

---

# ЈЕСУ ЛИ ЉУДИ РАЦИОНАЛНО ПРОДРУШТВЕНИ? НЕКИ ДОКАЗИ ИЗ БИХЕЈВИОРАЛНЕ ТЕОРИЈЕ ИГАРА

---

УДК 316.62:316.334.2

Оригинални научни рад

Милован Декић

Факултет политичких наука, Универзитет у Београду

---

Сажетак

---

У раду наводим налазе из бихејвиоралних експеримената који доказују да у темељу људског продруштвеног понашања леже рационалне преференције. Најпре појашњавам шта рационалност подразумева, а потом и како се рационалне преференције односе са вредностима и нормама у генерисању друштвене сарадње. Потом наводим могуће одговоре бихејвиоралним економистима који тврде да су људи често ирационални. Затим упућујем на поуке које доносиоци јавних политика могу да извуку из утврђене људске продруштвене рационалности. На крају, уместо закључка, дајем приказ објашњења које је Адам Смит понудио по питању порекла и природе људске продруштвености.

**Кључне речи:** рационалност, продруштвеност, друштвене норме, вредности, преференције, хеуристике, моралност, бихејвиорална економија

---

## Увод

Људи су кооперативна, продруштвена врста. Поред прирођене способности за акумулирање културе, то је друга ствар која их чини аномалијом у животињском свету (Henrich & Henrich, 2007; Bowles & Gintis, 2011; Field, 2001). Свакодневна искуства, али експериментални налази (лабораторијски али и теренски и крос-културни), доказују да су људи спремни да поступају фер и

правично и када се сусретну са потпуним странцима, када су сусрети потпуно анонимни и када су ти сусрети једнократни, односно без могућности да се понове (*one-shot interactions*). Поред тога, налази експерименталне економије потврђују да су људи спремни да буду први који ће отпочети сарадњу са другима, да имају аверзију према неегалитарним исходима, да воле да виде да други правично деле (новац и друга добра), да хоће да награде оне који поступају правично, али и да по цену излагања себе материјалном трошку казне оне који поступају неправично и преварантски (Chaudhuri, 2009).

Под појмом продруштвености подразумевам склоност људи да сарађују једни са другима, и помажу једни другима, не само из интереса унапређења личне материјалне користи, већ и из разлога што: (1) су људи интринсично заинтересовани за унапређење благостања других, (2) хоће да се придржавају друштвених норми, и (3) хоће да се понашају етички зато што сматрају да је то исправно по себи. Алтруизам је посебан вид продруштвености који се односи на онај чин којим актер који га изводи увећава фитнес организма ка коме је чин (или дело) усмерен. Питање на које одговор нуди овај текст је: да ли је ирационално или рационално гајити продруштвене преференције и понашати се у складу са њима – анонимно давати новац у добротворне сврхе, гласати на изборима, уступити место непознатој особи у превозу, оставити напојницу у ресторану у који се можда никада нећемо вратити, застати и помоћи особи на чијем аутомобилу је пукла гума на аутопуту, казнити онога ко се не понаша у складу са општеприхваћеним правилима понашања, жртвовати се за саборца на бојном пољу, гасити клима уређаје лети када се то саветује јер долази до преоптерећења електричне мреже? Да ли је ирационално или рационално понашати се морално, у складу са врлинама какве су честитост, верност, правичност или поузданост?

За почетак, узмимо за пример концепт који социолози, још од Емила Диркема, јако често воле да употребљавају у објашњењу људског понашања – друштвене норме.

Разбијање интегралног модела људског понашања вероватно је најславније изведено у делу Макса Вебера. По њему, као што је познато, друштвено делање (које социологија треба да разуме и тиме узрочно објасни) може бити: (1) циљнорационално, (2) вредноснорационално, (3) афективно (емоционално), и (4) тради-

ционално (Вебер, 1976, 16-17). С обзиром да се Вебер није потрудио да детаљније спецификује шта се тачно подразумева под овим типовима делања, чиме би нам олакшао посао класификовања неких уобичајених понашања (какво је гласање, рецимо), Јон Елстер је ову четвороделну поделу упростио у дводелну. По њему (Elster, 1991, 243), две су основне мотивације у људском понашању: рационалност и друштвене норме. Рационалност одговара Веберовој употреби инструменталне рационалности, док су друштвене норме, по Елстеру, у стању да обухвате преостала три облика делања као своју последицу: и норме су као и вредности неконсеквенцијалне природе, оне се преносе преко традиције, а главни механизам њихове колективне репродукције су емоције (какве су срамота и кривица). Питање следи: да ли је нормирано понашање рационално, или ирационално – како то тврди Јон Елстер?

Наравно, одговор зависи од тога како схватамо и дефинишемо појам рационалност. У овом раду ја се ослањам на класично схватање рационалности које каже да рационалност подразумева само конзистентност преференција и ништа више. Видећемо да рационалност не подразумева нити да су преференције себичне нити да оне морају да унапређују лично благостање. Затим ћемо видети неке експерименталне доказе из бихејвиоралне теорије игара који показују како у основи људског продруштвеног понашања леже конзистентне преференције које реагују на екстерне подстицаје. Ово је један значајан корак напред у односу на класичну социолошку теорију јер ћемо видети (само неке<sup>1</sup>) механизме који на микро нивоу генеришу макро феномен какав је друштвена норма. Након тога ћу, на основу постојеће литературе, понудити неке могуће одговоре бихејвиоралним економистима који тврде да је људско понашање ирационално. Тим налазима се критикује неадекватност мера путем којих су бихејвиорални економисти констатовали људску ирационалност, и предлаже ширење схватања рационалности, са потребом да се узме њена неформална, еволутивна и еколошка димензија. Затим ћемо видети неке могуће импликације људске продруштвене рационалности по доносиоце јавних политика. На крају, уместо закључка, видећемо како је Адам Смит објашњавао људску моралну динамику. Овај део је уједно и позив економисти-

1 Микромеханику настанка и одржања друштвених норми, кроз призму теорије игара, детаљно сам обрадио у свом раду „Зашто су људи алтруисти? Микро темељи друштвених норми и људске продруштвености у еволутивном контексту“. Радна верзија.

ма да се озбиљније позабаве људском продруштвеном страном, јер је она, по једном од оснивача економије као науке, пресудна за разумевање људског понашања.

Понашати се „ирационално“ је заправо рационално

*Рационалност подразумева само  
конзистентност преференција*

Рационални појединац је појединац који има конзистентне преференције. Без ове претпоставке друштвена наука губи постојано тло, јер је модел рационалног делатника једина теорија развијена у њеној историји која успешно служи предвиђању људског понашања у читавом низу друштвених ситуација (његова употребљивост је доказана и у биолошком моделирању животињског понашања). Теорија рационалног понашања је развијена у радовима фон Нојмана и Моргенстерна [1944] и Леонарда Севица [1954], а њени зачеци се налазе у радовима Церемија Бентама и Ђезара Бекарије.

Најбољу дефиницију рационалног делатника понудио је Херберт Гинтис (Gintis, 2009, ch. 1). По њему, модел рационалног делатника подразумева да “појединац има преференције које осликавају његове жеље и размене између ових жеља, и да појединци максимизирају своју корисност бирајући из скупа делања, при чему су ограничени доступним информацијама, материјалним ресурсима и временом, когнитивним и физичким капацитетима. Избор је завистан и од веровања која се тичу вероватноћа остваривања различитих стања природе, расподеле учесталости типова појединаца са којима је делатник у интеракцији, и релативне ефикасности различитих типова делања” (Gintis, 2006, 97). С обзиром да се појам рационалности, за разлику од оваквог одређења, употребљава и на прилично неодговарајуће начине, Гинтис је за теорију изведену из овог модела рационалног делатника предложио назив теорија веровања-преференција-ограничења (*beliefs-preferences-constraints theory*, БПЦ модел).

Ово одређење је сама срж теорије рационалног избора. Међутим, њој се додају бројне друге карактеристике које ина-

че не стоје, па са њима треба рашчистити на самом почетку. Најпре, рационални делатник не мора да буде себичан (*self-regarding behavior*), тј. да брине само о својим материјалним исплатама из интеракције, и да се труди да их увећа по сваку цену. Доказано је да су једини људи који се овако заиста понашају социопате и људи са оштећењима на мозгу (Koenigs et al., 2007), и да су углавном смештени по затворима (Mealey, 1995). Дакле, подједнако је рационално уживати у куповини и испијању скупог шампањца за себе, али и у анонимном поклањању новца у добротворне сврхе. Рационално је бринути о својим исплатама из интеракције, али је рационално водити рачуна и о томе какве ће последице наша одлука имати по другу страну. Такозваним преференцијама које се тичу других (*other-regarding preferences*) и њиховом рационалношћу детаљно ћу се бавити у другом делу овог рада. Друго, делатник никада не поступа са циљем да увећа своју корисност. Међутим, као што ћемо видети, из тога не следи да доношење одлуке не може бити моделирано “као да” је усмерено на увећање корисности, односно као оптимизирање функције преференција. Понекад, људи могу радити неке ствари зато што верују да је то исправно (без обзира на последице), а некада могу радити неке ствари како би избегли осећај кривице који би им се јавио уколико их не би урадили (рецимо, оставили напојницу конобару у кафани). Укратко, када актер бира неку опцију, он се понаша као да жели да жели да максимизира вредност функције корисности (Binmore, 2009, 14), а корисност се заправо повећава услед конзистентности преференција (Gintis, 2009, 6), односно је њихова логичка последица. Треће, рационалност не подразумева нити потпуну информисаност, нити то да веровања морају да буду тачна и ажурирана у складу са информацијама (ibid, 235). На крају, рационалност не подразумева ни то да људски избори морају да увећавају благостање. Порок, какав је рецимо пушење, је савршено рационалан из разлога што не крши конзистентност преференција.

