

UNIVERZITET U BEOGRADU

FILOZOFSKI FAKULTET

Igor R. Živanović

**BIOLOŠKE OSNOVE MORALA: EGOIZAM,  
ALTRUIZAM I SAMOOBMANJIVANJE**

doktorska disertacija

Beograd, 2015

UNIVERSITY OF BELGRADE

FACULTY OF PHILOSOPHY

Igor R. Živanović

**THE BIOLOGICAL BASIS OF MORALITY:  
EGOISM, ALTRUISM AND SELF-DECEPTION**

Doctoral dissertation

Belgrade, 2015

## **Informacije o mentoru i članovima komisije**

Mentor:

dr Eva Kamerer, docent, Filozofski fakultet Univerziteta u Beogradu

Članovi komisije:

dr Živan Lazović, redovni profesor, Filozofski fakultet Univerziteta u Beogradu

dr Aleksandar Dobrijević, docent, Filozofski fakultet Univerziteta u Beogradu

dr Biljana Stojković, vanredni profesor, Biološki fakultet Univerziteta u Beogradu

Datum odbrane disertacije

# **Biološke osnove morala: egoizam, altruizam i samoobmanjivanje**

## Rezime

*Ključne reči:* altruizam, moral, evolucija, prirodna selekcija, jedinica selekcije, empatija, psihopatija, samoobmanjivanje, moralni realizam, evolucionari antirealizam.

Osnovni cilj ove disertacije je uspostavljanje veze između etike i evolucione biologije. Mada postoji gotovo opšti konsenzus da su biologija i ljudska evolucionarna istorija imale nekakvog uticaja na oblikovanje naše moralne sposobnosti, ova veza se obično smatra zagonetnom. Jedan od razloga za to je evolucionarna nestabilnost morala.

Centralno pitanje koje se u disertaciji postavlja je zašto imamo moralne pojmove, verovanja i sposobnost moralnog rasuđivanja, kao i zašto su naši međusobni odnosi regulisani moralnim pravilima. Veza između morala i biologije, tj. etike i evolucione teorije, uspostaviću posredno preko moralne psihologije. Preciznije, ova veza će biti ustanovljena preko psihološkog fenomena samoobmanjivanja. Iako je ovaj fenomen dugo poznat, on je pretežno bio predmet proučavanja psihologije i filozofije duha, a tek nedavno je dobio naučnu artikulaciju u kontekstu evolucione teorije. Za filozofiju verovatno najvažnija posledica ovakvog pristupa je specifična metaetička pozicija koja je poznata kao evolucionari antirealizam. Uprkos činjenici da je većina filozofskih argumenata o evolucionom poretku morala adaptacionistička, u ovom radu se moralna sposobnost sagledava kao spandrel, odnosno kao nusproizvod prirodne selekcije.

Prvo poglavlje ove doktorske disertacije se bavi konceptualnim problemima koji se vezuju za biološku i psihološku interpretaciju egoizma (sebičnosti) i altruizma, a zatim objašnjava vezu između altruizma i morala. Zapravo, u ovom poglavlju se uspostavlja veza između altruizma i moralne sposobnosti koja uključuje moralno ponašanje, formiranje i održavanje moralnih verovanja i moralno suđenje. Osim toga, jedan deo prvog poglavlja posvećen je objašnjenju nekih karakteristika moralnih sudova.

Drugo poglavlje se bavi evolucijom altruizma i saradnje, kao i problemom jedinica selekcije. U ovom poglavlju ću kritički preispitati standardne modele za evoluciju altruizma: grupnu selekciju, genski selekcionizam, selekciju u srodničkom krugu (s osvrtom na problem usvajanja i Vestermarkov efekat), direktni reciprocitet (recipročni altruizam), indirektni reciprocitet i selekciju na više nivoa. Na osnovu eksplanatorne kompleksnosti, odbaciću selekciju na više nivoa i za potrebe ovog rada ću se oslanjati uglavnom na teoriju inkluzivne adaptivne vrednosti i direktni reciprocitet. Ovi eksplanatorni modeli će biti dovoljni da se objasni evolucija empatije o kojoj će biti reči u narednom poglavlju.

U trećem poglavlju se bavim emocijama i empatijom, koje smatram neposrednim uzrocima psihološkog altruizma i važnim činiocima moralne sposobnosti. Posebna pažnja će biti posvećena Preston/de Valovom integrativnom evolucionom modelu empatije, u čijoj osnovi se nalazi mehanizam zajedničkog kodiranja percepcije i akcije čiji neurobiološki supstrat predstavlja sistem neurona ogledala. Kroz analizu eksperimentalne evidencije i podataka sa terena o altruističkom ponašanju i pomaganju kod dece i šimpanzi, objasniću zašto je empatija značajna za moral. Ipak, iako su primati sposobni da se ponašaju altruistički i verovatno imaju takozvane gradivne blokove morala, njima ipak nedostaje potpuno razvijen fenomen moralnosti. Dakle, moralna sposobnost je specijes specifična karakteristika ljudi.

Deficiti u emocionalnoj empatiji su tesno povezani s problemima u razumevanju moralnih pojmova i odstupanjima u moralnom ponašanju. Ovi emocionalni deficiti i nevolje s moralom su evidentne kod psihopata. Tako, u četvrtom poglavlju fokus istraživanja će biti na osobama s psihopatskim karakteristikama, njihovoj neurobiologiji i psihologiji, razumevanju emocija i jezika, i posebno na evoluciji psihopatskih karakteristika. Predstaviću tri evolucionarna modela psihopatije: selekciju zavisnu od učestalosti gena, fenotipsku plastičnost i mutaciono opterećenje. Ako su psihopatske karakteristike uzrokovane evoluiranim neuronskim mehanizmima, onda su prema evolucionom objašnjenju psihopatske karakteristike *prave funkcije* tih mehanizama. S obzirom na to, psihopatija nije poremećaj ličnosti, već evoluirana životna strategija. U završnom delu ovog poglavlja uvešću distinkciju između moralnih i konvencionalnih pravila. Ovo je važna distinkcija, jer psihopate za razliku od četvorogodišnjaka ne mogu da razlikuju ova pravila. Problemima psihopatije ću se vratiti još jednom u narednom poglavlju koje se bavi samoobmanjivanjem.

Samoobmanjivanje i interpersonalno obmanjivanje se specijes specifični fenomeni. U petom poglavlju, detaljno ću se baviti konceptualnim i evolucionim problemima samoobmanjivanja. Glavna meta kritike biće Triversova naučna teorija samoobmanjivanja. Prema navedenoj teoriji, samoobmanjivanje je evoluiralo da bi poboljšalo obmanjivanje drugih. Kada se samoobmanjivanje tumači po modelu interpersonalnog obmanjivanja, kao u Triversovoj evolucionoj teoriji, pojavljuju se dva paradoksa. Jedan je dinamički i odnosi se na sam proces samoobmanjivanja. Budući da su obmanjivač i obmanuti identični, čini se da je nemoguće da se osoba svesno, namerno i uspešno samoobmanjuje. Drugi je statički i odnosi se na nemoguće stanje svesti osobe koja se samoobmanjuje, koji simultano ima protivrečna verovanja. Iako može da se čini da samoobmanjivanje osigurava egoističku efikasnost, to ne mora da bude slučaj. Neke studije pokazuju da je samoobmanjivanje evoluiralo kao potpora strategijama koje uključuju kratkoročne žrtve zarad ostvarenja dugoročnih dobiti, tako što suspenduje ili prikriva impulse za obmanjivanje i agresiju. Druge studije pokazuju da je samoobmanjivanje evoluiralo da bi održavalo mentalnu koherenciju i konzistentnost sistema verovanja. Ovi nalazi su u suprotnosti s Triversovom hipotezom. Osim toga, nedostatak samoobmanjivanja je u skladu s nalazima o disfunkciji mehanizma za inhibiciju nasilja kod psihopata, kao i impulsivnošću, obmanjivanjem i sklonošću ka preuzimanju rizika, što su karakteristike ove populacije.

U završnom poglavlju razmotriću dva evolucionarna argumenta protiv moralnog realizma. Ovi argumenti imaju prethodnike u Mekijevoj *Teoriji greške* i *Harmanovom izazovu*, tj. eksplanatornoj irelevanciji moralnih činjenica. Prvi argument je metafizički i nazvao sam ga *argument filogenetske kontingencije*. Jednostavno rečeno, ovaj argument tvrdi sledeće: da je naša evolucionarna prošlost bila drugačija, onda bi naša moralna sposobnost, moralni pojmovi i moralna verovanja bili drugačiji. Drugi argument je epistemološki i zasnovan je na Nozikovoj koncepciji osetljivosti i sigurnom praćenju istine. Ovim argumentom se tvrdi da moralna sposobnost nije selektovana za praćenje moralnih istina i da moralna verovanja nisu osetljiva na istinu. Konačno, pokazaću na koji način samoobmanjivanje, koje je po definiciji neosetljivo na istinu, može da pruži dodatni oslonac i pomogne boljoj artikulaciji navedenih argumenata.

Naučna oblast: filozofija

Uža naučna oblast:

UDK broj:

# **The Biological Basis of Morality: Egoism, Altruism, and Self-deception**

## Summary

*Key words:* altruism, morality, evolution, natural selection, the unit of selection, empathy, psychopathy, self-deception, moral realism, evolutionary anti-realism.

The main objective of this dissertation is to establish link between ethics and evolutionary biology. Although there is almost a general consensus among philosophers that biology and human evolutionary history had some role in shaping our moral capacity (behavior and judgment), this relation is usually considered puzzling. One of the reasons for this is evolutionary instability of morality.

The central issue this dissertation raises is why do we have moral concepts, beliefs and moral reasoning ability, and why are interpersonal and social relations maintained by moral rules. I will establish relation between morality and biology, i.e. ethics and evolutionary theory, indirectly via moral psychology. More specifically, this relation is established through psychological phenomenon of self-deception. While this phenomenon has long been known, it was mostly the subject of psychology and philosophy of mind, and just recently got scientific articulation in the context of the evolutionary theory. Probably the most significant philosophical implication of this approach is specific metaethical position known as *evolutionary anti-realism*. Despite the fact that the most of philosophical arguments on evolutionary origins of morality are adaptationist, in this dissertation moral capacity is considered to be a spandrel, the byproduct of natural selection.

The first chapter of this doctoral dissertation deals with conceptual issues of the biological and psychological interpretations of egoism (selfishness) and altruism, and subsequently, it explains relation between altruism and morality. To be more precise, it relates altruism to moral capacity which includes moral behavior, forming and maintaining moral

beliefs, and moral judgments. In addition, one part of this chapter I devote to the explanation of some characteristics of moral judgments.

The main subjects of the second chapter are the evolution of altruism and cooperation and the related problem of the units of selection. I critically examine the standard models for the evolution of altruism: group selection, gene's eye view, kin selection (with remarks on adoption and Westermarck effect), direct reciprocity (i.e. reciprocal altruism), indirect reciprocity, and Sober and Wilson's model of multilevel selection. I will dismiss multilevel selection theory on the grounds of its explanatory complexity and for the purposes of this paper I will mainly rely on inclusive fitness theory and direct reciprocity. Those explanatory models would suffice to explain the evolution of empathy which will be discussed in the next chapter.

So, the third chapter is on emotions and empathy which I take to be proximate causes of psychological altruism and important ingredients of moral capacity. Here, special attention is paid to Preston and de Waal's integrated evolutionary model of empathy, with perception-action mechanism in its core, and mirror neuron system as its neurobiological substrate. Through the analysis of experimental and field data on children's and chimpanzees altruistic and helping behavior I will explain why empathy is important for morality. Nevertheless, although primates other than humans are capable of acting altruistically and may have so called building blocks of morality, they lack full blown morality. Consequently, moral capacity is a species specific characteristic.

Deficits in emotional empathy are closely related to the problems regarding difficulties in comprehension of moral concepts and aberrations in moral behavior. Those emotional deficits and troubles with morality are evident in psychopaths. Hence, in the fourth chapter I will focus to individuals with psychopathic characteristics, their neurobiology and psychology, understanding of emotions and language, and especially to the evolution of psychopathic traits. I will present three evolutionary models of psychopathy: frequency dependent selection, contingent shifts and mutation load. If psychopathic characteristics are causally dependent on evolved neural mechanisms, then evolutionary explanation may render psychopathic traits *proper functions* of those mechanisms. Taking this into consideration, psychopathy is not personality disorder, but evolved life strategy. In the final section of this chapter I will introduce distinction between



moral and conventional rules. This is an important distinction, for psychopaths in contrast to four-year old children cannot differentiate between these rules. I'll return to the problems of psychopathy once again in the next chapter where I discuss self-deception.

Self-deception as well as interpersonal deception is species specific phenomenon. In the fifth chapter I examine in detail conceptual and evolutionary issues of self-deception. The main target of critical assessment will be Trivers' scientific theory of self-deception. According to this theory, self-deception evolved to enhance other-deception. When self-deception is understood in terms of interpersonal deception, as in Trivers' evolutionary model, two paradoxes emerge. One is dynamic and it concerns the process of self-deception, for in the case of self-deception deceiver and deceived are indistinguishable. Hence, it seems impossible for a person to consciously, intentionally and successfully be self-deceived. The other is so called static paradox and it concerns the impossible mental state of self-deceived individual, simultaneously having contradictory beliefs. Although it might appear that self-deception ensures egoistic efficiency, that doesn't have to be the case. Some studies demonstrate that it could evolve to support strategies which incorporate short-term sacrifices to achieve long-term benefits by suspending or concealing the impulses for deception and aggression. The other studies show how self-deception could evolve to maintain mental coherence and the consistency of belief system. These findings contradict Trivers' self-deception as function of deception hypothesis. Accordingly, deficiency of self-deception is in line with findings on violence inhibition mechanism dysfunction in psychopaths. It also accords well with impulsive, deceiving, and risk prone behavior find in population with psychopathic traits.

In the final chapter I discuss two forms of evolutionary debunking arguments. These arguments have precursors in Mackie's *Moral error theory* and *Harman's challenge*, i.e. the explanatory irrelevance of moral facts. The first argument is metaphysical, and I call it *the argument of phylogenetic contingency*. To put it simple, this argument claims that if our evolutionary past had been different, then our moral capacity, moral concepts and moral beliefs, would have been different as well. The other argument is epistemological, and it is based on the *Nozickean* conceptions of *sensitivity* and *truth tracking*. This argument claims that, when it comes to moral capacity, in the evolutionary past there was no selection for tracking moral truths. As a result, moral beliefs are insensitive to truth. Finally, I show how conception of self-

deception, which is by definition insensitive to truth, could augment these arguments and help their better articulation.

# SADRŽAJ

UVOD.....	14
1 EGOIZAM, ALTRUIZAM I MORAL.....	25
1.1 Egoizam i altruizam: biološka i psihološka interpretacija .....	25
1.2 Altruizam, moral i karakteristike moralnih sudova.....	30
2 EVOLUCIJA ALTRUIZMA I PROBLEM JEDINICA SELEKCIJE .....	41
2.1 Paradoks altruizma .....	41
2.2 Grupna selekcija.....	42
2.3 Grupna selekcija u 20. veku .....	45
2.4 Genski selekcionizam i evolucija iz perspektive gena .....	46
2.5 Teorija inkluzivne adaptivne vrednosti – selekcija u srodničkom krugu.....	51
2.5.1 Prepoznavanje srodnika .....	55
2.5.2 Problem usvajanja.....	59
2.5.3 Vestermarkov efekat i problem izbegavanja incesta .....	60
2.6 Direktni reciprocitet i evolucija saradnje .....	62
2.7 Indirektni reciprocitet i altruističko kažnjavanje.....	72
2.8 Selekcija na više nivoa: Soberov i Vilsonov model.....	78
3 EMOCIJE I EMPATIJA.....	87
3.1 Šta su emocije?.....	87
3.2 Empatija .....	89
3.3 Evolucionni model empatije.....	91
3.4 Model babuške .....	94
3.4.1 Emocionalna zaraza .....	95
3.4.2 Saosećajna zabrinutost .....	97
3.4.3 Afektivno zauzimanje perspektive.....	98
3.5 Afektivno zauzimanje perspektive i altruističko pomaganje kod dece i šimpanzi .....	100
3.6 Imaju li primati gradivne blokove morala? .....	103
3.7 Neurobiologija empatičkih kapaciteta.....	106
4 PSIHO PATIJA.....	116
4.1 Šta je psihopatija?.....	116

4.2	Neposredni uzroci psihopatije.....	118
4.3	Emocije i psihopatija.....	120
4.4	Evolucija psihopatije.....	123
4.5	Šta deca mogu, a psihopate ne mogu?.....	129
5	SAMOOBMANJIVANJE: KONCEPTUALNI PROBLEMI I EVOLUCIJA.....	135
5.1	Interpersonalno obmanjivanje, laganje i dovođenje u zabludu.....	136
5.2	Samoizazvano obmanjivanje i samoobmanjivanje.....	139
5.3	Paradoksi samoobmanjivanja.....	142
5.3.1	Dinamički ili strateški paradoks.....	142
5.3.2	Statički (doksatički) paradoks.....	143
5.4	Intencionalističko i divizionističko objašnjenje samoobmanjivanja.....	143
5.4.1	Mentalno raščlanjivanje.....	144
5.4.2	Vremensko raščlanjivanje.....	146
5.5	Neintencionalističko i nedivizionističko tumačenje samoobmanjivanja.....	148
5.5.1	Samoobmanjivanje i psihička tenzija.....	148
5.5.2	Samoobmanjivanje i pragmatičko testiranje hipoteza.....	150
5.6	Evolucija samoobmanjivanja: Triversov model.....	156
5.7	Šta nije u redu s Triversovom teorijom?.....	162
5.8	Alternativna evolucionarna teorija.....	163
5.9	Triversova hipoteza i psihopatija.....	165
6	EVOLUCIJA I ETIKA.....	173
6.1	Normativne implikacije teorije evolucije.....	173
6.2	Hjumov zakon.....	176
6.3	Naturalistička greška i test otvorenog pitanja.....	185
6.4	Teorija greške: argument na osnovu neobičnosti.....	192
6.5	Harmanov izazov: argument eksplanatorne irelevancije.....	195
6.6	Prizivanje duhova, Napoleoneove pilule i plovidba na Bermude.....	202
6.6.1	Argument filogenetske kontingencije.....	203
6.6.2	Argument na osnovu neosetljivosti.....	210
6.6.3	Napoleonove pilule.....	215
6.6.4	Darvinistička dilema.....	220
	ZAKLJUČAK.....	225

Navedena literatura .....	230
Biografija .....	258
Izjava o autorstvu .....	260
Izjava o istovetnosti štampane i elektronske verzije doktorskog rada .....	261
Izjava o korišćenju .....	262

## UVOD

Prvu polovinu 20. veka obeležilo je oštro protivljenje svakom pokušaju biologizacije etike. Razlozi za to bili su dvostruki. S jedne strane, oni su istorijski. Svako dovođenje u vezu biologije i etike, kao i pominjanje *evolucione etike* povezivalo se sa socijaldarvinizmom, pseudonaučnom praksom eugenike, totalitarnim ideologijama i koncentracionim logorima. S druge strane, postojali su i teorijsko-filozofski razlozi za odbacivanje evolucione etike. Ovi razlozi se odnose na stanovište prema kome se zadiranjem jedne prirodne nauke u ono što je bilo isključivo područje istraživanja filozofije narušava *autonomija* moralne teorije (Rawls 1999a). Osim toga, postojali su naizgled nerešivi logički problemi, *Hjumov zakon* i *naturalistička greška*, koji su predstavljali nepremostivu prepreku etičkom naturalizmu (Hume 1960, Moore 1963).

Tako je bilo sve dok se, sedamdesetih godina prošlog veka, jedan biolog nije odvažio da naruši ovu autonomiju: „Naučnici i humanisti bi trebalo zajedno da razmotre mogućnost da je došlo vreme da se etika privremeno uzme iz ruku filozofa i biologizuje“ (Wilson 1980, 287). Razlog za biologizaciju etike je činjenica da su emocionalni mehanizmi koji su uključeni u moralno rasuđivanje rezultat delovanja evolucionog procesa prirodne selekcije, tako da se ova jednostavna činjenica mora uzeti u obzir u objašnjenjima moralnog ponašanja, usvajanja moralnih verovanja i donošenja moralnih sudova. Vilson (Edward O. Wilson) je smatrao da se filozofi, jednako kao i obični smrtnici, oslanjaju na ove *emocionalne* kapacitete kada žele da *intuicijom* dođu do *deontoloških* merila onog što je dobro ili zlo (Wilson 1980). Vilsonov poziv na biologizaciju etike naišao je na osudu i kritike, s jedne strane zato što je isključivim upućivanjem na emocije i intuiciju zanemario čitavu jednu klasu filozofa, a stavljajući emocije, intuiciju i deontologiju u isti kontekst, pokazao nepoznavanje istorije filozofije i slabu osetljivost za filozofske probleme, dok je s druge strane njegov ambiciozni poduhvat ocenjen kao ideološki sumnjiv (Rose, Lewontin, and Kamin 1990).

Vilsonov preveliki poduhvat biologizacije etike je daleko od toga da bude realizovan, a zbog specifičnosti problema kojima se etika bavi, verovatno nikada ni neće biti potpuno sproveden u delo. Međutim, iz toga ne sledi da je etika potpuno imuna na uticaj koji teorija evolucije na nju može da ima. U ovom radu se zastupa stanovište prema kome je taj uticaj

prevashodno *negativan*. Preciznije, funkcija evolucionih premisa u etičkim argumentima i objašnjenjima pruža osnovu za *podrivanje* (eng. debunking) zdravorazumske metaetičke pozicije i moralnog realizma (naturalizma).<sup>1</sup>

Filip Kičer (Philip Kitcher), jedan od filozofa koji je kritički nastrojen prema Vilsonovom projektu predložio je četiri načina za biologizaciju etike (Kitcher 1994). Ti načini su sledeći:

1. Teorija evolucije može da objasni moralnu psihologiju čoveka, odnosno da objasni na koji način ljudi usvajaju moralne pojmove i kako donose moralne sudove. Ona može da objasni kako su naši daleki preci u nekoj predačkoj populaciji u nepovoljnim uslovima sredine došli do toga da misle u moralnim terminima.
2. Teorija evolucije može da posluži kao osnova (da pruži kriterijum) za ograničenje ili proširenje naših moralnih principa, tako što nam otkriva nove činjenice o ljudskoj prirodi.
3. Teorija evolucije je ključ metaetike i može nam pružiti standard za određivanje metafizičkog statusa moralnih svojstava, odnosno može dati konačni odgovor o objektivnosti morala.
4. Teorija evolucije može da bude više nego deskriptivna naučna teorija, ona može biti normativna.

U ovoj disertaciji akcenat je stavljen na prva tri od četiri moguća načina biologizacije etike. Najpre, teorija evolucije može da pruži objašnjenje *zašto* ljudi uopšte usvajaju moralne pojmove i donose moralne sudove. Ovaj nivo objašnjenja je strogo biološki i odnosi se na evolucionu prošlost čoveka – na život zavisne društvenosti u zajednicama manjeg obima utemeljenim na srodničkim vezama – i na pitanje kako je u takvim uslovima moglo da dođe do evolucije (selekcije) ljudskih moralnih sposobnosti. Na drugom mestu, teorija evolucije može da posluži kao polazište za reviziju usvojenih moralnih stanovišta, verovanja i principa, tako što nam obezbeđuje nova saznanja o različitim aspektima prirode ljudskih bića. Ovaj tip objašnjenja funkcioniše na dva nivoa, i to tako što kombinuje neurofiziološka objašnjenja s psihološkim

---

<sup>1</sup> Mada moralni realizam nije jedinstvena filozofska pozicija, antirealisti koji se oslanjaju na evolucionu paradigmu smatraju da evolucionu hipotezu o poreklu moralne sposobnosti, načelno podriva moralni realizam ili podriva njegove pojedine varijante (Joyce 2014a). Za različite varijante moralnog realizma i preglede ove filozofske pozicije vidi na primer (Boyd 1997, Railton 1997, Sayre-McCord 1986, Shafer-Landau 2003, Smith 1991).

objašnjenjima kognitivnih i bihevioralnih manifestacija. Na kraju, pretpostavlja se da teorija evolucije može da ima ključnu ulogu u metaetičkim razmatranjima time što obezbeđuje kriterijume za određivanje metafizičkog statusa moralnih svojstava i opravdanosti moralnih verovanja. Objašnjenje koje se odnosi na pružanje načelnog odgovora o objektivnosti morala i opravdanosti moralnih verovanja je strogo filozofsko. Mada je Kičer bio sumnjičav kada je reč o ovoj strategiji, ispostavilo se da je ona filozofski najplodonosnija. Eventualne normativne implikacije teorije evolucije, što je četvrta tačka Kičerovog zapažanja o biologizaciji etike, nisu predmet ovog rada.

Važno je napomenuti i sledeće. Iako su u filozofskim i naučnim teorijama koje pokušavaju da pruže odgovor na pitanje o poreklu morala najzastupljenija adaptacionistička objašnjenja, mogući su i drugačiji eksplanatorni modeli koji se oslanjaju na radove Stivena Dž. Gulda (Stephen J. Gould) i Ričarda Luontina (Richard Lewontin), odnosno na kritiku *adaptacionističkog programa*. Prema mišljenju ovih autora, za adaptacionizam su određujući stavovi da je (1) prirodna selekcija dominantan mehanizam evolucije, i (2) da su sve fenotipske karakteristike adaptacije koje predstavljaju optimalna rešenja za posebne uslove životne sredine (Gould and Lewontin 1979). Nasuprot tome, oni zastupaju stav da neke fenotipske karakteristike nisu adaptacije, već sporedni proizvodi procesa selekcije. Za takav tip karakteristike oni koriste engleski termin *spandrel*, pozajmljen iz arhitekture, koji označava trouglasti umetak između lučnih krakova. Tako, na primer, iako je postojala selekcija za formiranje donje vilice (lat. *Mandibula*) kod čoveka, brada nije adaptacija, već *spandrel* – nusproizvod. Moguće je, dakle, da je moral u smislu sposobnosti za moralno suđenje i ponašanje evoluirao kao nusproizvod drugih gotovo očigledno adaptivnih kognitivnih mehanizama koji obezbeđuju saradnju i ispomaganje. Na drugom mestu, moguće je da je moral *eksaptacija*, tj. osobina koja je proširena u populaciji zahvaljujući tome što je vršila neku funkciju koja se razlikuje od njene aktuelne funkcije (Gould and Vrba 1982). Na primer, osnovna adaptivna funkcija perja kod ptica je termoregulacija, a to što perje omogućava i letenje je eksaptacija ove osnovne funkcije. Moguće je da su kognitivne i bihevioralne funkcije koje su obezbeđivale brigu o potomcima, saradnju i ispomaganje u srodničkom krugu, a koje su opet gotovo očigledno adaptivne, dobile novu funkciju ili je njihova funkcija proširena.



U ovoj disertaciji se zastupa stanovište prema kome je moral, odnosno moralna sposobnost nusproizvod procesa selekcije. Radi jednostavnosti ću često koristiti pojam *moralna sposobnost*, da bih s jedne strane označio određene bihevioralne obrasce koje prosuđujemo kao moralne, i da bih, s druge strane, ukazao na sposobnost usvajanja moralnih verovanja i moralnog suđenja. Pri tome, treba imati na umu da ne mislim da nekakva moralna sposobnost s jedinstvenom funkcijom postoji. Pre je reč o skupu različitih emocionalnih i kognitivnih sposobnosti od kojih su neke svakako adaptacije.

Osnovni cilj ove doktorske disertacije jeste uspostavljanje veze između etike i evolucione biologije, odnosno između ljudskih bihevioralnih obrazaca, koje obično nazivamo moralno ponašanje i evolucione prošlosti čoveka. Mada među etičarima postoji gotovo opšti konsenzus da biologija i evoluciona prošlost čoveka igraju nekakvu ulogu u oblikovanju moralnog rasuđivanja, ova veza se obično smatra zagonetnom, a razlog za to može biti *evoluciona nestabilnost* morala (Bracanović 2007, Singer 2011, Ruse 2006b). Centralno pitanje koje se u ovom radu postavlja je *zašto* imamo moralne pojmove, verovanja i sposobnost moralnog rasuđivanja, kao i zbog čega su međuljudske relacije uređene određenim moralnim pravilima. Odgovor na ovo pitanje konstituiše ono što se u kontekstu teorije evolucije naziva *krajnjim objašnjenjem*, a usmereno je na otkrivanje evolucionog porekla određenog ponašanja ili neke druge fenotipske odlike, kao i na to da li se i kako ono odražava na adaptivnu vrednost organizma. Na drugom mestu, ova disertacija pruža odgovor na pitanje o *neposrednim uzrocima* koji stoje u pozadini usvajanja moralnih verovanja i moralnog rasuđivanja. Tako je veza između morala i biologije, tj. između etike i evolucione teorije, u ovom radu uspostavljena posredno, preko (moralne) psihologije. Preciznije, ova veza je uspostavljena preko jednog psihološkog fenomena, a to je samoobmanjivanje. Iako je ovaj fenomen dugo poznat, on je uglavnom bio predmet istraživanja psihologije i filozofije duha, a tek nedavno je dobio naučnu artikulaciju i u kontekstu teorije evolucije.

Filozofski verovatno najvažnija implikacija ovakvog pristupa je da tumačenje evolucionog porekla morala preko samoobmanjivanja povlači za sobom specifičnu metaetičku poziciju koja je poznata kao *evolucioni antirealizam*. Evoluciona teorija verovatno ne može da nam pruži osnovu za opravdanje moralnih verovanja, ali može da nam posluži kao *univerzalna kiselina* koja „razjeda skoro svaki tradicionalni pojam, ostavljajući većinu starih obeležja

prepoznatljivim, ali temeljno preobraženih“ (Dennett 1996, 63). Dakle, da ponovimo, njena funkcija u kontekstu metaetičkih razmatranja može da bude negativna. U ovoj disertaciji posebna pažnja je posvećena toj funkciji.

Doktorska disertacija će se sastojati iz šest poglavlja, od kojih će svako biti tematski raščlanjeno na odgovarajući broj potpoglavlja. Na osnovu Mejerove (Ernst Mayr) distinkcije između *neposrednih* i *krajnjih uzroka*, u prvom poglavlju će biti napravljena distinkcija između bihevioralnog, psihološkog (neurobiološkog) i biološkog, odnosno evolucionog razumevanja altruizma i sebičnosti (egoizma) (Mayr 1961). Mada je uobičajeno da se ovi fenomeni posmatraju iz dve perspektive, psihološke i biološke, važnost uvođenja ove tročlane podele, koja uključuje jednostavne deskripcije određenih aspekata ponašanja, smatrao sam neophodnom, s obzirom na to da neko ponašanje na bihevioralnom planu može delovati kao altruističko, ali da su njegovi neposredni i krajnji uzroci zapravo sebični. U ovom poglavlju će, takođe, biti istaknuto da direktna međuzavisnost između biološkog i psihološkog altruizma ne postoji, da biološki altruizma nije nužan uslov za pojavljivanje psihološkog, tj. da ne postoji direktna kauzalna veza između opaženog načina ponašanja, intencionalnih stanja koja mogu da budu pokretači takvog ponašanja i evolucionih efekata koje to ponašanje ima. I sebičnost i altruizam uzeti u psihološkom smislu mogu da imaju posledice koje su u biološkom smislu sebične ili altruističke. Kako nije dovoljno ponašati se altruistički da bismo jedno biće kakvo je čovek zvali moralnim, već su važni i neki psihički činoci kao što su motivi, namere i želje, i s obzirom na to da se veza između morala i biologije u ovoj disertaciji uspostavlja preko moralne psihologije, u nastavku prvog poglavlja biće uspostavljena veza između altruizma i morala. U istom kontekstu će biti razmotrene karakteristike moralnih sudova, s posebnim osvrtom na važnost razumevanja zabrana.

Kako su evoluciono-biološke i filozofske rasprave o načinu na koji je biološki i bihevioralni altruizam mogao da evoluirao, ujedno i rasprave o jedinici selekcije na koju mehanizam prirodne selekcije deluje, to će biti predmet drugog poglavlja. Jedinica selekcije je entitet na koji prirodna selekcija deluje, a spor oko jedinica selekcije se odnosi na neslaganja o tome koji je to entitet. Tačnije rečeno, on se odnosi na pitanje iz kog razloga je neka osobina evoluirala, da li zato što je korisna za gen, jedinku ili grupu. Iako je uglavnom prihvaćen stav da

je jedinica selekcije gen, tj. individua (individualni fenotip organizma), ima argumenata u prilog tome da bi jedinica selekcije mogla da bude grupa ili da prirodna selekcija deluje na više nivoa.

Za evolucioniste altruizam predstavlja paradoks, jer prirodna selekcija favorizuje osobine koje uvećavaju šanse jedinke da preživi i da ostavi potomke, a ponašanje koje vodi povećanju adaptivne vrednosti organizma se obično označava kao sebično. S druge strane, altruizam se definiše kao oblik ponašanja kojim jedinka uvećava adaptivnu vrednost druge jedinke na račun svoje sopstvene. Dakle, ako prirodna selekcija deluje protiv altruista, očekivali bismo da u prirodi ne nailazimo na altruistične jedinke. Međutim, na osnovu empirijske evidencije znamo da u prirodi ima altruista. To je ono što se tehnički označava kao *paradoks altruizma*.

U prvo vreme je izgledalo da koncepcija grupne selekcije nudi rešenje za ovaj paradoks. Prema ovoj koncepciji, altruističko ponašanje, kao što je žrtvovanje reproduktivne funkcije kod pčela, evoluiralo je zato što obezbeđuje dobrobit grupe. Ali logika koja stoji iza ovog modela selekcije pokazala se kao sumnjiva, jer su grupe altruista osetljive na tzv. *subverziju iznutra*: ako se u populaciji altruista pojavi samo jedan mutant ili migrant koji se ponaša egoistično, može se očekivati da će egoizam za nekoliko generacija da nadvlada altruizam. Pretpostavka ovakvog rasuđivanja je da je gen jedinica selekcije (to je tzv. genski selekcionizam), a da je dominantno svojstvo gena sebičnost (Dawkins 2006 [1976]). Sebičnost u ovom kontekstu znači sposobnost širenja sopstvenih kopija u genskom fondu. Organizmi se, pri tom, vide samo kao *nosioci* koji služe tome da gene prenesu u narednu generaciju i njihovo ponašanje se tumači polazeći od ovog navodnog biološkog cilja.

Suočeni sa empirijskim podacima koji pokazuju da je altruističko ponašanje rasprostranjeno u prirodi, zastupnici genskog selekcionizma nastoje da pokažu da u specifičnim okolnostima gen najbolje može da postigne svoje sebične ciljeve tako što će na bihevioralnom planu da podstiče izvestan oblik altruizma. Jedan od takvih oblika je altruizam prema srodnicima: altruizam je mogao da evoluiru u *srodničkom krugu* zahvaljujući tome što srodnici poseduju određeni broj istih gena i u tom slučaju evolucija altruizma zavisi od stepena srodstva između jedinki date populacije (Hamilton 1964a, b, 1963). Kako ljudske zajednice uključuju relacije koje prevazilaze srodničke veze, problem altruizma među nesrodnicima može da se reši pozivanjem na koncept reciprociteta. *Recipročni altruizam* ili *direktni reciprocitet* funkcioniše

po mehanizmu razmene usluga, a uspeh strategije *milo za drago* u *Ponovljenoj dilemi zatvorenika* bilo je važno potkrepljenje ove teorije (Trivers 1971, Axelrod 1984). Modeli razvijeni u teoriji igara su se pokazali kao značajni pomoćni instrumenti za objašnjenje važnosti recipročnih relacija i saradnje u prirodnim uslovima. Međutim, povećanje broja individua koje učestvuju u recipročnim interakcijama vodi kolapsu kooperacije (Boyd and Richerson 1988). Takođe, filozofi su u recipročnom altruizmu prepoznali samo formu *sofisticiranog egoizma* ili *prosvećenog ličnog interesa* (Sesardic 1995). U ovom poglavlju će takođe biti razmotren model indirektnog reciprociteta i altruističkog kažnjavanja koji podrazumeva da je individua spremna da kazni drugu individuu zbog kršenja nekog ustanovljenog pravila postupanja, iako joj to nanosi štetu, a ne donosi joj nikakvu direktnu korist (Alexander 2009 [1987]). Važnost altruističkog kažnjavanja za evoluciju morala se ogleda u tome što se kroz činove kažnjavanja ispoljava snažna potreba za retribucijom podstaknuta emocijama, kao i osećaj za pravdu. Prekršilac *zaslužuje* da bude kažnjen bez obzira na to da li onaj ko kažnjava ima bilo kakvu materijalnu korist od toga. Osećaj za pravdu, egalitarizam i težnja ka retribuciji koji dolaze do izražaja u eksperimentima igre ultimatum mogli bi da budu deo biološkog ustrojstva čoveka (Fehr, Bernhard, and Rockenbach 2008, Fehr and Gächter 2002).

Konačno, kada je reč o objašnjenju evolucije altruizma, teorija grupne selekcije je ponovo oživela u radovima Dejvida S. Vilsona (David S. Wilson) i Eliota Sobera (Elliott Sober) koji su predložili da u kontekstu grupe altruisti mogu da imaju manju adaptivnu vrednost od egoista, ali da grupe altruista mogu da imaju veću adaptivnu vrednost nego grupe egoista (Sober and Wilson 1998, Wilson and Sober 1994).

