I prvi i drugi dio jednadžbe je uvijek manji od **0** (za svaki r : $0 \leq r \leq$

$$\begin{aligned}\frac{\partial \mathbf{OM}}{\partial \mathbf{x}} &= -rp^2 - (1-r)(1-p) - rp(1-p) - rp(1-p) \\ &= -rp^2 - (1-r)(1-p) - 2rp(1-p)\end{aligned}$$

1, i svaki p : $0.5 < p < 1$), te zajedno doprinose negativnoj stopi promjene. Treći dio dodatno povećava negativnu stopu promjene, i to za apsolutni iznos: $|-2rp(1-p)|$ koji je uvijek veći od apsolutnog iznosa negativne stope promjene parcijalne derivacije **IM** po **x**: $|-r(1-p)|$ jer je $2p > 1$ *per definitionem*. Sada smo i dokazali da povećanje **x** ima veći negativni efekt na funkciju koristi **OM** nego na **IM** te tako povećanjem **x** pada privlačnost opcije **OM** prema **IM**.

Ako bi se **y** i **x** povećali za isti iznos $\Delta \mathbf{y} = \Delta \mathbf{x}$, konačni efekt na privlačnost opcija **OM** i **IM** bio bi sljedeći:

$$\begin{aligned}\text{Kako je } 2rp - p + 1 \text{ uvijek veće od nule (za svaki } r: 0 \leq r \leq 1, \\ (\mathbf{OM}_y - \mathbf{IM}_y) - (\mathbf{OM}_x - \mathbf{IM}_x) = rp^2 - [(rp^2 - 3rp + r + p - 1) - (rp - r)] = 2rp - p + 1\end{aligned}$$

i svaki p : $0.5 < p < 1$), jednako povećanje **y** i **x** rezultirat će povećanjem privlačnosti **OM** u odnosu na **IM**. Vidimo da je efekt promjene **y** veći nego kod jednake promjene **x**. To je naravno i posve logično jer u ovom slučaju vrijednost ekstra-profita koji može ostvariti **OM**, a jednaka je **y-x**, ostaje nepromijenjena, dok vrijednost ekstra-profita koji može ostvariti **IM**, a jednaka je $1-x$, pada zbog rasta **x**.

Zaključak

Ovaj tekst ne pokušava u cijelosti srušiti Gauthierovu koncepciju “morala temeljenog na dogovoru”, već samo pokušava ispraviti neke pogreške koje je Gauthier počinio u koncipiranju i formuliranju svojeg računa koristi. Smatram da ova ispravljena verzija ima jednaku snagu i relevantnost kao što je trebao imati i originalni Gauthierov dokaz, a lišena je balasta nejasne koncepcije, nedovoljne razrađenosti i proizvoljne parcijalnosti. Ukratko, u svojem dokazu Gauthier je počinio nekoliko grešaka:

1. Pogriješio je u definiranju varijabli računa koristi (uvodeći nepotrebnu varijablu q);
2. Nije prepoznao dvostruki moment identificiranja prilikom interakcije dvije **OM** osobe;
3. Nije prepoznao izostanak problema identifikacije kod **IM** osoba (i konzekventnu jednostrukost identifikacije kod interakcije **OM-IM**);
4. Nije prepoznao upotrebljivost vjerojatnosti ispravne identifikacije (p) kao mjere translucenosti, niti je istaknuo ograničenja njene domene ($0.5 < p < 1$);
5. Predložio je pogrešnu i nepotrebnu “mjeru translucenosti” (p/q) ne objasnivši pritom u potpunosti logičke temelje takvog prijedloga;
6. Posve logički neopravdano (iako etički donekle shvatljivo) u račun koristi nije unio sve, nego samo dio mogućih situacija koje se mogu javiti prilikom interakcije osobe, koja je odabrala sklonost **OM**, s drugim osobama u društvu.

Naravno da je pitanje smislenosti cijelog takvog pothvata iznalaženja matematičkog i interesno-racionalnog objašnjenja za problem svojevoljnog pridržavanja društvenih moralnih pravila i dalje otvoreno za raspravu. Također je jasno da je Gauthierov račun po svojoj naravi prilično redukcionistički. Vrlo je lako naći i “slabe točke” unutar samog modela (poput relativno proizvoljnog pojma translucenosti). Ipak, Gauthierov doprinos u vidu računa moralnosti predstavlja interesantno promišljanje problema poštivanja društvenih/moralnih pravila (te motivacije iza takvog djelovanja) unutar neoklasične paradigme racionalnog izbora. Kako danas takva perspektiva apsolutno dominira u našem razumijevanju ekonomskih odnosa, posve je smisleno da se njome pokušavaju objasniti i razni aspekti društvenih odnosa, u ovom slučaju moralna pravila.