Први услов конзистентности преференција је *комплетност*. Он подразумева способност делатника да успостави некакав поредак између алтернатива које су му на располагању. На пример, уколико су му на располагању алтернативе X и Y, комплетност подразумева да је делатник у стању да се одреди у односу на њих, тако да или X преферира у односу на Y ( $X \geq Y$ ), или да Y преферира у односу на X ( $Y \geq X$ ), или да буде индиферентан између њих ( $X \sim Y$ ). Други услов је *транзитивност*. То значи да ако делатник

има на располагању три алтернативе, X, Y и Z, и X преферира у односу на Y, а Y у односу на Z, онда мора да преферира и X у односу на Z. Трећи услов је *независност од ирелеватних алтернатива*. То значи да релативна привлачност два избора појединца не зависи од осталих избора који су му на располагању. На пример, уколико он, кад једе у ресторану, преферира бело месо у односу на свињске медаљоне, онда чињеница да се у ресторану служе и ћевапи не може да утиче на то да он преферира и изабере свињске медаљоне.

Главна поента реченог је да уколико је људска одлука конзистентна, њен исход је аутоматски рационално понашање. По Бинмору, рационалност и конзистентност су синоними (Binmore, 2009, 4). Гинтис овако закључује: “Када је однос преференција  $\geq$  комплетан, транзитиван, и независан од ирелеватних алтернатива, ми га називамо конзистентним. Уколико је однос преференција  $\geq$  конзистентан, онда увек постоји функција преференција за појединца који се понаша тако као да максимизира ову функцију преференција из скупа A на који је његов избор ограничен”. Из овога следи теорема: *бинарни однос  $\geq$  из ограниченог скупа исплата A може бити представљен функцијом преференција у: A–P ако и само ако је  $\geq$  конзистентан* (Gintis, 2009, 6).

На крају, тврдња да су преференције конзистентне јер тиме увећавају општу способност организма, може звучати као класично функционалистичко “причање приче”. Међутим, неуронаука нуди неке доказе да то и није тек празна прича. У домену продруштвеног понашања (за које ћемо у деловима који следе видети тестове који сведоче да је у питању рационално понашање), неуролошка истраживања су показала да се приликом испољавања алтруистичког акта у мозгу активирају центри за награђивање. То није само у случају ефекта “топлог сјаја” (*warm glow*), којим је Џејмс Андреони показао да људи извлаче осећај користи из самог чина приватног давања (Andreoni, 1990), већ и у случају анонимног донирања јавним добрима. Харбоу и сарадници су коришћењем фМРИ технике (*functional magnetic resonance imaging*) доказали и да опорезивање (не заборавите да је плаћање пореза обавеза грађана) и давање новца у хуманитарне сврхе, производе исте неуролошке одговоре (добро осећање из центара за награђивање) као што га производе и присвајање новца у личне сврхе и давање услед *warm glow* ефекта (Harbaugh, 2007). Подсетимо се једне реченице Адама Смита из *Теорије моралних осећања*, где он тврди да су “ред и напредак друштва нешто што човеку природно одговара и

о чему са уживањем размишља” (Smit, 2009, 82). Сада имамо и неуролошке потврде тога (Fehr & Rockenbach, 2004; Morischima et al., 2012; Waytz et al., 2012; Chang et al., 2013).

Следеће интересно питање је и зашто би људи својим изборима уопште максимизирали било шта? Према групи аутора (El Mouden et al., 2012), природна селекција делује на просечним последицама одређених понашања; из тог разлога, иако она минимизира просечне трошкове чињења грешака, она не може да их елиминише у потпуности. Међутим, док год се испоставља да рационално понашање генерише друштвено оптимална понашања (или увећава инклузивни фитнес организма који изводи понашање), може се очекивати да природна селекција фаворизује рационално одлучивање.

### *Вредности, друштвене норме и рационалност*

Најпознатији заступник схватања да су друштвене норме извор ирационалности је Јон Елстер. Како смо видели на почетку, он тврди да друштвене норме треба разумети као извор људских мотивација који је потпуно одвојен од рационалности. Нешто касније, он нуди трихотомну теорију, где су рационалност, норме и емоције главни узроци делања (Elster, 1994). У оба случаја, идеја је иста: “... људи који су покренути друштвеним нормама су ирационални” (ибид, 28). Иако је овим својим класификацијама покушао да уведе какав-такав аналитички смислен ред у већ постојеће типологије мотивација, мислим да је у томе више одмогао него што је помогао. Његов чувени пример је са ношењем црнине на сахрани. По њему, када људи носе црнину на сахрани, то нема никакав инструментални значај (Elster, 2007: 353). Да ли та тврдња побија да је ношење црнине рационални избор? Верујем да не, па ћу сада покушати да покажем да је главни узрок читаве забуне око друштвених норми потпуно первертирано схватање рационалности и рационалног избора.

Улога друштвених норми у еволуцији људске врсте је преношење моралних вредности кроз генерације, углавном путем механизма културне трансмисије и конформизма. За разлику од неких инструменталних веровања, која саветују како најбоље постићи неки циљ, норме, у виду вредности, дефинишу шта могу бити пожељни циљеви, какви су правичност, поштење, честитост или оданост. Ови циљеви у људском понашању оперишу као циљ-

љеви сами по себи, или као врста дужности у Кантовском смислу: исправно поступање је само себи циљ. (Тако човек из Елстеровог примера носи црнину зато што сматра да је тако исправно, јер можда тиме исказује поштовање свом преминулом пријатељу.) Питање је, може ли се овакво поступање објаснити као рационално?

Најбољи начин за тестирање рационалности поступања у складу са нормама и вредностима је провера њиховог реаговања на промене у окружењу. За почетак, треба нагласити да је исправно поступање преференција: као што неко преферира чоколаду, или садашњи тренутак у односу на будући, тако неко може преферирати исправно и честито поступање. Другим речима, вредности су део функције преференција, што значи да се оне могу размењивати (*trade-off*) са другим објектима који се желе. Из тога произилази важна претпоставка. Као што су људи склони томе да када дође до повећања цене чоколаде, купују мање чоколаде, исто би требало да важи и за врлине: када цена испољавања врлине скочи, она ће се мање испољавати. Да ли је ово тачно?

Ури Гнизи је то пробао да сазна (Gneezy, 2005). Он разматра облик лагања који је посебно важан за економске односе: онај у коме лаж може да повећа исплату за лажова а смањи исплату за другу страну. Друга карактеристика његовог експеримента је да истражује лагање у ситуацији у којој нема казне за исто. Укратко, у експерименту, названом Игра јефтине приче (*Cheap Talk Game*), се истражује какав утицај на лагање могу имати различите материјалне последице до којих се може стићи лагањем.

Постоје два новчана исхода: А и Б. Само играч 1 зна новчани садржај сваке опције (корпе), а правила игре су позната обојици учесника. Играч 1 шаље играчу 2 једну од следеће две поруке: (1) “Опција А ће ти донети више пара него опција Б”, и (2) “Опција Б ће ти донети више новца него опција А”. Након што ово чује, играч 2 бира опцију коју жели. 82 процената играча 1 је изјавило да су сигурни да ће им играч 2 веровати и да ће изабрати оно што им они препоруче (у стварности, 78% им је веровало). Ово је важно јер то значи да су они били свесни тога да ће њихова лаж имати утицаја на стварно одлучивање, односно да ће бити релевантно по стварне исплате и једне и друге стране. Какви су били резултати? У сесији 1, када је играч 1 од лагања могао да заради 1\$ а играч 2 да изгуби исто 1\$, лагало је 36% учесника. У сесији 2, када је играч 1 од лагања могао да заради 1\$ а играч 2 да изгуби 10\$, лагало је свега 17% учесника (више него дупло мање). У сесији 3, када је



играч 1 од лагања могао да заради 10\$ а играч 2 да изгуби исто 10\$, лагало је чак 52% учесника.

Шта ови резултати говоре? У основи, то какви ће бити исходи лагања утиче на одлуку да ли ће се лагати. Људи размишљају о томе шта добијају, а шта могу да изгубе лагањем. Још важније, људима је важно колико они добијају лагањем, али им је важно и колико други губе њиховим лагањем (иако санкције нема!) Уколико је пак добитак осетно велики, и упркос знатном губитку друге стране, људи ће ипак лагати.

Гнизи је извео и допунско истраживање, са упитницима. У њима је осликао стварну ситуацију у којој продавац аутомобила има прилику да слаже потенцијалног купца (не пријави му да пумпа за уље не ради добро), и тиме прода ауто за 1200\$. Уколико му каже за квар, мораће да му спусти цену за 250\$ (колико кошта поправка пумпе). На крају, продавац је прећутао квар и продао ауто за 1200\$. Људи су требали да кажу шта мисле, да ли је понашање продавца било: потпуно правично, правично, неправично, веома неправично. У другој верзији питања, разлика у цени за пумпу је била 1000\$ уместо 250\$. Да ли су се разликовали и одговори испитаника? Јесу, и то драстично. У ситуацији са 250\$, 70% испитаника је одговорило са неправично, а 18% са веома неправично. У ситуацији са 1000\$, 32% је одговорило са неправично, а 66% са веома неправично. У коначном, трећем питању, проблем у аутомобилу није била пумпа за уље, већ кочнице. Испитаници су одбили да одговоре на то питање јер су сматрали да је увредљиво. И овај експеримент доказује да људи мисле да је мање правично да се лаже како се исплате за другу страну смањују. Генерално, то значи да људи нису нити “свеци” нити преваранти по сваку цену. Способни су и за једно и за друго, а све зависи од тога како перципирају сам процес интеракције. Другим речима, унутрашње побуде нису апсолути, већ су о процесу одлучивања важне и њихове последице.