U narednom, trećem poglavlju, biće razmotreni psihološki (emocionalni i kognitivni) i neurobiološki aspekti empatije, kao i njen značaj za evoluciju psihološkog altruizma i morala. Empatija je proizvod prirodne selekcije i pojavljuje se u filogenetskom kontinuumu, što znači da je ova sposobnost prisutna i kod drugih sisara, pre svega primata, tako da se o njoj ne može govoriti kao o specijes specifičnoj karakteristici ljudi (de Waal 1996, 2008, de Waal et al. 2006, Flack and De Waal 2000, Preston and de Waal 2002b). U ovom poglavlju će biti razmotrena eksperimentalna evidencija dobijena iz proučavanja ponašanja dece i primata, čiji bihevioralni obrasci uključuju neke aspekte za koje se tvrdi da predstavljaju osnovne elemente ili *gradivne*

*blokove morala. Osim toga, posebna pažnja će biti posvećena integrativnom evolucionom modelu empatije koji je zasnovan na mehanizmu zajedničkog kodiranja percepcije i akcije.*

Teorija zajedničkog kodiranja percepcije i akcije, na kojoj je evolucioni model empatije utemeljen je funkcionalna i bez predikcija o zajedničkom neurobiološkom supstratu (Van Der Wel, Sebanz, and Knoblich 2013), ali nova istraživanja u oblasti neurobiologije pokazuju da takav supstrat verovatno postoji i da ga čine *neuroni ogledala* (Iacoboni 2009, Rizzolatti and Craighero 2005). Svojstva neurona ogledala i njihova važnost za evoluciju empatije takođe će biti objašnjeni u ovom poglavlju.

Potporna za stanovište da emocije i empatija imaju važnu funkciju u evoluciji morala dolazi iz proučavanja osoba s psihopatskim karakteristikama, koje ne iskušavaju emocionalnu empatiju, krivicu i kajanje, o čemu će biti reči u posebnom, četvrtom poglavlju. Psihopatija predstavlja trajni poremećaj ličnosti koji se, osim antisocijalnim ponašanjem, karakteriše i brojnim emocionalnim deficitima. Osobe s psihopatskim karakteristikama čine posebnu klasu individua koja je prisutna u svim kulturama i čiji članovi nikada nisu iskusili (emocionalnu) empatiju, kao i niz drugih, takozvanih prosocijalnih emocija. Psihopate su potpuno neosetljive na uznemirenost i patnje drugih što ih čini klasom savršenih društvenih predatora. Psihopate mogu da razumeju emocije na konceptualnom (kognitivnom) nivou, mogu čak uspešno da ih oponašaju, ali ne mogu da saosećaju, tj. ne mogu da budu uvučene u emocionalni svet druge osobe. Zato se često kaže da ove osobe „znaju reči, ali ne i muziku“, čime se referira na navedene emocionalne nedostatke, ali i na njihovu sposobnost da ih vešto maskiraju (Blair et al. 2006).

U ovom poglavlju će biti razmotreni neposredni uzroci psihopatije koji se najčešće vezuju za anatomske ili funkcionalne nedostatke u ventromedijalnom prefrontalnom korteksu, orbitofrontalnom korteksu i amigdali (Blair and White 2013, Blair et al. 2006, Hare 1998). Osim toga, biće razmotrena tri evolucionna modela psihopatije. Najpoznatiji evolucioni model psihopatije kombinuje dva evolucionna mehanizma – (1) negativnu selekciju zavisnu od učestalosti gena i (2) evoluciono-razvojni mehanizam uslovne adaptacije (fenotipska plastičnost) (Mealey 1995). Prema ovom modelu, psihopatija nije poremećaj, već poseban varijetet ljudske vrste. Drugačije evoluciono objašnjenje psihopatije se oslanja na koncepciju *fenotipske*

*plastičnosti* i ono se u literaturi obično označava kao *pomeranje kontingenata* ili *uslovna adaptacija* (Buss 2009, Giudice, Ellis, and Shirliff 2011, Glenn, Kurzban, and Raine 2011). I konačno, u trećem evolucionom modelu poreklo psihopatije se objašnjava kao posledica *mutacionog opterećenja*: psihopatija bi bila rezultat kumulativnog delovanja mnoštva mikromutacija koje nepovoljno utiču na razviće i funkcionisanje mozga. U ovom poglavlju će biti objašnjena i razlika između *moralnih* i *konvencionalnih* pravila, kao i to da psihopate, za razliku od četvorogodišnjaka, ne mogu da naprave ovu razliku, niti mogu da razumeju šta jedno moralno pravilo čini *moralnim*.

Predmet petog poglavlja je samoobmanjivanje. U ovom poglavlju će detaljno biti razmotrene filozofske i evolucione interpretacije samoobmanjivanja. Osim toga, biće ukazano na koji način samoobmanjivanje može da ima udela u moralnoj motivaciji. Tumačenje samoobmanjivanja po modelu obmanjivanja je primamljivo, ali takav pristup rezultira paradoksima koji se odnose na (1) sam čin samoobmanjivanja (dinamički ili strateški paradoks – osoba *S* bi morala da obmanjuje sebe znajući da se obmanjuje) i (2) na doksatičko stanje do koga samoobmanjivanje dovodi ako je uspešno (statički ili doksatički paradoks – osoba *S* istovremeno veruje da *p* i da *ne-p*). Filozofske interpretacije samoobmanjivanja obično se dele u dve struje: intencionalističku i neintencionalističku (Talbot 1995, 30 – 31). Intencionalisti su uglavnom divizionisti i pretpostavljaju da bismo u objašnjenju morali da se pozovemo na funkcionalnu ili temporalnu raščlanjenost uma. Neintencionalisti su obično i nedivizionisti i pretpostavljaju da objašnjenje samoobmanjivanja ne mora da uključuje bilo kakvo deljenje uma i da može da se objasni pozivanjem na kauzalne procese. Kada je reč o divizionistima, stupanj mentalnog raščlanjivanja varira od strogog, homunkularističkog pristupa (Rorty), do umerenog (Pears, Davidson), gde su delovi shvaćeni samo kao funkcionalno nezavisni podsistemi. Zastupnici temporalnog raščlanjivanja (Bermudez, Sorensen) smatraju da osoba može intencionalno da se dovede u stanje samoobmane, da ima protivrečna verovanja, ali da ova verovanja moraju da budu vremenski udaljena.

Prema neintencionalistima (Mele, Scott-Kakures), da bi se samoobmanjivanje na adekvatan način objasnilo, neophodno je da uslovi intencionalnosti i dvostrukog verovanja budu oslabljeni ili odbačeni. Prema ovom deflacionom modelu objašnjenja, dovoljno je da osoba koja se samoobmanjuje želi ili da je podstaknuta određenom emocijom da usvoji verovanje koje je

pogrešno. Emocije i želje utiču na formiranje i održanje verovanja, što znači da je jedan broj subjektivih verovanja formiran na epistemički nepouzdan način. Ova objašnjenja se uglavnom oslanjaju na psihološka istraživanja o uticaju motivacije na usvajanje verovanja (Kunda 1990, Trope and Liberman 1996, Trope, Gervy, and Liberman 1997). Prema ovim modelima, samoobmanjivanje je proizvod uobičajenih kognitivnih procesa: naime, uobičajeni kognitivni mehanizmi koji procesiraju informacije vode ne samo istinitim verovanjima, već i usvajanju pogrešnih verovanja.

Kada je reč o evoluciji samoobmanjivanja, biće razmotren osnovni Triversov (Robert Trivers) model koji taj fenomen tumači kao rezultat evolucione trke u naoružanju između obmanjivača i obmanutog (Trivers 2002, Von Hippel and Trivers 2011a). Prema ovom modelu, samoobmanjivanje je ofanzivna strategija čija je funkcija efikasnije obmanjivanje među jedinkama. Ova koncepcija je u neku ruku sintetička i kombinuje različite, često konceptualno i metodološki inkompatibilne pristupe, a autor ne mari preterano za nekonzistentnosti koje iz toga slede. Osim ovog, osnovnog, modela biće razmotrena i alternativna teorija koja se oslanja na neurofiziološka istraživanja, specijalizaciju moždanih hemisfera i kliničko stanje koje je poznato kao *anosognozija* (Ramachandran 1996, Levy 2007). Prema ovom objašnjenju, funkcija samoobmanjivanja je održanje konzistentnosti i koherentnosti sistema verovanja, kao i kontinuiteta ponašanja individue. Oštrica kritike će biti usmerena na *Triversovu teoriju*, a kao osnovu argumentacije ću koristiti saznanja dobijena iz proučavanja psihopatije, tako da će u ovom poglavlju još jednom biti reči i o specifičnostima osoba s psihopatskim karakteristikama.

Tumačenje odnosa etike i biologije preko samoobmanjivanja povlači za sobom specifičnu metaetičku poziciju koja je poznata kao *evolucioni antirealizam*: evoluciono objašnjenje uzroka koji stoje u pozadini naših moralnih verovanja podriva svaki razlog da ta verovanja smatramo epistemički opravdanim. O tome će biti reči u završnom, šestom poglavlju. Najpre će biti razmotreni klasični argumenti za odbacivanje naturalizma u etici: *Hjumov zakon* i *naturalistička greška*. Na drugom mestu, biće razmotrena dva argumenta koja su imala odlučujuću ulogu u raščišćavanju prostora za evoluciono podriivanje moralnog realizma (naturalizma). Reč je o *teoriji greške* Džona Mekija i argumentu eksplanatorne irelevancije moralnih činjenica, koji je poznatiji kao *Harmanov izazov*. Takođe, biće identifikovane i kritički razmotrene dva argumenta ili pre dve grupe argumenata koji imaju antirealističke zaključke: *argument filogenetske*

*kontingencije i argument na osnovu neosetljivosti.* Pojednostavljeno rečeno, prvim od ovih argumenta se tvrdi sledeće. Kako evolucionni proces nije ciljno orijentisan, da je evolucionarna putanja ljudske vrste bila drugačija, moralna pravila bi takođe bila drugačija. Drugi od ovih argumenta se oslanja na epistemološku koncepciju sigurnog praćenja istine i princip osetljivosti, kao i na teoriju pouzdanosti procesa. Ovi argumenti tvrde, otprilike, sledeće: ako su kognitivne i moralne sposobnosti proizvod prirodne selekcije ili nekog drugog evolucionog procesa i ako postoje moralne činjenice, onda ove kognitivne i moralne sposobnosti nisu evoluirale da prate moralne istine i to je razlog da se mogućnost moralnog znanja dovede u sumnju. Na kraju, pokazaću kakvu funkciju koncepcija samoobmanjivanja može da ima u ovim argumentima.



# 1 EGOIZAM, ALTRUIZAM I MORAL

## 1.1 Egoizam i altruizam: biološka i psihološka interpretacija

Termin altruizam je u teorijsku biologiju uveo Holdejn (Haldane), ali ga nije definisao, a Vilson (E. O. Wilson) ga je proglasio centralnim problemom u sociobiološkim razmatranjima (Haldane 1932, Wilson 1980). Rasprave o altruizmu i sebičnosti u biologiji i filozofiji biologije često su bile obeležene pojmovnom i terminološkom zbrkom (Sober 1988, Wilson 1992, Rosenberg 1992). Odsustvo razlikovanja psihološkog i evolucionog značenja ovih pojmova uzrokovalo je da nivoi objašnjenja ne budu jasno razgraničeni. Za diferencijaciju različitih nivoa objašnjenja neophodno je na umu imati distinkciju između krajnje i neposredne uzročnosti (Mayr 1961, 1998). *Neposredni uzroci* mogu biti fiziološki, psihološki ili sredinski.<sup>2</sup> Kauzalno objašnjenje u ovom slučaju počinje pitanjem „Kako?“. Uzroci koji se odnose gotovo isključivo na genetičke dispozicije organizma i njegovu evolucionu prošlost nazivaju se *krajnjim uzrocima*. Kauzalno objašnjenje u ovom slučaju počinje pitanjem „Zašto?“, a odgovor se dobija istorijskom rekonstrukcijom evolucione prošlosti organizma. Svako potpuno objašnjenje ponašanja organizma, trebalo bi da obuhvata oba skupa uzroka. Iako, distinkcija između neposrednih i krajnjih uzroka nije dovoljno precizna, jer, na primer, strogo razgraničava evolucione i razvojne činioce, ona je važna i biće nam od značaja u daljoj analizi.

U kontekstu evolucione biologije altruizam i sebičnost se uglavnom definišu preko efekata koje takvi načini ponašanja imaju po adaptivnu vrednost organizma. Za sebičnost se često pretpostavlja da ima povoljne, dok se altruizam vezuje za nepovoljne posledice po preživljavanje i reprodukciju i opaža se kao vrsta autodestruktivnog ponašanja (Rosenberg 1992). Tako je altruizam obično definisan kao onaj način ponašanja kojim jedinka uvećava adaptivnu vrednost druge jedinke na račun svoje sopstvene. S druge strane, sebičnost je onaj način ponašanja kojim jedinka obezbeđuje maksimum sopstvenog preživljavanja i reprodukcije.

---

<sup>2</sup> Neposredni uzroci ponašanja onih organizama koji nemaju nervni sistem su fiziološki ili sredinski, što ne isključuje mogućnost da neposredni uzroci ponašanja jedinke sa sofisticiranim nervnim sistemom budu fiziološki.

Nije teško zapaziti da se u prethodnim definicijama ne koriste nikakvi termini koji bi referirali na intencionalna stanja jedinke, budući da jedinka, da bi se ponašala altruistički (ili sebično) u biološkom smislu, uopšte ne mora da ima intencionalna stanja. Ona ne mora da ima verovanja, želje i motive koji su usmereni na njenu ili na dobrobiti druge jedinke. I ne samo to, ona uopšte ne mora da ima nervni sistem. Čak ne mora da bude višećelijski ili jednoćelijski organizam; virusi mogu da se ponašaju altruistički (Sober and Wilson 1998), a oni su, kako šaljivo primećuje Danijel Denet (Daniel C. Dennett) samo „lanac DNK (ili RNK) sa stavom“ (Dennett 2013). Stoga je odmah jasno da evolucionim altruizam ne treba izjednačavati s altruizmom kako se on tumači u svakodnevnom ili psihološkom kontekstu. Psihološki altruizam nužno uključuje određena intencionalna stanja, želje i motive kao neposredne uzroke ponašanja, koji su nedvosmisleno i na neinstrumentalan način usmereni ka dobrobiti druge jedinke (Sober 1988). Kada je reč o prisustvu intencionalnih stanja, isto važi i za psihološku sebičnost, koju ne treba mešati s *psihološkim egoizmom*. Grubo govoreći, jedinka je sebična u psihološkom smislu ako ima želje i motive koji su orijentisani na sopstvenu dobrobit. S druge strane, psihološki egoizam je učenje o (ljudskoj) motivaciji prema kome *svi* motivi koje jedinka ima mogu biti redukovani na motive orijentisane na sopstvenu dobrobit.<sup>3</sup>

U svakodnevnom i psihološkom kontekstu obično smatramo da altruizam i sebičnost imaju neke veze s interesima individue koji nisu nužno reproduktivni (Rosenberg 1992, Sober 1988).<sup>4</sup> Tako, kada govorimo o sebičnosti, ili pre o egoizmu u ovom kontekstu (jer i jedinka koja nema um može da bude sebična, ali samo ona koja ima sofisticirane kognitivne kapacitete može da bude egoista), pod njim obično podrazumevamo da individua 1) u svim ili u supstancijalnoj većini situacija ima želje i motive koji se isključivo odnose na zadovoljenje sopstvenih interesa, 2) da u svim ili u supstancijalnoj većini situacija nastoji da postupa tako da njeni postupci vode isključivo ostvarenju ili maksimaciji njenih ličnih interesa, i/ili da 3) u svim ili u supstancijalnoj

---

<sup>3</sup> U kontekstu ove teorije, ako osoba propusti da učini nešto što vodi maksimaciji njenih interesa, takav postupak se tumači kao *akratički*. Jedan postupak je akratički ako osoba *S* propušta da učini *x* u onim situacijama kada, 1) zna da činjenje *x* vodi maksimaciji njenih interesa, 2) želi da učini *x* i 3) nema spoljašnjih prepreka da učini *x*. Iako je individua znala šta doprinosi maksimaciji njenih interesa i uprkos tome što je želela maksimum koristi za sebe, ovo je propustila da učini jer nije imala dovoljno jaku volju da ostvari svoje želje.

<sup>4</sup> Kao što ćemo videti kasnije ima autora, kao što je recimo Ričard Aleksander, koji smatraju da se svi interesi u krajnjoj konsekvenci mogu redukovati na reproduktivne.

većini situacija zahteva od drugih individua da njihovi postupci isključivo vode ostvarenju ili maksimaciji ličnih interesa individue koja taj zahtev postavlja, čime se egoizam transformiše u ličnu diktaturu.

Na sličan način možemo da definišemo i altruizam. Pod altruizmom obično podrazumevamo da individua 1) u svim ili supstancijalnoj većini situacija ima želje i motive koji su orijentisani na zadovoljenje interesa neke druge osobe ili osoba, 2) da u svim ili u supstancijalnoj većini situacija nastoji da postupa tako da njeni postupci vode isključivo ostvarenju ili maksimaciji interesa neke druge osobe ili osoba, i/ili 3) da u svim ili u supstancijalnoj većini situacija zahteva (smatra ispravnim ili racionalnim) da sve individue imaju želje i motive koji su orijentisani na dobrobit druge osobe ili osoba i da nastoje da postupaju tako da njihovi postupci isključivo vode ostvarenju ili maksimaciji ličnih interesa neke druge osobe ili osoba.

Za razliku od autora koji psihološki altruizam i egoizam (sebičnost) smatraju bihevioralnim dispozicijama, a ne aktuelnim načinima ponašanja i definišu ih gotovo isključivo na osnovu intencionalnih stanja (motiva i želja), a nezavisno od posledica koje proizvode (Sober 1988, 1994), u definicijama koje predlažem aktuelno ponašanje i njegove posledice su važni. Sober pravi razliku između umerenog i ekstremnog egoizma, s jedne strane, i umerenog i ekstremnog altruizma, s druge strane, a na osnovu toga kakav preferencijlni status ova četiri moguća stanovišta daju ličnom interesu. Prema njegovom mišljenju, egoizam i altruizam kao bihevioralne dispozicije znače da altruisti imaju dispoziciju da žrtvuju sopstvene interese kada se ovi sukobljavaju s interesima drugih, ali je sasvim moguće da vode računa o sebi i sopstvenim interesima, dok egoisti imaju takve dispozicije da u slučajevima sukoba interesa daju prednost ličnom interesu, ali i oni mogu da vode računa o drugima i njihovim interesima. Izuzetak su samo *ekstremni egoisti* koji druge posmatraju isključivo kao sredstva za zadovoljenje vlastitih ciljeva. Soberu to omogućava da egoizam i altruizam u pojedinim slučajevima, kao što je roditeljska briga, posmatra kao funkcionalno ekvivalentne. Na primer, roditelj koji nahrani dete isključivo zato što ga iritira dečji plač je u psihološkom smislu egoista, dok je onaj roditelj koji nahrani dete jer mu je zaista stalo do dobrobiti deteta u psihološkom smislu altruista, ali u bihevioralnom smislu oni ne mogu da se razlikuju. U bihevioralnom smislu oba postupka su altruistička. Roditeljska briga je primer evoluciono uspešnog ponašanja, koje je u bihevioralnom

smislu altruističko, dok je u biološkom uvek sebično. Međutim, to ne implicira da je takvo ponašanje nužno uzrokovano neposrednim mehanizmima koji su sebični, jer može biti uzrokovano i mehanizmima koji su autentično altruistički (Sober 1988, 1994, Wilson 1992). Pojemovni parovi evoluciono uspešno/neuspešno, kao i biološki sebično/altruističko ne odgovaraju pojmovnom paru psihološki sebično/altruističko.

Stavljanje naglaska na to kako se individua zaista ponaša i kakve posledice iz toga slede važno je i iz sledećeg razloga. Ako osoba ima altruističke želje i motive, ali se oni retko ili nikada ne realizuju na adekvatan način, tako da nemaju altruističke bihevioralne posledice, takvo ponašanje nije altruističko bez obzira na to kakvi ga motivi uzrokuju, a osoba koja se ponaša na taj način nije altruista, uprkos svojim željama i motivima. Ali, ako su intencionalna stanja jedino s čim se računa kada se neko karakteriše kao altruista, onda izgleda da bi prema Soberovom mišljenju ovakva osoba ipak bila altruista bez obzira na štetu koju možda čini. Zbog toga mislim da definicije koje aktuelno ponašanje stavljaju u drugi plan nisu najadekvatnije. Ukoliko neko nekoga udavi u najboljoj nameri da ga nauči da pliva, da li je on altruista? Mislim da nije. Iako njegovi postupci možda nisu sebični, oni su za recipijenta štetni. Na taj način altruistički motivi mogu da uzrokuju postupke koji su zaista altruistički, ali i one koji su štetni, iako možda nisu sebični, jer ni na koji način ne maksimizuju interese individue. Takođe, sebični ili egoistički motivi mogu za posledicu da imaju sebične, ali i altruističke postupke. (Ako neko nekoga udavi učeći ga da pliva, takav postupak može da bude protumačen kao sebičan u biološkom smislu, jer jedinka može da poveća svoju adaptivnu vrednost ne samo tako što će činiti nešto što joj donosi direktnu korist već i tako što, nezavisno od motiva, drugima nanosi štetu ili se uzdržava da čini nešto što drugim jedinkama koristi.)

Iako je uobičajeno da se altruizam i sebičnost posmatraju iz dve perspektive, psihološke i biološke, ove pojmove možemo razložiti na tri komponente, bihevioralnu, psihološku i biološku.<sup>5</sup> Bihevioralna komponenta bi se odnosila na ono što opažamo kada posmatramo ponašanje određene jedinice. Imajući na umu dihotomiju između neposredne i krajnje uzročnosti, psihološka komponenta bi se odnosila na intencionalna stanja, kao što su želje i motivi, koji su neposredni uzroci ponašanja. Ali, u jednom širem smislu, odnosila bi se i na fiziološke ili

---

<sup>5</sup> Sličnu klasifikaciju predlaže Džojcs (Joyce 2006).

neurofiziološke mehanizme koji mogu da determinišu ponašanje. Evoluciono-biološka komponenta odnosi se na efekte koje ponašanje ima na adaptivnu vrednost organizma, odnosno, ticala bi se krajnjih uzroka ponašanja. Dok je bihejvioralna komponenta jednostavna deskripcija načina na koji se jedinka ponaša, psihološka i biološka konstituišu objašnjenje bihejvioralnih obrazaca kompleksnog biološkog organizma s nervnim sistemom (ili bez njega). Ako prihvatimo ovu klasifikaciju, onda ponašanje može da bude:

1. U bihejvioralnom smislu altruističko, u psihološkom altruističko i biološki altruističko.
2. U bihejvioralnom smislu altruističko, u psihološkom altruističko i biološkom sebično.
3. U bihejvioralnom smislu altruističko, u psihološkom sebično i biološkom altruističko.
4. U bihejvioralnom smislu altruističko, u psihološkom sebično i biološkom sebično.
5. U bihejvioralnom smislu sebično, u psihološkom altruističko i biološki altruističko.
6. U bihejvioralnom smislu sebično, u psihološkom altruističko i biološki sebično.
7. U bihejvioralnom smislu sebično, u psihološkom sebično i biološki altruističko.
8. U bihejvioralnom smislu sebično, u psihološkom sebično i biološki sebično.

Radi jednostavnosti i preglednosti to može da se predstavi u tabeli:

*Tabela 1*

	Bihejvioralno	Psihološki	Biološki
1.	A	A	A
2.	A	A	S
3.	A	S	A
4.	A	S	S
5.	S	A	A
6.	S	A	S
7.	S	S	A
8.	S	S	S

Iz ove sheme ne sledi da postoji međuzavisnost između psihološkog i biološkog altruizma, da je postojanje biološkog altruizma nužan uslov za pojavljivanje psihološkog, tj. da postoji direktna kauzalna veza između opaženog načina ponašanja, intencionalnih stanja koja mogu da budu pokretači takvog ponašanja i evolucionih efekata koje to ponašanje ima. To znači da biološki i psihološki altruizam mogu da se pojavljuju nezavisno jedan od drugog. Na primer,

ponašanje pčela radilica je u bihejvioralnom smislu altruističko. Psihološki smisao je u ovom slučaju prazan skup, dok je u biološkom smislu njihovo ponašanje sebično. Sebičnost u psihološkom smislu može da ima posledice koje su u biološkom smislu sebične ili altruističke, takođe altruizam u psihološkom smislu može da ima posledice koje su u biološkom smislu sebične ili altruističke.

U narednim odeljcima usredsrediću se uglavnom na biološke aspekte sebičnosti (egoizma) i altruizma i njihovu evoluciju. Ali pre nego što se usredsredim na to, želeo bih da kažem nekoliko reči o vezi između altruizma i morala, nekim karakteristikama moralnih sudova i tome šta je to što je neophodno da bismo jedan biološki entitet kakav je čovek nazivali moralnim.

## **1.2 Altruizam, moral i karakteristike moralnih sudova**

Evolucioni status morala je u najmanju ruku neobičan. Evolucione teorije morala, koje moral tumače u adaptacionističkom ključu, insistiraju na tome da je on biološka adaptacija čija je osnovna funkcija obezbeđivanje efikasne saradnje, koja dugoročno posmatrano, može da ima značajne evolucione prednosti za jedinku u odnosu na nesaradnju (Ruse 1995a, 2002, 2006a, Joyce 2006). To znači da moral u funkciji saradnje doprinosi tome da one jedinke koje svoje ponašanje usklađuju s određenim pravilima (normativnim standardima) i koje su sposobne da ta pravila internalizuju prolaze bolje, tj. imaju višu adaptivnu vrednost, u odnosu na jedinke čiji bihejvioralni obrasci odstupaju od takvih pravila.

Da bismo postigli „altruizam“, mi smo altruistični! Da bi učinila da saradujemo zbog bioloških ciljeva, evolucija nas je ispunila mislima o ispravnom i pogrešnom, potrebom da pomognemo bližnjima, i tako dalje. Očigledno je da nismo potpuno nesebični. Uistinu, zahvaljujući nadmetanju za reprodukciju, naša uobičajena dispozicija je da vodimo računa o sebi. Ipak, u našem biološkom interesu je da saradujemo. Na taj način, imamo evoluirane urođene mentalne dispozicije [...] koje nas podstiču da saradujemo, u ime ove stvari koju nazivamo moral. Naš moralni osećaj je adaptacija, ne manje nego što su to ruke i oči, zubi i noge (Ruse 1995a, 230).

Rus (Michael Ruse) ne misli da ovo znači prihvatanje slepog determinizma. Za razliku od mrava ili robota koji nemaju izbora da li će se ponašati na ovaj ili onaj način, već im ponašanjem upravlja predefinisani algoritam, bilo da je reč o instinktu ili softveru, ljudi su samo podstaknuti

da se ponašaju moralno, ali ne i predodređeni da se tako ponašaju. Moral ili pre moralni osećaj koji je permanentno ugrađen u ljude je ekonomičan način za postizanje biološkog altruizma i saradnje (Ruse 2002). „Da bi nas učinila „altruistima“ u metaforičkom biološkom smislu, biologija nas je učinila altruistima u doslovnom, moralnom smislu“ (Ruse 2006a, 238). Kako prirodna selekcija retko ispostavlja optimalna rešenja za adaptivne probleme, ni moralni osećaj niti moralna pravila nisu takva vrsta rešenja; ona su ekonomična, fleksibilna i nedovoljno precizna, ali sasvim dovoljna za funkcionisanje individue u svakodnevnom socijalnom kontekstu.

Na sličan načina razmišlja i Džojls, koji moral, takođe, smatra rezultatom delovanja prirodne selekcije čija je funkcija da obezbedi uspeh društvene saradnje, koja bi bila ugrožena kada bi zavisila od individualnih želja i stavova.

Pretpostavite da postoji carstvo postupaka od takve važnosti da priroda nije želela da praktični uspeh zavisi od krhke čudi obične ljudske praktične inteligencije. To carstvo može, na primer, da se odnosi na određene oblike kooperativnog ponašanja... Dobrobiti kooperacije – na primer, poboljšanje reputacije – su po pravilu dugoročne vrednosti, i samo biti svestan tih vrednosti i želeći ove dugoročne prednosti ne garantuje da će cilj biti postignut, kao što ni želja za dugovečnošću ne garantuje da će se osoba odreći masne hrane. Hipoteza je onda da je prirodna selekcija optirala za poseban motivacioni mehanizam za ovo carstvo: moralnu savest (Joyce 2006, 111).

Dok je za jednu grupu autora moral adaptacija, za neke druge je nusproizvod procesa selekcije – *spandrel* (Gould and Lewontin 1979). Kao što je već rečeno u uvodnom delu, engleski termin *spandrel* pozajmljen je iz arhitekture i označava trouglasti umetak između lučnih krakova. Na primer, iako je postojala selekcija za formiranje donje vilice (lat. *Mandibula*) kod čoveka, brada nije adaptacija, već *spandrel* – nusproizvod. Slično važi i za crvenu boju krvi. Prema evolucionim teorijama koje moral tumače kao sporedni proizvod prirodne selekcije, on je emergentno svojstvo drugih neruobioloških, emocionalnih i kognitivnih mehanizama od kojih su bar neki očigledno adaptivni (Prinz 2009).

Ovoj poziciji se suprotstavlja nemogućnost empirijskog razlikovanja adaptacija i nusproizvoda na osnovu teškoće da se tačno pronikne u to koji su evolucionari pritisci delovali na jedinke u predačkim populacijama (Joyce 2014b). Džojsov argument je zanimljiv, ali je analogija

koju pravi prilično nategnuta. Naime on tvrdi da prirodnu selekciju zaista možemo da uporedimo s arhitektom, ali da nije nezamislivo da njegova preokupacija ne bude projektovanje funkcionalnih nosećih stubova s lučnim krakovima, već trouglastih umetaka. U tom slučaju finalna građevina se ne bi razlikovala od zdanja koje bi nastalo da je njegov prevashodni interes bio pravljenje stubova. Arhitektura i prirodna selekcija zaista imaju jednu zajedničku osobinu, a to je da ispostavljaju funkcionalna rešenja pod ograničenjima koja nameće arhitektonski ambijent, odnosno uslovi sredine. Arhitekta koji bi u kompleksnom urbanom okruženju bio fokusiran na projektovanje trouglastih umetaka umesto funkcionalnih nosećih stubova bi bio veoma neobičan tip arhitekta i verovatno veoma neuspešan, a slično može da se kaže i za prirodnu selekciju. Prirodna selekcija koja bi bila „zainteresovana“ isključivo za boje i oblike, a ne za funkcionalne organske strukture, koje mogu da odgovore na izazove nepovoljnih uslova životne sredine, bila bi neobična i verovatno neuspešna evolucionarna sila.

Jedan drugačiji pristup evoluciji morala polazi od pretpostavke da je prirodna selekcija jedinke u predačkim populacijama najpre učinila ograničenim altruistima, da bi ih zatim opremila kapacitetom za normativno upravljanje svojim postupcima (Kitcher 2006b). Takve jedinke, iako sopstvene interese uglavnom stavljaju ispred interesa drugih, ponekada mogu da vode računa i o potrebama i interesima drugih individua, što ih osposobljava za život u zajednicama malog obima. Neke od ovih individua su pored ograničenog altruizma usvojile i kapacitet za normativno upravljanje svojim ponašanjem i bile su sposobne da ponašanje i stavove oblikuju i prilagođavaju prema određenim društvenim pravilima, koja su zatim transformisana u proto-moral – skup pravila ili zapovesti čija se funkcija iscrpljuje u regulisanju načina na koji se prave koalicije, definišu odnosi lojalnosti i sistem osвете. Prema ovom modelu, moral je *invencija*. Ova elementarna pravila zatim dolaze pod uticaj kulturne grupne selekcije, koja deluje pre po lamarkističkim nego po darvinističkim standardima.<sup>6</sup>

U ovom radu se zastupa pozicija prema kojoj moral nije adaptacija, već je sporedni proizvod neruobioloških, emocionalnih i kognitivnih mehanizama, kao što je empatija, od kojih su bar neki očigledno adaptivni. Ovi mehanizmi su poduprti sposobnošću individue da se samoobmanjuje, tj. da usvaja i održava (pogrešna) verovanja pod uticajem različitih emocija.

---

<sup>6</sup> Detaljnije o evoluciji kulture u (Richerson and Boyd 2006).



Pokazaće se da samoobmanjivanje ima važnu ulogu u objektivaciji morala. Filozofski pokušaji da se moral zasnuje na potpunoj racionalnosti su verovatno bili pogrešni, jer kako stvari stoje moral i racionalnost su u zavadi. Temeljiti normativnu etiku na onome što bi po pretpostavci racionalne, autonomne i nepristrasne osobe prihvatile je uzaludno, jer takve osobe nikada nisu postojale. Nije teško pokazati da bi potpuno racionalni subjekt zapravo bio krajnje nemoralan. Poznati računar HAL-9000 iz romana Artura Klarka (Arthur C. Clarke) *Odiseja u svemiru 2001* prolazi *Tjuringov test* i potpuno je racionalan, ali bismo teško mogli da kažemo da je moralan, s obzirom na njegove postupke. Moguće je da je HAL zapravo savršeni elektronski psihopata, a psihopate, na šta ću u radu da ukažem, uz ostale emocionalne deficite, ne mogu da se samoobmanjuju.<sup>7</sup>

Kao što je evolucionini status morala neobičan, odnos egoizma i altruizma prema moralu je složen. Mada iz egoističkih motiva bar ponekada slede altruistički postupci, obično se smatra da je psihološki altruizam nužan uslov morala, a da je pozicija egoizma donekle paradoksalna i teško odbranjiva (Williams 1976). To što iz egoističkih motiva bar ponekad slede altruistički postupci te motive ne čini ništa manje egoističnim, kao što te postupke ne čini ništa manje altruističkim, ali teško možemo da govorimo o altruističkim motivima ukoliko učinjeni postupci koji su posledica njihovog delovanja nisu takođe altruistički. Međutim, može se dogoditi da motivi individue budu altruistički, ali da postupci i posledice postupaka ne budu takvi. Ti postupci ne moraju da budu egoistički, ali mogu da budu skopčani s ponašanjem koje ne smatramo moralnim. Altruistički motivi, iako nužni, nisu dovoljni za moral.

Takođe, samo altruističko ponašanje, ma koji da su njegovi uzroci, nije dovoljno da bismo jedno biće nazivali moralnim. Slepí miševi vampiri, koji predaju deo sadržaja svog želuca drugoj jedinki koja nije uspela da dođe do obroka ponašaju se altruistički, ali njihovo ponašanje nije moralno, i oni nisu moralna bića. Takođe, poznato je da neke vrste ptica, foka i majmuna, uključujući i šimpanze, ponekada usvajaju mlade koji su bez roditelja. Takvo ponašanje može da se protumači kao altruističko, ali ni ono nije moralno.

---

<sup>7</sup> Hal ne može da se samoobmanjuje jer mu nedostaju odgovarajući emocionalni kapaciteti neophodni za samoobmanjivanje.

Takođe, prepoznavanje, razumevanje ili posedovanje osećaja za određene društvene regularnosti i njihovo poštovanje nije dovoljno da bismo govorili o moralnim bićima. Na primer, šimpanze mogu da prepoznaju određene društvene pravilnosti i da ih poštuju, tj. da poštuju ono što je prihvaćeno ili neprihvaćeno u datom socijalnom, hijerarhijski uređenom kontekstu. Oni mogu da razumeju da određeni bihevioralni obrasci nisu dopušteni i da ih alfa mužjak sankcioniše. Međutim, sposobnost razumevanja onoga što je prihvaćeno ili neprihvaćeno, dopustivo ili nedopustivo je prilično udaljeno od moralne sposobnosti prosuđivanja nečega kao prihvatljivog ili neprihvatljivog, dopustivog ili nedopustivog.

Jedan od najvažnijih aspekata morala, shvaćenog ne samo kao specifičnog oblika ponašanja, već i kao sposobnosti za donošenje moralnih sudova, jeste sposobnost razumevanja zabrana (Joyce 2006). U evolucionoj biologiji zamisao da zabrane i njihovo razumevanje imaju važnu funkciju u socijalnom i moralnom kontekstu nije nova. Vilijams ističe važnost razumevanja zabrana za adaptivnu vrednost jedinke i evoluciju kognitivnih sposobnosti (Williams 1996 [1966]).<sup>8</sup> Uzmimo, donekle modifikovan, primer koji koristi Vilijams. Pretpostavimo da u nekoj predačkoj populaciji dečake Freda i Barnija roditelji svakodnevno upozoravaju da ne prilaze reci, da se ne igraju vatrom i ne ulaze u pećinu sabljozubog tigra. I dok Fred s lakoćom razume zabrane koje mu roditelji postavljaju, ne izlaže se opasnostima i izbegava roditeljsko kažnjavanje, Barni je drugačiji. Barni ima teškoća u razumevanju zabrana, a ono što je zabranjeno čini mu se još privlačnijim. Budući da živi u neprijateljskom okruženju, iskustva koja ima nisu uvek prijatna. Zbog nerazumevanja zabrana često se izlaže opasnostima i kaznama. Ali, jednog dana, nakon što je posetio jazbinu sabljozubog tigra i sam je izbegao da bude kažnjen, ali iz razloga koji su sasvim drugačiji od onih koji se odnose na razumevanje zabrana. Dakle, sposobnost razumevanja zabrana pruža značajne adaptivne prednosti.