I doista, ako je kasni kapitalizam stvarno doveo do erozije morala u društvu, uslijed slabljenja utjecaja religije i ideologije te raspada nekadašnjih “zajednica” (što je sve djelovalo kao zapreka vladavini golog ekonomskog interesa), onda je možda i jedini način za očuvanje moralnosti njeno “sidrenje” u čvrstu podlogu upravo takve interesne motivacije. Kao što narodna mudrost kaže: “ako ga ne možeš pobijediti, pridruži mu se”.

Ipak, da bismo uopće mogli raspravljati o ideji interesno utemeljenog morala, potrebno je prvo imati ispravno definiran i razrađen račun moralnosti. Nadam se da će ove ispravke doprinijeti kvalitetnijoj budućoj raspravi o tom pitanju.

Literatura

- Berridge, G.R., Keens-Soper, M. i Otte, T.G. 2005.: *Diplomatska teorija od Machiavellija do Kissingera*, Fakultet političkih znanosti - Biblioteka Politička misao, Zagreb.
- Friend, C. M., 2001.: Trust And The Presumption Of Translucency, *Social Theory And Practice*, (27), 1.
- Gauthier, David, 1986.: *Morals by Agreement*, Clarendon Press, Oxford, 1986.
- Hobbes, Thomas, 1991.: *Levijatan* (1. sv.), Gradina, Niš.
- Ostrom, Vincent, Ostrom, Elinor, 1971.: Public Choice: A Different Approach to the Study of Public Administration, *Public Administration Review*, (31) ožujak-travanj: 203-216.
- Rawls, John, 1999.: *A Theory of Justice. Revised Edition*, Harvard University Press, Cambridge, Massachusetts.
- Scott, J., 2000.: Rational Choice Theory, u Browning, G., Halcli, A., Webster, F.: *Understanding Contemporary Society: Theories of The Present*, Sage Publications.
- Ward, Hugh, 2005.: Racionalni izbor, u: Marsh, David, Stoker, Gerry, *Teorije i metode u političkoj znanosti*, Fakultet političkih znanosti - Biblioteka Politička misao, Zagreb: 61-85.
- [www.en.wikipedia.org/ David Gauthier](http://www.en.wikipedia.org/David_Gauthier)
- [www.en.wikipedia.org/ Rational choice theory](http://www.en.wikipedia.org/Rational_choice_theory)

MORALS BY AGREEMENT? A CRITIQUE OF DAVID GAUTHIER

MARKO RADENović*

Summary:

*After a brief overview of the Rational Choice theory and its basic concepts, the author focuses on David Gauthier's application of the same. In his book *Morals by Agreement*, philosopher David Gauthier addressed the classical problem of individual compliance with the moral norms and rules of the society, albeit from a novel perspective. By re-defining it as a problem of contract/agreement adherence, Gauthier offered both an alternative to Adam Smith's "invisible hand" of the free market, as well as to Hobbesian "visible foot" of the Leviathan. Instead of relying on market forces or coercion by sovereign, Gauthier proposes the internalization of the propensity towards the agreement compliance based solely on the principle of the rational pursuit of self-interest by the individual. The essence of his proposal is a "calculus of utility", which demonstrates that, under certain conditions (considered by Gauthier to be relatively modest), it will more profitable to choose disposition to act as if a person were a "constrained" (by morality) maximizer, rather than a "straightforward" one. Unfortunately, Gauthier himself abandons the morally neutral position, embodied in the concept of*

the self-interested compliance with the moral rules of the society, and fails (even declines!) to consider all the potential gains and losses. That, however, renders his calculus completely useless. In this article, the author attempts to rectify Gauthier's mistake, while simultaneously proposing significant changes to the structure of the calculus itself. Such revised calculus of utility could then serve as a basis for the "morals by agreement".

Keywords:

rational choice, David Gauthier, prisoner's dilemma, morals by agreement, straightforward maximizer, constrained maximizer, contract respectin.

* Fourth year student of political science at the Faculty of Political Science in Zagreb.
e-mail: radenovicm@yahoo.com