Једна од главних теорија која је у стању да објасни динамику сарадње у друштвеним дилемама попут дилеме са јавним добрим је теорија снажног реципроцитета (Fehr & Fischbacher, 2005). Један од њених главних налаза је да увођење могућности кажњавања фрирајдера, значајно доприноси повећању нивоа сарадње у играма. Оно што је њен посебан допринос је објашњење узрока кажњавања, посебно у случају нестратешког кажњавања: наиме, зашто би ико желео да кажњава фрирајдере када то мора да плати, а кажњавање изводи у последњој рунди игре са ограниченим бро-

јем интеракција? Постоје два могућа објашњења. Једно узрок види у напрасном налету емоције као што је бес. Друго, плаузибилније и ухватљивије објашњење, је оно које каже да људи кажњавају фрирајдере због тога што крше норму условне сарадње. Друго објашњење је додатно потврђено у тзв. играма у којима кажњава трећа страна (*third-party punishment games*) (Fehr & Fichbacher, 2004).

Како је у овом случају могуће објаснити поступање у складу са нормама? Да и у овом случају понашање следи логику рационалности, показали су у свом експерименту Андерсон и Патерман (Anderson & Putterman, 2006). Субјекти у њиховом експерименту играју пет једнократних игара добровољног давања, у две фазе. У првој фази, они се анонимно групишу и слободно бирају колико желе да допринесу јавном добру. У другој фази, сваки субјекат добија информацију колико су остали дали, и која је цена сваког долара који он уложу у умањење суме којом на крају игре располажу остали субјекти из његове групе. На основу ових информација, субјект бира новчани ниво кажњавања осталих субјеката.

Прва фаза почиње класичном игром јавних добара. Сваки субјекат (по троје њих у гурпи) добија по 10 долара, од којих бира колико ће приложити у јавни фонд (који је овде јавно добро). У овом експерименту, број којим се јавни фонд умножава је 0.6. Групни фонд се онда дели субјектима на равне части. Наравно, у егоистичком интересу је сваког члана да у фонд не приложи ништа, већ да задржи својих 10 долара и онда добије и део онога што су остали приложили. У другој фази, свако добија информације колико су остали дали, и (свако посебно) свој лични трошак кажњавања. Вредности трошка кажњавања су били: 0, 30, 60, 90 и 120 центи у првом случају; 0, 5, 10, 20 и 30 центи у другом; и 30, 40, 50, 60 и 70 центи у трећем. Структура трошкова је додељивана насумице, током пет наредних рунди игре. Различити скупови трошкова кажњавања су изабрани да ли ће се доћи до истих квалитативних резултата са различитим вредностима.

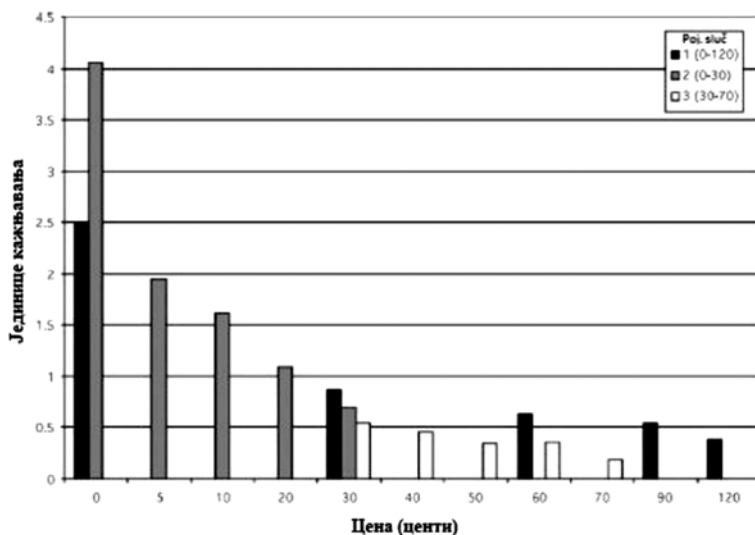
Субјекти потом бирају са колико јединица кажњавања желе да казне остала два члана групе (уколико то заслужују), при чему су ограничени стањем на свом рачуну. Када сви донесу своје одлуке, сваки члан групе добија своју новчану “пресуду” за игру у првој рунди.<sup>2</sup>

---

2 Како изгледа математичка структура коначног стања на рачуну играча може се видети у Christopher Anderson & Louis Putterman, „Do non-strategic sanctions

Циљ експеримента је да покаже да ли се кажњавање подвргава логици закона потражње, односно да ли трошење за кажњавање зависи од његове цене. Налази су сумирани у графикону 1. Као што се ту може видети, ниво кажњавања (схваћен као потрошња) је подложен променама у цени: што цена више расте, то потражња за кажњавањем опада. При цени од 5 центи, купује се 1.95 јединица кажњавања; при цени од 10 купује се 1.62. Кад цена порасте за 60%, потражња пада на 0.49 јединица. При цени од 1.20\$, потражња је само 0.38. Укратко, потраживање за кажњавањем је негативно повезано са ценом коју онај ко треба да га купи треба да плати. То значи да се кажњавање, утемељено у нормама, може анализирати и моделирати путем теорије игара, што је само доказ да је у питању преференција која, као и свака друга, улази у функцију преференција појединца.

Графикон 1: Просечан број јединица кажњавања потраживан при свакој цени кажњавања (Извор: Cristopher Anderson & Louis Putterman, „Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism“, *Games and Economic Behavior* 54 (1): 1-24.)



Верујем да ови примери доказују барем две важне ствари. Најпре то да се *вредности* (које се налазе у срцу друштвених

obey the law of demand? The demand for punishment in the voluntary contribution mechanism“, *Games and Economic Behavior* 54 (1), p. 5.

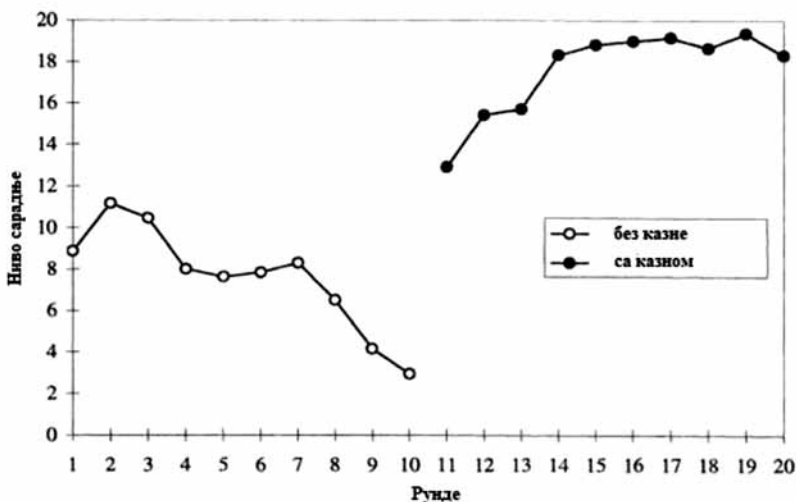
норми) *функционишу на бази преференција*, које је могуће математички моделирати путем функције преференција коју појединац максимизира. Односно, то значи да је *људско одношење према вредностима рационално, као што је рационално одношење човека према било ком другом добру, рецимо, чоколади*. Ово моделирање се може читати као доказ да је *Ното sociologicus* Емила Диркема и Талкота Парсонса непостојећи феномен у стварном животу, и да га као такав и треба одбацити у научним објашњењима. Теорија игара, која се углавном одбацује у социологији, омогућава знатно прецизније и тачније предвиђање понашања у ситуацијама у којима су присутне друштвене норме тако што успева да их интегрише у, емпиријски јако потковану, теорију људског одлучивања.

### *Да ли је алтруизам рационалан?*

Као што сам напоменуо, од осталих животињских врста људе одваја једна важна карактеристика: њихов колективни живот одликује се знатно високим степеном сарадње између великог броја генетски неповезаних појединаца. Главним узроком ове сарадње може се сматрати људски алтруизам. Алтруизам се може одредити као скупа радња, или жртвовање једне стране која увећава економске добити друге стране. Специфичније, особа је алтруистична уколико њена корисност расте са увећањем благостања других особа. Алтруизам је значајан феномен економског живота једног друштва. У Америци, 2011. године добровољна давања су износила 298.3 милијарди долара, што је повећање од 3.9% у односу на 2010. годину. Највећи извор давања су појединци (73%), а укупно 88% домаћинстава учествује у добровољним давањима.

Да ли је алтруизам рационалан? На ово питање се може одговорити организовањем интеракције у склопу Игре јавних добара. Она је од посебног значаја јер најбоље од свих експерименталних окружења омогућава истраживање централног проблема друштвене науке, а то је одговор на питање како обезбедити оптималну снабдевеност неког јавног добра. У игри, која се углавном понавља у 10 рунди, сваки играч у групи, која има више од два члана, добија одређену суму новца. Свако од њих треба да уложи један део новца у јавни фонд, а преостали део задржава за себе. Део уложен у јавни фонд се на крају улагања мултипликује одређеним бројем. На крају се укупан износ распоређује играчима. Одмах је

јасно да је у интересу сваког појединца да јавном фонду (или, рецимо, добру) не приложи ништа: да све своје задржи за себе, а да добије део умножене суме коју су остали приложили. Међутим, како се субјекти који играју ову игру понашају?