Moralna pravila formulisana u obliku zabrana imaju značajnih prednosti u odnosu na pozitivno formulisane zapovesti (Singer 2011). Najpre, zabrane je lako poštovati i kompatibilne su s uobičajenim tokom ljudskog života. Zatim, pravila u formi zabrana je lako naučiti, a kada su jednom naučena omogućavaju da se manje-više lako donese ispravna odluka u situacijama kada

---

<sup>8</sup> Vilijams govori o verbalnim instrukcijama, ali primeri verbalnih instrukcija koje daje su svi formulisani u obliku zabrana. Vidi (Williams 1996 [1966], 15).

je rasuđivanje otežano ili kada postoje vremenska ograničenja koja zahtevaju da se odluka donese brzo. Moralna pravila u takvom kontekstu funkcionišu kao „zatvarači razgovora“ koji onemogućavaju da se proces preispitivanja, opravdanja i odlučivanja odvija u beskonačnost (Dennett 1996). Konačno, postojanje zabrana dosta olakšava normalnu komunikaciju. Ukoliko znamo da postoji pravilo koje zabranjuje laganje i ako iz iskustva znamo da se supstancijalna većina osoba u supstancijalnoj većini situacija pridržava ove zabrane, onda se time konstituiše osnova za poverenje i u dovoljnoj meri gubi osnova za sumnju u ono što nam se saopštava. Ljudi načelno nisu sumnjičavi i lakovernost je uobičajena epistemološka pozicija u svakodnevnom saobraćanju.

Poštovanje zabrana se razlikuje od uzdržavanja ili izbegavanja da se nešto čini samo zato što to izaziva averzivnu emocionalnu reakciju kod individue. Ako se individua uzdržava da čini  $X$ , ili ne želi da čini  $X$ , jer  $X$  u njoj proizvodi nelagodu, to još ne znači da se ona ponaša moralno ili da donosi moralni sud o  $X$ . Zabrane se razlikuju od inhibicija, mada je moguće da su neke prirodne inhibicije s pojavom jezika transformisane u zabrane, kao što je to verovatno slučaj s izbegavanjem stupanja u seksualne odnose s bliskim osobama i srodnicima (Sesardic 2005). Takođe, moguće je da je moralnoj zabrani ubistva i povređivanja drugih, evoluciono prethodila načelna odbojnost prema nanošenju telesnih povreda iz neposredne blizine (vidi na primer Cushman, Young, and Hauser 2006, Greene et al. 2009, Cushman et al. 2012). Mada se sadržaj nekih prirodnih inhibicija podudara s konceptualnim sadržajem zabrana, inhibicije se razlikuju u odnosu na zabrane, bar u tome što se *zahtev* za poštovanjem zabrana ne obazire na individualnu emocionalnu reakciju koju osoba ima u susretu s nekim stimulusom. Time se ne tvrdi da su emocije irelevantne za donošenje moralnih sudova, naprotiv. Poznato je da ništa ne pobuđuje toliko emocija kao rasprava o nekoj moralno relevantnoj temi. Samo psihopate i filozofi mogu o ubistvu, genocidu, ropstvu, incestu ili ubijanju novorođenih beba da razgovaraju hladnokrvno i da pružaju argumente koji opravdavaju ove postupke, jednako kao i one koji ih osuđuju (Haidt and Bjorklund 2008). Da se *zahtev* za poštovanjem zabrana ne obazire na individualnu emocionalnu reakciju osoba znači samo to da, na primer, moralna zabrana ubistva važi i za one individue kojima ubistvo ili nanošenje povreda nije averzivno i koje žele da nekoga povrede ili ubiju. Slično važi i za laganje, kršenje obećanja, itd. S druge strane, čak i ako bi individua činila  $Y$ , koje inače može biti vredno moralnog uvažavanja, samo zato što je empatična, ima altruističke

motive, želje, itd., ni tada joj ne bismo pripisivali moralne karakteristike i sposobnost moralnog suđenja (Joyce 2006). Čak i kada bismo zamislili klasu individua čiji su članovi potpuno benevolentni, empatični i altruistički, oni ne bi bili moralni. Istini za volju, takvoj klasi individua moral možda ne bi bio ni potreban. Osim toga broj zabrana se ne podudara s brojem inhibicija. Na primer, iako ne postoji prirodna inhibicija za stupanje u seksualne odnose izvan braka, ipak postoji moralna zabrana preljube.

Emocije su važne, ali se sposobnost za moralno suđenje i značenje moralnih sudova ne iscrpljuje u individualnim emocionalnim reakcijama. Dakle, moralne zabrane ne zavise od bilo čijih želja, osećanja ili subjektivnih stavova. Tvrditi nešto kao što je „Ne ubij, ubistvo je pogrešno“, znači nešto više od trivijalnog „Buuu, ubistvo!“, „Ubijanje mi se ne sviđa“, „Ne želim ubistvo“ ili „Ne odobravam ubistvo“. Prosuđivati nešto kao moralno pogrešno ili ispravno nije puko izražavanje emocija, stavova ukusa, ili bilo koje druge vrste ličnih stavova. Time ne sugerišem da je neophodno napuštanje nekognitivizma. Zapravo, emocije su uključene u donošenje moralnih sudova, pogotovo onih deontoloških i konstituenti su sposobnosti za moralno suđenje (Greene 2007). Emocije su takođe konstituenti semantičkog statusa moralnih sudova. Takvi sudovi najčešće imaju emocionalnu dimenziju značenja, između ostalog i zbog toga što određene reči koje se u njima pojavljuju kod normalnih (emocionalno kompetentnih) individua izazivaju različite emocionalne reakcije. Reči kao što su ubistvo, silovanje, krađa, preljuba, bol, itd. ne pobuđuju istu emocionalnu i neurofiziološku reakciju kao reči sto, stolica, šolja ili olovka. To nije ništa novo. Još je Stivenson (Charles L. Stevenson) pisao da određeni etički termini imaju emotivno značenje koje predstavlja „tendenciju reči, nastalu kroz istoriju njene upotrebe, da proizvede afektivnu reakciju u ljudima. To je neposredna aura osećanja koja lebdi oko reči“ (Stevenson 1937, 23).

Nakon suda „ubistvo je pogrešno“, kompetentan govornik koji razume sve termine upotrebljene u sudu ne može da izgovori nešto kao „ubistvo je pogrešno, ali...“. Da parafraziram Mura, ubistvo je pogrešno i tu je kraj te stvari. Za razliku od Hobsovih (Thomas Hobbes) prirodnih zakona koji obavezuju samo na izvesnu želju da budu ostvareni, moralna pravila, imperativi i zabrane zapovedaju kategorički, mada kategoričnost o kojoj je reč nije apsolutna i toliko stroga da ne trpi nikakve izuzetke. Danas, ni oni filozofi koji rade u okviru Kantove deontološke paradigme ne smatraju da je zabrana laganja takve prirode da bi bila obavezujuća u

svim okolnostima (Korsgaard 1986). Hipotetički imperativi, bar onako kako ih je Kant (Immanuel Kant) definisao, tiču se odnosa sredstava za ostvarenje ciljeva koji su definisani različitim željama, dok se kategorički imperativ tumači kao objektivno važeći nezavisno od želja i postavljenih ciljeva (Kant 1979). Kategoričnost moralnih zapovesti nije Kantova invencija, već samo precizna filozofska artikulacija posebnog mentalnog fenomena s određenim neurobiološkim supstratom (Joyce 2001). Može se reći da ljudi brinu o sopstvenom potomstvu jer to *žele* da čine. Ali, budući da ova želja može da bude suspendovana, ona nije dovoljna da bi se briga (u biološkom smislu) obezbedila. Zato im je pored te želje, kao njen pojačivač, potrebno snažno izraženo osećanje da je to nešto što *moraju* da čine. Ljudi jednostavno osećaju da je voditi računa o potomstvu nešto što *treba* da čine.

Polazeći od uobičajene pretpostavke da svi imperativi mogu da se klasifikuju kao kategorički ili hipotetički, Filipa Fut (Philippa Foot) dovodi u pitanje kategoričnost moralnih zapovesti (Foot 1972).<sup>9</sup> Kategorički imperativi imaju sledeće karakteristike: 1) oni zapovedaju bezuslovno; 2) na nužan način pružaju razloge za postupanje; 3) njihovo poštovanje je neizbežno; i 4) to su moralni imperativi.<sup>10</sup> S druge strane, hipotetičke imperativne karakteriše to što: 1) zapovedaju samo uslovno; 2) na kontingentan način pružaju razloge za postupanje; 3) mogu se izbeći; i na kraju, 4) to nisu moralni imperativi. Fut uočava da postoji posebna klasa zapovesti koje se nalaze u graničnom području između kategoričkih i hipotetičkih imperativa čija specifična svojstva mogu da uzdrmaju stav da su moralni imperativi kategorički. To su zapovesti etikecije ili elitnih klubova. Ove zapovesti su, kako misli Fut, neizbežne (kategoričke), ali nisu moralne. Međutim kategoričnost pravila etikecije može da se dovede u pitanje. Naime, pravila etikecije pripadaju klasi konvencionalnih pravila, a konvencionalna pravila ljudi prosuđuju kao ona koja mogu da se ukinu ili suspenduju, a njihovo važenje ne smatraju objektivnim niti njihovo poštovanje smatraju obavezujućim na onaj način kako su obavezujuća moralna pravila koje se obično prosuđuju kao objektivna.<sup>11</sup>

---

<sup>9</sup> Pre nje je to učinio Haršanji (John C. Harsanyi). Vidi (Harsanyi 1958).

<sup>10</sup> Strogo govoreći kategorički imperativ može biti samo jedan, a može imati onoliko formulacija koliko je maksima moguće univerzalizovati. Za različite formulacije kategoričkog imperativa vidi (Kant 2008).

<sup>11</sup> O razlici između moralnih i konvencionalnih pravila više će biti reči u jednom od kasnijih poglavlja.

Na drugom mestu, ova autorka zapaža da neke zapovesti mogu da budu moralne, ali da na kontingentan način pružaju razloge za postupanje, iz čega onda sledi da njihovo poštovanje može da se izbegne i da je to potpuno racionalan kurs delovanja.

Činjenica je da čovek koji odbija moral jer ne vidi razlog da se povinuje njegovim pravilima može biti optužen za nitkovluk, ali ne i za nekonzistentnost. Niti će njegovi postupci nužno biti iracionalni. Iracionalni postupci su oni kojima čovek na neki način osujećuje svoje svrhe, kada čini ono što je procenjeno da je za njega nepovoljno ili što osujećuje njegove ciljeve. Nemoralnost ne uključuje *nužno* nijednu od tih stvari. Očigledno je da normativni karakter moralnog suda ne garantuje da on ima snagu pružanja razloga (Foot 1972, 310).

Konačno, Fut zapaža da pojedini imperativi mogu da budu moralni, ali i nužni za ostvarenje neke želje. Pri tome treba imati na umu da želje nisu uvek usmerene na zadovoljenje potreba, ciljeva ili interesa osobe o čijoj želji je reč, već da ima želja koje su na autentičan način orijentisane na dobrobit druge osobe. Ljudi mogu da imaju želje koje su autentično altruističke. Ovaj stav implicira prihvatanje jedne vrste umerenog moralnog eksternalizma, koji bi mogao da se sažme na sledeći način: moralni imperativi ponekada, ali ne i uvek ispostavljaju motive ili razloge za postupanje, a ponekada je neophodno da se moralnim pravilima pridruži i želja. One osobe koje ne žele da se ponašaju moralno ili koje nemaju nikakvu dispoziciju da se tako ponašaju, nikakav skup filozofskih razloga neće uveriti u to da treba tako da se ponašaju (Harsanyi 1958). Dakle, ako osoba prosuđuje određeni postupak kao pogrešan, onda će time bar u minimalnom smislu biti motivisana da ne učini taj postupak, tj. neće želeći da ga učini (James 2011). Pri tome ne treba zaboraviti da na snazi ostaje da moralne zabrane načelno ne zavise od individualnih želja. Nečija želja da počini ubistvo, neće ukinuti zabranu ubistva niti može ubistvo učiniti moralno ispravnim. Reč je samo o tome da moralna zabrana ubistva nikoga ko želi da počini ubistvo neće sprečiti u tome, ako je ta želja dovoljno jaka, a želja da se povinuje ovoj zabrani odsutna ili srazmerno slaba. Ali, svakodnevni tok moralnog života je prilično trivijalan i ljudi uglavnom nemaju snažno izraženu želju da ubijaju jedni druge (up. Ruse 2006a).

Jedno od osnovnih načela pravde je da svako treba da dobije ono što zaslužuje, zato je poštovanje ili kršenje moralnih pravila blisko povezano s konceptima zasluge i kazne. Za nekoga ko poštuje moralna pravila mislimo da *zaslužuje* da bude pohvaljen, dok neko ko ih krši *zaslužuje* da bude kažnjen. Zajednica koja bi imala pravila koja definišu ono što je dobro i loše,

ispravno i pogrešno, pravedno i nepravedno, i u kojoj bi ono što je loše, pogrešno i nepravedno bilo moralno zabranjeno, ali čiji članovi ne bi uopšte imala koncept zasluge i kazne, koji bi bili indiferentni prema prekršiocima ovih normi ili bi propuštali da ih opomenu, izvrgnu poruzi, kazne, itd., bila bi prilično neobična (Joyce 2006). Takva zajednica bi podsećala na Platonov karikaturalni opis demokratije, državnog uređenja u kome oni koji su osuđeni na smrt šetaju slobodni. S druge strane, osobe koje svoje ponašanje usklađuju s moralnim pravilima samo zato jer očekuju da za to budu nagrađene ili zato što strahuju da bi mogle da budu kažnjene, ne smatramo moralnim. Kao što se poreklo i suština morala ne iscrpljuje u psihološkom altruizmu, emocionalnim reakcijama individue i sl., isto tako moral ne može da se redukuje na strah od kazne ili očekivanje nagrade. Ono što je važno jeste *sposobnost* individue da postupa u skladu sa sudovima o tome šta *treba* činiti ili pre ne činiti. Time nije isključeno da je za počinjanje moralnim pravilima neizbežna neka vrsta intrapersonalne gratifikacije, kao što je za emocionalno kompetentnu individuu neizbežan neki oblik intrasubjektivnog sankcionisanja za kršenje moralnih pravila.

Tako dolazimo do toga da su moralna bića ona koja su sposobna da iskuse niz moralnih emocija od kojih su krivica i kajanje zbog kršenja moralnih zabrana samo neke od mnogih. Istraživači su utvrdili da je moral u većini, ako ne i u svim, zajednicama fokusiran oko sledećih pet kategorija: 1) *briga za dobrobit/povređivanje*; 2) *pravičnost/reciprocitet*; 3) *unutargrupna kohezija/lojalnost*; 4) *autoritet/poštovanje*; i 5) *čistota/svetost* (Haidt and Joseph 2007, Haidt and Bjorklund 2008).<sup>12</sup> Za svaku od ovih kategorija vezane su određene emocije. Tako je *briga za dobrobit/povređivanje* povezana sa *saosećanjem* (saosećajna zabrinutost), tj. osetljivošću na signale patnje i bola, posebno kod dece (potomstva), srodnika i ranjivih grupa. *Pravičnost/reciprocitet* je povezana s *ljutnjom* i *zahvalnošću*, s emocionalnim reakcijama koje se povezuju sa strategijom *milo-za-drago* u recipročnim interakcijama i negativnim reakcijama na one koji ne uzvraćaju usluge, kao i *krivicom* za propuštanje da se učestvuje u reciprocijaciji.<sup>13</sup> *Unutargrupna kohezija/lojalnost* se povezuje s *ponosom*, *osećajem pripadanja* i *besom* koji je usmeren ka izdajnicima. Za *autoritet/poštovanje* se vezuju emocije koje se odnose na poštovanje

---

<sup>12</sup> Ovi autori ove kategorije nazivaju urođenim intuicijama, ali za potrebe ovog rada njihov status nije od preteranog značaja. Uostalom, već je rečeno da se u ovom radu ne zastupa nativistički pristup evoluciji morala.

<sup>13</sup> Detaljnije o ljutnji i krivici u (Prinz and Nichols 2010).

statusnih hijerarhija, kao što su osećaj *dominacije* i *podređenosti*, kao i *ljutnja* zbog propuštanja da se učine odgovarajući znaci poštovanja. Dok se za prethodne kategorije mogu pronaći odgovarajući ekvivalenti među drugim primatima, odnos prema kategoriji kakva je *čistota/svetost* je specifično ljudski, kao i emocija koja se za nju vezuje, a to je *gađenje*. Gađenje je specijes specifična emocija i neophodna je za objašnjenje velikog broja pravila koja se vezuju za hranu, seks, menstrualni ciklus i odlaganje leševa.

Konačno, različita afektivna stanja i želje koje osoba ima utiču na proces usvajanja i održanja verovanja, čineći taj proces pristrasnim. To znači da je jedan broj verovanja koje individua ima i koji ne mora da bude mali formiran na epistemički nepouzdan način. Ovakav način usvajanja i održanja verovanja vodi samoobmanjivanju koje grubo možemo da definišemo na sledeći način. Osoba se samoobmanjuje kada usvaja ili održava određeno pogrešno verovanje, uprkos protivevidenciji koju ona ignoriše, zanemaruje, pristrasno tumači ili propušta da uvidi, služeći se različitim strategijama za svakodnevno testiranje hipoteza. O ovome će više reči biti kasnije. U ovom radu se zastupa stanovište prema kome je samoobmanjivanje važan „sastojak“ morala, tj. tvrdi se da postoji korelacija između sposobnosti za samoobmanjivanje i moralne kompetencije individue. Ona osobe kod kojih je kapacitet za samoobmanjivanje oslabljen ili odsutan imaju teškoća da normalno funkcionišu u socijalnom i moralnom kontekstu; one ne mogu da budu kompetentne moralne sudije. Ako samoobmanjivanje dolazi u stupnjevima, tako da se neke individue manje, a druge više samoobmanjuju, onda će one individue koje se najmanje samoobmanjuju, uz pretpostavku da imaju još neke emocionalne deficite, biti i one koje su najnemoralnije.

Do sada je postalo jasno da vezu između biologije, altruizma i morala nameravam da uspostavim posredno, preko moralne psihologije. Uzimajući u obzir da nema konkluzivnih dokaza da biološko ustrojstvo jedinke (geni) direktno determiniše bilo koji oblik ponašanja, pa ni moralno ponašanje ili sposobnost za moralno suđenje, ovaj pristup mi se čini sasvim prihvatljivim. Kako su rasprave u kontekstu filozofije, biologije i filozofije biologije o načinu na koji je biološki i bihejvioralni altruizam mogao da evoluiru ujedno i rasprave o jedinici selekcije na koju mehanizam prirodne selekcije deluje, to će biti predmet sledećeg poglavlja, dok će o neurobiološkim i psihološkim mehanizmima koji omogućavaju evoluciju psihološkog altruizma biti reči u posebnim poglavljima koja se bave empatijom i posledicama njenog nedostatka.



## 2 EVOLUCIJA ALTRUIZMA I PROBLEM JEDINICA SELEKCIJE

### 2.1 Paradoks altruizma

Za evolucione biologe altruizam u biološkom smislu predstavlja paradoks, jer prirodna selekcija favorizuje osobine koje uvećavaju šanse jedinke da preživi i ostavi potomke, a ponašanje koje vodi povećanju adaptivne vrednosti organizma, tj. uvećavanju ovih šansi, u biološkom kontekstu se obično označava kao sebično. Da bi se jedinka ponašala sebično nije neophodno da se ponaša intencionalno, sebičnost je tehnički termin i ne referira na motivacione ili sofisticirane kognitivne mehanizme. S druge strane, u istom kontekstu, altruizam se definiše kao oblik ponašanja kojim jedinka uvećava adaptivnu vrednost druge jedinke na račun svoje sopstvene. Ni altruizam u ovom kontekstu ne pretpostavlja postojanje složenih motivacionih i kognitivnih mehanizama. Ako je prirodna selekcija osnovni mehanizam evolucije, onda altruizam ne bi mogao da evoluirati, jer altruistične jedinke ili ne bi ostavljale potomstvo ili bi broj preživelih potomaka koji ostavljaju bio srazmerno manji u odnosu na broj preživelih potomaka sebičnih jedinki. Konačni rezultat bi bilo eliminisanje altruističkih jedinki iz populacije. Dakle, ako prirodna selekcija deluje protiv altruista, očekivali bismo da u prirodi ne nailazimo na altruistične jedinke. Međutim, na osnovu empirijske evidencije znamo da u prirodi ima altruista. Ili, nešto preciznije rečeno, određene bihevioralne obrasce na koje nailazimo u prirodi prepoznajemo kao altruističke. To je ono što se tehnički označava kao *paradoks altruizma*. Zapravo, nije reč o pravom logičkom paradoksu, već o diskrepanciji koja postoji između teorijskog predviđanja i empirijske evidencije. Kako je prirodna selekcija evolucionim mehanizam na koji se pozivamo kada pokušavamo da objasnimo kompleksne osobine organizama, čini se da polazeći od teorije evolucije ne možemo da rešimo ovaj paradoks.

Međutim, kada je reč o tome da paradoks altruizma nije pravi logički paradoks, izgleda da polazeći od Soberove distinkcije između „selekcije nečega“ i „selekcije za nešto“ i razumevanja *selekcije za* kao kauzalne relacije između posedovanja određene karakteristike i povećanja adaptivne vrednosti (Sober 1993), prethodni argument ipak može da se formuliše u obliku pravog logičkog paradoksa.

1. Ako postoji selekcija za altruizam, onda altruističko ponašanje uzrokuje povećanje adaptivne vrednosti organizma.
2. Altruističko ponašanje (po definiciji) ne uzrokuje povećanje adaptivne vrednosti organizma.
3. Dakle, selekcija za altruizam ne postoji.
4. Ako selekcija za altruizam ne postoji, onda altruizam neće biti selektovan i u populaciji neće biti altruista.
5. Ali, u populaciji ima altruista.
6. Dakle, altruizam je selektovan i postoji selekcija za altruizam.

Tačke tri i šest daju kontradikciju, a moguća rešenja da se protivrečnost prevaziđe jesu ili da se odbaci mehanizam prirodne selekcije, ili da se pruži evoluciono objašnjenje altruizma koje je kompatibilno s ovim mehanizmom. Odbacivanje principa selekcije je neprihvatljivo, jer teorija evolucije putem prirodne selekcije načelno ima veliku prediktivnu i eksplanatornu moć, tako da je kao alternativa ostalo jedino da se pruži evoluciono objašnjenje altruizma koje je u skladu s ovim principom. Ispostavilo se da to nije nimalo lak zadatak.

## 2.2 Grupna selekcija

U kojoj meri je altruizam kao empirijska činjenica potkopavao teoriju evolucije vidi se iz Darvinovog odnosa prema tom problemu. Za Darvina je postojanje sterilnih kasti među socijalnim insektima iz reda opnokrilaca (lat. *Hymenoptera*), kao i sterilnih hibrida, predstavljalo „posebnu teškoću“. Posebna teškoća se očitovala u tome što je Darwin u prvi mah bio uveren da bi izostanak evolucionog objašnjenja za postojanje sterilnih kasti insekata, koje čitav život posvećuju dobrobiti drugih jedinki (matici i njenim potomcima) na račun sopstvenog reproduktivnog uspeha, opovrgavao teoriju evolucije putem prirodne selekcije u celini. „[...] ograničiću se ovde samo na jednu posebnu teškoću, koja mi se u prvi mah učinila nesavladivom i zapravo pogubnom za čitavu teoriju. Ovde mislim na postojanje neutera ili sterilnih ženki u zajednicama insekata“ (Darwin 2009b, 175).

Kako sterilnost poništava mogućnost da jedinka ostavi potomke, ona shodno tome, umanjuje adaptivnu vrednost organizma koja je određena preko individualnog preživljavanja i

reprodukcije. Darwin je morao da ponudi objašnjenje koje bi bilo konzistentno s principom prirodne selekcije koji predviđa da će jedinke s adaptivnom vrednošću nižom u odnosu na adaptivnu vrednost drugih jedinki iste populacije, u jednom trenutku, ma koliko on bio dalek, morati da nestanu. U prvo vreme je izgledalo da model grupne selekcije nudi rešenje. Prema ovoj koncepciji, altruističko ponašanje, kao što je već opisano žrtvovanje reproduktivne funkcije kod socijalnih insekata, evoluiralo je zato što obezbeđuje dobrobit grupe. Ovim je načinjen pomak od individualne selekcije, koja je glavni eksplanatorni model i okosnica Darwinovog dela, ka grupnoj selekciji.

Darwin ističe da objašnjenje evolucionog procesa koji je vodio do steriliteta određenih jedinki u kolonijama pčela (pčelinjim društvima) zaista predstavlja teškoću, ali da ona nije mnogo veća od teškoće da se objasni bilo koja druga neobična karakteristika (modifikacija strukture). Sterilitet radilica nije toliko neuobičajen koliko bi se u prvom trenutku moglo činiti, „jer se može pokazati da neki insekti i druge životinje u prirodi ponekada postanu sterilne; i ako su takvi insekti socijalni i ako je bilo korisno za zajednicu da godišnje jedan broj jedinki bude sposoban za rad, ali nesposoban za reprodukciju, onda ne vidim posebnu teškoću da je do toga došlo prirodnom selekcijom“ (Darwin 2009b, 176).

Osim toga, Darwin se poziva na model grupne selekcije prilikom objašnjenja činjenice da medonosne pčele imaju nazubljenu žaoku koja se prilikom uboda odvaja od tela zajedno s delom abdomena što uzrokuje smrt jedinke. Takva karakteristika može biti selektovana ako koristi pčelinjem društvu uprkos tome što uzrokuje smrt nekolicine njenih članova (Darwin 2009b). Dakle, on se u *Postanku vrsta* oslanja na koncepciju grupne selekcije bar u dva slučaja. Jedan je radi objašnjenja evolucije sterilnih jedinki, a drugi da bi objasnio autodestruktivne fenotipske karakteristike medonosnih pčela.

Zanimljivo je da Darwin u *Postanku vrsta* tvrdi da se „posebna teškoća“ umanjuje ili nestaje ako se mehanizam prirodne selekcije primeni na porodicu, a ne samo na pojedinačni organizam (Darwin 2009b), što je dalo povoda nekim teoretičarima da ona mesta na kojima se Darwin poziva na grupnu selekciju tumače kao da je reč o rudimentarnoj formi selekcije u srodničkom krugu. Kao razlog za nedovršenost i nedorečenost Darwinovih argumenata oni navode to što Darwin još nije imao valjanu teoriju nasleđivanja (Ruse 2006a). Tako, Rus ističe da

u Darwinovom delu, bar kada je reč o *Postanku vrsta*, nema govora o grupnoj selekciji, već o nečemu nalik selekciji u srodničkom krugu, ali da Darwin nije imao ni dovoljno empirijskih podataka ni razvijen konceptualni aparat da bi argument mogao da bude definisan onako kako je to mnogo kasnije učinio Hamilton (Ruse 1980). Međutim, izgleda da je Rusov argument anahron i neopravdan (Borrello 2005), jer kada Darwin piše o adaptacijama koje su rezultat delovanja prirodne selekcije na nivou grupe, on ne pravi jasnu distinkciju između grupa srodnih i nesrodnih jedinki, između onoga što je korisno za bilo koju grupu i onoga što je korisno za grupu srodnika. Nivoi objašnjenja nisu jasno diferencirani, tako da je tekst otvoren za različita tumačenja. Zastupnici genskog selekcionizma i selekcije u srodničkom krugu mogu da ga tumače kao da govori o selekciji u srodničkom krugu, dok oni kojima je koncepcija grupne selekcije prihvatljiva mogu da ga tumače u tom kontekstu.

Međutim, utisak da Darwin prosocijalno ponašanje tumači pozivanjem na grupnu selekciju teško može da se izbegne ako na umu imamo njegovu studiju o poreklu čoveka gde je čitava argumentacija o evoluciji moralnih instinkata zasnovana na modelu grupne selekcije (Darwin 1981). Darwin modelu grupne selekcije pristupa s izvesnom oprežnošću. Kada je reč o evoluciji pojedinih karakteristika striktno socijalnih životinja, kakvi su socijalni insekti, prirodna selekcija direktno deluje na grupnom, a tek posredno na individualnom nivou. Međutim, kada je reč o sisarima, teško je zaključiti da je bilo koja karakteristika evoluirala isključivo za dobrobit grupe, iako neke od tih karakteristika posredno koriste i grupi. Jedan izuzetak od tog pravila je evolucija moralnog osećaja kod ljudi. Ali, Darwin najpre uočava da je veoma sumnjivo da u kontekstu jedne populacije adaptivna vrednost altruističkih jedinki koje raspolazu izraženim moralnim osećajem bude veća od adaptivne vrednosti sebičnih jedinki koje takav osećaj nemaju ili ga imaju u manjoj meri. Prirodna selekcija koja deluje na individualnom nivou bi takve jedinke eliminisala iz populacije.

Onaj ko je bio spreman da žrtvuje svoj život ... pre nego da izda svoje drugove često uopšte nije ostavljao potomstvo koje bi nasledilo njegovu plemenitu prirodu. Najhrabriji ljudi, koji su uvek bili spremni da se u ratu nađu u prvim redovima i koji su slobodno rizikovali život za druge, u proseku su ginuli više nego drugi ljudi. Zato se čini malo verovatnim ... da se broj ljudi koji su bili obdareni takvim vrlinama ili dostignuti nivo njihove izvrsnosti mogao povećati putem prirodne selekcije, odnosno preživljavanjem najsposobnijih (Darwin 1981, 163).

Ali, ako se kao jedinica selekcije uzme grupa, to jest ako prirodna selekcija deluje na diferencijalno preživljavanje grupa, onda se perspektiva menja i žrtvovanje sopstvene adaptivne vrednosti postaje razumljivo i lako objašnjivo. Iako altruističko ponašanje umanjuje adaptivnu vrednost altruista, ono uvećava adaptivnu vrednost grupa altruističkih jedinki u odnosu na grupe čiji su članovi sebični.

Ne sme se zaboraviti da visok nivo moralnosti daje malu ili nikakvu prednost onim individuama i njihovoj deci u odnosu prema drugim ljudima istog plemena, a da će ipak moralni napredak (napredak u moralnim standardima) i uvećanje broja obdarenih ljudi sigurno dati ogromnu prednost jednog plemena nad drugim. Nema sumnje da će pleme čiji su mnogi članovi nadahnuti snažnim patriotizmom, vernošću, pokoravanjem, hrabrošću i saosećanjem, koji su uvek spremni da pomognu jedni drugima i da se žrtvuju za zajedničko dobro, odnositi pobjedu nad većinom drugih plemena; a to će biti prirodna selekcija. Oduvek su u svetu jedna plemena potiskivala druga; a kako je moral jedan činilac njihovog uspeha, nivo moralnosti i broj nadarenih ljudi će uvek težiti da se uzdiže i uvećava (Darwin 1981, 166).

Ovi argumenti upućuju na to da je Darwin, iako možda s oklevanjem, prihvatao koncepciju grupne selekcije bar kada je reč o socijalnoj evoluciji, tako da teško može da se govori isključivo o blagom odstupanju od uobičajenog metodološkog pristupa, kako tvrdi Rus (Ruse 1980).

### **2.3 Grupna selekcija u 20. veku**

Teorija grupne selekcije, čiji je glavni zagovornik u 20. veku bio Vero Kopner Vin-Edwards (Vero Copner Wynne-Edwards), zaobilazi ili uspeva da reši tragediju zajedničkog dobra (eng. *the tragedy of commons*) koja sledi iz individualističkog pristupa prirodnoj selekciji (Hardin 1968). U ekonomskom smislu, zajednička dobra predstavljaju iscrpljive resurse kod kojih svaka dodatna potrošnja umanjuje količinu resursa koja ostaje na raspolaganju budućim potrošačima. Tragedija zajedničkog dobra je zasnovana na staroj Maltusovoj (Thomas Robert Malthus) zamisli da uvećanje populacije vodi iscrpljivanju resursa i katastrofalnim ishodima. Potencijalno rešenje ovog problema je takođe maltusovsko: regulacija brojnosti populacije uzdržavanjem od reprodukcije. Ako bi jedinke mogle samostalno da regulišu brojnost populacije, time bi se obezbedio stabilan odnos između broja članova populacije i raspoloživih resursa, a tragedija zajedničkog dobra bi bila izbegnuta. Prema Vin-Edwardsu životinje su adaptirane da

samostalno kontrolišu gustinu populacije, a mehanizmi za regulaciju gustine populacije deluju automatski tako što se broj jedinki prilagođava s obzirom na količinu i dostupnost resursa neophodnih za samoodržanje (Wynne-Edwards 1962, 1963).

Prema Vin-Edvardsovom modelu, prirodna selekcija deluje na grupe jedinki i proizvodi adaptacije na nivou grupe. Ove adaptacije nisu samo agregat adaptacija pojedinačnih organizama, već se odnose na grupe kao koherentne celine koje su i krajnji korisnici ovih adaptacija. Jednostavnije rečeno, postoje adaptacije koje podstiču evolucionu uspeh grupa koje ih poseduju, ali umanjuju adaptivnu vrednost njenih pojedinih članova, što bi značilo da prirodna selekcija favorizuje evoluciju određenih oblika autodestruktivnog ponašanja ukoliko takvo ponašanje doprinosi prosperitetu grupe. Budući da je unutargrupna kompeticija integralni deo ove teorije, osnovna teškoća jeste to što selekcija istovremeno deluje na dva nivoa sa suprotstavljenim efektima, a konflikt između selekcije koja deluje na individualnom nivou i selekcije na nivou grupe uvek se razrešava u korist grupne selekcije, što je zaključak koji Vin-Edwards izvodi na osnovu terenskih istraživanja (Wynne-Edwards 1962, 1963). Međutim, nije bilo dovoljno reći da je određeno autodestruktivno ponašanje evoluiralo jer koristi nekom višem nivou biološke hijerarhije, već je bilo potrebno pokazati na koji način takvo ponašanje najpre koristi individui.

## 2.4 Genski selekcionizam i evolucija iz perspektive gena

Zagovornici grupne selekcije su smatrali da altruističko ponašanje može da evoluiru u populaciji tako što prirodna selekcija u pojedinim slučajevima favorizuje one karakteristike koje su štetne za organizam, ali koje koriste grupi što je bilo u suprotnosti sa brojnim saznanjima o tome kako prirodna selekcija deluje. Glavni argument protiv grupne selekcije jeste da su grupe altruističnih jedinki osetljive na ono što je Dokins nazvao subverzijom iznutra.<sup>14</sup> Argument se oslanja na koncepciju evoluciono stabilne strategije (ESS): ako jednu strategiju usvoji dovoljno

---

<sup>14</sup> Ovaj argument su pre Dokinsa formulisali Mejnard Smit (John Maynard Smith) i Vilijams (George Williams) (Maynard Smith 1964, 1976, Williams 1996 [1966]). Istini za volju, Dokins ni na jednom mestu, ni u *Sebičnom genu* ni u *Proširenom fenotipu* nije upotrebio formulaciju „subverzija iznutra“ (eng. *subversion from within*). Dokins se koristi terminima *izdaja*, (eng. *treachery from within*) (Dawkins 2006 [1976]), odnosno *promena* (eng. *change from within*) (Dawkins 1999 [1982]).

veliki broj članova grupe onda takva strategija ne može da bude nadvladana alternativnom strategijom. Ispostavilo se da altruizam nije ESS i da relativno lako može da bude nadvladana. Pretpostavimo da se neka grupa  $G$  sastoji od jedinki  $A$  koje su altruistične. Sada pretpostavimo da se u  $G$  pojavi mutacija  $S$  koja čini da se mali broj sebičnih mutanata brže i uspešnije razmnožava i ostavlja više preživelog potomstva nego jedinke  $A$ . Kako se adaptivna vrednost jedinke meri stopom preživljavanja i reprodukcije u odnosu na druge jedinke iste populacije, adaptivna vrednost jedinki  $S$  će shodno tome biti veća nego adaptivna vrednost jedinki  $A$ , tako da će za nekoliko generacija jedinke  $S$  istisnuti jedinke  $A$  iz populacije.

Vilijamsova kritika grupne selekcije, koja je poznata i kao *Vilijamsova revolucija*, pokazala je da grupe nisu transgeneracijski stabilne da bi mogle da budu jedinice selekcije, kao i da posledica selektivnog procesa ne mogu da budu adaptacije na nivou grupe, jer je selekcija koja deluje na tom nivou suviše slaba da bi proizvodila adaptacije (Williams 1996 [1966]). On ne poriče teorijsku mogućnost grupne selekcije, već njenu empirijsku efikasnost.<sup>15</sup> Vilijams na više mesta ističe da je ono što bi moglo da izgleda kao adaptacija na nivou grupe, zapravo samo statistička suma individualnih adaptacija (Williams 1996 [1966]). Međutim, sumiranjem individualnih adaptacija ne dobijamo ništa što bi moglo da se tumači kao adaptacija grupe, jer ono što se obično smatra adaptacijom na nivou grupe nije agregat adaptacija njenih pojedinačnih članova (Okasha 2008).

Prema mišljenju Vilijamsa i drugih zastupnika genskog selekcionizma, zagovornici grupne selekcije narušavaju jedno od temeljnih naučnih načela – načelo parsimonije (Williams 1996 [1966]). Načelo parsimonije je posledica praktične primene *Okamove oštrice*: između dva ili više alternativnih objašnjenja istog skupa podataka uvek biraj ono koje je jednostavnije. I kada je reč o evolucionim objašnjenjima složenih bioloških karakteristika, kakva je altruizam, kompleksna objašnjenja moraju da ustupe mesto jednostavnijim eksplanatornim modelima, a najjednostavnija objašnjenja su obično ona koja se pozivaju na selekciju gena u populaciji. Svako ko između dva eksplanatorna modela koja objašnjavaju isti skup podataka bira ono koje je složenije, izlaže se opasnosti da bude posečen *Okamovom oštricom*.<sup>16</sup>

---

<sup>15</sup> Do sličnih zaključaka došao je i Mejnard Smit (Maynard Smith 1964, 1976).

<sup>16</sup> Sober je skeptičan kada je reč o univerzalnoj primeni *Okamove oštrice*. Vidi (Sober 2000).

Osim što insistira na doslednoj primeni *Okamove oštrice*, Vilijamsov argument u prilog tezi da je gen osnovna jedinica selekcije jednostavan je primer eliminativne indukcije (Williams 1996 [1966]). Da bi neki entitet imao status jedinice selekcije on najpre mora da zadovolji određene uslove. Ti uslovi su dugovečnost, preciznost u pravljenju kopija i reproduktivni potencijal.<sup>17</sup> Vilijams najpre utvrđuje koji entiteti ne zadovoljavaju ove uslove. To su svakako grupe, ali ni fenotipovi ni genotipovi ne mogu da ih ispune. Fenotipovi kao specifične kombinacije gena i uslova sredine nestaju sa okončanjem životnog ciklusa. S druge strane, za organizme koji se razmnožavaju seksualnim putem, pravila mejoze i rekombinacija deluju na genotipove jednako kao i smrt. Ono što nakon eliminacije ostaje jesu mali, kvalitativno stabilni i potencijalno besmrtni delovi hromozoma koji su nosioci naslednih informacija i koje nazivamo geni.<sup>18</sup> Preciznije, gen nije nosilac nasledne informacije, on je upravo ta informacija. Onako kako ga definiše Vilijams, gen je „bilo koja nasledna informacija za koju postoji selekcija“ (Williams 1996 [1966], 25). Budući da ne pretpostavlja postojanje bilo kog organskog (empirijskog) entiteta na koji pojam gen referira, ova definicija je omogućila da se o genima raspravlja potpuno apstraktno (Hull 1980, 2001). Naravno, kada, recimo, molekularni biolozi govore o genima, oni nesumnjivo referiraju na određenu molekularnu strukturu. Ali, da pojam gen, u jedinom biološkom kontekstu koji nam je za sada poznat, referira na određenu organsku strukturu kakva je deo lanca DNK, predstavlja kontingentnu istinu (Godfrey-Smith 2000). U nekoj protivčinjeničkoj situaciji u kojoj ne bi bilo ovih organskih jedinica nasleđivanja, ali bi postojala nasledna varijacija u adaptivnoj vrednosti, proces evolucije putem prirodne selekcije bi mogao da ima odgovarajuće učinke, a pojam gen bi u tom slučaju imao sasvim drugačiji objekt referencije (Lewontin 1970, Godfrey-Smith 2000).

---

<sup>17</sup> Ove uslove Vilijams nije jasno naznačio, ali se oni mogu identifikovati u argumentu. Njihov konačni oblik dugujemo Dokinsu (Dawkins 1999 [1982], 2006 [1976]).

<sup>18</sup> Pojam „gen“ je u biologiju uveo Vilhelm Johansen da bi opisao osnovnu jedinicu nasleđivanja. Ovaj pojam je označavao nepoznati entitet za koji se pretpostavljalo da postoji u polnim ćelijama i da određuje posebne karakteristike organizma. „Reč gen je oslobođena bilo koje pretpostavke; ona izražava samo očiglednu činjenicu da su... mnoge karakteristike organizma sadržane u gametima...“ (Johannsen 1909, navedeno u Keller 2000) Zapravo Vilijams nije bio prvi koji je na evoluciju gledao iz pozicije gena, jer je pre njega to učinio Hamilton, a zamisao je već bila prisutna kod populacionih genetičara iz prve polovine 20. veka. Hamilton je pokazao na koji način altruizam može da evoluiru, a da krajnji korisnici ove adaptacije budu geni, a ne pojedinačni organizmi ili grupe organizama. O tome će više biti reči u narednom odeljku.



Dokins je umesto pojma gen koji je određen inherentnom sposobnošću preciznog pravljenja sopstvenih kopija i dugovečnošću u teoriju uveo opštiji pojam – *replikator* – koji ne referira nužno na neki određeni biološki entitet (Dawkins 1999 [1982], 2006 [1976]).<sup>19</sup> Iako su svi geni kao biološki entiteti replikator, ne sledi da su svi replikator gen, jer replikator može biti bilo šta s inherentnom sposobnošću pravljenja sopstvenih kopija (Dawkins 1999 [1982]). Dominantno adaptivno svojstvo gena, to jest replikatora, je obezbeđivanje preživljavanja pravljenjem sopstvenih kopija.<sup>20</sup> Ovo ne predstavlja ništa drugo do *sebičnost* i u ovom kontekstu se odnosi na svojstvo gena koje mu omogućava širenje u populaciji uprkos nepovoljnim uslovima sredine.<sup>21</sup> Čini se da je sebičnost kao specifična strategija preživljavanja analitički sadržana u samom pojmu replikatora.

Iako su svi geni replikator, većina replikatora, naročito oni koji se nalaze u telesnim ćelijama i koji se kopiraju prilikom mitoze, završava u evolucionom ćorsokaku (eng. *deadend replicators*): oni se ne prenose u narednu generaciju i nestaju zajedno s organizmom u kome se nalaze. Dokins je zainteresovan za posebnu vrstu replikatora, za *aktivne replikatore germinativne linije* koji kauzalno mogu da određuju sopstvenu sudbinu, jer su kauzalno odgovorni za produkciju odgovarajućih fenotipova koji im olakšavaju (ili čine mogućim) da budu preneti u narednu generaciju (Dawkins 1999 [1982], Lloyd 2007, Okasha 2008). Kako su ovi replikator prave jedinice selekcije zbog kojih postoje adaptacije, sebičnost replikatora se odražava i na bihevioralni obrazac organizma čija je funkcija svedena na to da bude efemerni *nosilac* (eng.

---

<sup>19</sup> Dokinsov pristup problemu jedinice selekcije može se sagledati u dve ravni, empirijskoj i konceptualnoj. Empirijska bi se u raspravama o jedinici selekcije odnosila na dihotomiju *grupa/organizam*, dok bi se konceptualna ravan odnosila na dihotomiju *organizam/gen*. Dokins je prevashodno zainteresovan za konceptualnu ravan, odnosno za pitanje šta znači biti jedinica prirodne selekcije ili na šta treba da mislimo kada govorimo o jedinicama selekcije (Dawkins 1999 [1982]).

<sup>20</sup> Nisu replikator ti koji preživljavaju ili ne, već njihove kopije. Vidi (Lloyd 2007).

<sup>21</sup> Pod uslovima sredine podrazumeva se najpre prirodno okruženje organizma, ali što je u ovom kontekstu još važnije uslovi sredine uključuju i druge gene na istom lokusu koji čine prirodno okruženje gena. Ovo drugo je nestandardno razumevanje prirodne sredine.

*vehicle*) potencijalno besmrtnih replikatora od generacije do generacije.<sup>22</sup> Svi replikator koji se nalaze u jednom nosiocu dele zajedničku sudbinu.

Stanovište da su geni (replikatori) jedine prave jedinice selekcije naišao je na niz kritika, a verovatno najvažnija među njima je ona da su geni nevidljivi za prirodnu selekciju i da prirodna selekcija operiše s organizmima kao koherentnim celinama koje stupaju u direktan odnos s uslovima sredine (interaktori), odnosno s različitim fenotipovima (Brandon 1982, Gould 1992). Geni su nevidljivi za selekciju jer ih fenotipovi „zaklanjaju“: ako neki organizam ima određenu fenotipsku karakteristiku koja objašnjava i predviđa stopu preživljavanja i reprodukcije tog organizma, onda dodavanje gena koji kodiraju tu osobinu u jednačinu ništa ne menja u objašnjenju i predviđanju (Brandon 1982). Fenotipovi suspenduju genotipove, a samim tim i gene iz razmatranja. Međutim, ovo neće uvek biti slučaj, jer ako zamislimo tri jedinke s genotipovima *AA*, *Aa* i *aa*, tako da se *AA* i *Aa* fenotipski ne razlikuju i ostavljaju jednak broj potomaka, dok se razlikuju od jedinke *aa* koja ostavlja manje preživelog potomstva u odnosu na *AA* i *Aa*, adaptivna vrednost jedinki *AA* i *Aa* će se ipak razlikovati, jer dok *AA* nikada ne ostavljaju potomke koji su *aa* s nižom stopom preživljavanja i reprodukcije, *Aa* ponekada ostavljaju potomke čiji je genotip *aa* (Sober and Wilson 1994).

Zastupnici genskog selekcionizma uglavnom zaobilaze navedeni prigovor tvrdnjom da je rasprava o dihotomiji *gen/organizam*, tj. *replikator/interaktor*, deplasirana jer se radi samo o različitim perspektivama kada je reč o problemu jedinice selekcije, tako da je pojedinačni organizam neposredni, dok je gen krajnji entitet na koji prirodna selekcija deluje (Dawkins 1999 [1982], Sterelny and Kitcher 1988).

Uprkos sebičnosti kao inherentnoj karakteristici replikatora, oni u specifičnim okolnostima i kroz niz složenih interakcija sa drugim replikatorima (genima) i uslovima sredine najbolje mogu da ostvare svoje ciljeve tako što, na bihevioralnom planu, podstiču izvestan oblik

---

<sup>22</sup> Hal je umesto pojma nosioca uveo pojam *interaktora* koji se odnosi na one entitete koji kao „koherentne celine direktno stupaju u interakciju sa sredinom, na takav način da čine replikaciju diferencijalnom“ (Hull 1980, 318). Iako se pojmovi nosioca i interaktora često smatraju identičnim i uzajamno zamenjivim kao što to čine Sober i Vilson (Sober and Wilson 1994), oni ipak nisu posve ekvivalentni (Hull 2001). Dok su za Dokinsa geni uvek replikator, a nikada nosioci, za Hala su geni uvek replikator, ali su ponekada i interaktori.

altruizma (Dawkins 2006 [1976]). Možemo da uočimo dva nivoa na kojima se altruizam *naizgled* ispoljava.<sup>23</sup> Najpre, prirodna selekcija će davati prednost onim replikatorima čija ekspresija nema negativnih efekata na organizme koji su njihovi nosioci, što znači da su replikatori prinuđeni da međusobno sarađuju i vode računa o telima u kojima se nalaze, jer se u suprotnom neće naći u narednoj generaciji. Na drugom mestu, pojavu altruizma možemo da očekujemo ako su replikatori zainteresovani za preživljavanje onih nosilaca u kojima se sami ne nalaze (Dawkins 1999 [1982]). Izgleda da postoje najmanje dva slučaja u kojima su replikatori za to zainteresovani. Prvo, budući da preživljavaju kopije replikatora, a ne sam replikator, može se očekivati da se one nalaze i u onim telima u kojima se sam replikator ne nalazi i eksplanatorni model koji bi objašnjavao ovu pojavu altruizma bi bio model selekcije u srodničkom krugu. Takođe, u slučajevima simbioze, parazitizma i uzajamnosti replikator može da ima interesa od preživljavanja onih organizama u kojima se ne nalaze čak ni njegove kopije i adekvatan eksplanatorni model za evoluciju altruizma u ovom slučaju bi bio onaj koji se poziva na koncept reciprociteta. U narednim odeljcima biće više reči o ovim eksplanatornim modelima.

## **2.5 Teorija inkluzivne adaptivne vrednosti – selekcija u srodničkom krugu**

Iako se teorija selekcije u srodničkom krugu najčešće vezuje za Hamiltonov članak o evoluciji društvenog ponašanja (Hamilton 1964a, b), ona ima dugu istoriju koja seže sve do Darvina (Dugatkin 2007). U *Postanku vrsta* kao rešenje „posebne teškoće“ s kojom se suočavala teorija prirodne selekcije, tj. postojanje sterilnih kasti kod socijalnih insekata, Darwin uočava da ako se mehanizam prirodne selekcije primeni na porodicu, a ne samo na individuu, onda se teškoća o kojoj je reč umanjuje ili čak nestaje (Darwin 2009b).

---

<sup>23</sup> Vilijamsova logika i je bila neumoljiva i kada je reč o istraživanjima vezanim za jedinicu selekcije dovela je do gotovo potpunog pomeranja težišta s grupa i pojedinačnih organizama na gen. Dokinsova popularizacija genskog selekcionizma je tome značajno doprinela. S napuštanjem grupne selekcije i pomeranjem težišta na gen činilo se da su se umanjili izgledi da se pruži zadovoljavajuće objašnjenje altruizma koje ne bi referiralo natrag na interese gena. Takva vrsta povratne referencije zapravo vodi redukciji altruizma na poseban oblik genetičke sebičnosti. Kada je reč o biološkom kontekstu, tamo gde nema žrtvovanja sopstvene adaptivne vrednosti, nema ni smisla govoriti o altruizmu.

Darvin sličan princip predlaže i za rešavanje postojanja altruizma u ljudskim zajednicama (Darwin 1981). Ako altruistična jedinka ostavi potomke i ako je altruizam nasledna varijacija, onda je racionalno očekivati da će i njegovo potomstvo pokazivati istu fenotipsku karakteristiku. Kako altruizam u biološkom smislu po definiciji uključuje žrtvovanje sopstvene adaptivne vrednosti, te shodno tome verovatno neostavljanje direktnih potomaka, čak i tada, ako je jedinka imala srodnike, ova fenotipska karakteristika može biti očuvana i preneti u sledeću generaciju. Iako je tumačenje Darwinovih uvida kao anticipacije selekcije u srodničkom krugu možda anahrono, na šta sam već ranije ukazao, ipak se ne može izbeći utisak da je Darwin zaista naslutio ono što je kasnije Holdejn anegdotski objasnio, a zatim Hamilton formalno dokazao.

Altruističko ponašanje može da evoluiru među srodnicima zbog prirodne sklonosti ljudi ka porodičnom životu (Haldane 1932). Mada nikada nije do kraja formalizovao argument selekcije u srodničkom krugu, Holdejn je ipak dovoljno jasno definisao elementarne pretpostavke (uslove) za evoluciju altruizma na osnovu kojih će Hamilton kasnije razviti sopstveni argument. Te pretpostavke su sledeće: 1) altruistički činovi moraju da budu usmereni ka povećanju adaptivne vrednosti direktnih potomaka ili bliskih rođaka i 2) da bi se gen za altruizam proširio, populacija mora da bude dovoljno mala zajednica bliskih srodnika.

Holdejinov argument je predstavljen u formi misaonog eksperimenta (Haldane 1955). Pretpostavimo da se jednog sunčanog popodneva šetate pored reke s najboljim prijateljem i da ugledate kako se dete davi. Pretpostavimo da posedujete gen za altruizam koji kodira vaš mozak na takav način da pomažete davljenicima, dok vaš prijatelj ne poseduje takav gen. Zatim, pretpostavimo da postoje šanse da se prilikom spasavanja i sami udavite. Ako detetu priskočite u pomoć rizikujući da izgubite život spasavajući ga, time umanjujete sopstvene šanse za preživljavanje i reprodukciju, ali delimično uvećavate njegove. Ako ga ne spasite, a prilikom spasavanja se i sami udavite, u tom slučaju život davljenika, vaš život i pokušaj spasavanja su doslovno pali u vodu. Ako s detetom niste genetički povezani (identični po poreklu), izgleda da onda nemate mnogo razloga da rizikujete, jer ako se udavite gen za altruizam će zauvek biti izgubljen. Međutim, ako je davljenik vaše dete, odnosno brat ili sestra, (još je bolje da to bude dvoje dece ili brat i sestra) onda su šanse da i ono poseduje isti gen  $\frac{1}{2}$ , jer je koeficijent srodstva između roditelja i dece, odnosno između braće i sestara upravo toliki; tada ćete imati genetički interes da rizikujete. S opadanjem koeficijenta srodstva, rastu izgledi da će potencijalna šteta od

spasavanja biti veća od moguće koristi. Dakle, gen za altruističko spasavanje davljenika ima najviše šansi da se proširi u uslovima zavisne društvenosti gde su svi ili bar supstancijalna većina članova zajednice bliski srodnici. U onim slučajevima gde je koeficijent srodstva zanemarljivo mali, najracionalniji kurs delovanja, posmatrano iz pozicije gena, bi bio ostaviti dete da se udavi ili da se samo nekako izvuče. Izgleda da je majka priroda zaista stara zla veštica (Williams 1993).

Sledeći Holdejna, Hamilton je pretpostavio da altruizam može da evoluiru ako su recipijenti altruističnog postupka bliski rođaci (Hamilton 1963, 1964a). Prema Hamiltonovoj teoriji, da bi gen za altruizam bio selektovan nije nužno da natprosečno povećava adaptivnu vrednost organizma koji je njegov nosilac, ako se time značajno umanjuje adaptivna vrednost genetički srodnih jedinki koji su potencijalni nosioci kopije istog gena. S druge strane, gen za altruizam može da bude selektovan i jedinka može da poveća svoju adaptivnu vrednost uprkos šteti koju sama trpi, ako altruizmom doprinosi povećanju adaptivne vrednosti drugih nosilaca kopije istog gena. Argument je zasnovan na proceni verovatnoće da su dve jedinke u srodstvu i odnosu štete i koristi od altruističkog čina. Rezultat njegovih analiza je ono što je danas poznato kao Hamiltonovo pravilo:  $rb > c$ .

U Hamiltonovom pravilu, kojim se specificuju uslovi za evoluciju altruizma u srodničkom krugu,  $b$  je dobrobit za recipijenta altruističnog postupka,  $c$  je šteta koju trpi jedinka koja se ponaša altruistički, dok je  $r$  koeficijent srodstva. Verovatnoća da dve jedinke imaju kopiju istog gena (iznad proseka u populaciji) predstavlja koeficijent srodstva. Koeficijent srodstva predstavlja diskontni parametar, koji umanjuje vrednost koristi od altruističkog čina (Krebs 2011).

Ako zamislimo da postoji gen za altruizam, onda iz perspektive gena nije od naročitog značaja da li takva vrsta ponašanja koristi jedinki koja gen poseduje, već da li koristi samom genu. Gen za altruizam može da se proširi u populaciji ako jedinka svojim ponašanjem doprinosi povećanju frekvencije tog gena u genskom fondu, a to može da se dogodi ako je altruističko ponašanje usmereno prema onim jedinkama za koje postoji određeni stepen verovatnoće da poseduju kopiju istog gena. Kako jedinka koja postupa altruistički trpi štetu od takvog postupanja njoj se isplati da bude altruistična samo ako postoje izgledi da će njen postupak

povećati šanse preživljavanja i reprodukcije one jedinke koja joj je genetički identična po poreklu, odnosno koja s jedinkom koja postupa altruistički ima jedan deo zajedničkih gena.

Da bismo uočili zašto se isplati ulagati u srodnike, možda je najbolje poslužiti se primerom socijalnih insekata iz roda *Hymenoptera* koji se razmnožavaju haplodiploidno (Hamilton 1964b). Kod medonosne pčele, na primer, trutovi (mužjaci) se razvijaju iz neoplođenih jaja i haploidni su, dok se radilice (ženke) razvijaju iz oplođenih jaja i diploidne su. Diploidna matica proizvodi (mejoza) i polaže haploidna jaja, dok je mužjak haploidan. Budući da je mužjak haploidan čitav set njegovih hromozoma koje je nasledio od matice se nalazi i u gametima. Ovaj set hromozoma će preneti na kćeri. S majčine strane kćeri će naslediti  $\frac{1}{2}$  gena, jednako koliko i sinovi. Koeficijent srodstva matice sa sinovima i kćerima je  $r = \frac{1}{2}$ , ali koeficijent srodstva među radilicama (sestrama) je značajno veći, jer s majčine strane nasleđuju jednu polovinu gena, dok s očeve strane nasleđuju čitav genom:  $\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1 = \frac{3}{4}$ . U genetičkom interesu radilica je da pomažu matici da proizvodi sestre ( $r = \frac{3}{4}$ ) umesto da same imaju potomstvo s koeficijentom srodstva  $r = \frac{1}{2}$ .

Tabela 2

	<b>Matica</b>	<b>Otac</b>	<b>Sestra</b>	<b>Brat</b>	<b>Sin</b>	<b>Kćer</b>
<b>Ženka</b>	0.5	0.5	0.75	0.25	0.5	0.5
<b>Mužjak</b>	1	0	0.5	0.5	0	1

Uobičajena pogrešna interpretacija *Hamiltonovog pravila* jeste ona prema kojoj ovo pravilo funkcioniše nalik nekakvom načelu distributivne pravde u srodničkom krugu, tako da se altruizam distribuira proporcionalno koeficijentu srodstva (McKenzie Alexander 2008). Izgleda da ovakvu vrstu pogrešne interpretacije predlažu Stevens, Kušman i Hauser (Stevens, Cushman, and Hauser 2005). Strogo govoreći, *Hamiltonovo pravilo* samo specifikuje uslove pod kojima altruizam može da evoluirati, a ne kako se on distribuira među srodnicima. Iako bi se moglo očekivati da s opadanjem ili rastom stepena srodstva, proporcionalno opada ili raste i stepen altruizma koji jedinka pokazuje prema srodniku, ova veza ipak nije nužna. Naime, ne sledi da će individua u svim slučajevima davati preferencijalni status rođacima, kao što ne sledi da će individua da pokazuje jednak nivo altruizma prema individuama s jednakim koeficijentom

srodstva.<sup>24</sup> Ono što sledi jeste da srazmerno rastu, odnosno opadanju koeficijentu srodstva, rastu ili opadaju šanse za evoluciju altruizma. Jedinke, kao što ćemo videti, nemaju nekakv detektor kojim bi mogle *precizno* da utvrde ko su im srodnici, a još manje precizno mogu da utvrde koeficijent srodstva.<sup>25</sup>

### 2.5.1 *Prepoznavanje srodnika*

Za evoluciju altruizma procesom selekcije u srodničkom krugu sposobnosti jedinke da prepozna s kojim jedinkama je u srodstvu je važan, mada ne i nužan preduslov. Sposobnost prepoznavanja je relaciono svojstvo, a prepoznavanje je trijadna relacija koja uključuje pošiljaoca signala, signal koji je medijator između pošiljaoca i primaoca, i primaoca koji uočava i tumači signal pošiljaoca (Halpin 1991, Tucić 2003, Pfennig 2002). Prepoznavanje je značajno, jer omogućava jedinkama da naprave distinkciju između srodnika i nesrodnih jedinki, kao i između bližih i daljih srodnika. Životinje, uključujući i ljude, nemaju nekakav pouzdani biološki algoritam na osnovu koga mogu precizno da identifikuju jedinke s kojima su u srodstvu, već se u prepoznavanju oslanjaju na neposredne signale (obeleživače) i heurističke metode. Posledica uspešnog prepoznavanja je razlikovanje srodnika od nesrodnih jedinki, a zatim i razlika u bihevioralnom obrascu koji prati identifikaciju. Jedinke koje su uspešnije u prepoznavanju i usmeravanju pomoći ka srodnicima, umesto ka nesrodnim jedinkama ima veće šanse da njihovi geni budu preneti u narednu generaciju. Neposredni mehanizmi za prepoznavanje srodnika na koji se jedinke obično oslanjaju odnose se na neposrednu udaljenost, prisnost (familijarnost), fenotipsko prepoznavanje i prepoznavanje genskih alela (Cartwright 2000, Halpin 1991, Krebs

---

<sup>24</sup> Ipak treba imati na umu da je korelacija između biološkog srodstva i altruističkog ponašanja kod ljudi potvrđena eksperimentalno u različitim kulturnim kontekstima (Madsen et al. 2007). U kontrolisanim uslovima ljudi su zaista spremni da se žrtvuju više za bliže srodnike nego za dalje. Takođe, spremniji su da se žrtvuju više za srodnike nego za nesrodne individue ili grupe. Preciznije, ljudi su spremni da duže trpe bol koji je izazvan neprijatnom izometrijskom vežbom, ako će za to biti nagrađen neki njihov blizak rođak ili oni sami, za razliku od toga koliko su dugo spremni da trpe bol ukoliko nagrada ide daljem rođaku, prijatelju (nesrodniku) ili nekoj dobrotvornoj organizaciji.

<sup>25</sup> Takođe, u evolucionoj ravni ne vrede sve jedinke s jednakim koeficijentom srodstva jednako. Srodnici koji su na pragu seksualne zrelosti ili vrhuncu reproduktivne moći za jedinku vrede više od onih koji su suviše mladi ili suviše stari da bi se reprodukovali. Naglasak je na budućoj reprodukciji, a ne na prošloj. Vidi (Fisher 1930).

2011). Takođe, kada je reč o ljudima, prepoznavanje srodnika zavisi i od podučavanja ili davanja instrukcija (Dunbar 2009). Ovde ću se usredsrediti na prva tri mehanizma.

**Neposredna udaljenost ili prepoznavanje na osnovu prostornih odnosa.** Ovaj način prepoznavanja se odnosi na korelaciju koja postoji između teritorijalne bliskosti i srodstva. Ako se jedinke rađaju i žive u porodičnim grupama na određenoj lokaciji, zatim tu provode jedan vremenski period, a nakon disperzije nastavljaju da žive u blizini lokacije na kojoj su rođeni, za takve jedinke postoji velika verovatnoća da su joj neposredni susedi srodnici (Cartwright 2000, Halpin 1991, Krebs 2011). Kako je u ovom slučaju identifikacija srodnika potpuno nasumična, budući da jedinka prema svim drugim jedinkama u neposrednoj blizini pokazuje identičan obrazac ponašanja, ovakav način prepoznavanja je krajnje nepouzdan. Iako nije reč o pravom prepoznavanju, već o prostorno posredovanom nepotizmu (Halpin 1991), altruističkim postupanjem prema onim jedinkama koje su u neposrednoj blizini, ipak se uvećavaju šanse da se altruistički postupa prema srodnicima. Kako je identifikacija srodnika na ovaj način nasumična i nepouzdana, često može dolaziti do lažnog prepoznavanja (Pfennig 2002). Ukoliko učestalo dolazi do pogrešne identifikacije nesrodnika kao srodnika, to jedinki koja altruistički postupa značajno može da umanjí adaptivnu vrednosti. Ovo bi moglo da bude kompenzovano ako jedinka deli teritoriju s velikim brojem srodnika. Povremene pogrešne identifikacije nesrodnika kao srodnika (eng. *false positives*) bi s druge strane mogle da vode proširenju (evoluciji) altruizma i saradnje izvan srodničkog kruga. Osim toga, može se očekivati da one jedinke koje učestalo greše u identifikaciji budu izložene eksploataciji i izumiranju.

**Fenotipsko prepoznavanje.** Jedinke koje koriste ovaj mehanizam prepoznavanja srodnika oslanjaju se na signale koji su sasvim izvesno povezani s genetičkim srodstvom, ali ne nužno i s prethodnim boravkom na istoj lokaciji ili zajedničkim odrastanjem (Halpin 1991). Prepoznavanje će biti najuspešnije ako je utemeljeno na prethodno naučenim fenotipskim karakteristikama koje su direktna ekspresija genoma. U tom slučaju jedinka može da generalizuje prethodno iskustvo i da sve druge jedinke koje poseduju istu oznaku prepoznaje kao srodnike procenom genetičke sličnosti između sopstvenih karakteristika i karakteristika druge jedinke.



Uobičajeni način prepoznavanja srodnika među životinjama je na osnovu hemijskih signala koje životinje šalju i koje druge životinje mogu da prepoznaju. Slanje hemijskih signala je načelno jeftino, to jest ne predstavlja veliki energetska trošak za jedinku, jer je reč o hemijskim supstancama koje su uglavnom sporedni proizvod metabolizma, a imunološki sistem životinja ih lako prepoznaje i tumači (Tucić 2003, Pfennig 2002). Najčešće je reč o prepoznavanju specifičnog mirisa što objašnjava pojavu njuškanja među mnogim životinjama (Cartwright 2000, Krebs 2011). Takođe, među sisarima, važnu ulogu u prepoznavanju imaju i vizuelni i auditivni signali (Halpin 1991). Kada je reč o ljudima, uočavanje i prepoznavanje “porodične sličnosti” jedan je od standardnih načina za prepoznavanje srodnika na osnovu vizuelnih fenotipskih karakteristika, preciznije sličnosti u izgledu lica, boji očiju, kose, itd. (Krebs 2011, Halpin 1991). Tako na primer, majke mogu da prepoznaju fotografije svoje bebe među fotografijama drugih beba dan nakon rođenja (Halpin 1991).

Osim toga, glikoproteini, koje kodiraju veoma varijabilni geni glavnog histokompatibilnog kompleksa (MHC), imaju najmanje dve funkcije (Dugatkin 2000, Tucić 2003). Najpre, oni funkcionišu kao obeleživači identiteta i njihova osnovna funkcija se odnosi na imuni odgovor organizma – oni omogućavaju organizmu da nešto prepozna kao svoje ili tuđe i da se izbori s patogenima. Osim toga, oni imaju funkciju u prepoznavanju srodnika. Zbog velike varijabilnosti ovog skupa gena, veoma mali broj individua ima potpuno isti MHC, ali bliski srodnici pokazuju daleko veću sličnost u strukturi MHC nego one jedinke koje nisu u srodstvu. Ove razlike u MHC jedinke mogu da detektuju čulom mirisa i one utiču na izbor seksualnog partnera. Naime, jedinke izbegavaju da se ukrštaju s jedinkama koje imaju sličan MHC što vodi izbegavanju inbridinga (incesta) i potencijalno nepovoljnih genetičkih efekata na fenotip potomaka.

Seksualne preferencije uzrokovane razlikama u strukturi MHC najpre su otkrivene i potvrđene u populacijama glodara (Dugatkin 2000, Tucić 2003). Ali, to ne implicira da je mogućnost prepoznavanja razlike u kombinaciji MHC putem mirisa ljudima potpuno van domašaja. Zapravo, prepoznavanje mirisa ima značajnu ulogu u prepoznavanju srodnika i izboru seksualnih partnera čak i kod ljudi (Halpin 1991, Pfennig 2002). Tako na primer, majke mogu da prepoznaju miris (odeće) svog novorođenčeta 24 - 48 sati posle porođaja, iako su u kontaktu s bebom bile samo jedan sat ili manje, što očevi ne mogu. Takođe, bebe mogu da prepoznaju miris

svoje majke. Čak i osobe koje nisu u srodstvu mogu da upare majke i bebe samo na osnovu mirisa. Osim toga, roditelji mogu da prepoznaju miris i svoje starije dece u odnosu na miris dece drugih roditelja, a mogu i sopstvenu decu međusobno da razlikuju po mirisu. Na isti način deca mogu da prepoznaju svoju braću i sestre. U eksperimentu koji je postao poznat kao *Eksperiment znojavih majica* utvrđeno je da žene prosuđuju miris muškarca kao prijatniji i seksualno privlačniji, ako je razlika u MHC veća (Wedekind et al. 1995). Načelno, to znači da ljudi doživljavaju miris srodnika kao neprijatan i da ih ne doživljavaju kao potencijalne partnere za parenje, što uz *Vestermarkov efekat*, o kome ću nešto više reći kasnije, predstavlja prirodnu branu protiv incesta (Cartwright 2000).

**Prisnost ili prepoznavanje na osnovu udruživanja (asocijacije).** Ovaj, među kičmenjacima najčešće korišćeni mehanizam identifikacije srodnika, temelji se na činjenici da u prirodnim uslovima postoji manje više pouzdana korelacija između teritorijalne bliskosti, učestalosti interakcija i fenotipske sličnosti, a kod ljudi se manifestuje i snažnim emotivnim vezama među osobama koje su zajedno odrasle i zajedno provele detinjstvo (Korchmaros and Kenny 2001, 2006, Krebs 2011, Halpin 1991). Osnovna razlika između ovog načina prepoznavanja, s jedne strane, i identifikacije srodnika na osnovu prostornih odnosa, s druge strane, sastoji se u tome što jedinke uče da prepoznaju srodnike na osnovu oznaka koje mogu imati genetičko ili sredinsko poreklo, a zatim ono naučeno koriste u kontekstima koji se razlikuju od konteksta u kome su prvobitno stekle iskustvo (Halpin 1991). Uzimajući u obzir evolucionu prošlost čoveka i to da su predačke populacije ljudi živele u srodnički tesno povezanim zajednicama i imajući u vidu da su i savremene ljudske zajednice uglavnom organizovane na porodičnim osnovama, osobe s kojima se zajedno odrasta i s kojima se razvijaju snažne emotivne veze su najčešće srodnici (Korchmaros and Kenny 2001, 2006). Afektivna privrženost osobama koje se nalaze u neposrednoj blizini, s kojima se provodi rano detinjstvo, s kojima se učestalo stupa u interakcije i koje su međusobno slične zbog izloženosti sličnim sredinskim uslovima može da posluži kao još jedno heurističko sredstvo za identifikaciju srodnika i kao medijator za evoluciju altruizma (Korchmaros and Kenny 2006, Dunbar 2009). Oni koji su u neposrednoj blizini, koji su poznati i s kojima je individua emocionalno povezana su verovatno srodnici. Ali, kako se među decom koja zajedno odrastaju razvijaju snažne emocionalne veze bilo da su u srodstvu ili ne, to ima za posledicu da se i prema nesrodnim jedinkama ispoljavaju obrasci

altruističkog ponašanja. Na taj način može da dođe do evolucije altruizma koji prevazilazi granice srodničkog kruga.

Ako sada ostavimo po strani problem prepoznavanja, postoje dva važna fenomena koja su naizgled suprotstavljena predviđanjima Hamiltonove teorije. Jedan je usvajanje jedinki izvan srodničkog kruga. Ako se iz pozicije gena ne isplati ulagati u one jedinke koje nisu srodnici, onda je usvajanje van srodničkog kruga s kojim se susrećemo u ljudskim (i ne samo ljudskim) zajednicama u najmanju ruku veoma neobična pojava. S druge strane, iz Hamiltonove teorije sledi da bi za jedinku bilo najbolje da se ukršta u srodničkom krugu i da na taj način poveća inkluzivnu adaptivnu vrednost, ali ovo predviđanje se kosi s činjenicom izbegavanja inbridinga među životinjama, kao i s postojanjem konvencionalnih pravila koja zabranjuju stupanje u seksualne odnose sa srodnicima (tabu incesta). O ovim problemima će biti reči u naredna dva odeljka.

### ***2.5.2 Problem usvajanja***

Ako napravimo distinkciju između biološke i društvene kategorije srodstva, onda možemo da zapazimo na koji način se emocionalna privrženost u srodničkom krugu (što je biološka kategorija) i altruizam proširuju izvan njega, najpre na društvenu kategoriju srodstva, a zatim i na druge međuljudske relacije koje uključuju afektivnu privrženost. Možda je to najbolje ilustrovati primerom. Usvajanje nesrodnih jedinki je, kako se čini, potpuno altruistički čin jer je investiranje izvan srodničkog kruga i bez šansi za reciprocijaciju iz evolucione perspektive uglavnom pogubno. Na taj način, usvajanje nesrodnih jedinki je obrazac ponašanja koji bi bar u nekim slučajevima (pod uslovom da zanemarimo mogućnost konstruisanja nategnutih egoističkih objašnjenja) mogao da bude protumačen kao altruistički u bihejvioralnom, psihološkom i biološkom smislu. Međutim ispostavlja se da je ovaj primer, kao i društvena kategorija srodstva, više nego problematičan, iako sam čin usvajanja nije način postupanja koji je karakterističan isključivo za ljudska bića. Sa usvajanjem se susrećemo među mnogim životinjama: pticama (Brown 1998, Redondo, Tortosa, and deReyna 1995), severnim morskim lavovima (Riedman and Le Boeuf 1982), crnočelim titi majmunima i crvenim drekavcima (Agoramoorthy and Rudran 1992, Cäsar and Young 2008), a čak je zabeležen jedan slučaj među različitim rodovima majmuna, kapucinera i marmozeta (Izar et al. 2006). Mada se pretpostavljalo

da je usvajanje mladunaca među šimpanzama retko, istraživači su uspjeli da zabeleže 18 slučajeva tokom 27 godina, a ono što je zanimljivije, jeste podatak da je gotovo u polovini zapaženih slučajeva usvojitelj mužjak (Boesch et al. 2010). Međutim, usvajanje među životinjama je pre izuzetak nego pravilo; ponekada se zapravo radi samo o izolovanim slučajevima, a usvajanje izvan srodničkog kruga je prava retkost. Do slučajeva usvajanja izvan srodničkog kruga dolazi pre svega zbog teškoća u identifikaciji srodnika ili zbog toga što jedinka koja usvaja može od toga da ima neku sekundarnu reproduktivnu korist.

Moglo bi se pomisliti da praksa usvajanja u ljudskim zajednicama prkosi predviđanjima Hamiltonove teorije, ali izgleda da to nije slučaj. Proučavanjem ovih praksi u društvima malog obima došlo se do podataka da većina hranitelja i osoba koje usvajaju decu dolazi iz srodničkog kruga, tako da nije bilo teško zaključiti da je udeo usvajanja izvan srodničkog kruga u evolucionoj prošlosti čoveka bio zanemarljiv (Silk 1987). Zapravo, usvajanje potpunih stranaca je novina karakteristična za visoko razvijena industrijalizovana društva i najverovatnije je uzrokovana kulturnim standardima koji su prilično udaljeni od svojih bioloških korena (Barnard 2004, Salmon 2005). Osim toga, deca su posebno osetljiva kategorija i njihovo usvajanje izvan srodničkog kruga i odrastanje bez prisustva ili u prisustvu samo jednog biološkog roditelja povezano je s većom verovatnoćom diskriminacije, zanemarivanja i zlostavljanja sa smrtnim ishodom (Daly and Wilson 2000). Iako je ovaj zaključak naišao na osporavanja zbog navodnih metodoloških propusta i pristrasne interpretacije podataka (Buller 2005), ipak se čini da priče o zlim maćehama i očusima nisu samo plod puke mašte (Daly and Wilson 2001, Tooley et al. 2004).