Графикон 2: Просечна сарадња током времена у условима са и без могућности кажњавања (Преузето из: Ernst Fehr & Simon Gächter, „Cooperation and Punishment in Public Goods Experiments“, *The American Economic Review* 90 (4), p. 989)

На графикону 2 се види динамика сарадње закључно са рундом 10. Дакле, у почетку знатан део играча жели да сарађује. Међутим, како рунде пролазе, на крају којих се сазнаје ко је колико приложио, све већи број одустаје од сарадње и све мање прилаже. Како ово објаснити? Као што сам већ напоменуо, по налазима Бруно Фреја сви експериментални налази говоре да група увек садржи око 30% људи који не желе да дају ништа (дакле, фрирајдера). Ова мала група фрирајдера довољна је да сломи целокупан поредак сарадње. Јер људи нису безусловни алтруисти: они желе да сарађују само под условом да и други сарађују (таквих је углавном око 50%). Када они примете да су наивичине, који на својој грбачи носе један део егоиста, и они престају да сарађују, чиме колективно делање пропада.

У другом делу графикона 2, ситуација се мења јер се у том тренутку уводи опција да фрирајдери могу бити кажњени.

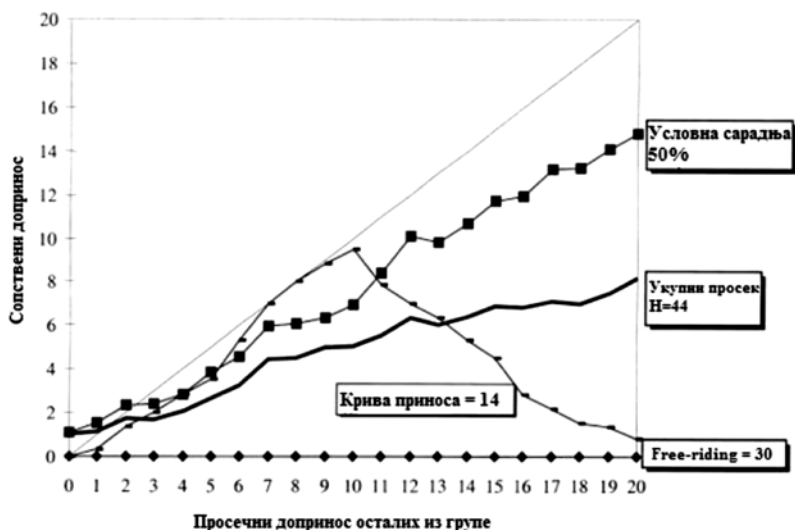
Како смо видели, људи су склони кажњавању из три разлога: (1) због налета емоција (каква је бес, јер су испали наивчине), (2) због алтруизма и норми сарадње, и (3) због изградње репутације (тзв. стратешко кажњавање, које у овом случају не ради јер не може да објасни кажњавање у последњој рунди када се игра завршава, и више нема разлога за изградњу репутације, чиме трошкови кажњавања превазилазе добити).

Динамика сарадње се још боље може сагледати на графику 3. На њему се види да су чак и људи који су спремни да сарађују од почетка, дакле непобитно алтруисти, ипак рационални јер своју стратегију планирају на бази веровања о томе колико други дају. У експерименту из којег је преузет графикон, јавило се 14% субјеката који су се понашали по тзв. криви приноса (*hump-curve*): они су давали мање од просечног доприноса пуних алтруиста (Fischbacher, Simon Gächter, & Ernst Fehr, 2001). У експерименту је било и 30% безусловних фрирајдера. С обзиром да постепено свако схвата да остали изневеравају њихова веровања и дају мање него што би требало, то их наводи да и они сами дају мање. Оваква первертирана динамика веровања на крају доводи до слома колективног делања, иако је знатна већина субјеката (64%) била спремна на сарадњу.

Дакле, као што показују ови примери, алтруизам није безуслован, и то да ли ће колективно делања бити успешно у великој мери зависи од динамике односа и веровања алтруиста и фрирајдера у популацији. Оно што графикон 2 показује, то је да вероватноћа успеха сарадње расте у условима у којима је присутна могућност да алтруисти казне оне који не желе да сарађују ни по коју цену.

Међутим, најбољи доказ тврдње да је алтруистичко понашање рационално може се наћи у експерименту Џејмса Андреонија и Џона Милера (Andreoni & Miller, 2002). Они су рационалност алтруистичког понашања истраживали путем Игре диктатора. У тој игри, диктатор добија одређену суму новца (буџет) који може да расподелити између себе и примаоца. С обзиром да је он диктатор, прималац мора да прихвати суму коју год да диктатор понуди. У пракси ове игре, новац се често дели пола-пола, па се игра узима као савршени пример постојања неог типа подруштвених и алтруистичких ставова. Андреони и Милер се питају да ли се ово алтруистичко понашање може објаснити кроз лични интерес, у смислу постојања криве корисности путем које се то понашање може

претставити. Они су проверавали да ли је то понашање у складу са генерализованим аксиомом откривених пеференција (GARP). Уколико јесте онда се може закључити даје алтруистички избор генерисан помоћу функције корисности.

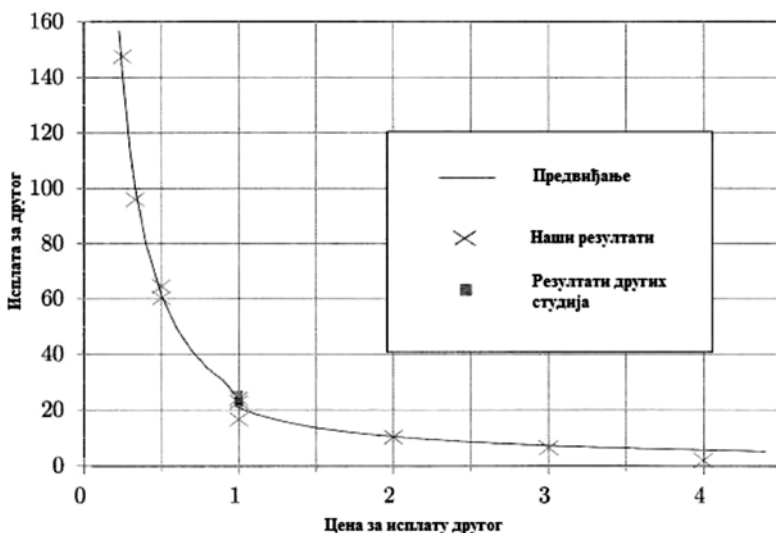


Графикон 3: Ниво просечног личног доприноса по основу нивоа сваког просечног доприноса других чланова (Преузето из: Urs Fischbacher, Simon Gächter, & Ernst Fehr, “Are people conditionally cooperative? Evidence from a public goods experiment“, *Economics Letters* 71(3), p. 400)

Замислимо да се диктатору каже да за ће за сваку јединицу коју проследи примаоцу морати да плати неку цену (као врста пореза). Цена, или порез, може да износи 5. Дакле, за сваку јединицу коју проследи примаоцу, диктатор ће морати да плати додатних 5 јединица пореза. Ако узмемо у обзир и буџет (од рецимо 100 јединица), главно питање постаје да ли постоји функција преференција коју диктатор максимизира. Ако постоји, то значи да је и алтруистичко понашање подједнако рационално као класично потрошачко понашање, у којем потрошач води рачуна о цени роба.

Цене по јединици су у њиховој верзији игре биле 0.25, 0.33, 0.5, 1, 2, 3, 4, а укупан број јединица на располагању диктаторима је био 40, 60, 75, 8 и 100. Они су открили да је само 18 играча од 176 прекршило ГАРП аксиом. Да су њихови избори били

генерисани насумице, за очекивати је било да би њих 78 до 95% прекршило тај аксиом. Из овога се закључује да је алтруистичке изборе могуће захватити функцијом корисности. На графикону бр. 4 приказани су налази истраживања.



Графикон 4: Плаћање диктатора са сумом од 100 (Преузето из: James Andreoni & John Miller, „Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism“, *Econometrica* 70 (2), p. 747)

## Бихејвиорални економисти против рационалности

Читава зграда бихејвиоралне економије, или истраживачког програма хеуристика и пристрасности, почевши од налаза Данијела Канемана и Ејмоса Тверског из 1970-их година, саздана је на идеји да су људи систематски склони грешкама у мишљењу и ирационалном понашању, и, штавише, да су “предвидљиво ирационални” (Аријели, 2008). Иако њен циљ није да у потпуности замени стандардни модел економске рационалности, већ да исти само продуби и ојача у његовим претпоставкама и, самим тим, његовој тачности (Но, Lim & Camerer, 2006), њене теорије се често интерпретирају као потпуно одбацивање стандардног модела (Wilkin-



son & Claes, 2012). Да ли бихејвиорална економија заиста доказује људску ирационалност? Шта уопште доказује тврдња Пола Словика да људи нису “логични већ пре *психологични*” (Slovic, 1995)?

По становишту Џонатана Коена, људска рационалност, у смислу способности дедуктивног и пробабилистичког резоновања, не може се тестирати и доказивати експериментално (Cohen, 1981). По њему, склоност људи ка чињењу логичких грешака у прорачунима, или оно што су бихејвиорални економисти склони да назову ирационалношћу, по њему су само “грешке у корацима” (*performance errors*), које пре доказују интелигенцију и образовање људи него њихову ирационалност.