### **2.5.3 Vestermarkov efekat i problem izbegavanja incesta**

Zajedničko provođenje najranijeg detinjstva (aproksimativno negde od rođenja pa do desete godine života) ima za posledicu pojavu *Vestermarkovog efekta* (Wolf 2005). Reč je o prirodnoj seksualnoj averziji među osobama koje su zajedno provele najranije detinjstvo. Budući da su osobe koje zajedno provode najranije detinjstvo najčešće srodnici, ovaj fenomen može da se protumači kao pravilo izbegavanja incesta (inbridinga) u ljudskim zajednicama koje nije utemeljeno na kulturnim normama (*tabu*). Ovaj fenomen je dobro dokumentovan u istraživanjima običaja koji je poznat kao *Mala snaha* ili *Mala nevesta* (Wolf 2005). *Mala snaha*

predstavlja vrstu ugovorenog braka gde devojčicu u ranom detinjstvu usvaja porodica za čijeg sina će se ona kasnije udati, tako da budući supružnici odrastaju zajedno. Brakovi sklopljeni na ovaj način su u biološkom i socijalnom smislu daleko manje uspešni od onih koji se sklapaju među odraslim osobama, jer ih karakteriše visok stepen neplodnosti, bračnog neverstva i razvoda. *Vestermarkov efekat*, poduprt seksualnim preferencijama uzrokovanim razlikama u strukturi MHC, konstituše biološku inhibiciju koja je najverovatnije uzrokovala naknadnu moralnu zabranu stupanja u seksualne odnose s bliskim rođacima.

Ovakvo razumevanje odnosa inhibicije i zabrane suočilo se s prigovorom Bernarda Vilijamsa koji ističe da se u ovom slučaju konceptualni sadržaj zabrane ne podudara sa sadržajem inhibicije (Williams 1995). Kako sadržaj inhibicije nije predstavljen na adekvatan način u konceptualnom sadržaju zabrane, jer se prvo odnosi na averziju prema stupanju u seksualne odnose s osobama s kojima se provodi detinjstvo, a drugo na stupanje u seksualne odnose s bliskim srodnicima, kauzalni odnos može da bude doveden u pitanje. Međutim, ova nepodudarnost nestaje ili postaje irelevantna ako imamo na umu da prirodna selekcija ne isporučuje optimalna, već jednostavna i nepouzdana, ali dovoljno efikasna rešenja za adaptivne probleme. Prirodna selekcija je slepa i ne zna za bilo kakve kategorije, pa ni za kategoriju srodstva. U objektivnoj ravni u kojoj deluje prirodna selekcija averzija nije usmerena na kategoriju srodstva nego na ono što je u neposrednoj blizini čime se efikasno umanjuje verovatnoća inbridinga. S druge strane, ona u subjektivnoj ravni može da se ispolji na najjednostavniji mogući način tako što oni koji prepoznaju ovu averziju, da bi je učinili smislenom, referiraju na kategoriju bliskog srodstva (otac, majka, brat, sestra), umesto da se pozivaju na neodređeni pojam zajedničkog provođenja najranijeg detinjstva (Sesardic 2005).

Drugačiju vrstu kritike *Vestermarkovog efekta* upućuje Prinz (Prinz 2007). Najpre, on smatra da evidencija koja se navodi kao primer *Vestermarkovog efekta* (*Mala snaha*) nije pouzdana i da neuspeh maloletničkih brakova može pre da se pripíše lošem socioekonomskom statusu, nego izbegavanju incesta. Na drugom mestu, zabrana incesta u mnogim kulturama se ne odnosi samo na osobe s kojima se deli domaćinstvo, već i na dalje srodnike, što koncepcija *Vestermarkovog efekta* ne može da objasni. I konačno, moguće je da je rizik od potencijalnih negativnih posledica inbridinga zapravo mnogo manji nego što se pretpostavlja. Na primer, u društvima malog obima genetička raznolikost je veoma mala, ali su njihovi članovi ipak zdravi.

Umesto *Vestermarkovog efekta*, Princ predlaže koncepciju prirodne sklonosti ka *egzogamiji*, koja je s njim kompatibilna, ali ima veću eksplanatornu moć i konzistentna je s bihevioralnim obrascima šimpanzi i bonoboa, naših najbližih evolucionih srodnika: s postizanjem seksualne zrelosti, razvija se seksualna sklonost ka osobama izvan porodičnog kruga. Egzogamija ima najmanje dve biološke koristi: 1.) štiti od potencijalnih negativnih posledica inbridinga (ma koliko da su one male) i 2.) plasiranje gena izvan srodničkog (porodičnog) kruga može da bude izlaz iz genetičkog ćorsokaka i šansa za povećanje adaptivne vrednosti, bar u onim situacijama kada porodici prete bolesti, nesreće, elementarne nepogode, itd.

Mada je u različitim kulturama i epohama bila konceptualizovana na različite načine, zabrana incesta je gotovo univerzalna. Pravila koja regulišu stupanje u seksualne odnose sa srođnicima su uvek uslovljena koeficijentom srođstva, kojim se određuje granica važenja zabrane, i u nekim kulturama su rigidna, dok su u drugim veoma fleksibilna. Uglavnom, incest je osetljiva tema o kojoj se nerado govori. Moguće je da su negativne posledice inbridinga zaista male, ali podaci kojima za sada raspolažemo na osnovu nekoliko sprovedenih istraživanja ukazuju da je zastupljenost autozomnih recesivnih poremećaja, urođenih deformiteta, iznenadne smrti deteta, teških i blagih slučajeva mentalne zaostalosti kod potomaka incestuoznih parova znatno viša nego kod neincestuoznih (Bittles 2004). Pored genetičkih uzroka, ne bi trebalo zanemariti ni uticaj sledećih činilaca: majka može biti suviše mlada i nedovoljno polno razvijena, oba roditelja mogu biti suviše stari, jedan ili oba roditelja mogu imati neki fizički i/ili mentalni poremećaj. Osim toga, loš socioekonomski status, kao i neki specifični nepovoljni sredinski i društveni uslovi uključujući pokušaj abortusa, mogu doprineti nepovoljnim razvojnim ishodima. Uprkos uticaju drugih činilaca i uprkos činjenici da su istraživanja kojima raspolažemo nepotpuna i ponekada metodološki sporna, studije (metaanalize) koje osporavaju pouzdanost podataka o nepovoljnim posledicama inbridinga bi trebalo uzeti s rezervom.

## **2.6 Direktni reciprocitet i evolucija saradnje**

Iako je Robert Trivers (Robert Trivers) bio prvi koji je koncept recipročnog altruizma naučno (teorijski precizno) definisao, Dejvid Hjum (David Hume) je prvi jasno uočio važnost principa reciprociteta za regulaciju društvenih odnosa. U *Raspravi o ljudskoj prirodi* on piše da

drugima činimo usluge bez ikakve prisnosti, već samo zato što očekujemo da nam usluga bude uzvraćena (Hume 1960). Ovakva vrsta postupanja ima bar dva nedostatka. Prvi nedostatak bi bio taj da je onaj ko je napravio prvi korak i učinio uslugu u strateškom zaostatku u odnosu na onoga ko uslugu prima. Hjum je bio uveren da će on, time što mu se čini usluga, biti podstaknut da uslugu uzvрати, jer uviđa da bi mogao da trpi posledice od drugačijeg postupanja. Uprkos tome što je Hjum mislio da se na principu reciprociteta temelje dobri međuljudski odnosi i da on može da posluži kao valjano regulatorno načelo u bezličnim relacijama, Darwin je bio uveren da je, u moralnom pogledu, pomaganje u očekivanju da pomoć bude uzvraćena utemeljeno u niskim pobudama (Darwin 1981), a izgleda da je sličnog mišljenja bio i Džordž Vilijams (Williams 1996 [1966]). U svakom slučaju, koncept reciprociteta kao regulatornog načela nije bio nepoznat istraživačima pre Triversove analize.<sup>26</sup>

Mehanizam na osnovu koga funkcioniše recipročni altruizam je mehanizam razmene usluga na obostranu korist: ja činim tebi, ako ti činiš meni i obrnuto (Trivers 2006). Subjekt (donor) i recipijent, u slučaju recipročnog altruizma, imaju koristi od altruističkog postupka, ali nije nužno da motivacioni mehanizam koji stoji iza postupka bude altruistički. Dovoljno je samo to da se individui u evolucionom smislu *isplati* da postupa altruistički, ako očekuje da će joj u budućnosti biti uzvraćeno istom merom, odnosno ako postoje izgledi da će ukoliko uskrati pomoć, i njoj pomoć da bude uskraćena. Nužan uslov za evoluciju recipročnog altruizma je da donor (ona jedinka koja pomaže) trpi malu štetu, a da recipijent (ona jedinka kojoj se pomaže) ima veliku korist od altruističkog postupka (Trivers 1971). Ostali neophodni uslovi su da interakcije jedinki budu učestale, da je u ponovljene interakcije uključen mali broj jedinki i da budu simetrične. Simetričnost podrazumeva ne samo da recipijenti ponekada budu i donori, već da i teret saradnje, tj. pružanja i uzvrćanja pomoći, bude ravnomerno raspoređen (uravnotežen).

Za razliku od Hamiltonovog modela selekcije u srodničkom krugu kojim je objašnjeno na koji način altruizam može da evoluiru u zajednicama bliskih srodnika, pozivanje na reciprocitet

---

<sup>26</sup> Primer pozivanja na reciprocitet koji najviše iznenađuje nalazimo u Kantovoj etici dužnosti. Kant u *Zasnivanju metafizike morala* tvrdi da bi volja koja bi donela odluku da ne pomogne drugome u nevolji protivrećila samoj sebi, „pošto se ipak mogu desiti neki takvi slučajevi u kojima će dotičnom biti potrebna ljubav ili saućešće drugih ljudi, i u kojima bi on na osnovu jednog takvog prirodnog zakona, poniklom iz njegove vlastite volje, lišio sama sebe svake nade u tuđu pomoć koju pri tom želi“ (Kant 2008, 65).

imalo je značajnu eksplanatornu snagu kada je reč o objašnjenju saradnje među individuama koje nisu u srodstvu. Princip reciprociteta omogućava jedinkama da altruistički postupaju i prema onim individuama s kojima nisu identične po poreklu. Za evoluciju recipročnog altruizma nije bitno da li jedinke dele kopije istih gena, niti da li su iste vrste, čak ni da li imaju sofisticirane kognitivne kapacitete koji bi im omogućili da na precizan način vode evidenciju prethodnih kontakata.<sup>27</sup>

Deljenje hrane među slepim miševima vampirima koji se hrane krvlju domaćih i divljih životinja jedan je od često navođenih primera recipročnog altruizma među sisarima (Wilkinson 1988, 1990). Budući da ne mogu da prežive dugo bez hrane, one jedinke koje nisu uspele da dođu do obroka (često mladunci), moraju da se oslone na pomoć drugih članova kolonije. Kako ove životinje mogu da popiju mnogo više krvi nego što im je potrebno, one izgladnelim članovima kolonije predaju deo sadržaja svog želuca. Iako se većina razmena odvija na relaciji majka – mladunče, ili među bliskim srođnicima, jedan deo razmena odvija se među nesrodnim jedinkama iste populacije. Ustanovljeno je da među nesrodnim jedinkama koje se međusobno ispomažu postoji „drugarski“ sistem u kome su dve jedinke dugogodišnji partneri u međusobnim razmenama, čime je osigurano da usluga bude uzvraćena. Ovi slepi miševi imaju sasvim dobro pamćenje i mogu da se međusobno prepoznaju na osnovu mirisnih i auditivnih signala. Zbog specifičnog načina organizacije kolonije oni mogu da identifikuju varalice i da im uskrate pomoć u budućnosti.

Kao primere recipročnog altruizma, osim signala upozorenja kod ptica i simbioze riba čistača i njihovih domaćina predatora, Trivers (1971) navodi i nekoliko oblika ljudskog ponašanja. Ti oblici ponašanja su pomaganje u nevolji, deljenje hrane, pomaganje bolesnima, ranjenima, onima koji su suviše mladi ili stari da se brinu o sebi, deljenje oruđa, kao i deljenje

---

<sup>27</sup> Ima i drugačijih mišljenja. Nedostatak dovoljno sofisticiranih kognitivnih kapaciteta navodi se kao objašnjenje zbog čega je recipročni altruizam dokumentovan samo kod relativno malog broja vrsta koje takve kapacitete imaju (Stevens, Cushman, and Hauser 2005). U pitanje je doveden i primer sa slepim miševima vampirima, budući da se veliki broj recipročne razmene obroka odvija među srođnicima, tako da bi evolucija ove vrste ponašanja mogla da bude objašnjena selekcijom u srodničkom krugu. Za odgovor na ovaj prigovor vidi (Carter and Wilkinson 2013).



znanja. Prema njegovim rečima, svi ovi oblici ponašanja zadovoljavaju kriterijum asimetričnosti između štete i koristi koja je neophodna za evoluciju recipročnog altruizma.

Kao što je rečeno, da bi recipročni altruizam postao efektivan, neophodno je da interakcije individua budu učestale, jer su jednokratne interakcije, u kojima postoji sukob interesa, veliki izazov za varanje. U jednokratnim interakcijama za svaku jedinku je pojedinačno bolje je da pokuša da izigra drugu, to jest da ne uzvraća pomoć ako joj je pomoć jednom pružena. Na primer, ako jedinka  $A$  bište jedinku  $B$  u vremenskom intervalu  $t$ , onda je u interesu jedinke  $B$  da uskrati bištenje jedinki  $A$  u vremenskom intervalu  $t_1$  i da vreme koje bi koristila za bištenje iskoristi, recimo, za traženje hrane i parenje. Postojanje vremenskog jaza između  $t$  i  $t_1$ , čini da je  $A$ , ako prva bište, u strateškom zaostatku u odnosu na  $B$ ; jednokratne interakcije su prilika za eksploataciju. Ako pretpostavimo da jedinka  $A$ , na neki način zna da bi mogla da postane žrtva eksploatacije, onda bi u njenom najboljem interesu bilo da ne bište jedinku  $B$ . Na taj način, proces reciprocijacije uopšte ne može ni da otpočne.

Da bismo to ilustrovali poslužićemo se jednostavnim analitičkim modelom koji je poznat kao *Dilema zatvorenika* i već je uobičajeno sredstvo za analizu situacija u kojima postoji sukob interesa. Ukoliko su interakcije jedinki jednokratne, kao u jednokratnoj *Dilemi zatvorenika* u kojoj je dominantna strategija (racionalnih subjekta) uvek nesaradnja, izgledi za evoluciju altruizma i saradnju su veoma mali. Ako zamislimo dijadnu relaciju u kojoj svaka jedinka zastupa isključivo vlastiti interes, ishod njihovog odlučivanja će biti suboptimalan i za njih pojedinačno i za grupu kao celinu, u odnosu na ishod koji bi bio ostvaren da nisu zastupali isključivo sopstveni interes

*Dilemu zatvorenika* u njenom uobičajenom obliku možemo formalno izraziti na sledeći način (Axelrod 1984):

1. U igri učestvuju dva igrača i imaju dva alternativna izbora – saradnja (C) ili nesaradnja (D).
2. Svaki od igrača mora da načini izbor bez prethodnog znanja o izboru njegovog saigrača.
3. Ako jedan od igrača izabere C, a drugi D, onaj koji je izabrao D dobija više.

4. Ako oba igrača izaberu C ishod je bolji nego u slučaju da su oba izabrala D.
5. Mogući ishodi za svakog igrača su sledeći: R (nagrada) za međusobnu saradnju, S (naivčina) za jednostranu saradnju, T (iskušenje) za jednostranu nesaradnju i P (kazna) za međusobnu saradnju.
6. Vrednosti isplata su:  $R = 3$ ,  $S = 0$ ,  $T = 5$ ,  $P = 1$ .
7. Odnos preferencija mogućih ishoda i odgovarajućih isplata je:  $S < P < R < T$  i važi da je:  $2R > S + T$ .

Tabele 3 i 4

	C	D
C	R R	S T
D	T S	P P

	C	D
C	3 3	0 5
D	5 0	1 1

U tabelama 3 i 4 prikazani su ishodi igre i njihovi numerički ekvivalenti, odnosno isplate igre. Odmah postaje uočljivo da je dominantna strategija za oba igrača nesaradnja, jer ukoliko jedan igra kooperativno, a drugi ne, onaj koji je igrao nekooperativno dobija **5 poena**, za razliku od onog koji je igrao kooperativno i nije dobio ništa (**0 poena**). Ukoliko jedan od igrača igra nekooperativno, onom drugom se i u ovom slučaju više isplati da igra nekooperativno, jer u svakom slučaju dobija **1 poen** što je više nego ništa (**0 poena**) koliko bi dobio da je igrao kooperativno. Međutim, kada bi oba igrača igrala kooperativno, svaki od njih bi dobio **3 poena**. Kako nijedan od igrača neće igrati dominirane strategije, ovakav ishod igre je van njihovog domašaja.

Dominantne strategije **D,D** i odgovarajući ishod igre **P,P**, konstituišu *Nešov ekvilibrijum*, odnosno, pozicija oba igrača je takva da ni jedan od njih ne može jednostrano da promeni strategiju, a da pri tome ne dovede sebe u nepovoljan položaj. Ovaj ishod igre je stabilan. Kada uočimo da je ishod igre **R,R** kada su strategije **C,C** *Pareto optimalan* (nijedan od igrača ne može da poboljša svoj položaj, a da ne pogorša položaj drugog), problem postaje očigledan (Axelrod and Hamilton 1981). Međutim, optimalan ishod igre nije stabilan, budući da jednostranom promenom strategije jedan od igrača svoj položaj može da učini boljim.

Ovako osmišljena situacija ilustruje sukob koji postoji između individualne i kolektivne racionalnosti, ali jednostavnom transformacijom može da se pokaže na koji način je ona značajna za evoluciju recipročnog altruizma i saradnje. Dovoljno je samo da broj poena zamenimo brojem potencijalnih potomaka i da zatim zapazimo da jedinke koje bezuslovno saraduju i pomažu drugima gube u evolucionoj utakmici. Takođe, možemo da zapazimo da je iz evolucione perspektive obostrana saradnja poželjnija od obostrane nesaradnje, jer oni koji ne saraduju imaju manje potomaka od onih koji saraduju. Problem se sastoji u tome što je, kako se čini, saradnja van domašaja jedinki ako „pretpostavljaju“ da su upletene u jednokratne interakcije u kojima je dominantna strategija nesaradnja.

U jednokratnoj *Dilemi zatvorenika* se pretpostavlja da su se učesnici u igri sreli jednom i da nema izgleda da će se ikada ponovo sresti. Međutim, Trivers ističe da bar kada je reč o ljudima, izuzimajući ubistvo, jednokratnih interakcija jednostavno nema, da su ljudi uvek upleteni u niz interakcija, bez obzira na to koliko su one kratkotrajne (Trivers 2006). Ako se interakcije jedinki ponavljaju, to jest ako su takve da se varalicama može uzvratiti istom merom tako što će im se naneti šteta ili tako što će im se u nekoj narednoj interakciji uskratiti pomoć, onda bi prirodna selekcija mogla da deluje protiv varalica, pod pretpostavkom da je potencijalni dobitak od neiskorišćenih altruističkih činova za varalice veći od potencijalnog gubitka za uzvratanje pomoći (Trivers 1971). U tom smislu, *Ponovljena dilema zatvorenika* je mnogo bolji analitički model, koji na realniji način može da prikaže socijalne odnose.

Kada se *Dilema zatvorenika* ponovi, mogućnost ponovljenih interakcija utiče na evoluciju saradnje, jer izbor strategije u jednom vremenskom periodu može da se naknadno reflektuje na to koju će strategiju izabrati druga strana u igri, što se zatim reflektuje na izbor budućih strategija oba saigrača. U ponovljenim interakcijama značajan činilac za evoluciju saradnje je vreme. Ukoliko su igrači zainteresovani za budućnost, ona može baciti senku na sadašnjost uslovljavajući do izvesne mere sadašnje izbore. Međutim, budućnost se obično manje vrednuje u odnosu na sadašnjost, tako da se isplata svakog narednog poteza vrednuje manje od isplate sadašnjeg. Da bi se ovo izrazilo, isplate možemo posmatrati kumulativno u vremenu tako da svaki naredni potez predstavlja neki razlomak sadašnjeg. Ovaj razlomak predstavlja diskontni parametar ili senku budućnosti (Axelrod 1984). Na primer, zamislimo igru dužine tri, neka su strategije (C,C), (C,C), (C,C) sa odgovarajućim ishodom i isplatama. Kako je malo verovatno da

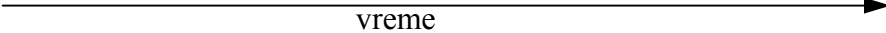
će individue vrednovati budućnost jednako kao i sadašnjost, pretpostavimo da je diskontni parametar  $\delta = \frac{1}{2}$ , isplata je onda  $3 + (\frac{1}{2})3 + (\frac{1}{2})^2 3 = 5,25$ . Kada bi diskontni parametar bio  $\delta = \frac{3}{4}$ , onda bi isplata bila oko 7, a kada bi bio  $\delta = 1$ , odnosno kada bi se svaki naredni potez vrednovao jednako kao i prethodni, isplata za svakog igrača bi bila 9. Što je senka budućnosti veća, veće su i šanse za saradnju. Ako je broj interakcija u koje individue stupaju dovoljno visok i ako na raspolaganju imaju dovoljno vremena, to onda može da utiče na evoluciju saradnje i stabilnost kooperativnih poduhvata.

Kako u *Ponovljenoj dilemi zatvorenika* nema dominantne strategije, Robert Axelrod (Robert Axelrod) je osamdesetih godina prošlog veka organizovao niz računarskih turnira testirajući uspešnost različitih strategija (Axelrod 1984). Strategija koja se pokazala najuspešnijom bila je istovremeno i najjednostavnija - *milo za drago* (eng. *Tit-for-Tat, TFT*). *Milo za drago* je uslovno kooperativna strategija, što znači da igrač u prvom potezu igra kooperativno, a zatim ponavlja svaki naredni potez svog saigrača. Pokazalo se da ova strategija može da „nadigra“ nekooperativne strategije i da sakupi više poena ukoliko je senka budućnosti dovoljno velika, tj. ako na raspolaganju imamo dovoljno dug vremenski period (Axelrod 1984). Budući da je za efektivnost recipročnog altruizma i strategije *milo za drago* ključan činilac vreme, ova strategija ima tu osobinu da, ukoliko je dat dovoljno dug vremenski period u kome bi jedinke mogle da ostvare veliki ili neodređeni broj uzajamnih kontakata, može da iznudi kooperativno ili altruističko ponašanje one strane koja se prethodno ponašala drugačije.

Pretpostavimo da u igri učestvuju dve individue **A** i **B**, i pretpostavimo da **A** primenjuje TFT, dok **B** uvek igra nekooperativno, ALLD. U tom slučaju **A** može biti izigrana samo jednom, dok u svim narednim interakcijama uzvraća jednakom merom. Nesaradnja se kažnjava nesaradnjom. Ali, ako i **B** primenjuje TFT, situacija se potpuno menja. Uparena sa samom sobom TFT se ponaša kooperativno u svim interakcijama. U ovom slučaju saradnja se nagrađuje saradnjom.

Tabela 5

<b>A/tft</b>	C	D	D	D	D	D	D	D
<b>B/alld</b>	D	D	D	D	D	D	D	D



<b>A/tft</b>	C	C	C	C	C	C	C	C
<b>B/tft</b>	C	C	C	C	C	C	C	C

Ako pretpostavimo da se neka prirodna populacija sastoji od više od dve individue i ukoliko se dovoljan broj individua rukovodi strategijom TFT, onda saradnja može odneti prevagu nad nesaradnjom. U slučaju kada populaciju čine uglavnom one jedinke koje ne saraduju, reciprocijacija bi mogla da odnese prevagu nad nekooperativnim strategijama ako bi došlo do asortativnog okupljanja (pravljenja kooperativnih klastera) onih jedinki koje saraduju (Axelrod and Hamilton 1981). Značaj ove strategije je između ostalog i u tome što ako na raspolaganju imamo dovoljno dug vremenski period, ona može indukovati kooperativno ponašanje druge strane. Rezultati Akselrodovih turnira naišli su na širok odjek u naučnoj javnosti, jer su na eksperimentalan način pokazali da princip reciprociteta, koji dolazi do izražaja u ponovljenim interakcijama, može uspešno da uređuje interpersonalne odnose. Trivers je ove rezultate video kao potkrepljenje svoje teorije recipročnog altruizma (Trivers 2002).

Iako se činilo da princip reciprociteta uz selekciju u srodničkom krugu može da posluži kao valjano sredstvo za tumačenje socijalnih odnosa, model *Ponovljene dileme zatvorenika* i strategija *milo za drago* naišli su na niz prigovora. Ovde izdvajam tri. Najpre, evolucija saradnje u *Ponovljenoj dilemi zatvorenika* je osetljiva na argument inverzne indukcije. Zatim, strategija *milo za drago* je osetljiva na takozvani šum u komunikaciji (eng. *Noise*) i konačno povećanje broja individua koji učestvuju u sistemu reciprocijacije u *Ponovljenoj dilemi zatvorenika* vodi kolapsu saradnje.

Pođimo redom. *Ponovljena dilema zatvorenika* poseduje značajan teorijski nedostatak, bar kada je reč o međuljudskim interakcijama, naime osetljiva je na argument inverzne indukcije

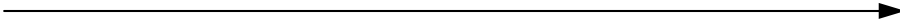
(Kavka 1986, Rosenberg 1992). Pretpostavimo da dve jedinke učestvuju u *Ponovljenoj dilemi zatvorenika* s konačnim brojem interakcija  $n$ , što je jedinkama poznato. Pod ovom pretpostavkom, poslednja interakcija u ponovljenoj dilemi zatvorenika s poznatim brojem interakcija, jeste zapravo *Dilema zatvorenika* bez ponavljanja. Tako da ako individue znaju da postoji  $n$  rundi igre, onda takođe znaju da postoji sekvenca  $n - 1$  u kojoj je dominantan potez nekooperacija. Inverznom indukcijom ( $N'' \rightarrow N' \rightarrow N$ ) lako dolazimo do toga da je racionalno igrati nekooperativno već u prvom potezu. Time je *Ponovljena dilema zatvorenika* redukovana na jednokratnu igru s nesaradnjom kao dominantnom strategijom.

Da bi proces inverzne indukcije otpočeo, individue ne moraju da znaju tačan broj mogućih interakcija, već njihovu *gornju granicu*. Budući da tu granicu ne znamo ni u društvu u kome živimo, nemamo mnogo razloga da prihvatimo da bi individue u nekoj predačkoj populaciji zadovoljavale ovako strog uslov (Hampton 1986, Kavka 1986). Osim toga, eksperimentalna evidencija pokazuje da ljudi do izvesne mere saraduju u *Ponovljenoj dilemi zatvorenika* i onda kada im je poznata gornja granica mogućih interakcija (Andreoni and Miller 1993).

Čak i ako odbacimo argument inverzne indukcije kao irelevantan, ostaje da je strategija *milo za drago* osetljiva na „šum u komunikaciji“. Ako dve individue učestvuju u reciprocijaciji, ali nakon izvesnog broja rundi jedna od individua pogrešno protumači potez druge strane kao nesaradnju i pokušaj eksploatacije, na koju *milo za drago* uvek odgovara nesaradnjom, onda se prethodno ustanovljeni kooperativni odnosi transformišu u neprekidni niz nesaradnje. (Šum u komunikaciji je u tabeli obeležen crveno.)

Tabela 6

<b>A/tft</b>	C	C	C	C	D	D	D	D
<b>B/tft</b>	C	C	C	D	D	D	D	D


  
vreme

Da bi se ovaj problem prevazišao teoretičari igara su razvili druge strategije u kojima je osnovna strategija *milo za drago* obogaćena (oplemenjena) velikodušnošću ili kajanjem, što

znači da igrači na nesaradnju ne odgovaraju odmah nesaradnjom (Wu and Axelrod 1995). Ove strategije, koje uglavnom predstavljaju modifikaciju strategije *milo za drago*, pod određenim uslovima mogu da je nadigraju (Nowak and Sigmund 1993, Nowak, May, and Sigmund 1995, Nowak and Sigmund 1992). Uprkos izvesnim nedostacima koji mogu da poremete sistem reciprocijacije, princip reciprociteta se pokazao kao važan činilac za evoluciju saradnje.<sup>28</sup>

Budući da su se Akselrodovi turniri, kao i Triversova koncepcija recipročnog altruizma, fokusirali na dijadne relacije, verovatno jedan od najznačajnijih prigovora formulisali su Bojd i Ričerson (Robert Boyd, Peter J. Richerson): povećanje broja individua koje učestvuju u recipročnim interakcijama vodi kolapsu kooperacije (Boyd and Richerson 1988). Ovi autori pretpostavljaju da se saradnja bilo koje dve individue iz grupe pozitivno odražava na svakog pojedinačnog člana grupe. Ukoliko bilo koja individua u grupi kuju čini  $n$  individua, gde je  $n > 2$ , promeni strategiju s **D** na **C**, svi članovi grupe od toga imaju koristi. Trivers napada ovu pretpostavku i pita se zbog čega ako jedinka *A* bište jedinku *B*, a zatim *B* bište *A*, od toga svi moraju da imaju koristi (Trivers 2006). Prema njegovom mišljenju, ovi autori su stvorili artificalni svet u kome one jedinke koje ne saraduju automatski povećavaju adaptivnu vrednost na račun razmene usluga onih jedinki koje saraduju, što ne mora da bude tačno. On, takođe, naglašava da i u kontekstu velikih grupa odnosi uglavnom ostaju dijadni ili trijadni, tako da ne moraju da se reflektuju na čitavu populaciju. Prema Triversovom mišljenju, Bojdova i Ričersonova kritika je neosnovana, jer je model koji su ovi autori predložili krajnje nerealističan.

Osim toga, tvrdi se da recipročni altruizam nije altruizam ni u biološkom ni u psihološkom smislu. Prvo, u biološkom smislu nije reč o altruizmu, jer ako smo altruizam definisali kao onu vrstu ponašanja koja povećava adaptivnu vrednost recipijenta na račun adaptivne vrednosti jedinke koja se ponaša altruistički, onda u slučaju recipročnog altruizma takvog žrtvovanja adaptivne vrednosti nema. Na drugom mestu, prigovara se da nije reč o

---

<sup>28</sup> Zajednički naziv za ove strategije je „okidač strategije“, jer jedan ili eventualno dva nekooperativna poteza iniciraju (deluju kao okidač) za istu reakciju. Za iscrpan spisak modifikacija vidi (Dugatkin 1998). Ričard Džojz misli da neke od ovih novorazvijenih strategija zapravo nisu modifikacije strategije *milo za drago*. Prema njegovom mišljenju takva strategija je recimo PAVLOV, koja je osetljiva na učenje i rukovodi se principom *pridrđavaj se istog poteza ako dobijaš, promeni potez ako gubiš* i nemilosrdna je u eksploataciji kada otkrije безусловnu saradnju (Joyce 2006).

altruizmu u psihološkom smislu, jer ne postoji odgovarajući kauzalni lanac između motiva i postupka. Preciznije, nema altruističkih motiva koji su u psihološkom smislu neposredni uzroci odgovarajućeg ponašanja. U prilog ovoj tezi govori i činjenica da se u onim slučajevima u kojima je ishod *Dileme zatvorenika* obostrana saradnja individue dodatno podstaknute da se ponašaju kooperativno neurofiziološkim procesima koji učestvuju u doživljavanju zadovoljstva (Rilling et al. 2002). Na taj način, posledica saradnje je lično zadovoljstvo. To implicira da je ona vrsta ponašanja koja se obično naziva recipročnim altruizmom sebična i u biološkom i u psihološkom smislu. Iz ovih razloga, jedan deo istraživača uopšte izbegava da govori o recipročnom altruizmu, već radije govore samo o reciprocitetu (Krebs 2011, Sesardic 1995, Sober 1988).

## 2.7 Indirektni reciprocitet i altruističko kažnjavanje

Bojdova i Ričersonova kritika nije uzimala u obzir ono što je Trivers nazvao uopštenim, a Aleksander (Richard Alexander) indirektnim reciprocitetom (Trivers 1971, Alexander 1985, Alexander 2009 [1987]). Direktni reciprocitet funkcioniše po principu razmene usluga tako što jedinka *A* pomogne jedinki *B*, a zatim *B* pomogne *A*, a korist od interakcije je obostrana. Za razliku od direktnog reciprociteta gde su u reciprocijaciju uključene uglavnom dve jedinke, tako da se od recipijenta očekuje da donoru uzvрати uslugu, u slučaju indirektnog reciprociteta u reciprocijaciju je uključena treća strana koja pažljivo i kontinuirano procenjuje pouzdanost partnera u recipročnim relacijama i u zavisnosti od procene uzvraća uslugu u ime recipijenta ili kažnjava stranu koja ne saraduje. Evolucija indirektnog reciprociteta u mnogome zavisi od prisustva zainteresovanih posmatrača (Alexander 1985). Kao i za evoluciju direktnog reciprociteta, za evoluciju indirektnog reciprociteta je značajno da interakcije jedinki budu učestale. Najvažniji činioci za evoluciju indirektnog reciprociteta su kompleksna interaktivna svojstva, kao što su društveni status, reputacija i mogućnost nagrađivanja i kažnjavanja, koja mogu da se pojave samo u prisustvu zainteresovanih posmatrača (Alexander 1985, Alexander 2009 [1987]).

Na identičan način kao što je direktni reciprocitet (recipročni altruizam) samo odloženo zadovoljenje vlastitih (genetičkih) interesa zaodeno u bihejvioralni obrazac altruizma, indirektni reciprocitet se, takođe, zasniva na zainteresovanosti jedinke za sopstvene životne



interese, koji se u *krajnjoj* konsekvenci mogu redukovati na reproduktivne. Iz filozofske perspektive Aleksanderova teorija interesa, izvedena iz evolucione teorije životnih istorija, može delovati kontroverzno, jer se u filozofskom kontekstu ovi interesi obično definišu preko određenih intencionalnih stanja. Ali, prema Aleksanderu, interesi ne moraju da budu definisani u kategorijama svesnog ili intencionalnog, što potvrđuje biološko proučavanje životnih interesa različitih organizama koji ili nemaju dovoljno razvijen ili uopšte nemaju nervni sistem da bi mogli da imaju bilo kakvu svest o tim interesima (Alexander 1985, Alexander 2009 [1987]). U takvim slučajevima može se govoriti o interesima organizma, ako su oni određeni u evolucionim terminima preživljavanja i reprodukcije, kao što se može reći da takvi organizmi slede vlastite interese kada se hrane i razmnožavaju, a da o tome ne znaju ništa. Ni čovek nije izuzetak. Ljudi veruju da teže boljem društvenom položaju, simbolima društvenog statusa, dobroj reputaciji, itd. iz razloga koje ne dovode u vezu s povećanjem adaptivne vrednosti, a u stvari to čine jer im takvo postupanje obezbeđuje 1) pristup većem broju potencijalnih partnera za parenje i posledično 2) više potencijalnih začeća.

Mada se može činiti da je indirektni reciprocitet način ponašanja koji je karakterističan isključivo za organizame sa sofisticiranim kognitivnim kapacitetima, kao što su teorija uma, pamćenje, sposobnost apstrakcije, primena presedana, zatim jezik koji omogućava protok informacija o recipročnim interakcijama različitih jedinki i verbalnu formulaciju pravila, Aleksander ne isključuje mogućnost da obrasci indirektnog reciprociteta budu otkriveni i kod nekih drugih društvenih životinja: kitova, pasa, mačaka i nekih primata (Alexander 2009 [1987], Nowak and Sigmund 2005).<sup>29</sup> Za sada su obrasci indirektnog reciprociteta u eksperimentalnim uslovima registrovani u međuvrsnim uzajamnim odnosima riba čistača i njihovih klijenata, kao i kod pacova (Bshary and Grutter 2006, Rutte and Taborsky 2007), dok je prisustvo indirektnog reciprociteta (pre svega altruističkog kažnjavanja) kod šimpanzi i dalje sporno (de Waal 1991, Riedl et al. 2012).

Možemo identifikovati dva, odnosno tri, načina na koji indirektni reciprocitet funkcioniše (Alexander 2009 [1987], Sigmund 2013). Jedan je da jedinka *A* pomogne jedinki *B*, zatim *B*

---

<sup>29</sup> Teorija uma predstavlja sposobnost individue da prepozna da druga individua ima intencionalna stanja, kao i da ta intencionalna stanja mogu da se razlikuju od njenih (Baron-Cohen, Leslie, and Frith 1985).

pomogne jedinki *C*, a *C* vrati uslugu jedinki *A*. Međutim, ovakvi ciklusi lako mogu da budu prekinuti jer su strategije zasnovane na načelu „budi dobar prema drugima, ako su drugi dobri prema tebi“ nestabilnije u odnosu na strategije zasnovane na principu „budi dobar prema onima koji su dobri prema drugima“ (Boyd and Richerson 1989). Drugi način, koji ima više izgleda za uspeh, je da *A* pomogne *B*, *C* pomogne *A* jer je opazio da *A* pomaže *B*, a zatim *A* pomogne *C*. U ovom slučaju indirektni reciprocitet prethodi direktnom. *C* je prosudio *A* kao pouzdanog partnera za saradnju, jer je *A* pomažući *B* stekao dobru reputaciju. Pri tome treba imati na umu da sticanje dobre ili loše reputacije ne zavisi nužno od direktnog opažanja interakcija među jedinkama, već može da se uspostavi putem slobodnog protoka informacija, ili da stvari nazovemo pravim imenom – ogovaranjem.<sup>30</sup> Osim prethodna dva, Sigmund identifikuje i treći način funkcionisanja indirektnog reciprociteta koji ne zavisi od reputacije i ponovljenih interakcija (Sigmund 2013). Tako, na primer, *A* može da pomogne *B*, zatim *B* pomogne *C*, *C* pomogne *D*, itd., kao kada nekome pridržavamo vrata supermarketa pošto je neko prethodno pridržao nama vrata. Utvrđeno je da su osobe koje su u vremenskom intervalu koji neposredno prethodi prilikama da se nekome pomogne i same bile recipijenti altruističkog čina spremnije da pomognu, za razliku od drugih osoba. Sličan obrazac ponašanja u eksperimentalnim uslovima pokazuju i pacovi (Rutte and Taborsky 2007). Druga strana ovog mehanizma je ta da se povređuju osobe koje ni na koji način nisu bile uključene u neko prethodno negativno iskustvo (Nowak and Sigmund 2005, Sigmund 2013).