Као што сам већ нагласио, људски мозак је, као и сваки други људски орган, обликован процесом еволуције људске врсте. Он је, са својим пратећим функцијама и способностима, адаптација, тј. последица просека ситуација кроз које је прошао у процесу еволуције и којима се у том процесу прилагодио. Из тог разлога, његове способности не би требали објашњавати одвојено од природе проблема које он морао да решава у историји врсте<sup>3</sup>, па самим тим ни одвојено од структуре окружења у којем он и данас оперише и доноси одлуке. Ово је најјасније нагласио Херберт Сајмон у својој концепцији “ограничене рационалности” (Simon, 1956). По његовом схватању, истраживање рационалности не може се свести само на проналажење когнитивних фаличности људског резоновања и одлучивања; сам процес одлучивања мора се изучавати у склопу контекста шире структуре у којој се одлучивање дешава. Другим речима, прилично је лако одлуку прогласити ирационалном уколико се она дешава у окружењу којем мозак природно није прилагођен. Већина тих “неприлагођених” околности креира се управо у лабораторијским условима.

Герд Гигерензер је, на трагу Херберта Сајмона, људску рационалност покушао да разуме као *еколошку рационалност*, супротстављајући њену интуитивну природу покушајима да се она

3 Из овог става су Ден Спербер и Уго Мерсије извели тзв. аргументативну теорију људског резоновања. По тој теорији функција резоновања није да тражи (логичку) истину већ да омогући људима да победе у расправи. Дакле, по овој теорији, систематске склоности људи кад пристрасностима нису фаличности ума већ његово природно стање. Види Dan Sperber & Hugo Mercier, „Why Do Humans Reason? Arguments for Argumentative Theory“, in: *Behavioral and Brain Sciences* 34: 57-111. Стивен Пинкер је записао да је људски мозак „ди-зајниран за фитнес а не за истину. Истина је некад адаптивна а некад није.“ Види Steven Pinker, *How The Mind Works*, W.W. Norton Company, 1997, str. 305.

тестира путем закона формалне логике (Gigerenzer & Selten, 2001). По њему, људска интуиција је, због свог еволутивног порекла, садржајно богатија од формалне логике. Иако та интуитивна испуњеност људе може навести да начине логичку грешку у конструисаном окружењу, она им јако често омогућава и то да, у стварном свету, када немају пуно информација, ипак донесу најбољу одлуку. У једном експерименту (Gigerenzer & Goldstein, 1996), Гигерензер и Данијел Голдстеин су показали како колективно незнање може произвести тачнију одлуку у односу на ситуацију у којој људи имају пуно информација о проблему који треба решити. За тачну одлуку у првом случају није био задужен некакав компјутерски прорачун вероватноћа на основу свих релевантних информација, већ слеђење једноставне хеуристике препознавања која каже: “Ако препознаш име једног града, а не и другог, закључи да град који си препознао има више становника”. Овакав начин одлучивања је рационалан из разлога што је прикупљање додатних информација понекад јако тешко, а често може бити и јако скупо.

По бихејвиоралним економистима, људска склоност доношењу одлука на основу тога како су неки избори представљени, или уоквирени (*framing*), типичан је пример ирационалности. На пример, доказано је да када људима који би требали да иду на операцију то саопштите на следећи начин: “Имате 90% шансе да преживите”, они су склонили да на операцију оду у том случају, а у знатно мањој мери када им то саопштите на следећи начин: “Имате 10% шансе да умрете”. Ово је ирационално јер је објективна вероватноћа иста у оба случаја, а људи ипак различито одлучују у те две ситуације.

Други пример је тзв. “проблем Линде” који служи као доказ постојања систематске грешке у виду *зablуде конјункције* (Tversky & Kahneman, 1982). По Канеману и Тверском, одлука људи у експерименту да тврде да је већа вероватноћа да је Линда “службеница у банци и активисткиња феминистичког покрета”, је ирационална јер очигледно крши законе логике. Конјункција два догађаја (попут горе наведеног) логички не може бити већи скуп од једног његовог дела (“Линда је службеница у банци”). Међутим, из овог експеримента Гигерензер и Хертвиг не изводе закључак да су људи ирационални (Hertwig & Gigerenzer, 1990). Људи се у овој, као и у свакој другој ситуацији, поводе за интуицијом, па употреба логичких норми које су следеће на садржај те интуиције, могу заве-

сти људе на погрешно размишљање. Да је управо ово био проблем у експерименту Канемана и Тверског (употреба речи *и* и *вероватноћа*), доказали су Хертвиг и Гигерензер који су у својој верзији експеримента уместо речи “вероватно” користили израз “колико”. “Постоји стотине особа које су као Линда. Колики број од њих јесу (1) службенице у банци, (2) службенице у банци и активисткиње у феминистичком покрету?” У овој верзији, људи нису починили заблуду коњункције, чиме су доказали да је у првобитном експерименту до изражаја дошла пре интелигентна конверзацијска интуиција него систематски дефект у мишљењу.

Главни проблем бихејвиоралне економије, по Гигерензеру (Gigerenzer, 2005) је то што њени концепти не воде никуда, а нарочито не добром предвиђању. Неки од њих су толико неупотребљиви да су у стању да објасне два супротна феномена. На пример, хеуристика репрезентативности се таутолошки користи у објашњењу два супротстављена феномена какви су заблуда загрејане руке (*hot hand fallacy*) (Gilovich, Tversky, Vallone 1985) и коцкарска заблуда (*gambler's fallacy*) (Tversky & Kahneman, 1974). Измишљање оваквих “хеуристика” је корак уназад у науци јер оне као такве не нуде никакав нови предиктивни и процесни модел. (“Теорију једино можеш победити другом теоријом.” Gigerenzer, 2005, 51) Хеуристике, као когнитивни процеси, нису нова грешка, већ само објашњење пристрасности или грешке. Тако схваћене, оне могу обезбедити процесне моделе феномена за које још увек није попуњено добро објашњење. На пример, тзв. пристрасност накнадне памети, или знао-сам-све-време (*hindsight bias*), је само успутни производ адаптивног механизма надоградње меморије. Процесни модел који је успео да захвати овај механизам, РАФТ (Hoffrage et al, 2005), успешан је до 75% у предвиђању када ће се јавити ова пристрасност, када ће се појавити у обрнутом облику, или када ће изостати. Тиме је такође доказано пристрасност накнадне памети није систематска диспозиција већ пре последица интеракције између когнитивног система и структуре окружења.

Један од типичних примера ирационалности и тзв. нестабилних преференција у бихејвиоралној литератури је и склоност актера да одлуку о томе како ће се у некој игри понашати усмеравају у складу са тим како се игра назива. Доказано је, на пример, да субјекти више сарађују када им експериментатор саопшти да играју Игру заједнице (*Community Game*), него када им саопшти

да играју Вол Стрит Игру (*Wal Street Game*) (Lieberman et al, 2004; Kay et al, 2003). Да ли је ово заиста ирационално понашање, и да ли крши конзистентност преференција? По Херберту Гинтису једна од незаобилазних претпоставки људског понашања је ситуациони карактер преференција (Gintis, 2013). По овом схватању преференција, у одређеним ситуацијама људи се могу понашати у складу са нормама које се уобичајено поштују у тим ситуацијама. Ова жеља за поштовањем норми се посматра као мета-преференција. Базични индивидуални механизам репродукције друштвених норми су очекивања људи који се налазе у нормираној ситуацији. Тако, неки људи могу поштовати друштвену норму зато што други од њих очекују да је они испоштују. Уколико докажемо да се људи у ситуацијама прожетим нормама понашају у складу са очекивањима, независно од сопствене бриге за благостање других, то значи да очекивања других везана за њихово понашања имају утицај на њихово реално понашање у играма.

Овакав доказ може се наћи у експериментима Дане, Кејна и Дауса (Dana et al, 2006). Њихов чланак носи симболичан назив: оно што не знаш ме не може повредити: скуп (али миран) бег из игре диктатора. У две варијанте експеримента Игре диктатора, они су показали да су људи склони томе да избегну ситуације у којима ће се од њих очекивати да се понашају у складу са нормама. У једној варијанти, субјекти су могли да уђу у игру и поделе суму од 10\$ како год пожелеле (јер су диктатори) или да избегну игру и оду кући са 9\$. Субјекти су бирали другу опцију, иако су у првој могли да прођу боље (са свих 10\$ у џепу). Интерпретација је та да су се трудили да избегну ситуацију у којој ће, очекивањима других, бити подстакнути да направе више фер расподелу. Ситуација слична овој се може видети на улици, када неко заобиђе просјака управо како би заобишао притисак да се понаша у складу са очекивањима и опште прихваћеним нормама.

Још један пример из бихејвиоралне економије је временска неконзистентност, односно присуство високе стопе дисконтовања времена код људи. За почетак треба разумети да чињеница да овакви типови понашања воде ка субоптималним животним изборима (секс без заштите, преједање и сл.) не подразумева и њихову ирационалност. Као што смо видели, рационалност не подразумева и благостање.

Што се тиче дисконтовања времена, оно се може показати као рационално уколико успемо да га моделирамо кроз мето-

дологију веровања-преференција-ограничења. То се може постићи једноставним ширењем простора избора. Наиме, уколико се простор временских преференција састоји само од пара: награда-тренутак преузимања награде, онда преференције могу деловати као временски неконзистентне. Међутим, уколико се простор избора прошири, добијамо другачију слику. У условима проширеног простора, са три елемента: награда-садашњи тренутак-тренутак преузимања награде, понашање постаје конзистентно. То доказују модели хиперболичког и квази-хиперболичког дисконтовања (Ainslie & Haslam, 1992).

На крају, Сегал (Segal, 1987) и Макина (Machina, 1987) показали су на који начин је могуће Елсбергов (Ellsberg, 1961) и Алисов (Allais, 1953) парадокс моделирати као очекивану корисност са нелинеарним тежинама.