Nagrađivanje, kažnjavanje i obmanjivanje su tri pojavna oblika indirektnog reciprociteta koji imaju važnu ulogu u evoluciji saradnje (možda je bolje reći da je podupiru, ali nisu ni nužni ni dovoljni uslovi za njenu evoluciju) (Alexander 1985, Alexander 2009 [1987]). *Nagrađivanje* evoluiralo tako što jedinka *C*, opažajući uspešnu reciprocijaciju između *A* i *B*, odlučuje da pomogne jedinki *B* u očekivanju da će *B* i njoj da uzvрати pomoć. S druge strane, ako *B* ne uzvрати pomoć, nakon što joj je jedinka *A* pomogla, *C* može da uskrati pomoć jedinki *B* na osnovu zapažanja da je *B* varalica i nepouzdan partner za saradnju.<sup>31</sup> *Kažnjavanje* evoluiralo tako što *C* opazi da je *A* povredio *B*, a zatim nanosi povredu *A* zbog očekivanja da bi *A* (ili neko drugi

---

<sup>30</sup> Svakako, informacije mogu da budu pogrešne i nepouzdana, zbog čega je značajno kontinuirano procenjivanje učesnika u recipročnim interakcijama.

<sup>31</sup> Ovakva vrsta uskraćivanja pomoći pre se može tumačiti kao neka vrsta kažnjavanja.

ugledajući se na  $A$ ) mogao da povredi i njega ako to ne učini. *Obmanjivanje* (i sposobnost otkrivanja obmanjivanja) evoluirao kao posledica trke u naoružanju među jedinkama koje se nadmeću u evolucionoj utakmici tako što jedinka  $A_1$  najpre samo naizgled pomogne jedinki  $B$ , a  $C_1$  pomogne  $A_1$ , pogrešno verujući da će  $A_1$  da pomogne njoj. Pažljivi posmatrač  $C_2$  otkriva da  $A_1$  vara i uskraćuje joj pomoć, izbegava je ili je kažnjava, ali je  $A_2$  bolji obmanjivač i uspeva da izigra  $C_2$ , što zatim  $C_3$  prepozna kao varanje, itd. O evolucionom objašnjenju obmanjivanja i vezi između obmanjivanja i samoobmanjivanja, kao i važnosti samoobmanjivanja za evoluciju saradnje i morala biće više reči u petom poglavlju. Ovde je dovoljno reći da se u standardnom evolucionom modelu koji je predložio Trivers samoobmanjivanje pojavljuje kao funkcija efikasnijeg obmanjivanja drugih jedinki (Trivers 2002, 2010, Von Hippel and Trivers 2011a).

Za koncepciju kažnjavanja o kojoj je ovde reč u literaturi se mogu pronaći dva naziva: moralistička agresija i altruističko kažnjavanje.<sup>32</sup> Termin „altruističko“ u pojmu „altruističko kažnjavanje“, koji je uobičajen u raspravama o indirektnom reciprocitetu, znači da jedinka koja kažnjava trpi štetu zarad koristi druge jedinke ili očuvanja socijalne harmonije. Konkretnije, iz Aleksanderove definicije altruističkog kažnjavanja sledi da je kažnjavanje u bihevioralnom smislu altruističko, ali je genetički sebično, budući da su jedinke koje kažnjavaju mogle da evoluiraju samo ako je kažnjavanje na neki način vodilo njihovom reproduktivnom uspehu. To je mogao da bude slučaj ako su jedinke koje kažnjavaju prekršioce neke prihvaćene norme, iako nisu direktno zainteresovana strana, prepoznate kao vredni partneri za parenje, jer ako makar i naizgled vode računa o interesima drugih i socijalnoj harmoniji, možda bi još više vodili računa o interesima svojih potomaka.

Kako altruističko kažnjavanje uključuje ne samo nanošenje štete, povređivanje, ekskomunikaciju<sup>33</sup>, stigmatizaciju ili izvrgavanje poruzi onoga ko se kažnjava, već štetu trpi i onaj ko kažnjava, jedinke se nalaze pred evolucionom dilemom da li da kažnjavaju ili ne. Jedinki

---

<sup>32</sup> Kažnjavanje u ovom kontekstu nema jedinstvenu funkciju i ne treba ga mešati s pojmom kažnjavanja kakav nalazimo u filozofskim tekstovima koji se bave pravnim i moralnim statusom kazne. U evolucionom kontekstu se ne postavlja pitanje moralnog opravdanja kazne, već se pita *zašto* ljudi (i možda još neki primati) upražnjavaju određenu aktivnost, kakva je kažnjavanje, koja im nije od direktne koristi.

<sup>33</sup> Zanimljivo je da neurofiziološka reakcija u slučaju ostrakizma slična neurofiziološkoj reakciji kao u slučaju fizičkog bola (Eisenberger, Lieberman, and Williams 2003).

nije u interesu da kažnjava ako to za nju znači gubitak energije, moguće izlaganje povredama ili utrošak vremena koje bi moglo da se iskoristi za ostvarenje nekih sebičnih ciljeva, kakvi su recimo hranjenje, parenje ili briga o potomstvu. Ali, upravo jedinka koja može sebi da priušti „luksuz“ da se izlaže opasnostima i troškovima u vremenu i energiji za rešavanje unutargrupnih konflikata može izgledati kao dobar partner za parenje i da sebi obezbedi veći broj potencijalnih začeca (Zahavi 1995). S druge strane, jedinka mora moći da na adekvatan način identifikuje ko je varalica i ko krši neku normu. Osim toga, zainteresovani posmatrači moraju moći da identifikuju one koji kažnjavaju ili uskraćuju pomoć varalicama kao „dobre momke“, jer u suprotnom oni koji uskraćuju pomoć onima koji ne pomažu drugima, ne dele hranu, itd., mogu da u neopreznim očima posmatrača postanu oni koji *neopravdano* uskraćuju pomoć, steknu lošu reputaciju i sami postanu oni koji se kažnjavaju (Nowak and Sigmund 2005, Sigmund 2013).

Ako se sada okrenemo *neposrednim* uzrocima altruističkog kažnjavanja, i problem formulišemo u terminima teorije racionalnog izbora, onda možemo da kažemo da kažnjavanje konstituiše problem javnog dobra drugog reda koji bi možda najjednostavnije mogao da se opiše Ezopovom basnom *Mišji savet* (Coleman 1990). U ovoj basni miševi dolaze do zaključka da je mačka odgovorna za smanjenje populacije miševa. Mačka predstavlja problem javnog dobra prvog reda i miševi jednoglasno odlučuju da ovaj problem reše tako što će okačiti mački zvono oko vrata kako bi svi mogli da je čuju kada nailazi. Time se konstituiše problem javnog dobra drugog reda, jer iako su svi saglasni da je u javnom interesu da se mački okači zvono oko vrata, nijedan od miševa nije podstaknut da to i učini. Primenjeno na probleme saradnje i indirektnog reciprociteta to znači da iako je u javnom interesu da se saradnja podupre kaznama za nesaradnju, nikome nije u interesu da kažnjava, jer potencijalna individualna šteta od kažnjavanja ne može da bude kompenzovana sumom korisnosti za grupu. Ovaj problem može da bude prevaziđen ako dovoljan broj individua ima dispoziciju da kažnjava altruistički, a takvu dispoziciju će imati ako je ona u evolucionoj prošlosti obezbeđivala neposredan ili posredan reproduktivni uspeh (Fehr and Gächter 2002, West, El Mouden, and Gardner 2011).

Fehr i Gächter (Ernst Fehr, Simon Gächter) su u eksperimentu javnog dobra svakog člana grupe od četiri osobe zadužili s 20 monetarnih jedinica (MJ), a zatim im je pružena prilika da investiraju čitavu sumu, jedan njen deo ili ništa u neki zajednički projekat. Ostatak od uloženog novca su mogli da zadrže za sebe. Učesnicima je garantovan povraćaj od 40 procenata od

njihove investicije, tako da ako bi svi uložili po 10 MJ, ukupan povraćaj za grupu bi bio 16 MJ, što uz 10MJ koje su uložili ukupno iznosi 56 MJ. Ova suma se zatim deli na jednake delove, tako da svaki učesnik u eksperimentu dobija po 14 MJ, što uz 10 MJ koje su im ostale od investicije čini ukupno 24 MJ za svakoga. Kada bi svi uložili sve, dobili bi po 32 MJ. Međutim, ako su interakcije anonimne, tj. ako ne postoji protok informacija među članovima grupe, onda je za svaku individuu bolje da uloži manje, jer je nesigurna da će bilo ko uložiti bilo šta, a ako neka od individua uloži više od nje ona u svakom slučaju prolazi bolje.

Ali ako se u eksperiment uvede mogućnost protoka informacija među članovima grupe, kao i mogućnost da se oni koji ne ulažu ništa ili ulažu manje od proseka kazne, situacija postaje drugačija. Kažnjavanje funkcioniše i pod eksperimentalnim ograničenjima gde se struktura grupe stalno menja i dve individue nikada ne stupaju u ponovljenu interakciju i to na sledeći način. Ako *A* odluči da kazni *B* jer je *B* uložila manje nego ostatak grupe, onda onaj ko kažnjava trpi štetu u MJ, ali onaj ko je kažnjen trpi još veću štetu. Da bi se kažnjavalo, mora da postoji disproporcija između štete koju trpi izvršilac kazne i štete za prekršioca, tako da šteta koju trpi prekršilac mora da nadmašuje štetu koju trpi izvršilac. Da kažnjavanje daje rezultate pokazuje činjenica da u onim eksperimentalnim uslovima u kojima je prisutna mogućnost kažnjavanja, oko 40 odsto učesnika na kraju eksperimenta ulaže svih 20 MJ, dok oko 80 odsto njih ulaže 15 MJ ili više. S druge strane, tamo gde nema kazne oko 60 odsto učesnika ne ulaže ništa, dok oko 75 odsto njih ulaže 5 MJ ili manje.

Neposredni mehanizam preko kog kažnjavanje deluje je snažna negativna emocionalna reakcija koju izaziva kršenje pravila (Fehr and Gächter 2002). Međutim, neki autori smatraju da ova reakcija, pre svega zavist i ljutnja, nije uzrokovana kršenjem pravila i nepravdom koja je nekome učinjena, već lošim ličnim ishodom (Pedersen, Kurzban, and McCullough 2013). Prema mišljenju ovih autora, ukoliko su zavist i ljutnja zbog lošeg ličnog ishoda značajni motivacioni mehanizmi u procesu altruističkog kažnjavanja, onda je teško uopšte govoriti o *altruističkom* kažnjavanju. Preciznije, tvrdi se da ne postoji adekvatan kauzalni lanac između motiva i postupaka koji bi dozvolio da se kažnjavanje okvalifikuje kao altruističko. Osim toga, neurofiziološka reakcija ukazuje da izvršioци kazne prilikom kažnjavanja osećaju zadovoljstvo

(de Quervain et al. 2004).<sup>34</sup> U slučajevima kada subjekt ima želju i priliku da stvarno, a ne samo simbolično kazni prekršioca neke ustanovljene norme aktivira se repato jedro (lat. *Nucleus caudatus*) i povećava prokrvljenost talamusa (lat. *Thalamus*), što su delovi subkortikalnih struktura mozga koji učestvuju u gratifikaciji. Ovaj podatak ukazuje na to da je kažnjavanje neposredno podstaknuto hedonističkim, dakle sebičnim, motivima. Imajući ovu činjenicu u vidu Trivers upozorava da navedena vrsta kažnjavanja može lako da se izvrgne u torturu, pogotovo ako izvršilac kazne takvo postupanje smatra moralno opravdanim (Trivers 2006). Ipak, jednako kao i u slučaju direktnog reciprociteta, otvoreno je pitanje da li je takva vrsta osećanja ličnog zadovoljstva, koje se javlja aktivacijom određenih struktura mozga, nespojiva s altruizmom.

U raspravama o altruističkom kažnjavanju veliku zabunu unosi to što se veoma lako prelazi s pitanja o krajnim uzrocima (*zašto?*) na pitanja o neposrednim uzrocima (*kako?*) ili se na pitanja o krajnim uzrocima odgovara u terminima neposredne uzročnosti (West, El Mouden, and Gardner 2011). Konačno, imajući na umu ove probleme, možemo da zaključimo da je ono što se obično naziva altruističkim kažnjavanjem u bihejvioralnom smislu zaista altruističko, ali je u psihološkom i biološkom smislu uglavnom sebično, tj. sebičnost ne može da se istisne iz objašnjenja.

## 2.8 Selekcija na više nivoa: Soberov i Vilsonov model

Iako se početkom druge polovine 20. veka, nakon *Vilijamsove revolucije*, činilo da je s koncepcijom grupne selekcije kao naučnom paradigmom završeno, biolog Dejvid Sloan Vilson (David Sloan Wilson) i filozof biologije Eliot Sober ponudili su rafiniranu verziju tradicionalnog modela grupne selekcije: selekciju koja deluje na više nivoa (Wilson and Sober 1994, Sober and Wilson 1998). Selekcija na više nivoa polazi od pretpostavke da potpuno objašnjenje evolucionog procesa, zahteva da se različiti nivoi selekcije uzmu u razmatranje. Kako različiti nivoi biološke hijerarhije – geni, hromozomi, ćelije, organi, organizmi, itd. – nisu nastali ni od kuda, već su i sami proizvod evolucionog procesa pretpostavlja se da je prirodna selekcija delovala na više nivoa, iako danas možda ne deluje na svim tim nivoima (Okasha 2006).

---

<sup>34</sup> Eksperiment koji su izveli ovi autori je sproveden na takav način da svako kažnjava kada smatra da mu je učinjena nepravda, dakle u dijadnim relacijama, tako da ovde izgleda uopšte nije reč o indirektnom reciprocitetu i altruističkom kažnjavanju.

Najpre, Sober i Vilson čine zaokret u razumevanju pojma jedinka. Uobičajeno razumevanje značenja pojma jedinka je pojedinačni višećelijski organizam. Ali, ono što iz jedne perspektive može da izgleda kao jedinka, iz drugačije perspektive može da se protumači kao grupa jedinki. Tako, ako se krećemo linijom biološke hijerarhije odozgo na dole, onda je pojedinačni višećelijski organizam koherentna i funkcionalna grupa ćelija koje su u ovom slučaju jedinke u odnosu na organizam, a ćelija je grupa ćelijskih organela, citoplazme, jedra i hromozoma, koje su jedinke u odnosu na ćelije, koje su sada grupe navedenih elemenata, itd. To znači da ono što na jednom nivou biološke hijerarhije nazivamo organizam, na drugom bi mogao da bude organ ili neki niži stupanj biološke organizacije. Kolonije socijalnih insekata su primer društvene organizacije u kojima grupa tradicionalno shvaćenih jedinki konstituiše entitet koji je, s obzirom na svoje karakteristike, „superorganizam“ i u kome jedinke zapravo funkcionišu kao organi (Wilson and Sober 1994).

Na drugom mestu, Soberov i Vilsonov model grupne selekcije zavisi od specifičnog razumevanja toga koje karakteristike čine grupu. U tradicionalnim modelima grupne selekcije insistiralo se na trajnosti, prostornoj i reproduktivnoj izolovanosti kao važnim karakteristikama grupe (Maynard Smith 1964, 1976). Nasuprot tome, u ovom modelu grupe su definisane isključivo preko efekata koje delovanje jedinki unutar grupe ima na njihovu adaptivnu vrednost, a ne preko trajanja ili odnosa s drugim grupama (Sober and Wilson 1998). Ovakva definicija grupa je suviše liberalna i omogućava da se svaka interakcija dve ili više individua, ma koliko ona bila kratkotrajna, tumači kao da je reč o grupi. Dalje, time je omogućeno da se redefinišu drugi modeli za evoluciju altruizma (modeli selekcije u srodničkom krugu i direktnog reciprociteta) i da se integrišu u model grupne selekcije, u kojoj su grupe ili srodnici ili parovi nesrodnih jedinki u međusobnim interakcijama. Važno je naglasiti da većina evolucionih biologa opire ovakvoj vrsti integracije. Kako Sober i Vilson ne odbacuje selekciju na nivou gena, u ovoj konceptualnoj shemi za genski selekcionizam rezervisano je mesto u objašnjenju intragenomskog konflikta, čiji je najbolji primer *mejotičko vođenje* – situacija u kojoj određeni geni krše uobičajena pravila mejoze i pojavljuju se u gametima sa zastupljenošću koja je veća od jedne polovine.

Pitanje koje se u ovoj teoriji postavlja nije da li grupe jedinki mogu da budu replikatori, već da li grupe mogu da budu nosioci prirodne selekcije ili interaktori (Wilson and Sober 1994,

Sober 2000). Da bi se utvrdilo na kom nivou deluje prirodna selekcija Wilson i Sober predlažu da se postavi sledeći niz pitanja: Da li se geni unutar jedinke razlikuju po adaptivnoj vrednosti?, Da li se jedinke u okviru grupe razlikuju po adaptivnoj vrednosti? i konačno, Da li se grupe razlikuju po adaptivnoj vrednosti (koja je merena individualnim preživljavanjem i reprodukcijom) u odnosu na druge grupe? (Wilson and Sober 1994). Ako je odgovor na prvo pitanje odričan, to znači da prirodna selekcija ne deluje na tom nivou, tj. da geni dele zajedničku sudbinu i da su deo istog nosioca. Zatim, sledi drugo pitanje i ako je odgovor na ovo pitanje odričan, to znači da jedinke dele zajedničku sudbinu i da su deo istog nosioca. To, takođe, znači da prirodna selekcija ne deluje na ovom nivou. Konačno, ako je odgovor na poslednje pitanje potvrđan, onda to znači da su grupe nosioci prirodne selekcije i da ona deluje na nivou grupe.

Da bismo lakše razumeli ovaj model selekcije, kao i zašto se on razlikuje od Vin-Edwardsovog modela grupne selekcije, važno je da se za trenutak fokusiramo na distinkciju između *selekcije na više nivoa [1]* i *selekcije na više nivoa [2]* (Damuth and Heisler 1988, 410). *Selekciju na više nivoa [1]* ima sledeće karakteristike: 1) koncepcija grupne selekcije se odnosi na posledice koje pripadanje grupi ima na adaptivnu vrednost *jedinke*; 2) adaptivna vrednost je svojstvo *jedinki*, 3) karakteristike su vrednosti koje se pripisuju *jedinkama*; 4) populacija se sastoji od *jedinki*, koje su organizovane u grupe; i 5) eksplicitni zaključci mogu da se izvedu samo o promenama u proporciji različitih vrsta *jedinki* u celoj populaciji (metapopulaciji). U *selekciji na više nivoa [2]*, 1) koncepcija *grupne selekcije* se odnosi na promene u frekvenciji različitih *grupa*; 2) adaptivna vrednost je svojstvo *grupa*; 3) karakteristike su vrednosti koje se pripisuju *grupama*; 4) populacija se sastoji od *grupa* koje su sačinjene od jedinki; 5) eksplicitni zaključci mogu da se izvedu samo o promenama u proporciji različitih *grupa* u populaciji. Dakle, u *selekciji na više nivoa [1]* u fokusu evolucionarne promene su jedinke koje su organizovane u grupe, dok su u *selekciji na više nivoa [2]* u fokusu evolucionarne promene grupe koje su sastavljene od jedinki. Soberov i Wilsonov model je primer *selekcije na više nivoa [1]*.

Razmotrimo najpre situaciju u kojoj ne dolazi do evolucije altruizma. Radi jednostavnosti Sober i Wilson pretpostavljaju da se adaptivna vrednost meri isključivo brojem potomaka, da je razmnožavanje aseksualno i da jedinke roditeljske generacije umiru nakon reprodukcije. Pretpostavimo da se populacija sastoji od altruista ( $A$ ) i sebičnih jedinki ( $S$ ), čiji je ukupan broj  $n$ . Frekvencija altruista je  $p$ , dok je frekvencija sebičnih jedinki  $(1 - p)$ , tako da je broj altruista



jednak  $np$ , a broj sebičnih  $n(1 - p)$ ; sve jedinke se razmnožavaju osnovnom stopom  $x$ . Kako je altruizam po definiciji ona vrsta ponašanja koja povećava adaptivnu vrednost drugih jedinki na račun sopstvene, prisustvo altruista u populaciji čini da se jedinke razmnožavaju u većoj meri i da imaju više potomaka  $b$  za šta altruisti trpe štetu  $c$ .

Adaptivnu vrednost altruista izračunavamo pomoću sledeće formule:

$$W_A = x - c + [b(np - 1)/(n - 1)]$$

Svaki altruista se razmnožava osnovnom stopom  $x$ , trpi štetu  $c$ , ali je i potencijalni recipijent korisnosti  $b$  koju dobija od drugih altruista  $np - 1$ . Korist od altruističkih postupaka je raspoređena na sve članove grupe, osim na altruistu  $(n - 1)$ , tako da je ukupna korisnost koju altruista ima  $b(np - 1)/(n - 1)$ .

Adaptivnu vrednost sebičnih jedinki izračunavamo pomoću sledeće formule:

$$W_S = x + [bnp/(n - 1)]$$

Sebične jedinke, za razliku altruističkih ne trpe štetu i razmnožavaju se osnovnom stopom  $x$ , ali su pri tome potencijalni recipijenti korisnosti  $b$  od altruističkih postupaka svih altruista  $np$ .

Pretpostavimo da je  $n = 200$ ;  $p = 0.5$ ;  $x = 10$ ;  $b = 5$  i  $c = 1$ . U tom slučaju adaptivna vrednost altruista iznosi oko 11.5 ( $W_A \approx 11.5$ ), dok adaptivna vrednost sebičnih jedinki iznosi oko 12.5 ( $W_S \approx 12.5$ ). Iako prisustvo altruista u grupi čini da svima raste adaptivna vrednost, sebične jedinke ipak prolaze bolje. Broj jedinki potomačke populacije dobija se pomoću sledeće formule:  $n' = n[pW_A + (1 - p)W_S]$  i iznosi 2400 jedinki. Frekvencija altruista u novoj populaciji dobija se na osnovu  $p' = npW_A/n'$  i iznosi oko 0.48, dok je frekvencija sebičnih jedinki  $(1 - p')$  oko 0.52. Dakle, prisustvo altruista u populaciji čini da frekvencija sebičnih jedinki raste. Istovremeno frekvencija altruista postepeno opada, tako da bi vremenom altruisti nestali iz populacije. Na taj način altruizam ne može da evoluiru.

Da bi objasnili na koji način altruizam može da evoluiru Sober i Vilson se oslanjaju na statistički fenomen koji je poznat kao *Simpsonov paradoks* koji dovodi u pitanje uobičajeno zdravorazumsko shvatanje da ono što važi za delove mora da važi i za celinu. Uzmimo sledeći

primer. Pretpostavimo da u nekoj instituciji kakva je recimo *Filozofski fakultet* na predlog uprave treba doneti neku važnu odluku većinom glasova. Glasanje se obavlja po odeljenjima, a odluka će biti usvojena ako većina članova na većini odeljenje bude podržala predlog. Iako se čini da kada bi većina članova na većini odeljenja podržala predlog, to značilo da ga je podržala većina članova fakulteta kao institucije, to ipak nije slučaj. Pretpostavimo da ova visokoškolska institucija ima ukupno 100 članova koji su raspoređeni na tri odeljenje *A*, *B* i *C*, koja imaju 15, 35 i 50 članova tim redom. Pretpostavimo da su na odeljenju *A* svi protiv, tako da za predlog nije glasao niko (0 glasova), na odeljenju *B* je većina za predlog (18) i na odeljenju *C* je više od polovine članova za (26). Međutim, iako je većina članova na većini odeljenja podržala predlog, većina članova cele institucije je protiv (54). Ovo je jednostavan primer *Simpsonovog paradoksa*. Sober i Vilson se oslanjaju na ovaj statistički fenomen da bi objasnili na koji način altruizam može da evoluiru među grupama uprkos tome što na individualnom planu u svakoj od grupa altruisti prolaze gore od sebičnih jedinki. Dakle, nije dovoljno da altruisti budu prisutni u populaciji, već je nužno da postoje grupe u kojima broj altruista varira.

Uzmimo primer koji koriste Sober i Vilson. Pretpostavimo kao i ranije da imamo populaciju od 200 jedinki od kojih polovinu čine altruisti, ali pretpostavimo sada da je populacija podeljena na dve grupe od po 100 jedinki. Udeo altruista u jednoj grupi je 20 odsto ( $p = 0.2$ ), dok je udeo altruista u drugoj grupi 80 odsto ( $p = 0.8$ ). Pod ostalim jednakim uslovima ( $x = 10$ ;  $b = 5$  i  $c = 1$ ) u grupi koja ima 20 odsto altruista adaptivna vrednost altruista bi bila  $W_A = 9.96$ , a sebičnih jedinki  $W_S = 11.01$ . U narednoj generaciji grupa bi imala 1080 jedinki, a frekvencija altruista bi pala na 0.184. U grupi u kojoj je frekvencija altruista  $p = 0.8$ , adaptivna vrednost altruista je  $W_A = 12.99$ , a sebičnih jedinki  $W_S = 14.04$ . U narednoj generaciji grupa bi imala 1320 jedinki, a frekvencija altruista bi pala na 0.787. Uočavamo da u obe grupe altruisti prolaze lošije od sebičnih jedinki, ali ako uporedimo frekvencije altruista u roditeljskoj i potomačkoj populaciji uočićemo da se frekvencija altruista u potomačkoj populaciji povećala. U roditeljskoj populaciji je iznosila  $[0.2(100) + 0.8(100)]/200 = 0.5$ , a u populaciji potomaka iznosi  $[0.184(1080) + 0.787(1320)]/2400 = 0.516$ . Ako bi se ove grupe izmešale i zbog ograničenog rasta populacije ponovo svele na 200, altruisti bi imali proporcionalno veću zastupljenost u populaciji. Ako bi se populacija ponovo podelila na dve grupe od po 100 jedinki altruisti bi imali proporcionalno veću

zastupljenost i u grupama što bi omogućilo njihovu dalju ekspanziju. Na taj način, Sober i Vilson pokazuju da ponašanje može da bude istovremeno evoluciono uspešno i altruističko.

Kada se posmatra iz perspektive teorije grupne selekcije, evolucija altruizma postaje manje zagonetna i naizgled lako objašnjiva: u kontekstu grupe altruisti mogu da imaju manju adaptivnu vrednost (merenu individualnim preživljavanjem i reprodukcijom) od sebičnih jedinki, ali grupe jedinki koje sadrže altruiste mogu da imaju veću adaptivnu vrednost (takođe merenu individualnim preživljavanjem i reprodukcijom) nego grupe sebičnih jedinki. To znači da altruizam može da evoluirá ako imamo populaciju grupa (metapopulaciju) u kojima varira zastupljenost altruista, čije prisustvo u grupama se reflektuje na adaptivnu vrednost članova grupe.

Važno je zapaziti da se u ovom modelu adaptivna vrednost uvek meri diferencijalnim preživljavanjem i reprodukcijom jedinki, a ne diferencijalnim preživljavanjem i reprodukcijom grupa jedinki. Tumačeći grupnu selekciju kao analognu individualnoj, Mejnard Smit je insistirao na tome da grupe kao posebni entiteti moraju da imaju svojstvo nasledne varijacije u adaptivnoj vrednosti da bi se kvalifikovale kao jedinice selekcije. Budući da grupe ne ostavljaju potomke, on je zaključio da one ne ispunjavaju uslov heritabilnosti, a bez heritabilnosti nema adaptacija na nivou grupe i da iz tog razloga one nisu jedinice selekcije. Ali, kako je u modelu grupne selekcije koji predlažu Sober i Vilson reč o *selekciji na više nivoa [1]*, gde u roditeljsko-potomačkom odnosu stoje jedinke, a ne grupe, uslov heritabilnosti ne mora da bude zadovoljen na nivou grupe, da bismo govorili o grupnoj selekciji ili bar tako misle zagovornici ovog pristupa (Sober 1987, Okasha 2001). Dovoljno je da pripadanje grupi ima efekata na adaptivnu vrednost jedinki. Okaša (Samir Okasha) naglašava da postoji asimetrija između uloge koju imaju grupe u ovom modelu grupne selekcije i uloge koju imaju individue u modelima individualne selekcije i da grupna selekcija, onako kako je predstavljena u Soberovom i Vilsonovom modelu, nije analogna individualnoj selekciji (Okasha 2001). Ali, prema Mejnardu Smitu, to bi značilo ili da ovi autori ne razumeju o čemu se u grupnoj selekciji radi ili da u Soberovom i Vilsonovom modelu uopšte nije reč o grupnoj selekciji (Maynard Smith 1987).

Drugačiji način da se uvidi šta u populaciji evoluirá, jeste da se izračuna adaptivna vrednost za svaki tip jedinke u grupama, a da se zatim izračuna *prosečna* adaptivna vrednost

jedinki među grupama. U tom slučaju prosečna adaptivna vrednost altruista (A) bi iznosila  $12.99 \times 0.8 + 9.96 \times 0.2 = 12.38$ , a sebičnih jedinki (S) bi bila  $11.01 \times 0.8 + 14.04 \times 0.2 = 11.62$ . Ovi načini izračunavanja evolucione promene su ekvivalentni, jer polaze od istih bioloških pretpostavki i dolaze do istih zaključaka o tome šta evoluiru u metapopulaciji, ali uprosečavanje dovodi do gubitka informacije o tome na kom nivou prirodna selekcija deluje (Wilson 2015). Kako je prosečna adaptivna vrednost altruista veća od prosečne adaptivne vrednosti sebičnih jedinki, altruizam se objašnjava kao ono što evoluiru i to selekcijom koja deluje na individualnom nivou. Ali, posmatrano iz genocentrične perspektive, ono što ima veću prosečnu adaptivnu vrednost i što evoluiru nije altruističko, već sebično. Prema Soberu i Vilsonu, ako se zanemari struktura grupe i ako se vodi računa isključivo o onome što evoluiru, altruizam se lako svodi na sebičnost, odnosno, evolucija altruizma se svodi na *contradictio in terminis*, a selekcija koja deluje na individualnom nivou se tumači kao sinonim za prirodnu selekciju (Sober and Wilson 1998, Okasha 2006). Ovakav način rasuđivanja ovi autori smatraju pogrešnim i označavaju ga terminom *greška uprosečavanja* (eng. averaging fallacy). Oni takođe smatraju da veliki broj kontroverzi koji okružuju koncepciju grupne selekcije može da bude izbegnut ukoliko istraživači budu izbegavali da prave ovu grešku (Sober and Wilson 1998).

U osvrtu na Vilsonov i Soberov model selekcije, a i ranije, u jeku kontroverze o grupnoj selekciji, sredinom sedamdesetih godina prošlog veka, Džon Mejnard Smit (John Maynard Smith) je istakao da je nesporazum između zagovornika selekcije na više nivoa i ortodoksnih darvinista, genskih selekcionista, semantički, tako da ne može da bude rešen u empirijskoj ravni (Maynard Smith 1976, 1998). S druge strane, Okaša smatra da nesporazum nije semantički, već konceptualni (Okasha 2008). Naime, reč je o tome koja konceptualna shema bolje opisuje svet. Nesumnjivo je da Sober i Vilson misle da konceptualna shema selekcije na više nivoa ima veću eksplanatornu moć. U nedavno objavljenoj knjizi o altruizmu Vilson ne odbacuje ni jedan od prethodno navedenih pristupa za izračunavanje evolucione promene, štaviše, bar na rečima ih smatra podjednako ispravnim evolucionim objašnjenjima, koja zaslužuju da koegzistiraju (Wilson 2015). Polazeći od toga da nema potpuno neprevodivih konceptualnih shema i uz upozorenje na neodređenost prevođenja, a kako bi se izbegli budući nesporazumi i kontroverze vezane za koncepciju grupne selekcije, Vilson preporučuje istraživačima da budu „multilingvalni“. Iako je ova preporuka velikodušna, većina istraživača koja se ne slaže s njim,

smatra da je njegova konceptualna shema redundantna i da je poznavanje jednog jezika sasvim dovoljno za evoluciono objašnjenje bioloških fenomena.

Nužni uslovi za evoluciju altruizma prema ovom modelu su sledeći: 1) mora postojati populacija grupa (metapopulacija); 2) grupe moraju da variraju u proporciji altruista; 3) grupe s altruistima moraju da imaju veću adaptivnu vrednost (merenu individualnim preživljavanjem i reprodukcijom) u odnosu na grupe isključivo sebičnih jedinki i 4) grupe jedno vreme moraju da budu izolovane, ali da se njihovo potomstvo u nekom trenutku meša i formira nove grupe (Sober and Wilson 1998). Osim toga, asortativno okupljanje, tj. ne-slučajno razvrstavanje u grupe, jeste još jedan nužan uslov evolucije altruizma prema ovom modelu (Okasha 2001, 2006). Otvoreno je pitanje koliko je takvo okupljanje učestalo u prirodnim uslovima (Maynard Smith 1987). Takođe, uspešnost ovog modela grupne selekcije u mnogome zavisi od pažljivo izabranih brojeva i neophodnosti da se grupe ujedinjuju, a zatim iznova formiraju kolonije u preciznim vremenskim intervalima u zavisnosti od intenziteta prirodne selekcije koja deluje na individualnom nivou, a koja favorizuje evoluciju sebičnosti.

Zapravo nije teško pronaći slučaj u kome altruizam ne evoluirao. Polazeći od osnovnog modela, pod ostalim jednakim uslovima, pretpostavimo da je nakon izvesnog broja generacija populacija svedena na 200 individua, da je frekvencija altruista opala na 0.1, a frekvencija sebičnih jedinki se povećala na 0.9. Pretpostavimo da se populacija подели na dve grupe od po 100 jedinki i da se altruisti asortativno okupljaju tako da su svi koncentrisani u jednoj od grupa. Na taj način imamo dve grupe od kojih je jedna potpuno sačinjena od sebičnih jedinki, dok u drugoj frekvencija altruističkih jedinki iznosi 0.2, a frekvencija sebičnih je 0.8. U prvoj grupi, 100 sebičnih jedinki se razmnožava osnovnom stopom bez pomoći altruista. Njihova adaptivna vrednost je 10 i u narednoj generaciji ih ima ukupno 1000, od kojih su sve sebične. U drugoj grupi adaptivna vrednost altruista iznosi 9.96, a sebičnih 11.01. Ukupan broj jedinki je 1080, a frekvencija altruista je opala na 0.184. Za razliku od prethodne situacije, ako sada uporedimo frekvencije altruista u roditeljskoj i potomačkoj populaciji, uočićemo da je frekvencija altruista u potomačkoj populaciji takođe opala sa 0.1 na 0.095. Konflikt između selekcije koja deluje na individualnom nivou i one koja deluje na nivou grupe rešava se u korist prve; nema Simpsonovog paradoksa i delovanjem selekcije na individualnom nivou altruisti će biti istisnuti iz populacije.

Važnost Soberovog i Vilsonovog modela je u tome što su ovi autori pokazali na koji način *biološki* altruizam načelno može da evoluiru ukoliko prirodna selekcija deluje na nivou koji je viši od individualnog, ali i dalje ostaje otvoreno pitanje u kojoj meri je ova teorija prihvatljiva i kolika je *zaista* sila prirodne selekcije koja deluje na nivou grupe. Većina evolucionih biologa misli da objašnjenja koja se pozivaju na više nivoe biološke hijerahije, kao na one na koje deluje prirodna selekcija, nisu plauzibilna, budući da već postoje jednostavniji eksplanatorni modeli. Kao što je već rečeno, oni ne osporavaju teorijsku mogućnost grupne selekcije, već njenu empirijsku efikasnost. Konačno za potrebe ovog rada pozivanje na koncepciju selekcije na više nivoa neće biti neophodno. Sva dalja evolucionarna objašnjenja koja se ovde pojavljuju zasnovana su na koncepcijama individualne selekcije, odnosno na teoriji inkluzivne adaptivne vrednosti i reciprociteta.

U narednom poglavlju baviću se evolucionim problemima koji se vezuju za mehanizme koji su neposredni uzroci altruističkog ponašanja. Tačnije, osnovni predmet narednog poglavlja će biti problem empatije u kontekstu *teze o kontinuitetu* kognitivnih, pa onda i moralnih sposobnosti čoveka i drugih sisara, pre svega drugih primata.

## 3 EMOCIJE I EMPATIJA

### 3.1 Šta su emocije?

Uzimajući racionalnost kao demarkacionu liniju između čoveka i drugih životinja istraživači, a filozofi pogotovo, su zanemarili važnost i pozitivne posledice koje emocije imaju na ponašanje ljudi; njihov značaj za uspostavljanje, održanje i jačanje socijalnih veza, kao i njihovu ulogu u altruističkoj motivaciji. Kroz istoriju filozofije, ako izuzmemo delo Dejvida Hjuma i Adama Smita (Adam Smith), emocije se gotovo poslovično pojavljuju kao negativni i ometajući činoci ljudske racionalnosti.<sup>35</sup> Tu situaciju precizno identifikuje Damasio (Antonio R. Damasio):

Termini rasuđivanje i odlučivanje obično impliciraju da donosilac odluke poseduje neku logičku strategiju za donošenje ispravnih zaključaka na osnovu koje bira odgovarajuću reakciju među različitim opcijama, kao i da su na delu pomoćni procesi koji su neophodni za rasuđivanje. Među potonje se obično ubrajaju pažnja i radna memorija, ali ni jednom rečju se ne pominju emocije ili osećanja, i gotovo ništa ne može da se čuje o mehanizmima pomoću kojih se raznovrstan repertoar opcija generiše za selekciju (Damasio 1994, 166).

Danas znamo da emocije imaju važnu adaptivnu funkciju i značajnu ulogu u rasuđivanju, donošenju odluka, usvajanju i održavanju verovanja i da njihovi efekti na ove procese ne moraju isključivo da budu negativni. Na primer, čak i usvajanje i održanje određenih (pogrešnih) verovanja pod uticajem emocija, što se obično označava kao usvajanje motivisanih verovanja, može da ima važnu funkciju za preživljavanje i reprodukciju organizma, kao i njegovo uspešno funkcionisanje u socijalnom i moralnom kontekstu. O ovim aspektima uticaja emocija na usvajanje i održanje verovanja biće više reči kasnije – u poglavlju koje se bavi problemom samoobmanjivanja. U ovom delu ću se fokusirati na vezu između emocija i ponašanja, kao i evoluciju empatije koja je neposredni mehanizam za ispoljavanje psihološkog altruizma, koji je,

---

<sup>35</sup> Teorije morala u 20. veku, kakve su Rolsova (John Rawls) ili Gotijeova (David Gauthier) nastoje da moral zasnuju isključivo na sposobnosti jednakih i autonomnih subjekata da složenim kognitivnim procesima donose racionalne odluke (Gauthier 1986, Rawls 1999b). Zato je teorija racionalnog izbora postala paradigmatična disciplina koja je nudila uzor za utemeljenje morala. To ne znači da ovi autori nisu bili upoznati sa stremljenjima u teorijskoj biologiji i sa psihološkim teorijama moralnog razvoja.

kao što je već rečeno, neophodan uslov morala. Prema argumentima koje nameravam da izložim, moralno rasuđivanje i ponašanje nije moguće bez značajnog uticaja emocija.

Prema jednoj od prihvaćenih definicija emocije predstavljaju složen skup interakcija između različitih subjektivnih i objektivnih činilaca koji je posredovan neurofiziološkim mehanizmima koji mogu da 1) dovedu do afektivnog iskustva kao što je osećanje uzbuđenja, zadovoljstva/nezadovoljstva; 2) da generišu kognitivne procese kao što su emocionalno relevantni perceptualni efekti, procene vrednosti, procesi etiketiranja; 3) aktiviraju brojna fiziološka usklađivanja s uslovima u kojima se uzbuđenje pojavljuje; i 4) vode ponašanju koje je često, ali ne i uvek, ekspresivno, ciljno orijentisano i adaptivno (Kleinginna and Kleinginna 1981, 355). U skladu s prethodnom definicijom emocije imaju tri relevantne dimenzije: a) subjektivno iskustvo, čija funkcija je samoregulacija – subjektivna fenomenalna stanja signaliziraju da jedinka ima potrebu, želju ili neko osećanje; b) ekspresiju, bilo da je reč o facijalnoj ekspresiji, položaju tela, slanju hemijskih signala ili vokalizaciji, čija funkcija je komunikacija i socijalna koordinacija; c) fiziološku reakciju, čija funkcija je održanje homeostaze i manifestuje se promenama u radu različitih organskih sistema (Buck 2014). Uzete zajedno ove dimenzije daju aproksimativnu osnovu za motivaciju.

Emocionalne reakcije su kod ljudi povezane s individualnim vrednostima, stanjima, ciljevima i životnim projektima i možemo da razlikujemo njihove dve funkcije u relaciji s altruizmom i empatijom (Batson 2011). Jedna je *informaciona*, intenzitet emocije koja se iskušava prema određenoj osobi, stanju ili događaju ukazuje na to koliko se oni vrednuju. S druge strane, emocije imaju *amplifikacionu* funkciju, to jest imaju sposobnost da potencijalnu motivaciju transformišu u aktualnu. Kada individua opaža da je neko u nevolji, intenzitet emocije koju oseća, na primer straha ili anksioznosti, informiše individuu da li i koliko tu osobu vrednuje, a zatim, na osnovu dobijene informacije, može da bude motivisana da reaguje ili da ostane pasivna. Pri tome, nije nužno da se ovaj proces odvija isključivo u kognitivnoj ravni. Po svemu sudeći, ljudi su evoluirali da funkcionišu u različitim socijalnim kontekstima ne oslanjajući se na kognitivne procese (Hoffman 1981).

U narednim odeljcima biće razmotreni psihološki (emocionalni i kognitivni) i neurobiološki aspekti empatije i njen značaj za evoluciju psihološkog altruizma i morala. U



istom kontekstu, posebna pažnja će biti posvećena integrativnom evolucionom modelu empatije koji je zasnovan na mehanizmu zajedničkog kodiranja percepcije i akcije. Biće razmotrena eksperimentalna evidencija dobijena iz proučavanja ponašanja dece i primata čiji bihevioralni obrasci uključuju neke aspekte za koje se tvrdi da predstavljaju osnovne elemente morala. Konačno, potpora za stanovište da emocije i empatija imaju važnu funkciju u evoluciji morala dolazi iz proučavanja osoba s psihopatskim karakteristikama, koje ne iskušavaju emocionalnu empatiju, krivicu i kajanje, o čemu će biti reči u posebnom poglavlju.

## 3.2 Empatija

Prema *Malom oksfordskom rečniku engleskog jezika*, empatija je „sposobnost da se razumeju i dele osećanja drugog“, dok je sažaljenje (odnosno saosećanje, eng. *sympathy*), koje se često meša s pojmom empatije definisano kao „osećanje žaljenja i tuge zbog nečije nesreće“. Iako s rečničkim definicijama ne možemo da odemo daleko kada je reč o filozofiji, one nam u ovom slučaju mogu biti od početne koristi. Činjenica je da o tome šta je empatija nema načelne saglasnosti i da se često naizgled slični, iako konceptualno različiti fenomeni nazivaju istim imenom, mada se možda pre može govoriti o različitim aspektima jedinstvenog fenomena koji imaju zajedničku funkcionalnu i neurobiološku osnovu. Bejtson izdvaja osam takvih fenomena koji se mogu naći u literaturi koja se bavi empatijom, i koji sežu od lične uznemirenosti i emocionalne zaraze do kognitivne empatije (Batson 2009, 2011).

Danas postoji opšti konsenzus o tome da je su empatički kapaciteti čoveka proizvod delovanja prirodne selekcije i da uključuju dve dimenzije – emocionalnu i kognitivnu (Darwin 1981, 2009a, Hoffman 2000, Preston and de Waal 2002b, de Waal 2008, Vaish, Carpenter, and Tomasello 2009, Batson 2011). Mada Darwin ne koristi termin empatija, već termin saosećanje (eng. *Sympathy*), nesumnjivo je da ga koristi u značenju koje danas obično vezujemo za pojam empatije (deljenje zajedničkih osećanja i stavljanje sebe u poziciju druge osobe, itd.). U prilog tome da kada piše o saosećanju, Darwin na umu ima empatiju, a ne sažaljenje, govori i činjenica da individue mogu da dele ne samo negativne emocije (bol, tugu, patnju, itd.) već i pozitivne (zadovoljstvo, radost, sreću, itd.).

Ne samo da su empatički kapaciteti proizvod prirodne selekcije, već se pojavljuju u filogenetskom kontinuumu, što znači da je empatija na različitim stupnjevima razvoja prisutna i kod drugih sisara, pre svega primata, tako da se o njoj ne može govoriti kao o specijes specifičnoj karakteristici ljudi (de Waal 1996, 2008, de Waal et al. 2006, Flack and De Waal 2000, Preston and de Waal 2002b). Teza o kontinuitetu, prema kojoj empatija zauzima centralno mesto između socijalnosti sisara uopšte i moralnosti ljudi, i čini jedan od gradivnih blokova morala, stoji nasuprot Hakslijevom, Vilijamsovom i Dokinsovom stanovištu o diskontinuitetu, prema kome iz surovosti evolucionog procesa, slede jednako surovi učinci. Uzmimo Hakslijevu poziciju za primer, prema kojoj u prirodi ne možemo da nađemo nikakve (moralne) uzore, jer priroda nije „škola vrline, već štab neprijatelja“ (Huxley 2009, 27). Nasuprot tome, teza o kontinuitetu zasnovana je na Darwinovom stavu o postojanju kontinuiteta u emocionalnim reakcijama i kognitivnim sposobnostima viših životinja i ljudi. Između ostalog, u prilog tezi o kontinuitetu govori i činjenica da su populacije velikih vretenastih neurona, koji povezuju regione mozga koji učestvuju u procesiranju emocija i empatičkih reakcija (prednji cingularni korteks i prednji insularni korteks), otkriveni samo kod velikih bezrepih majmuna i ljudi, kao i da je njihov broj kod starih ljudi najveći, a da srazmerno opada kod dece, gorila, bonoboa i šimpanzi (Decety 2011a).

Ipak, nisu svi uvereni da su empatički kapaciteti prisutni kod drugih životinja. Mada ne sumnja u evoluciono poreklo empatije, Bejtson je sumnjičav kada je reč o tome da (ne-ljudski) primati i druge životinje imaju dovoljno sofisticirane kognitivne kapacitete za empatiju (Batson 2011). Ovaj autor naglašava kognitivnu dimenziju empatije za koju je nužno da jedinka ne samo percipira drugu jedinku u nevolji, već i da na intrinzičan, tj. neinstrumentalan način vrednuje njenu dobrobit. Na taj način, da bismo uopšte govorili o empatiji neophodno je da jedinka opaža da druga jedinka ima neku potrebu ili da se nalazi u nevolji, zatim ovo opažanje mora da bude praćeno emocijom koja je orijentisana isključivo na drugu jedinku i ciljno orijentisanom željom da se potreba zadovolji ili nevolja ukloni. Za sada nema konkluzivnih dokaza da su druge životinje, uključujući šimpanze, sposobne za takvu vrstu vrednovanja.

Kao što postoji opšti konsenzus da je empatija proizvod delovanja prirodne selekcije, među teoretičarima gotovo da postoji opšti konsenzus da empatija može da motiviše altruističke postupke i da je važna za moralni razvoj individue (Hoffman 2000, Preston and de Waal 2002b,

de Waal 2008, Batson 2011, Decety 2011b). Prema ovom tumačenju, empatija je neposredni uzrok psihološkog altruizma, koji je, prema nekim stanovištima, nužan uslov morala (Williams 1976). Izuzetak od ovog stanovišta je Prinz (Jesse Prinz), koji, iako smatra da su moralna svojstva zavisna od emocionalnog odgovora koji određeni postupci izazivaju u posmatraču i priznaje važnost empatije za neke aspekte ljudskog postupanja, ipak insistira na tome da empatija kao automatski proces nije nužan uslov gotovo ni za jedan aspekt morala, a kao jedan od razloga navodi slabosti njenog motivacionog potencijala (Prinz 2011).

Prema stanovištu koje zastupam, emocije i empatija, kao specifična sposobnost neposrednog iskušavanja tuđeg emocionalnog iskustava, imaju posrednu funkciju u evoluciji morala. One omogućavaju altruističku motivaciju, ali i usmeravaju procese rasuđivanja i odlučivanja, kao i usvajanja i održanja verovanja.

### **3.3 Evolucionni model empatije**

Verovatno najrazvijeniji integrativni evolucionni model empatije predlažu Stefani Preston i Frans de Val (Preston and de Waal 2002b). Ovaj model je zasnovan na teoriji zajedničkog kodiranja percepcije i akcije. Prema ovoj teoriji, percepcija i akcija nisu dva potpuno izolovana aspekta ljudske psihologije, već opaženi događaji i odgovarajući postupci dele zajednički reprezentacioni domen (Prinz 1997, Van Der Wel, Sebanz, and Knoblich 2013). To znači da će dve individue  $A$  i  $B$ , ako  $A$  opaža kako  $B$  nešto čini, i  $A$  i  $B$  imati slične reprezentacije postupka. Prema teoriji zajedničkog kodiranja, reprezentacije postupaka mogu da se aktiviraju jednostavnim posmatranjem šta neko čini. Reč je o međusobno isprepletenim i međuzavisnim aspektima koji su zajednički kodirani tako da određeni opažaji mogu da indukuju određene akcije na osnovu sličnosti, ali i postupci mogu da oblikuju opažanje (Prinz 1997, Van Der Wel, Sebanz, and Knoblich 2013). Kada ovaj mehanizam primenimo na domen emocionalnih reakcija i sposobnosti za empatiju, onda možemo da kažemo da za dve jedinke  $A$  i  $B$  važi sledeće: ako  $A$  posmatra  $B$  i  $B$  je u stanju  $F$ , onda je u  $A$  automatski proizvedeno slično ili identično stanje. Ova dva stanja ne moraju nužno da budu identična, budući da jedinke ne moraju nužno da proživljavaju isto, već samo slično iskustvo. Na primer, opažanje da jedinka  $A$  trpi bol, i emocionalna reakcija jedinke  $B$  na njen bol, nije identična (po intenzitetu) bolu jedinke  $A$ , ali na

neurofiziološkom planu među njima postoji značajno podudaranje. Reč je, dakle, o razlici u intenzitetu, a ne u vrsti iskustva.

Prema modelu empatije koji je zasnovan na zajedničkom kodiranju percepcije i akcije, svi empatički procesi se oslanjaju na ovu osnovnu funkcionalnu organizaciju nervnog sistema, koja omogućava da se u posmatraču automatski i bez svesne kontrole aktiviraju neuronske reprezentacije koje su slične ili odgovarajuće onim reprezentacijama koje ima opažena individua (Preston and de Waal 2002b, de Waal 2008). Prema mišljenju pomenutih autora ovakva organizacija nervnog sistema ima nesumnjive pozitivne efekte na adaptivnu vrednost organizma. Najpre, ako percepcija i akcija dele zajednički reprezentacioni domen i odgovarajući neurofiziološki mehanizam, to znači da se iste reprezentacije koriste za kodiranje opaženog objekta (druge jedinke) i njegovog ponašanja, i onda možemo da pretpostavimo da bi evolucija ovakvog nervnog sistema bila ekonomičnija i efikasnija. Jedinica koja je dizajnirana tako da reaguje strahom na opažene signale straha druge jedinke ima više šansi za preživljavanje i reprodukciju od onih jedinki koje, da bi bile uplašene, moraju da se direktno suoče sa zastrašujućim stimulusima. U evolucionom smislu može da bude važno da percepcija usmerava i oblikuje ponašanje, pogotovo za jedinke koje žive u grupi. Na ovaj način jedinke mogu na jednostavno da koordinišu ponašanje. Dalje, zajedničko kodiranje percepcije i akcije omogućava lakšu komunikaciju i koordinaciju između majke i mladunčeta (deteta), pogotovo kada je reč o prepoznavanju i učenju na osnovu izraženih emocionalnih signala, a važno je i za kasniji razvoj emocionalne regulacije i kompetencije. Zapaženo je da rano razdvajanje majke i mladunčeta uzrokuje visok nivo stresa i dovodi do razvojnih poteškoća. Osim toga, emocionalne reakcije mladunaca (deteta) usmeravaju ponašanje roditelja, tako da na primer visok nivo uznemirenosti deteta vodi visokom nivou uznemirenosti roditelja, što uzrokuje nastojanje da se uznemirenost ukloni, npr. hranjenjem ili pomaganjem. Konačno, signali koji upućuju na uznemirenost su evoluirali da bi bili okončani. Ovi signali izazivaju uznemirenost kod svih jedinki, što omogućava da odgovarajuće reakcije ne budu vezane samo za odnos roditelj/dete, već i za srodnike, ali i za jedinke izvan srodničkog kruga.

Iako bi se moglo pomisliti da je prema ovom modelu zajedničko kodiranje percepcije i akcije evoluiralo delovanjem selekcije u srodničkom krugu, to nije zaključak do koga Preston i de Val dolaze. Mehanizam zajedničkog kodiranja percepcije i akcije je evoluirao jer je

„adaptivan za osnovne reakcije na uslove životne sredine i život u grupi“ (Preston and de Waal 2002b, 9). Iako ovaj mehanizam ima važnu funkciju u komunikaciji i koordinaciji majke i mladunčeta (deteta), u prepoznavanju i učenju na osnovu izraženih emocionalno signala, ova njegova funkcija se tumači kao *eksaptacija*, a ne adaptacija. To znači da je njegova osnovna funkcija kod altricijalnih vrsta – vrsta čiji mladi dospevaju na svet nedovoljno razvijeni i imaju potrebu za prolongiranom roditeljskom brigom – dobila novu namenu, kao što je potpora roditeljske brige i razvoj empatičkih kapaciteta (za pojam eksaptacije vidi Gould and Vrba 1982, Preston and de Waal 2002b). Mada mehanizam percepcije i akcije nije evoluirao u srodničkom krugu, empatija koja je zasnovana na ovom mehanizmu je evoluirala najpre u kontekstu brige o potomstvu, a zatim je verovatno došla pod uticaj selekcije u srodničkom krugu, a potom izmeštena i na recipročne relacije.

Prema modelu zajedničkog kodiranja percepcije i akcije, „empatija je izvorni prelingvistički oblik međusobnog povezivanja koji je tek sekundarno došao pod uticaj jezika i kulture“ (de Waal et al. 2006, 24). Na taj način, ovakvo razumevanje empatije dopušta da se pojam empatije primenjuje van domena ljudske psihologije i da se empatički kapaciteti pripisuju i onim organizmima koji nemaju jezik i kulturu, ali su evoluciono bliski ljudima. U ovaj model empatije su uključeni različiti psihološki slojevi i njihovi neurobiološki supstrati, od elementarnih emocionalnih reakcija na emocionalne reakcije drugih jedinki (emocionalna zaraza), do složenijih i kognitivno zahtevnih empatičkih reakcija na emocionalna stanja ili situaciju u kojoj se drugi nalazi, kao što je afektivno zauzimanje perspektive.

Hofman empatiju definiše kao emocionalnu *reakciju* koja u većoj meri odgovara situaciji u kojoj se nalazi druga individua, nego što odgovara sopstvenoj (Hoffman 2000). Za razliku od Hofmana, Preston i de Val empatiju definišu kao *proces* što im omogućava da daju jedinstveno objašnjenje za različite empatičke fenomene, od emocionalne zaraze do kognitivne empatije i pomaganja. Empatija je, tako, proces u kome opažanje emocionalnog ili situacionog stanja druge individue u posmatraču proizvodi emocionalno stanje koje je prikladnije za opaženu individuu nego za sopstveno pređašnje emocionalno stanje ili situaciju (Preston and de Waal 2002b). Uvođenje mogućnosti iščitavanja *situacionih signala* u definiciju je značajno, jer individua u odsustvu emocionalnih signala (facijalna ekspresija, vokalizacija, itd.) može da se osloni na informacije dobijene jednostavnim opažanjem situacije, koje mogu da generišu emocionalnu

reakciju posmatrača i odgovarajuće ponašanje. Eksperimentalna evidencija pokazuje da deca uzrasta 18 meseci mogu da saosećaju s nepoznatom osobom kojoj je učinjena šteta, čak i ako ona ne pokazuje nikakve znake uznemirenosti, o čemu će više reći biti kasnije (Vaish, Carpenter, and Tomasello 2009).

### 3.4 Model babuške

Paralelizam između filogenetskog pojavljivanja i razvojnog ispoljavanja empatije u čijoj osnovi je mehanizam zajedničkog kodiranja percepcije i akcije može slikovito da bude predstavljen modelom babuške (eng. *Russian doll model*), čiji spoljašnji slojevi oslikavaju visoko razvijene i kognitivno kompleksne empatičke sposobnosti, dok su unutrašnji slojevi bazični, emocionalni i bez ili s malo kognitivne kontrole (de Waal et al. 2006, de Waal 2008). De Val razlikuje tri takva sloja, koji se nadovezuju jedan na drugi, a u čijem središtu se nalaze mehanizam zajedničkog kodiranja percepcije i akcije. Ovi slojevi imaju svoje bihevioralne korelate vezane za oponašanje, budući da postoji paralelizam između oponašanja i empatije (Preston and de Waal 2002b, de Waal 2008). Prvi se odnosi na podudaranje stanja i rezultira ličnom uznemirenošću ili emocionalnom zarazom koja je najniži stupanj ispoljavanja empatije, a njen bihevioralni korelat vezan za oponašanje je motorna mimikrija. Na drugom mestu je saosećajna zabrinutost, prvi sloj empatije koji osim emocionalne uključuje i kognitivnu dimenziju, ona se ispoljava tešenjem, a njen bihevioralni korelat vezan za oponašanje je koordinacija u postizanju zajedničkog cilja. Treći sloj zauzima afektivno zauzimanje perspektive, zahteva visok nivo kognitivne obrade i ispoljava se ciljanim pomaganjem, altruističkim ponašanjem koje je usklađeno s potrebama druge jedinke; njen bihevioralni korelat vezan za oponašanje je emulacija, tj. učenje na osnovu posmatranja i primena naučenog u novim kontekstima.

Ova troslojna karakterizacija empatičkih kapaciteta je donekle podudarna s Hofmanovim petostepenim i preciznijim razumevanjem razvoja empatije (Hoffman 2000). Hofman razlikuje pet načina ispoljavanja i isto toliko stupnjeva razvoja empatije, a svaki od ovih stupnjeva i načina ispoljavanja je u kognitivnom smislu napredniji od prethodnog. Tako su 1) motorna mimikrija, 2) klasično uslovljavanje i 3) direktna asocijacija, primarni, nevoljni, automatski i prelingvistički načini ispoljavanja empatije, koji ukazuju da ljudi mogu da bez svesne kontrole reaguju na

emocionalno stanje druge osobe, dok su 4) posredovana asocijacija i 5) afektivno zauzimanje perspektive kognitivno zahtevniji načini ispoljavanja empatije. Pet stupnjeva razvoja empatije su sledeći: a) reaktivni plač novorođenčadi; b) egocentrična empatijska uznemirenost, u kojoj bebe reaguju na uznemirenost drugih kao da je reč o njihovoj sopstvenoj uznemirenosti, bez diferencijacije ja/drugi; c) kvaziegocentrična empatijska uznemirenost, u kojoj je diferencijacija ja/drugi nepotpuna; uznemirenost se doživljava kao tuđa, ali se zbog nedostatka teorije uma, načini pomaganja oblikuju shodno sopstvenim potrebama; d) stvarna empatijska uznemirenost, pojavljuje se oko 18. meseca starosti i koincidira s početkom prepoznavanja u ogledalu, a zahvaljujući pojavi teorije uma, pomoć se usklađuje s potrebama druge osobe; i e) empatija za tuđe iskustvo van neposredne situacije, odlikuje se razumevanjem da druge individue imaju živote koji načelno mogu biti srećni ili tužni. Prema Hofmanovom mišljenju, od treće razvojne faze (c), empatička uznemirenost se transformiše u saosećanje, odnosno u izvesnu kombinaciju empatije i saosećanja, a de Val govori o saosećanju zasnovanom na empatiji (de Waal 1996, Hoffman 2000).

### **3.4.1 Emocionalna zaraza**

Emocionalna zaraza je prvi stupanj empatičke uznemirenosti i u tesnoj je vezi s facijalnom ekspresijom i motornom mimikrijom. Prema modelu empatije koji je zasnovana na teoriji zajedničkog kodiranja percepcije i akcije, emocionalna zaraza i motorna mimikrija stoje u osnovi drugih empatičkih procesa (Preston and de Waal 2002b). Ovakva vrsta empatičkog reagovanja je prisutna kod ljudi jednako kao i kod drugih životinja i ide u prilog tezi o kontinuitetu.

Emocionalna uznemirenost jedne životinje u susretu s neprijatnim ili zastrašujućim stimulusima vodi automatskoj i sinhronizovanoj uznemirenosti drugih životinja u istoj grupi, a ovakva reakcija može biti važna za povećanje adaptivne vrednosti, jer životinje ne moraju da se direktno izlože izvoru inicijalne uznemirenosti (opasnosti) već automatski reaguju na emocionalne signale drugih jedinki (de Waal 2008). Kada je reč o ljudima, deca u najranijem uzrastu, već od rođenja, mogu da komuniciraju s roditeljima pomoću emocionalnih signala, da oponašaju različite izraze lica roditelja ili staratelja i reaguju na njihovu facijalnu ekspresiju i mišićni tonus, što je značajno za njihov budući emocionalni razvoj (na primer, postporođajna

depresija se negativno odražava na budući emocionalni razvoj deteta); ovaj mehanizam je automatski, nije pod svesnom kontrolom i verovatno je zasnovan na mehanizmu zajedničkog kodiranja i zajedničkim reprezentacijama (Hatfield, Cacioppo, and Rapson 1994, Hoffman 2000). S druge strane, odrasli automatski i kontinuirano tokom razgovora, oponašaju izraze lica, visinu glasa, položaje tela, pokrete i ciljno orijentisano ponašanje drugih, što utiče na njihovo subjektivno emocionalno iskustvo (Hatfield, Cacioppo, and Rapson 1994).

Jedan od primera emocionalne zaraze je reaktivni plač novorođenčadi (Hoffman 2000). Ljudska bića su prirodno programirana da reaguju na patnju i uznemirenost drugih, pa tako u porodilištima sve bebe plaču u glas kada jedna počne da plače, a zatim se postepeno umiruju. Reaktivni plač može biti protumačen i kao naučena reakcija zasnovana na uslovljavanju, ali Hofman misli da je verovatnije objašnjenje da je on biološki determinisan. Ovakvo automatsko reagovanje na tuđu uznemirenost prestaje negde oko šestog meseca starosti.<sup>36</sup>

Mada se obično smatra prvom empatičkom reakcijom kod ljudi, moguća su i alternativna objašnjenja. Jedno takvo objašnjenje predlaže Bejtson (Batson 2009). Prema ovom tumačenju, reaktivni plač novorođenčadi je kompetitivna strategija, a ne empatička reakcija; ona je bliska ponašanju mladunaca u ptičijem gnezdu: galami se da bi se dobilo više pažnje i obrok više. Uznemirenost mladunaca izaziva uznemirenost roditelja, koji teže da uznemirenost što pre okončaju, a najviše pažnje dobijaju oni koji su najuznemireniji na štetu onih koji izražavaju uznemirenost sniženim intenzitetom. Na taj način, uznemirenost mladunaca je u funkciji *egoističke efikasnosti*, tj. biološke sebičnosti, jer ima efekata na povećanje adaptivne vrednosti organizma i nije reakcija na uznemirenost druge jedinke. Čak i da je tako, uznemirenost mladunaca indukuje empatičku reakciju i altruističko ponašanje roditelja, koje opet s biološkog stanovišta može biti protumačeno kao sebično: s jedne strane, jer roditelji žele da prekinu sopstvenu uznemirenost, a s druge, jer pomoć pružena mladuncima doprinosi povećanju njihove adaptivne vrednosti. Ovaj argument se oslanja na činjenicu da svako ponašanje koje vodi povećanju adaptivne vrednosti organizma *uvek* može biti protumačeno kao sebično, bez obzira na njegove bihejvioralne posledice i iz tog razloga se može smatrati irelevantnim (Preston and de Waal 2002a, b). Kako su signali koji izazivaju uznemirenost evoluirali, jer drugi imaju

---

<sup>36</sup> Ne reaguje se samo na signale koji izazivaju uznemirenost. I smeh može da bude zarazan.



tendenciju da ih prekinu, čini se da u ovom slučaju sebičnost nije moguće istisnuti iz objašnjenja, ali je to sebičnost koja nije nespojiva s altruizmom.

S druge strane, u kontekstu modela empatije koji je zasnovan na zajedničkom kodiranju percepcije i akcije, emocionalna zaraza nije nužno pasivan proces, tj. jednostavna reakcija na uznemirenost druge jedinke, već jedinke mogu sopstvene emocionalne signale da koriste instrumentalno (de Waal 2008). Burna ekspresija uznemirenosti izaziva uznemirenost drugih jedinki koje zatim nastoje da je prekinu, tako da na taj način može da se utiče na njihovo ponašanje.<sup>37</sup>

Prema Bejtsonu, Preston i de Val prenaglašavaju ulogu zajedničkog reprezentacionog domena i motorne mimikrije za ostale empatičke procese (Batson 2009, 2011). Kao protivargument on navodi da je oponašanje ograničeno samo na neke postupke i da se ne oponaša uvek, što je svakako tačno, ali Bejtson prenebregava činjenicu da s razvojem prefrontalnog korteksa koji učestvuje u odlučivanju i kontroli emocija, oponašanje postaje kontrolisana aktivnost, kao i empatija u pojedinim slučajevima, tako da ovi procesi mogu da budu inhibirani delovanjem prefrontalnog korteksa (Preston and de Waal 2002b).

### **3.4.2 Saosećajna zabrinutost**

Saosećajna zabrinutost predstavlja drugi sloj u modelu babuške koji počiva na emocionalnoj zarazi i ličnoj uznemirenosti, ali uključuje viši stepen kognitivne kontrole i altruističku motivaciju. On donekle odgovara fazi (c) u Hofmanovoj klasifikaciji, u kojoj je diferencijacija između ja i drugi nepotpuna. Iako se na ovom stupnju uznemirenost doživljava kao tuđa, zbog nedostatka teorije uma, načini pomaganja se oblikuju shodno sopstvenim potrebama, iz tog razloga često mogu biti neadekvatni. Deca u ovoj fazi nastoje da uznemirenu individuu uteše na način na koji bi utešila sebe u sličnoj situaciji, maženjem, upućivanjem na

---

<sup>37</sup> Ako uznemirenost izaziva uznemirenost u posmatraču, on može da nastoji da je se oslobodi, a jedini način za to nije ukazivanje pomoći jedinki koja je inicirala uznemirenost. To može da se učini i tako što se udaljava od neprijatnog stimulusa (tj. od uznemirene jedinke) ili se na neprijatni stimulus reaguje agresijom u nastojanju da se on prekine (Hoffman 2000). Rezultat emocionalne zaraze može da bude i zlostavljanje u odnosima roditelj/dete, ne samo altruističko zadovoljenje potreba (Preston and de Waal 2002b).

osobu koja bi njih same utešila (na primer, sopstvena majka, a ne majka uznemirenog deteta, dodavanjem omiljene igračke i sl.) (Hoffman 2000).

Kada je reč o ponašanju primata, saosećajna zabrinutost se manifestuje na više načina kao što su odbrana od agresije, bištenje, deljenje hrane i tešenje, kome de Val posvećuje posebnu pažnju, budući da je reč o bihevioralnom obrascu koji je, prema njegovim rečima, svojstven isključivo ljudima i velikim bezrepm majmunima, dok je potpuno odsutan kod običnih majmuna (de Waal 1996, Flack and De Waal 2000, Preston and de Waal 2002a, de Waal et al. 2006, de Waal 2008). Ovaj obrazac ponašanja se odnosi na umirivanje jedinke koja je pretrpela agresiju ili je izašla poražena iz konflikta, ali još nema konkluzivnih dokaza da tešenje pouzdano vodi smanjenju uznemirenosti kod primata. U kontekstu teze o kontinuitetu, tešenje je filogenetski značajno za razumevanje empatije, jer uključuje sposobnost razlikovanja sebe i sopstvene situacije od drugog i situacije u kojoj se on nalazi, uz zadržavanje emocionalne povezanosti među jedinkama, za šta su, kako se čini, sposobni samo ljudi i njihovi najbliži evolucionari srodnici.

### ***3.4.3 Afektivno zauzimanje perspektive***

Afektivno zauzimanje perspektive predstavlja poslednji i kognitivno najsloženiji aspekt empatije u predloženom modelu i podudara se delimično s razvojnom fazom (d) u Hofmanovom modelu; kod dece se pojavljuje oko 18. meseca starosti i podudara se s početkom prepoznavanja u ogledalu, sposobnošću koju među primatima, imaju samo ljudi i veliki čovekoliki majmuni. U ovom periodu se pojavljuju znaci teorije uma što, između ostalog, za posledicu ima da deca mogu da razumeju da druge individue imaju potrebe koje se razlikuju od njihovih, tako da se pomoć usklađuje s potrebama druge osobe.

De Val koristi različite termine kako bi označio ovu vrstu empatije: atribucija (de Waal et al. 2006), zauzimanje perspektive (de Waal 2008, 2012), kognitivna empatija (Preston and de Waal 2002a). Takođe, ono što na jednom mestu naziva kognitivna empatija (de Waal et al. 2006), na drugom mestu naziva saosećajna zabrinutost (de Waal 2008, 2012), a ono što se smatra kognitivnom empatijom u jednom slučaju (Preston and de Waal 2002a), zauzima drugačije mesto u slojevima babuške od kognitivne empatije u drugom slučaju (de Waal et al. 2006), koja

se u kasnijim radovima naziva saosećajna zabrinutost (de Waal 2008, 2012). Ovakve promene u korišćenju terminologije za posledicu imaju stvaranje konceptualne zbrke.

U ovom radu o trećem sloju empatičkih kapaciteta govoriću kao o *afektivnom zauzimanju perspektive*, budući da De Val tvrdi da samo posmatranje situacije iz pozicije drugog i postavljanje sebe u njegovu poziciju nije empatija u pravom smislu reči, ako ovo zauzimanje perspektive nije praćeno odgovarajućom emocionalnom reakcijom s altruističkim motivacionim potencijalom (de Waal 2008). To ima smisla iz još jednog razloga. U slučaju emocionalne zaraze jedinka proizvodi uznemirenost u drugoj jedinki kako bi zadovoljila neke sopstvene adaptivne potrebe, kao što su potreba za hranom ili zaštitom. S druge strane, empatija koja je svedena isključivo na kognitivno zauzimanje perspektive može da vodi bihejvioralnim ispoljavanjima u kojima individua teži instrumentalnom izazivanju visokog nivoa uznemirenosti druge individue samo da bi posmatrala kako individua pati ili da bi doživela uzbuđenje zbog njene patnje (Preston and de Waal 2002b, Blair and White 2013).

Ako je zauzimanje perspektive udruženo s adekvatnom emocionalnom reakcijom, onda ovakav vid empatije može da uzrokuje autentičnu altruističku motivaciju i odgovarajuće ponašanje koje je usmereno ka zadovoljenju potreba jedinke kojoj je potrebna pomoć. U tom slučaju empatija predstavlja „evoluirani neposredni mehanizam za ciljno orijentisani altruizam“ (de Waal 2008, 291); za pružanje pomoći koja je podešena prema specifičnim potrebama jedinke u nevolji.

Alternativno, egoističko, objašnjenje ciljno orijentisanog altruističkog ponašanja u ovom kontekstu bi se odnosilo na to da jedinka koja pomaže očekuje spoljašnju nagradu ili naknadnu reciprocijaciju za svoje postupanje. Mada takvo objašnjenje ne može *a priori* biti isključeno, ono je malo verovatno, jer je uticaj takvih činilaca na evoluciju altruizma zanemarnjiv; u prirodnim uslovima spoljašnje nagrađivanje je neizvesno, najčešće veoma odloženo i slab je podsticaj za učenje (de Waal 2008). Kada je reč o ljudima u društvenom kontekstu, tačno je da ukazivanje pomoći omogućava da se dobije društveno priznanje ili izbegne kažnjavanje, ali to ne znači da je nagrađivanje krajnji cilj altruističke motivacije zasnovane na empatiji (afektivnom zauzimanju perspektive), već su takve nagrade sporedni proizvodi, to jest neintencionalne posledice i rezultati delovanja drugih motiva (Batson 2011). Ali, osim spoljašnjim nagradama osoba može

da teži i nekoj vrsti unutrašnje, emocionalne gratifikacije, čime bi se altruizam sveo na sebičnost. Međutim, kao i u slučaju emocionalne zaraze, to nije vrsta sebičnosti koja je nespojiva s altruizmom. Kako je u slučaju altruizma zasnovanom na empatiji takvu vrstu gratifikacije moguće dobiti samo preko zadovoljenja potreba ili pružanjem pomoći nekom drugom, ovakav oblik egoističkog objašnjenja ne podriva osnovni argument u prilog altruističkoj motivaciji (de Waal 2008).

Uprkos anegdotskim podacima da su za afektivno zauzimanje perspektive, osim ljudi, sposobne i neke druge životinje, kao što su delfini, kitovi, slonovi i veliki bezrepi majmuni, ipak nema konkluzivnih dokaza koji bi van osnovane sumnje potvrdile ovaj stav (de Waal 1996, 2008). S druge strane, eksperimentalna evidencija ukazuje da su deca u najranijem uzrastu (oko godinu dana starosti), već sposobna za afektivno zauzimanje perspektive, da su altruistički motivisana i da imaju jasno izraženu sklonost ka instrumentalnom pomaganju.

### **3.5 Afektivno zauzimanje perspektive i altruističko pomaganje kod dece i šimpanzi**

U nizu eksperimenata koje su izveli Warneken i Tomasello utvrđeno je da deca u najranijem uzrastu nastoje da pomognu osobi s kojom nisu u srodstvu, a kojoj je potrebna pomoć (Warneken and Tomasello 2006, 2007, 2009a, b). Načelno, predmet testiranja bio je da se utvrdi da li deca uzrasta između 14 i 18 meseci pomažu osobi da dohvati predmet koji je van domašaja, da li joj pomažu da ukloni prepreku koja ometa izvršenje nameravanog zadatka, da li ispravljaju grešku koju osoba čini i da li pomažu u izboru ispravnog bihevioralnog sredstva za ostvarenje određenog zadatka (Tomasello 2009). Rezultati ovih eksperimenata ukazuju na to da su deca već u ranom uzrastu altruistički motivisana, budući da pomažu nepoznatim osobama koje ne zahtevaju da im se pomogne i bez očekivanja nagrade. Štaviše (materijalno) nagrađivanje nema pozitivnog efekta na pomaganje. Iako bi se moglo pomisliti da nagrađivanje deluje podsticajno u dečjem uzrastu, zapaženo je upravo suprotno - nagrađivanje podriva intrinzičnu motivaciju da se pomogne, koju deca ispoljavaju u najranijim životnom periodu (Warneken and Tomasello 2009b). Deca koja su nagrađena za pruženu pomoć u kasnijim vremenskim intervalima su manje sklona da pomognu.