## Неке импликације изнетог по јавне политике

Истраживање и разумевање природе и логике људске рационалности није од значаја само за аутономну егзистенцију друштвене науке. (Наиме, по Карлу Поперу, принцип рационалности – или “делања прикладног ситуацији” – је нулти принцип друштвене науке, без кога она не може да постоји (Popper, 1994)).

Оно што је још важније је то да квалитет и ефикасност јавних политика увек зависе од тога да ли њихови креатори добро разумеју логику људског понашања (Shafir, 2013). Из тога следи да је адекватно тј. емпиријски утемељено разумевање праве природе људске рационалности прва претпоставка усвајања и спровођења оптималних и ефикасних јавних политика.

Већина јавних политика савремених држава скројена је на основу уверења да је човек *Homo economicus*. Основна карактеристика *Homo economicus* је егоизам, или прецизније тврдња да појединац само брине за себе (*self-regarding*) и да води рачуна само о својим личним материјалним исплатама из сарадње (у виду добара или новчаних добитака). По Дену Кејану, конвенционалну економску теорију, у чијем је срцу *Homo economicus* и која је стандардна апаратура за анализу и предлагање јавних политика, карактеришу четири главне тврдње: (1) актери су увек максимизира-

тори богатства (2) и карактеришу их хомогене преференције; (3) у колективном делању увек постоји јединствени еквилибријум, односно једна доминантна стратегија за актера (а то је фрирајдинг); (4) проблем колективног делања се превазилази увођењем екстерног система подстицаја (Kahan, 2005). Парадигматски пример ове теорије је *Логика колективног делања* Манкура Олсона [1965].

Недостатке тврдње под бројем 3 смо већ утврдили у претходним деловима. Тврдња под редним бројем 4 је такође нарочито важна за успешно креирање јавних политика. Наиме, стандардна економска претпоставка је да је делање актера увек покренуто преференцијама (тј. жељама) и ограничењима (каква су цене добара, буџет или време). С обзиром да су ограничења лакше мерљива од преференција, ова бинарна логика имала је и значајне импликације по дизајн јавних политика: променом система ограничења, тј. подстицаја, могуће је мењати и људско понашање. Рецимо, у сфери казних политика, као решење ова перспектива предлагала је повећавање трошкова криминалне радње: с обзиром да приликом чињења кривичног дела потенцијални прекршилац закона увек ради *cost-benefit* анализу, све што треба да урадимо да би га одвратили од те радње је да повећамо цену чињења злочина. Или, у сфери односа принципала и агента, ова перспектива породила је својеврсан „закон понашања“ (Gneezy et al., 2011): већа плата водиће ка повећању учинковитости радника на радном месту, а ученици плаћени за добре оцене добијаће све боље оцене у школи. У сфери друштвено корисног понашања влада иста логика: уколико људе мотивишемо материјалним наградама, они ће бити склонији да доприносе јавном добру (рецимо да се прикључе акцији добровољног давања крви). Међутим, овај начин размишљања, који занемарује природу људских преференција, скупо је плаћен. Оно што је у анализи занемарено то је потенцијал тзв. интринсичних мотивација које подразумевају да људи могу радити неке ствари једноставно зашто што им се оне свиђају, јер уживају у њима или зато што сматрају да је морално исправно радити их (Ryan & Deci, 2000). Политика екстринсичних, финансијских или правних интервенција може истиснути (*crowd out*) те интринсичне мотивације<sup>4</sup>. По Сему Боулсу четири су главна механизма одговорна за первертирани однос подстицаја и унутрашњих мотивација. Подстицаји могу (1) открити информације о ономе ко уводи подстицај, (2) фрејмовати (*framing*)

4 Темелјни прегледи налаза односа подстицаја и понашања могу се наћи у: Bowles & Polania-Reyes, 2009; Kamenica, 2011.

ситуацију у којој се доноси одлука при чему фраме сугерише шта је одговарајуће понашање у датој ситуацији, (3) угушити осећај аутономије код појединаца који не воле доминацију и контролу, и (4) утицати на процес којим појединци усвајају нове преференције (Bowles & Polanya-Reyes, 2009).

### Уместо закључка: лекција једног економисте о пореклу људске продруштвености

Развој бихејвиоралне теорије игара значајно је унапредио нашу способност да људске циљеве третирамо пре као ствар чињеница него логике, што је од суштинског значаја за разумевање људског понашања, како у дескриптивном, исто тако и у нормативном и предиктивном смислу. Један од историјски првих и најзначајнијих помака у том смеру можемо наћи код Адама Смита. У свом делу *Теорија моралних осећања* [1759] Смит је покушао да развије теорију морала (и то у правом смислу речи теорија): којих се то моралних начела људи придржавају у свакодневном понашању, какво је њихово порекло и који су механизми њихове репродукције (Smir, 2009, 70). Иако је Теорија знатно слабије позната од *Богатства народа* [1776], сам Смит је сматрао да је у питању дело које је далеко важније за разумевање функционисања људског друштва<sup>5</sup>.

Као и његов савременик Дејвид Хјум, и Смит је сматрао да је људско делање више производ инстинктивног и емотивног него рационалног и прорачунатог. Уводна реченица *Теорије* гласи: „Ма колико се претпостављало да је човек себичан, очигледно је да у његовој природи постоје нека начела која подстичу његову заинтересованост за срећу других људи, и која чине да му је њихова срећа потребна, иако он сам од тога нема ништа осим задовољства што такву срећу види“ (ibid, 3). По Смиту, осећај *симпатије* је кључни водич моралног размишљања и понашања. Како су људи по природи друштвена бића, „код сваког душевног стања коме је подложен човеков дух, емоције посматрача увек одговарају ономе што он, уживљавајући се у ситуацију, замишља да осећај онај који пати“. Ово саосећање има и неуролошки доказ: „Кад видимо

5 Док је последње, шесто издање *Теорије*, које се појавило 1790. године, било значајно допуњено у односу на претходна издања, последња два издања *Богатства народа* су садржала тек незнатне промене у односу на претходна.

да неко замахне да удари некога по руци или нози ми, природно, тог тренутка устукнемо и повучемо своју руку или ногу; а кад ударац падне, и ми у извесној мери осетимо повреду и бол удареног“ (ibid, 4). Смит наглашава и емоционалну димензију људске друштвености: „Када је стварала човека за живот у друштву, природа га је обдарила почетном и суштинском жељом да чини пријатности својој сабраћи, као и почетним и суштинским гнушањем према свему што их повређује. Научила га је да осећа задовољство због њиховог повољног и бол због њиховог неповољног става. Она је ово одобравање учинила нечим најласкавијим и најпријатнијим за њега, а њихово неодобравање нечим најмучнијим и нечим што највише вређа“ (ibid, 108). Другим речима, када људи повреду неко правило понашања они ће се осећати лоше, што је довољан знак да такво понашање треба избегавати у будућности.

Које је порекло ових „начела“ која човека подстичу да, поред својих, води рачуна и о исходима сарадње по друге људе? Ово питање је о суштинског значаја за разумевање природе људске рационалности и друштвености. По Смиту, морално расуђивање је производ друштвене интеракције, односно искуственог учења које у основи има посматрање и имитирање других људи. Смит добро примећује да су пристрасност, једностраност и кратковидост „извор половине свих поремећаја у нашем животу“ (ibid, 138). Међутим, посматрање понашања других људи човека доводи до формирања општих правила о томе шта је подобно и исправно, шта треба чинити а шта избегавати. Када схвати да „значајни други“ деле и оправдавају тај општи став, човек почиње да осећа задовољство због тога што чини оно што сви доживљавају као исправно. Од тог тренутка човек одлучује да не крши општеприхваћена правила понашања, јер би у супротном постао предмет неодобравања, презира, мржње и кажњавања. Са друге стране, поштовање правила ка њему привлачи осећања љубави, захвалности и дивљења од стране других људи. Зато, наједном, постаје у интересу човека да се понаша у складу са правилима, и, штавише, да непрестано тражи прилику где може да испољава општеприхваћено понашање. Другим речима, „опште правило ствара се тек пошто на основу искуства утврдимо да све акције извесне врсте, или све околности извесне врсте, редовно наилазе на одобравање или, такође редовно, на осуду ... То је начин на који се стварају правила моралности“ (ibid, 139). Та правила морала коригују наше идеје о томе шта треба радити. Она имају и потпору у емоцијама: уколико урадимо нешто



што је супротно правилима, када завршимо са тим чином и ставимо се у улогу других који нас хладнокрвно посматрају, ми добијамо напад гриже савести и кајања који нас узнемиравају и муче<sup>6</sup>.

По Смиту, човек је увек подељен да две личности: једну која дела, и другу која треба да просуђује да ли је понашање прве оправдано. Ова друга личност (нека врста огледала прве) носи назив „непристрасни посматрач“. Непристрасни посматрач је кључни механизам преко кога се репродукује колективна моралност. Он подразумева незаинтересованост, односно измештање себе на универзално стајалиште за које нам се чини да га деле све људи, и потом процењивање сопственог понашања са те удаљене тачке. Смит каже: „Ми никад не можемо спознати своја осећања и мотиве нити стећи неки суд о њима уколико себе не уклонимо из положаја у коме се природно налазимо и не настојимо да ствар посматрамо са извесне удаљености од нас самих. Но, то нисмо у стању да учинимо ни на који други начин до настојећи да на ствари гледамо очима других људи или онако како би ти други људи на њих највероватније гледали“ (Smit, 2009, 103).

Када непристрасни посматрач прође кроз процес биолошке и културне коеволуције на друштвеном нивоу, тј. неке врсте процеса „невидљиве руке“ у сфери морала, он постаје оличен у виду друштвених норми, традиције, вредности и, уопштено, једног дубоког, спонтаног моралног и културног поретка друштва, који надаље појединцима служи као неформална оријентација за њихова свакодневна делања. Ова врста поретка је извор онога што је Вернон Смит, попут Гигерензера, назвао „еколошка рационалност“: понашање је еколошки рационално у оној мери у којој се прилагодило структури окружења у којем оперише. Другим речима, смернице за „еколошки рационално“ делање извлаче се управо

6 Сјајно објашњење овог процеса понудио је Фридрих фон Хајек (који је иначе био инспирисан овим моделом размишљања карактеристичним за Шкотско просветитељство). Он каже: „Учење тога како се треба понашати је више *извор* него *резултат* увида, разума и разумевања. Човек није рођен мудар, рационалан и добар, већ је научен да такав постане. Није наш интелект створио наш морал; пре су људске интеракције вођене нашим моралом омогућиле раст разума и могућности повезаних са њим. Човек је постао интелегентањер је постојала традиција – она која лежи између инстинкта и разума – из које може да учи. А ова *традиција* није израсла из способности да се рационално интерпретирају уочене чињенице већ из навике одговора (*habit of responding*). Она је у основи говорила човеку шта би или не би требао да ради у одређеним околностима, а не шта мора да очекује да се деси.“ (Најек, 1988, 21-22)

из оног домена оног прећутног, имплицитног, често инстинктивног знања, које сам споменуо у трећем параграфу овог дела. Ово имплицитно знање је од пресудног значаја за функционисање како друштва, тако и самих људи: по речима Фридриха Хајека, када бисмо престали да радимо оно чему не знамо разлог, врло брзо бисмо убрзо били мртви (Науек, 1988, 68)<sup>7</sup>.

Смитова *Теорија моралних осећања* се као значајан извор увида за друштвену теорију и политичку филозофију открива током последњих неколико година. Она је од посебног значаја за усавршавање класичних модела економског понашања који самостално нису у стању да објасне бројне налазе економских експеримената какви су они са Игром диктатора, Игром јавних добара и Игром поверења. Без интегрисања теорије моралног понашања у објашњење и предвиђање економског понашања нема даљег унапређења друштвене теорије (Stevens, 2011). На крају, вреди споменути и покушај Ашрафа, Камерера и Ловенстина који су покушали да докажу да је Адам Смит, својим приступом људском понашању, заслужио да буде сматран првим бихејвиоралним економистом (Ashraf, Camerer, Loewenstein, 2005).

## Литература

Ainslie, George & Haslam, Nick, „Hiperbolic Discounting“, in George Loewenstein & Jon Elster (eds.), *Choice Over Time*, New York: Russel Sage Foundation, 1992.

Anderson, Cristopher & Putterman, Louis (2006), „Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism“, in: *Games and Economic Behavior* 54 (1): 1-24.

Andreoni, James (1990), „Impure Altruism and Donations to Public Goods: A Theory of Warm Glow Giving“, in: *Economic Journal* 100 (401): 464-477.

Andreoni, James & Miller, John (2002), „Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism“ in: *Econometrica* 70 (2): 737-753

Andreoni, James, William T. Harbaugh and Lise Vesterlund (2008), „Altruism in Experiments“, in: Steven N. Durlauf and Lawrence E. Blumer (eds.), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, 2008.

---

7 Хајекова детаљна схватања на тему спонтаног поретка и његових имплицитних правила могу се наћи у Хајек, 2002.

Allais, Maurice, (1953), „Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école Américaine“, in: *Econometrica* 21 (4): 503–546.

Аријели, Ден, *Нелогичан избор*, Алнари, Београд, 2008.

Ashraf, Nava, Camerer, Colin and Loewenstein, George, (2005) „Adam Smith, Behavioral Economist“, in: *Journal of Economic Perspectives* 19 (3): 131-145.

Becker, Gary (1968), “Crime and Punishment: An Economic Approach”, in: *The Journal of Political Economy* 76 (2): 169–217.

Binmore, Ken. *Rational Decisions*. Princeton: Princeton University Press, 2009.

Boyd, Robert & Richerson, Peter, *The Origin and Evolution of Cultures*. New York, Oxford University Press, 2005.

Boyd, Robert & Richerson, Peter, *Not By Genes Alone. How Culture Transformed Human Evolution*. Chicago, University of Chicago Press, 2005.

Bowles, Samuel & Gintis, Herbert, (2002) „Homo Reciprocans“, in: *Nature* 415: 125-128.

Bowles, Samuel & Gintis, Herbert, *A Cooperative Species. Human Reciprocity and Its Evolution*. Princeton, Princeton University Press, 2011.

Bowles, Samuel & Polania-Reyes, Sandra, (2009), “Economic Incentives and Social Preferences: Substitutes or Complements?”, in: *Journal of Economic Literature* 50 (2): 368-425.

Carpenter, Jeffrey, (2007), “The Demand for Punishment”, in: *Journal of Economic Behavior & Organization* 62(4): 522-542.

Chaudhuri, Ananish, *Experiments in Economics. Playing Fair With Money*. London, Routledge, 2009.

Chang, S., Gariépy, J-F., & Platt, M., (2013), „Neuronal Reference Frames for Social Decisions in Primate Frontal Cortex“, in: *Nature Neuroscience* 16 (2): 243-250.

Cohen, Jonathan, (1981), „Can Human Irrationality be Experimentally Demonstrated“, in: *Behavioral and Brain Sciences* 4 (3): 317-331.

Dana, Justin, Daylian M. Cain and Robyn M. Dawes, (2006), „What You Don’t Know Won’t Hurt Me: Costly (But Quiet) Exit in Dictator Games“, in: *Organizational Behavior and Human Decision Processes* 100: 193-201.

Dekić, Milovan, „Zašto su ljudi altruisti? Mikro temelji društvenih normi i ljudske pro društvenosti u evolutivnom kontekstu“. Radna verzija dostupna na: <http://milovandekic.files.wordpress.com/2013/12/mikroevolucija-altruizma.pdf>

El Mouden, Claire, Maxwell Burton-Chellew, Andy Gardner and Stuart West, “What do Humans Maximize?” in: Samir Okasha & Ken Binmore (eds.), *Evo-*

*lution and Rationality: Decisions, Cooperation and Strategic Behavior*, Cambridge, 2012.

Elster, Jon, „Jon Elster“, in: Richard Swedberg (ed.) *Sociology and Economics. Redefining Their Boundaries: Conversations with Economists and Sociologists*. Princeton: Princeton University Press, 1990.

Elster, Jon, (1994), „Rationality, Emotions and Social Norms“, in: *Synthese* 98 (1): 21-49.

Elster, Jon, *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press, 2007.

Ellsberg, Daniel, (1961), „Risk ambiguity and the Savage axioms“, in: *Quarterly Journal of Economics* 75: 643–649.

Fehr, Ernst & Gächter, Simon, (2000), „Cooperation and Punishment in Public Goods Experiments“, in: *The American Economic Review* 90(4): 980-994.

Fehr, Ernst & Gächter, Simon (2002), „Altruistic Punishment“, in: *Nature* 415 (6868): 137-140.

Fehr, Ernst & Fischbacher, Urs, (2003), „The Nature of Human Altruism“, in: *Nature* 425: 785-791.

Fehr, Ernst & Rockenbach, Bettina, (2004), „Human Altruism: Economic, Neural, and Evolutionary Perspectives“, in: *Current Opinion in Neurobiology* 14: 784–790.

Fehr, Ernst & Fischbacher, Urs, (2004), „Third-party punishment and social norms“, in: *Evolution and Human Behavior* 25 (2): 63-87.

Fehr, Ernst & Schmidt, Klaus, (2006), „The Economics of Fairness, Reciprocity, and Altruism – Experimental Evidence and New Theories“, in: Serge-Christophe Kolm and Jean Mercier Ythier (eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity, Foundations, Vol. 1*. North Holland: Elsevier, pp. 615-670.

Fehr, Ernst & Fischbacher, Urs, (2005), „The Economics of Strong Reciprocity“, in: Gintis, Herbert et al. (eds). *Moral Sentiments and Material Interests. Foundations of Cooperation in Economic Life*. Cambridge: The MIT Press, 2005.

Fehr, Ernst, Holger Herz and Tom Wilkening, (2013), “The Lure of Authority: Motivation and Incentive Effects of Power”, in: *American Economic Review* 103 (4): 1325-59.

Field, Alexander, *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory and the Origins of Reciprocity*. Ann Arbor, University of Michigan Press, 2001.

Fischbacher, Urs, Gächter, Simon & Fehr, Ernst, (2001), “Are people conditionally cooperative? Evidence from a public goods experiment“, in: *Economics Letters* 71(3): 397-404.

- Frank, Robert, *Passions Within Reason. The Strategic Role of the Emotions*. New York: W.W. Norton & Company, 1988.
- Frey, Bruno & Benz, Matthias, (2001), „Motivation Transfer Effect“. Working paper. Institute for Empirical Research in Economics, University of Zürich.
- Gigerenzer, Gerd and Goldstein, Daniel, (1996), “Reasoning the Fast and Frugal Way: Models of Bounded Rationality”, in: *Cognitive Science* 103 (4): 650-666.
- Gigerenzer, Gerd & Selten, Reinhard (eds.) *Bounded Rationality. The Adaptive Toolbox*. MIT Press, 2001.
- Gigerenzer, Gerd, “Is Mind Irrational or Ecologically Rational?”, in: Vernon Smith & Francesco Parisi (eds.), *The Law and Economics of Irrational Behavior*, Stanford: Stanford University Press, 2005.
- Gilovich, Thomas, Tversky, Amos & Vallone, R., (1985), “The Hot Hand in Basketball: On the Misperception of Random Sequences”, in: *Cognitive psychology* 3 (17): 295–314.
- Gintis, Herbert et al. (eds.), *Moral Sentiments and Material Interests. Foundations of Cooperation in Economic Life*. Cambridge: The MIT Press, 2005.
- Gintis, Herbert, Samuel Bowls, Robert Boyd, Ernst Fehr, “Moral Sentiments and Material Interests: Origins, Evidence and Consequences”, in: Gintis, Herbert et al. (eds.), *Moral Sentiments and Material Interests. Foundations of Cooperation in Economic Life*. Cambridge: The MIT Press, 2005.
- Gintis, Herbert, (2006), “Towards a Unity of the Human Behavioral Sciences”, in: *Papers. Revista de Sociología* 80: 97-122.
- Gintis, Herbert, (2007), “A Framework for the Unification of the Behavioral Sciences”, in: *Behavioral and Brain Sciences* 30 (1): 1-17.
- Gintis, Herbert, *The Bounds of Reason. Game Theory and the Unification of the Behavioral Sciences*. Princeton: Princeton University Press, 2009.
- Gintis, Herbert, (2013), “Altruistic Reciprocity”, in: Luigino Bruni & Stefano Zamagni (eds.), *Handbook on the Economics of Reciprocity and Social Enterprise*, Cheltenham: Edward Elgar Publishing, 2013.
- Gneezy, Uri, (2005), “Deception: The Role of Consequences”, in: *The American Economic Review* 95 (1): 384-394.
- Gneezy, Uri, Meier, Stephan, & Rey-Biel, Pedro, (2011), “When and Why Incentives (Don't) Work to Modify Behavior”, in: *Journal of Economic Perspectives* 25 (4): 191-210.
- Hajek, Fridrih, *Pravo, zakonodavstvo i sloboda*. Beograd/Podgorica: Službeni glasnik i CID, 2002.
- Harbaugh, W., Mayr, U., Burghart, D., (2007), “Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations”, in: *Science* 316 (5831): 1622–1625.

- Hayek, Friedrich, *The Fatal Conceit. The Errors of Socialism*. London: Routledge, 1988.
- Henrich, Joseph & Henrich, Natalie, *Why Humans Cooperate? An Evolutionary and Cultural Explanation*. Oxford University Press, 2007.
- Hertwig, Ralph & Gigerenzer, Gerd, (1999), „The ‘Conjunction Fallacy’ Revisited: How Intelligent Inferences Look Like Reasoning Errors“, in: *Journal of Behavioral Decision Making* 12: 275–305.
- Ho, T., Lim, N. & Camerer, F., (2006), „Modeling the Psychology of Consumer and Firm Behavior with Behavioral Economics“, in: *Journal of Marketing Research* 43 (3): 307-331.
- Hoffrage, Ulrich, Hertwig, Ralph & Gerd Gigerenzer, (2000) „Hindsight Bias: A By-Product of Knowledge Updating?“, in: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (3): 566-581.
- Hume, David (2000 [1739-40]). *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Machina, Mark, (1987) „Choice under Uncertainty: Problems Solved and Unsolved“, in: *Journal of Economic Perspectives* 1(1): 121-154.
- Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. and Fehr, E., (2012) „Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism“, in: *Neuron* 75: 73–79.
- Kahan, Dan, „The Logic of Reciprocity: Trust, Collective Action and Law“, in: Herbert Gintis et al., *Moral Sentiments and Material Interests: The Foundation of Cooperation in Economic Life*. Cambridge: The MIT Press, 2005.
- Kamenica, Emir, (2012) „Behavioral Economics and Psychology of Incentives“, in: *Annual Review of Economics* 4 (1): 427-452.
- Kay, A. C., & Ross, L., (2003), „The perceptual push: The interplay of implicit cues and explicit situational construal in the Prisoner’s Dilemma“, in: *Journal of Experimental Social Psychology* 39: 634-643.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A., (2007), „Damage to the prefrontal cortex increases utilitarian moral judgments“, in: *Nature* 446: 908-911.
- Liberman, V., Samuels, S.M. & Ross, L., (2004) „The name of the game: predictive power of reputations versus situational labels in determining prisoner’s dilemma game moves“, in: *Personality & Social Psychology Bulletin* 30 (9): 1175-85.
- Mealey, Linda., (1995), „The Sociobiology of Sociopathy: An Integrated Evolutionary Model“, in: *Behavioral and Brain Sciences* 18 (3): 523-541.
- Olson, Mancur, *Logika kolektivnog djelovanja*. Zagreb, Fakultet političkih znanosti, 2009.

Okasha, Samir & Binmore, Ken (eds.), *Evolution and Rationality: Decisions, Cooperation and Strategic Behavior*, Cambridge, 2012.

Ostrom, Elinor, „Social Dilemmas and Human Cooperation“, in: Ronald Noë, Jan A. R. A. M. van Hooff, and Peter Hammerstein (eds.), *Economics in nature: Social Dilemmas, Mate Choice and Biological Markets*, Cambridge: Cambridge University Press, pp. 23-42.

Ostrom, Elinor, „Toward a Behavioral Theory Linking Trust, Reciprocity and Reputation“, in: Elinor Ostrom & James Walker (eds.), *Trust and Reciprocity. Interdisciplinary Lessons from Experimental Research*. New York: Russel Sage Foundation, 2005.

Popper, Karl, “Models, Instruments, and Truth. The Status of the Rationality Principle in the Social Sciences”, in: Karl Popper, *The Myth of the Framework. In Defence of Science and Rationality*. London: Routledge, 1994.

Pinker, Steven, *How The Mind Works*, W. W. Norton & Company, 1997.

Rand, David, Alexander Peysakhovich, Gordon T. Kraft-Todd, George E. Newman, Owen Wurzbacher, Martin A. Nowak & Joshua D. Greene, (2013), „Social Heuristics Shape Intuitive Cooperation“, in: *Nature Communications* 5 (3677), 1-11.

Raphael, D.D., *The Impartial Spectator. Adam Smith’s Moral Philosophy*. Oxford: Oxford University Press, 2007.

Ryan, Richard & Deci, Edward, (2000), „Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions“, in: *Contemporary Educational Psychology* 25 (1): 54–67.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD., (2003), “The neural basis of economic decision-making in the Ultimatum Game”, in: *Science* 300 (5626): 1755-8.

Savage, Leonard, *The Foundations of Statistics*. John Wiley & Sons, Inc., 1954.

Segal, Uzi, (1987), „The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach“, in: *International Economic Review* 28 (1): 175-202.

Shafir, Eldar (ed.), *The Behavioral Foundations of Public Policy*. Princeton & Oxford: Princeton University Press, 2013.

Simon, Herbert, (1956), „Rational Choice and the Structure of the Environment“, in: *Psychological Review* 63 (2): 129-138.

Slovic, Paul, (1995), “The construction of preference”, in: *American Psychologist* 50: 364-371.

Smit, Adam, *Teorija moralnih osećanja*. Podgorica: CID, 2008.

Smith, Vernon, *Rationality In Economics. Constructivist and Ecological Forms*. Cambridge: Cambridge University Press, 2008.

Sperber, Dan & Hugo Mercier, (2011), „Why Do Humans Reason? Arguments for Argumentative Theory“, in: *Behavioral and Brain Sciences* 34: 57-111.

Stevens, Douglas, *Rediscovering Adam Smith. How The Theory of Moral Sentiments can explain emerging evidence in experimental economics*. Working paper. Adam Smith Research Foundation, 2011.

Swedberg, Richard, “Introduction”, in: Swedberg, Richard (ed.), *Economics and Sociology. Redefining Their Boundaries: Conversations with Economists and Sociologists*, Princeton: Princeton University Press, 1990.

Tversky, Amos & Kahneman, Daniel, (1974) „Judgment Under Uncertainty: Heuristics and Biases“, in: *Science* 185 (4157): 1124-1131.

Tversky, Amos and Kahneman, Daniel, (1982), “Judgments of and by representativeness”. In: D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press, 1982.

Veber, Maks, *Privreda i društvo. Tom I*. Beograd: Prosveta, 1976.

Von Neumann, John & Morgenstern, Oskar, *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

Waytz, A., Zaki, J., and Mitchell, J., (2012), „Response of Dorsomedial Prefrontal Cortex Predicts Altruistic Behavior“, in: *The Journal of Neuroscience* 32(22): 7646-7650.

Wilkinson, Nick & Klaes, Matthias, *An Introduction to Behavioral Economics. 2nd Edition*. London: Palgrave Macmillan, 2012.

**Milovan Dekić**

## ARE HUMANS RATIONALLY PROSOCIAL? SOME EVIDENCE FROM BEHAVIORAL GAME THEORY

---

### Summary

In the article I am offering findings from behavioral experiments that demonstrate that rational preferences are the foundation of human prosocial behavior. First of all I explain what rationality means, and then how rational preferences correlate with values and social norms in generating social cooperation. Then I review possible answers to behavioral economists who argue that humans are often irrational. Then I refer to some lessons policy makers could learn from demonstrated human prosocial rationality. In the final part, instead of a con-



clusion, I review the explanation that Adam Smith proposed regarding the origin and nature of human prosociality.

**Keywords:** rationality, prosociality, social norms, values, preferences, heuristics, morality, behavioral economics

---

---

\* Овај рад је примљен 21 фебруара 2015. године а прихваћен за штампу на састанку Редакције 20. марта 2015. године.