U zavisnosti od uzrasta, postoje značajne razlike u stepenu ispoljenog pomaganja, tako da deca između 18 meseci i dve godine starosti pomažu gotovo u svim zadacima, dok ona stara 14 meseci pomažu isključivo u zadacima koji se odnose na situacije u kojima je predmet van domašaja eksperimentatora (Warneken and Tomasello 2007). Deca s navršenih 18 meseci pomažu i u kompleksnijim zadacima kada se od njih očekuje da pre nego što dodaju predmet koji je van domašaja moraju da zaobiđu neku prepreku, dok deca uzrasta 20 meseci pomažu čak i ako moraju da prekinu igru zanimljivim igračkama (Warneken and Tomasello 2009a).

Mada bi moglo da se pomisli da je instrumentalno pomaganje relativno jednostavan zadatak, to ipak nije slučaj. Da bi individua *A* pomogla individui *B*, neophodno je da *A* najpre ispravno prepozna namere i želje individue *B*, tj. šta je ono što želi da postigne, i da zatim ova opažena intencionalna stanja s karakterističnim bihevioralnim ispoljavanjem kod *A* proizvedu odgovarajući motivacioni učinak i ponašanje koje je usklađeno s potrebama individue *B* (Warneken and Tomasello 2009a). Dakle, reč je o kompleksnom procesu prepoznavanja intencija i željenih ciljeva, kao i razlikovanja namernih od slučajnih postupaka, što deca uzrasta između 12 i 18 meseci mogu da savladaju.<sup>38</sup> Iako složen, ovaj proces je manje kompleksan od saradnje. Iz tog razloga, stupanj saradnje koji ispoljavaju deca na uzrastu od 14 meseci je veoma nizak, ona nikada ne saraduju međusobno, već isključivo kada je u saradnju uključena odrasla osoba s jasno postavljenim ciljevima i kada su zahtevi za koordinacijom postupaka mali (Warneken and Tomasello 2007).

Princ je skeptičan kada je reč o interpretaciji rezultata ovih istraživanja i nudi alternativno objašnjenje (Prinz 2007). On ističe da su deca do ovog uzrasta već imala dovoljno iskustva sa sličnim zadacima koje im roditelji često postavljaju kroz igru, tako da se možda pre radi o naučenom ponašanju, nego o biološki programiranom kapacitetu. Takođe, moguće je da se uopšte ne radi o dispoziciji da se pomogne već o posebnom slučaju opštije tendencije da se oponašaju drugi i da se neuspešno izvedeni ciljno orijentisani postupci okončavaju. Ovaj argument nije opovrgavajući, ako se na umu ima da prema modelu empatije koji je zasnovan na teoriji zajedničkog kodiranja percepcije i akcije postoji kontinuitet između motorne mimikrije i

---

<sup>38</sup> U eksperimentalnim uslovima eksperimentator slučajno ispusti predmet, a u kontrolnim uslovima ga namerno baci.

bihevioralnih obrazaca koji uključuju pomaganje.<sup>39</sup> Na drugom mestu, eksperimenti su bili tako organizovani da eksperimentator nikada nije tražio pomoć direktno, kao što to čine roditelji kada se igraju s decom, već je samo izražavao negodovanje zbog nemogućnosti da izvrši zadatak (Warneken and Tomasello 2007). Warneken i Tomasello smatraju da se ne može govoriti o naučenom ponašanju, jer su eksperimentalni uslovi deci bili nepoznati, situacije nove, a eksperimentatori nepoznate osobe (Warneken and Tomasello 2006). Oni insistiraju na tome da rezultati eksperimenata ukazuju na činjenicu da za pomaganje i slične oblike ponašanja postoji prirodna (biološka) dispozicija i da oni nisu deo naučenog kulturnog obrasca (Warneken and Tomasello 2009b).

Konačno, eksperimentalna evidencija pokazuje da deca uzrasta 18 meseci mogu da saosećaju s nepoznatom osobom kojoj je učinjena šteta, čak i ako ona ne pokazuje nikakve znake uznemirenosti (Vaish, Carpenter, and Tomasello 2009). U odsustvu emocionalnih signala deca mogu da saosećaju s osobom koja je povređena oslanjajući se na situacione signale i kompleksan kognitivni proces kao što je afektivno zauzimanje perspektive. Afektivno zauzimanje perspektive podrazumeva zaključivanje o emocionalnom stanju druge osobe stavljanjem na njeno mesto na čemu se zatim zasniva sopstvena emocionalna reakcija. Ponekada se o afektivnom zauzimanju perspektive govori kao o kognitivnoj empatiji za čije ispoljavanje je neophodno da individua ima razvijenu teoriju uma (Hoffman 2000, Preston and de Waal 2002b, Batson 2009, 2011). Ovi eksperimenti ukazuju na mogućnost da već dvogodišnjaci imaju razvijene sposobnosti za afektivno zauzimanje perspektive, kao i neke aspekte teorije uma, iako se ranije verovalo da se teorija uma kod dece razvija između treće i četvrte godine. Osim toga, ovi eksperimenti sugerišu da empatija ima motivacioni potencijal. Kada u eksperimentalnim uslovima eksperimentator grubo uništi neki predmet drugog eksperimentatora, za koji je ovaj prethodno pokazao da mu je do njega naročito stalo, deca u kasnijim interakcijama mnogo više pomažu onome ko je bio povređen (tj. pretrpeo nepravdu). Opažanje da je neko pretrpeo nepravdu indukuje saosećanje i povećava šanse da mu se kasnije pomogne.

---

<sup>39</sup> Iz toga da u osnovi empatičkog kapaciteta stoji mehanizam percepcije i akcije, ne sledi da su sve empatičke reakcije od njega *neposredno* zavisne.

Ispoljavanje altruizma kod dece je brzo, spontano i neselektivno, bez uzimanja u razmatranje evolucionih činilaca koji se obično navode kao potencijalna rešenja paradoksa altruizma (Warneken and Tomasello 2009a, 2007). Deca ne računaju koeficijent srodstva, ne obaziru se na recipročne razmene usluga i ne pomišljaju na konsekvence delovanja indirektnog reciprociteta. Budući da je altruističko pomaganje kod dece neselektivno čini se da takva vrsta altruizma nije mogla da evoluirati, jer bi neselektivni altruisti, bez intrinzičnih mehanizama za prepoznavanje srodnika i varalica u recipročnim odnosima bili eksploatisani i zbrisani delovanjem prirodne selekcije. Varneken i Tomasello smatraju da se ovi mehanizmi ne ispoljavaju u najranijem detinjstvu kada je srodnički krug prirodna sredina deteta i kada su šanse za varanje i eksploataciju male, tj. kada su razlozi za njihovo ispoljavanje odsutni. Ti mehanizmi su posledica kasnijeg razvoja, kada altruistički postupci postaju selektivni, pristrasni u odnosu na srodnike, prijatelje i osobe kojima je individua afektivno privržena (Warneken and Tomasello 2009a). Rezultati ovih istraživanja su u skladu s nalazima da su empatičke reakcije osetljive na efekat bliskosti, sličnosti i da se povinuju predviđanjima teorije inkluzivne adaptivne vrednosti (Preston and de Waal 2002b, de Waal 2008).

### **3.6 Imaju li primati gradivne blokove morala?**

Potporna da sklonost ka instrumentalnom pomaganju nije deo naučenog kulturnog obrasca dolazi iz proučavanja primata. U sličnim zadacima, šimpanze koje su odgajili ljudi u eksperimentalnim uslovima i kontaktu s eksperimentatorom pomažu na sličan način kao i deca (Warneken and Tomasello 2006, 2009a). Slično kao deca uzrasta 14 meseci, šimpanze pomažu u situacijama kada je predmet van domašaja eksperimentatora, ali ne i u situacijama kada je neophodno ukloniti neku prepreku. Međutim, šimpanze imaju teškoća kada je reč o međusobnoj saradnji, postizanju zajedničkog cilja i deljenju ostvarene dobiti iz saradnje (Warneken and Melis 2012). Izuzetak od ovog pravila su bonobo koji su znatno tolerantniji i kooperativniji od šimpanzi, što ih čini sličnijim ljudima. Šimpanze, za razliku od ljudi, nemaju koncept zajedničkog cilja niti deluju na osnovu zajedničkih intencija. Međusobna saradnja koja uključuje postizanje zajedničkog cilja i deljenje dobiti iz saradnje se ostvaruje retko, pod strogim eksperimentalnim uslovima i kada je jedinka uparena s tolerantnim partnerom. Izgleda da ovi primati situaciju koja je posmatrana iz ljudske perspektive kooperativna, doživljavaju kao kompetitivnu, a drugu jedinku isključivo kao sredstvo za ostvarenje sebičnog cilja. Ne samo da

se druga jedinka doživljava instrumentalno, već i u situacijama u kojima mogu da izaberu da sarađuju ili da zadatak izvrše sami, šimpanze biraju individualni kurs delovanja.

Uprkos tome, evidencija ukazuje da šimpanze u divljini sarađuju i pomažu jedni drugima na više različitih načina koji uključuju kooperativni lov, pravljenje koalicija, recipročnu razmenu usluga i deljenje hrane (Muller and Mitani 2005), a slična evidencija postoji i za šimpanze u zatočeništvu (de Waal 1991, 1996, Flack and De Waal 2000).<sup>40</sup> Problem predstavlja to što je dostupna evidencija uglavnom anegdotska, a rezultati istraživanja instrumentalnog pomaganja i saradnje među šimpanzama nisu konkluzivni i predmet su spora, budući da istraživači često spekuliraju, pripisujući ovim životinjama neka intencionalna stanja i sposobnosti koje oni možda ne poseduju (Warneken and Melis 2012). Flak i de Val navode da primati koriste sofisticirane metode za rešavanje konflikta interesa za koje je verovatno neophodno razumevanje namera i situacije u kojoj se nalazi druga jedinka, za koje tvrde da su gradivni blokovi morala (Flack and De Waal 2000). Ove metode uključuju recipročnu razmenu usluga i deljenje hrane za šta je, kako tvrde, neophodna fino podešena komunikacija o namerama i željama. Ostali načini rešavanja sukoba su sledeći: *pomirenje* – čin kojim se među jedinkama koje su bile u konfliktu, gestovima, grimasama ili različitim ritualima ponovo uspostavlja dobar odnos; *tešenje* – čin u kome se jedinka koja nije učestvovala u konfliktu udružuje s jedinkom koja je pretrpela agresiju; nepristrasna *intervencija* u slučaju konflikta – čin u kome jedinka koja nije bila uključena u konflikt, najčešće alfa mužjak, interveniše da bi razrešio konflikt; i *medijacija* – čin u kome jedinka koja nije bila uključena u konflikt posreduje u ponovnom uspostavljanju dobrih odnosa među jedinkama koje su se sukobile. Ovakva interpretacija ponašanja šimpanzi je predmet ozbiljnih kritika, a jedna od njih upućuje na antropomorfizam. Tvrđi se da autori koji na ovakav način tumače ponašanje ovih primata čine *grešku retrospekcije*, tj. karakteristici koju opažaju kod životinja pripisuju funkciju koja je ispunjena samo kod ljudi (Thierry 2000). Uz to, šimpanzama se često pripisuje da su altruistički motivisane da pomažu jedne drugima, da imaju osećaj za neku vrstu društvene regularnosti, kao i da vode računa o interesima zajednice (eng. *community concern*), čime su bezmalo učinjene republikancima (de Waal 1991, 1996, Flack and De Waal 2000, Preston and de Waal 2002b), ali eksperimentalna evidencija pokazuje da su one

---

<sup>40</sup> Važnost pravljenja koalicija za evoluciju psihološkog altruizma i morala posebno naglašava Filip Kičer (Kitcher 2006a).



prilično nezainteresovane jedne za druge i da nemaju motive koji su orijentisani isključivo ka drugom članu zajednice (Silk et al. 2005).

Mada postoje brojne sličnosti između ponašanja šimpanzi i ljudi, razlike u stepenu pomaganja i saradnje delimično mogu da budu objašnjene time što za razliku od šimpanzi deca mogu lakše da prepoznaju intencionalna stanja drugih osoba i da reaguju u skladu s njima, dok su ograničeni kognitivni kapaciteti šimpanzi prepreka takvom razumevanju i shodno tome spontanom i dobrovoljnom pomaganju. Anegdotski podaci (de Waal 1996) i eksperimentalna evidencija (Yamamoto, Humle, and Tanaka 2012) potvrđuju da šimpanze verovatno mogu da razumeju namere drugih kojima je neophodna pomoć i da deluju u skladu s njima. De Val navodi slučaj šimpanze iz čijeg ponašanja se može zaključiti da razume namere druge jedinke koja ne može da dođe do željenog predmeta i kojoj zatim nastoji da pomogne. Ali tu je reč o usamljenom primeru. U eksperimentalnim uslovima šimpanze pomažu jedne drugima kada je reč o transferu oruđa koje je drugoj jedinki potrebno da bi rešila određeni problem i ostvarila željeni cilj (da bi došla do namirnica), ali isključivo ako je jedinka kojoj je potrebna pomoć eksplicitno izrazila zahtev, najčešće pružanjem ruke, vokalizacijom, gledanjem u jedinku koja može da pruži pomoć, itd (Yamamoto, Humle, and Tanaka 2009). Zahtevanje je vrsta komunikacionog čina i jedinka pomaže kada je ovakav zahtev postavljen, ali zbog teškoća u razumevanju intencionalnih stanja druge jedinke, to gotovo nikada ne čini spontano i dobrovoljno. Na eksplicitno izražen zahtev, ovi primati uglavnom biraju odgovarajuće oruđe od više ponuđenih alternativa i dodaju ga drugoj jedinki u slučajevima kada mogu vizuelno da procene situaciju u kojoj se ona nalazi (Yamamoto, Humle, and Tanaka 2012).

Podaci iz primatologije idu u prilog shvatanju da su ljudski altruistički kapaciteti verovatno evoluirali kao deo šireg spektra socijalnih i altruističkih dispozicija primata, kao i da su evoluirali pre pojave kulture i transmisije bihevioralnih obrazaca putem verbalno formulisanih pravila, zapravo mnogo ranije nego što su se evolucione linije šimpanzi i ljudi razdvojile pre oko šest miliona godina (de Waal 1996, de Waal et al. 2006, Flack and De Waal 2000, Warneken and Tomasello 2009a). Ovi podaci takođe ukazuju i na značajne razlike između ljudi i drugih primata, ali i na to da ponašanje prethodi pravilima. Verovatno je da su neki uspešni oblici ponašanja, zajednički ljudima i drugim primatima, u kasnijoj evolucionoj istoriji i s pojavom jezika formulisani kao pravila. Činjenica da su uspešni oblici ponašanja upravo oni

koje danas nazivamo altruističkim i moralnim stvar je filogenetske kontingencije. U prilog tome da je moral specijes specifična karakteristika ljudi ide i to da primati nemaju osećaj gađenja, koji je važna emocija za donošenje (konvencionalnih) moralnih sudova koji se odnose na hranu, seks i odlaganje leševa. Konačno, činjenica je da primati imaju neke od gradivnih blokova morala, ali i to da su njihovi bihevioralni obrasci veoma udaljeni od ljudskog moralnog repertoara. Primati svakako ne mogu da usvajaju moralna verovanja, da naprave razliku između onoga što je dopušteno ili prihvaćeno u određenom socijalnom kontekstu, nasuprot onome što je dopustivo ili prihvatljivo, kao i da donose moralne sudove. Mada je kontinuitet nesporan, razlike su, takođe, očigledne.

### **3.7 Neurobiologija empatičkih kapaciteta**

Teorija zajedničkog kodiranja percepcije i akcije, na kojoj Preston i de Val temelje svoju teoriju empatije, je funkcionalna i bez predikcija o zajedničkom neurobiološkom supstratu (Van Der Wel, Sebanz, and Knoblich 2013), ali kako se čini, nova istraživanja u oblasti neurobiologije pokazuju da takav supstrat verovatno postoji i da ga čine *neuroni ogledala* čija svojstva i važnost za evoluciju empatije će upravo biti objašnjeni (Iacoboni 2009, Rizzolatti and Craighero 2005).

Otkriće neurona ogledala pružilo je neurobiološku potporu mehanizmu zajedničkog kodiranja percepcije i akcije. Funkcija neurona ogledala se najpre vezivala za prepoznavanje postupaka druge jedinke, da bi se zatim to područje proširilo na razumevanje namera, ciljeva, drugih intencionalnih stanja i sposobnosti oponašanja, što je dalje vodilo naglašavanju važnosti ovih neurona za evoluciju i razvoj empatije. Time se ne sugerise da su za evoluciju i razvoj empatije značajni isključivo oni regioni mozga za koje se pretpostavlja da sadrže neurone ogledala. Brojni drugi regioni koji su uključeni u opažanje i odgovore na negativne emocionalne signale koji uglavnom zahtevaju neodložnu reakciju, kao što su glad, bol, uznemirenost i strah, takođe su uključeni u doživljavanje empatije (Decety 2011b). To takođe ne znači da su neuroni ogledala izuzeti od učenja i da se uvek aktiviraju po biološki determinisanom automatizmu. Eksperimenti na makaki majmunima ukazuju da su neka svojstva neurona ogledala oblikovana iskustvom (Iacoboni 2009). Na primer, dugotrajno posmatranje upotrebe oruđa aktivira neurone ogledala iako inicijalno nisu bili aktivni prilikom posmatranja te aktivnosti. Ovaj nalaz možda ide u prilog tezi da su ovi neuroni odgovorni za rapidno širenje korišćenja oruđa i kulture u

ljudskim populacijama. Ima autora koji su uvereni da je evolucija ovih neurona dovela do velikog skoka u evoluciji čoveka i vodila ubrzanom razvoju kulture i civilizacije (Ramachandran 2011), ali takve interpretacije verovatno preneglašavaju njihovu stvarnu funkciju.

Neuroni ogledala su najpre otkriveni u regionu F5 (premotori korteks) mozga makaki majmuna. Njihovo osnovno svojstvo je da se aktiviraju u dva slučaja, najpre kada jedinka čini određeni *ciljno orijentisani* postupak (najčešće grabljenje), i zatim kada posmatra da druga jedinka čini isti postupak, čime se percepcija i akcija dovode u vezu (Rizzolatti and Craighero 2005, Iacoboni 2009). Kada jedinka makakija posmatra neki besmisleni postupak, neuroni ogledala se ne aktiviraju, što ukazuje na to da ciljevi jedinke imaju važnu ulogu u razumevanju njihove funkcije. Međutim, kod ljudi izgleda da dolazi do aktiviranja sistema neurona ogledala čak i u slučajevima pokreta koji nemaju neki poseban smisao (Sinigaglia and Rizzolatti 2011).

U slučaju majmuna, za proučavanje neurona na ćelijskom nivou se koriste invazivne tehnike koje su moralno nedopustive u proučavanju ljudi, tako da su dokazi o postojanju neurona sa navedenim svojstvima kod ljudi posredni, dobijeni skeniranjem, funkcionalnom magnetnom rezonancom ili proučavanjem osoba s moždanim oštećenjima (Iacoboni 2009). Dva regiona ljudskog mozga se povezuju s prisustvom neurona ogledala. To su donji deo frontalnog korteksa, obuhvatajući zadnji deo donje čeonice, i rostralni deo donjeg temenog režnja (Iacoboni and Dapretto 2006, Iacoboni 2009, Rizzolatti and Craighero 2005). Ovi delovi čine sistem neurona ogledala koji, kada im se pridruži vizuelni input koji dolazi iz zadnjeg dela gornje temporalne brazde, stvaraju neuronsko kolo zaduženo za oponašanje (Iacoboni and Dapretto 2006).

Mada su ovi neuroni podesni za oponašanje, pretpostavlja se da je njihova osnovna funkcija bila prepoznavanje i razumevanje postupaka (majmuni kod kojih je potvrđeno prisustvo ovih neurona su loši u oponašanju), a da su sekundarne funkcije učestvovanje u oponašanju i razumevanju namera drugih jedinki (Iacoboni 2009, Rizzolatti and Craighero 2005). Postupak se prepoznaje kada se isti skup neurona aktivira u posmatraču kao i u onome ko izvršava određenu radnju. Na sličan način dolazi do prepoznavanja namera. Posmatranje određenog postupka aktivira isti skup neurona u motornom korteksu posmatrača i izvršioca, zatim se namera koju posmatrač vezuje za taj postupak automatski reprezentuje u posmatraču stavljajući ga u istu

mentalnu poziciju u kojoj je izvršilac, nakon čega on istu nameru koju vezuje za izvršenje postupka u sopstvenom slučaju pripisuje izvršiocu (Gallese and Goldman 1998). Ukoliko ovo podudaranje uzrokuje povratnu neurofiziološku reakciju u posmatraču, koja nije inhibirana, tako da je prati odgovarajući bihevioralni obrazac, onda možemo da govorimo o oponašanju. Na taj način, svojstva neurona ogledala i pridruženih neurofizioloških mehanizama se podudaraju s osnovnim pretpostavkama teorije zajedničkog kodiranja percepcije i akcije.

Možemo da razlikujemo dve vrste neurona ogledala. Oni su a) strogo podudarni neuroni ogledala i b) „široko“ podudarni neuroni ogledala (Iacoboni 2009). Strogo podudarni neuroni ogledala se aktiviraju posmatranjem istog postupka za koji kodiraju motorni sistem. U tom smislu možemo da kažemo da ova klasa neurona ogledala učestvuju u jednostavnom prepoznavanju postupka i eventualnom oponašanju. Mada su neurofiziološki mehanizmi koji učestvuju u opažanju, prepoznavanju i reakcijama na ponašanje ili intencionalna stanja druge jedinke adaptivni (Decety 2011a), bilo bi maladaptivno kada bi sve percepcije postupaka bile mapirane u motornom korteksu i rezultirale kompulzivnim oponašanjem (Hickok 2009). Time bi bile onemogućene koordinisane i komplementarne aktivnosti i iz tog razloga neophodna nam je druga klasa neurona ogledala. Ova druga klasa neurona ogledala, široko podudarni neuroni ogledala, vezuju se za donji temeni režnjić (Rizzolatti and Craighero 2005) i aktiviraju se posmatranjem akcije koja je logički slična, ali ne i identična s akcijama za koje kodiraju motorički sistem. To znači da se ti neuroni aktiviraju kada jedinka *A* posmatra jedinku *B* kako čini neki ciljno orijentisani pokret (pruža ruku da bi zgrabila neki predmet, najčešće hranu), ali i kada posmatra neki postupak koji je logički povezan s prethodnim (npr. prinošenje hrane ustima). Fleksibilnost široko podudarnih neurona ogledala je karakteristika koja čini mogućim ne samo prepoznavanje i oponašanje, već i predviđanje i koordinaciju postupaka, saradnju i pojedine komplementarne aktivnosti (Iacoboni 2009, Rizzolatti and Craighero 2005)

Pretpostavlja se da ova klasa neurona ima važnu funkciju u prepoznavanju namera i emocija drugih individua, budući da učestvuje u kodiranju facijalne ekspresije, koja je indikator emocionalnog stanja individue (Iacoboni 2009). Emocije se prepoznaju na dva načina od kojih jedan uključuje kognitivnu obradu čulnih podataka, dok se drugi oslanja na automatsko deljenje zajedničkih reprezentacija na osnovu čulnog inputa. U prvom slučaju jedinka na kognitivnom planu razume šta druga jedinka oseća, bol, strah, zadovoljstvo, gađenje, itd., ali ne *saoseća* s

drugom jedinkom. U drugom slučaju senzorni input dovodi do direktnog podudaranja neuronskih reprezentacija kod osobe i posmatrača, tako da opažaj indukuje istu emociju kod posmatrača. Na taj način, emocija druge osobe se iskušava direktno i iz prvog lica. U jednom pomalo strogom smislu, empatija predstavlja sposobnost individue da direktno iskušava *identična* emocionalna stanja koje su karakteristična za situaciju i vezana za stanja neke druge individue. Ako dve individue u određenoj situaciji imaju identična emocionalna iskustva (ali različitog intenziteta), iako je situacija takva da bi bilo uobičajeno da samo jedna od njih ima odgovarajuće emocije, onda je reč o empatiji koju možemo da diferenciramo od *vikarnih emocija* (Paulus et al. 2013).

To je najlakše objasniti na primeru gađenja, jer ova emocija ima snažan komunikacioni potencijal i važnu adaptivnu funkciju, budući da u svom bazičnom obliku ukazuje na to koja namirnica može biti loša i opasna za jelo (Rizzolatti and Craighero 2005). Tako, kada jedna individua posmatra drugu koja pokazuje nedvosmislene izraze gađenja, kao što su na primer podizanje nozdrva, skupljanje obrva i širenje usta na dole s karakteristično izbačenim jezikom, i ona sama oseća gađenje. Izražavanje gađenja izaziva istu emociju u posmatraču, jer ova emocija i njeno opažanje kod drugih dele zajednički reprezentacioni domen i odgovarajući neurobiološki supstrat – klasu neurona ogledala u insularnom regionu kore velikog mozga, a oštećenje ovog regiona mozga dovodi do deficita u razumevanju ekspresije gađenja. Takođe, u skladu s teorijom empatije zasnovanoj na koncepciji zajedničkog kodiranja percepcije i akcije postoji evidencija da facijalna ekspresija emocija indukuje slične emocionalne reakcije u posmatraču. U prilog ovakvom razumevanju govori i podatak da se iste neuronske reprezentacije aktiviraju prilikom doživljavanja bola i opažanja drugih individua koje su u bolu (Decety 2011a).

Uprkos početnoj primamljivosti i jednostavnosti, objašnjenja koja se oslanjaju na karakteristike neurona ogledala nisu naišla na opšte odobravanje, niti su svi uvereni u njihovu eksplanatornu moć. Iako uvažavaju značaj otkrića neurona ogledala, kritičari tvrde da se ovoj klasi neurona često pripisuju pojedine funkcije za koje nema dovoljno eksperimentalne potvrde. Najpre se osporava njihova primarna funkcija, prepoznavanje postupaka, a zatim se na osnovu toga poriče i njihovo učešće u kompleksnijim kognitivnim sposobnostima kao što su prepoznavanje namera, oponašanje i empatija. Prepoznavanje postupaka se može objasniti bez pozivanja na zajednički reprezentacioni domen i neurone ogledala, jednostavnom asocijacijom određenog čulnog nadražaja s odgovarajućim regionom u mozgu, kao što je gornja temporalna

brazda koja je uključena u opažanje bioloških pokreta, ali koja nema motorička svojstva i za koju se ne vezuju neuroni ogledala (Hickok 2009). Takođe, jedinke su sposobne da prepoznaju postupke koje nikada ranije nisu učinile i koje nikada neće moći da učine jer ne pripadaju njihovom bihevioralnom repertoaru (npr. lajanje ili letenje), što ne ide u prilog teoriji zasnovanoj na sistemu neurona ogledala (Churchland 2011, Hickok 2009). Ali, ova kritika možda nije opravdana, jer počiva na neosnovanoj generalizaciji. Prema teoriji zajedničkog kodiranja, neuroni ogledala nemaju odlučujuću ulogu u prepoznavanju svih postupaka, već samo onih koji pripadaju bihevioralnom repertoaru jedinke (Rizzolatti and Sinigaglia forthcoming). Iz tog razloga moguće je da su i neke druge kritike koje se stavljaju na teret zastupnicima teorije zajedničkog kodiranja nedovoljno opravdane.

Kada je reč o prepoznavanju namera, iz teorije zajedničkog kodiranja percepcije i akcije sledi da je ova kognitivna funkcija na neurobiološkom planu veoma jednostavna, a zapravo se radi o složenom procesu o kome ima veoma malo podataka. Mehanizam prepoznavanja namera, kao i mehanizam prepoznavanja postupaka, može da se objasni asocijacijom čulnog inputa i motoričke realizacije pokreta, koja se zatim generalizuje na slične kontekste (Churchland 2011). Ako je to zaista slučaj, onda je sistem neurona ogledala u ekspanatornom smislu redundantan. Međutim, iako je naizgled jednostavan, asocijacija je u kognitivnom (verovatno i u evolucionom) smislu kompleksniji i svakako sporiji mehanizam (Rizzolatti and Sinigaglia forthcoming), tako da se ne vidi zbog čega bi prirodna selekcija favorizovala ovako kompleksan mehanizam, pogotovo za nešto što može da ima dalekosežne posledice po adaptivnu vrednost jedinke.

Istaknut je i prigovor da jedna namera može da se na bihevioralnom planu realizuje na mnoštvo različitih načina, kao što jedan postupak može da bude realizacija velikog broja najrazličitijih namera (Churchland 2011, Hickok 2009). Iako postupak ponekada može da bude adekvatan pokazatelj namera, to svakako ne mora uvek da bude slučaj, tako da posmatrajući postupke ne možemo uvek i sa sigurnošću da ustanovimo odgovarajuću intenciju. Ako posmatranje i izvršenje određenog ciljno orijentisanog postupka u funkcionalnom i neurobiološkom smislu dele zajednički reprezentacioni domen, tako da se u posmatraču automatski pojavljuju reprezentacije koje on vezuje za sopstvene namere koje zatim projektuje na drugu jedinku, onda se time tvrdi i da je proces prepoznavanja i razumevanja sopstvenih namera (samosvest), za koji se zna da je neurofiziološki kompleksan, direktan i krajnje

pojednostavljen (Churchland 2011). Saznanja iz psihologije govore u prilog tome da ljudi zapravo veoma malo znaju o svojim namerama, da im one nisu direktno dostupne i da su saopštenja o namerama uglavnom *ex post facto* racionalizacije. Na taj način automatsko projektovanje sopstvenih namera na druge ne može biti pouzdan način otkrivanja njihovih namera i predviđanja budućih postupaka.

Mada su ove primedbe na mestu, kada se govori o prepoznavanju i razumevanju tuđih namera čini se da se, jednako kao i u slučaju prepoznavanja postupaka, ne misli na precizno i pouzdano pripisivanje intencija bez mogućnosti greške, već o elementarnoj aproksimaciji mogućih namera i događaja koji mogu da slede iz njihove realizacije (Sinigaglia and Rizzolatti 2011). Tako, na primer, ako neko dohvati čašu s vodom, možemo da zaključimo da je žedan i da želi da pije vodu, iako možda želi da zalije cveće ili da je prospe. Zaključak da želi da je popije je najbolja aproksimacija za pripisivanje određene namere na osnovu postupka, ali nije izuzet od greške, pogotovo ne u artifičijelnim eksperimentalnim kontekstima.

Nekoliko metodoloških primedbi se upućuje proučavanju neurona ogledala kod ljudi. Najpre, kao što je već pomenuto, tvrdi se da neuroni ogledala učestvuju u kompleksnim kognitivnim procesima kao što su prepoznavanje i pripisivanje intencionalnih stanja, oponašanje i empatija, ali kod vrste kod koje je pouzdano dokazano da poseduje takve neurone ovi viši kognitivni procesi izostaju, dok kod ljudi nema konkluzivnih dokaza za postojanje neurona ogledala, ali ima viših kognitivnih sposobnosti (Hickok 2009). Deduktivni lanac koji kreće od prisustva neurona ogledala kod makaki majmuna, preko pretpostavke da ljudi poseduju neurone ogledala, pa do zaključaka o ponašanju ljudi, popunjen je brojnim pomoćnim hipotezama za koje ne postoji empirijska potvrda. Takođe, pretpostavlja se da su svojstva neurona ogledala koja postoje kod majmuna očuvana i kod ljudi, ali da su kod ljudi ovi neuroni dobili funkcije koje su eksaptacije u odnosu na njihovu primarnu funkciju, međutim svi ovi argumenti su spekulativni (Hickok 2009).

Iako se postojanje neurona ogledala najpre vezivalo samo za dva regiona mozga, u nekim istraživanjima se sugerše njihovo postojanje i u drugim moždanim područjima. Tako se, na primer, tvrdi da klasa neurona ogledala postoji u insularnom korteksu, ali i drugde (Rizzolatti and Craighero 2005, Churchland 2011). Proširivanje delokruga neurona ogledala pomaže da se

objasne pojedini fenomeni, ali i da se lakše odgovori na upućene kritike. Međutim, proširivanje broja područja velikog mozga u kojima se navodno otkrivaju neuroni ogledala slabi eksplanatornu moć objašnjenja zasnovanih na njihovom postojanju. Ako se neuroni ogledala pojavljuju u velikom broju regiona velikog mozga, onda je problem napraviti razliku između njih i drugih neurona. S druge strane, ako se na teorijske nedostatke i opovrgavajuće argumente uvek odgovara uvođenjem pomoćnih hipoteza, time se naučna teorija udaljava od uobičajenog naučnog postupka i približava dogmatskim učenjima.

Još jedan metodološki nedostatak se vezuje za istraživanja aktivacije neurona ogledala u slučajevima prepoznavanja postupaka, intencija i emocija, a to je cirkularnost (Decety 2011a). Većina studija koje se bave ovim problemima zanemaruje sve druge neuronske aktivnosti u korteksu velikog mozga i fokusira se samo na donju frontalnu gangliju i donji temeni režnjić, jer su to regioni za koje se veruje da su homologni regionu F5 kod makakija, tako da se pretpostavlja da se tu nalaze neuroni ogledala.<sup>41</sup> Zatim se svaka neurofiziološka aktivnost u ovim regionima tumači kao da je reč o aktivnostima neurona ogledala, pri čemu se prenebregava činjenica da aktivnost može da se pripiše drugim neuronima, koji su svakako brojniji od neurona ogledala. Ovakva tumačenja slabe poziciju onih teorija koje svoja objašnjenja zasnivaju na aktivnostima ovih nervnih ćelija.

Kada je reč o empatiji i ulozi neurona ogledala i zajedničkog kodiranja percepcije i akcije za prepoznavanje emocija, kao opovrgavajući argument se navodi da je obrada određenog broja bazičnih emocija vezana za amigdalnu, koja nije u sistemu neurona ogledala (Churchland 2011). Osim toga, osobe kojima je iz različitih kliničkih razloga amigdala odstranjena bilateralno, mogu da prepoznaju emocije kao što je na primer strah. S druge strane, osobe kod kojih je oštećen insularni korteks za koji se pretpostavlja da nije u sistemu neurona ogledala, ne osećaju empatiju i nemaju intenzivne emocionalne reakcije. Ova kritika prenebregava nalaze koji upućuju na to da je amigdala povezana sa sistemom neurona ogledala preko insularnog korteksa i važnost ove veze za empatičko reagovanje (Carr et al. 2003). Ovaj argument upućuje na to da bi svako podudaranje između direktnog iskušavanja i opažanja određenih emocija moglo da bude protumačeno kao da počiva na zajedničkim neuronskim reprezentacijama, iako one ne moraju da

---

<sup>41</sup> Zapravo ni paralelizam između ovih regiona nije neupitan. Vidi (Hickok 2009).



pripadaju sistemu neurona ogledala (Decety and Meyer 2008). Međutim, ovakvo premeštanje težišta sa neurona ogledala na druge regione mozga koji podupiru delovanje ovih neurona i s kojima su funkcionalno povezani unosi konceptualnu zbrku i zahteva posebnu opreznost (Decety 2011a).

Konačno, nekoliko neuroloških poremećaja ne idu u prilog teoriji empatije koja počiva na zajedničkom kodiranju percepcije i akcije. Ako neuroni ogledala na neki način učestvuju u kodiranju facijalne ekspresije (Iacoboni and Dapretto 2006, Iacoboni 2009), onda individue s *urođenim* nedostacima facijalne ekspresije, kakav je *Mebijusov sindrom*, koji se karakteriše obostranom paralizom lica, ne bi trebalo da mogu da prepoznaju facijalnu ekspresiju emocija drugih osoba. Međutim, ove osobe, iako same ne mogu da komuniciraju služeći se facijalnom ekspresijom, bez problema prepoznaju emocionalne signale na licima drugih individua (Decety 2011a).

Drugi primer se odnosi na sposobnost prepoznavanja facijalne ekspresije vezane za doživljavanje bola koja je značajna za razvoj empatičkih kapaciteta. Prema definiciji *Međunarodne asocijacije za proučavanje bola* (IASP), bol je neprijatno čulno i emocionalno iskustvo povezano sa stvarnim ili potencijalnim oštećenjem tkiva. Sposobnost da se doživi i izrazi bol ima važne posledice po adaptivnu vrednost organizma. Bol najpre upozorava jedinku koja oseća bol da s njenim organizmom nešto nije u redu i da je potrebno preduzeti određene radnje kako bi se bol ili njegov uzrok uklonio. Na drugom mestu, on signalizira drugim jedinkama da je jedinki koja trpi bol potrebna pomoć. Mada ima dokaza da se iste neuronske reprezentacije aktiviraju prilikom doživljavanja bola i opažanja drugih individua koje su u bolu, uloga sistema neurona ogledala u ovom procesu, koji je nesumnjivo važan za empatiju, nije potpuno jasna, imajući u vidu da je podudaranje otkriveno u regionima koji se obično ne vezuju za ovaj sistem (Decety 2011a).

Shodno predviđanjima teorije zajedničkog kodiranja, opažanje facijalne ekspresije određenog emocionalnog stanja je okidač za reprezentacije identičnih ili sličnih emocionalnih stanja u posmatraču, tako da ukoliko ove reprezentacije nisu inhibirane, onda se može očekivati pojava emocionalne reakcije kod posmatrača. Ako je tako, onda je racionalno očekivati da su individue koje nikada nisu imale iskustvo određenog emocionalnog stanja neosetljive na

ekspresiju tog stanja kod drugih individua. Međutim, ovo predviđanje se kosi s empirijskim podacima dobijenim iz proučavanja osoba s urođenom neosetljivošću na bol (Danziger, Faillenot, and Peyron 2009). Ovo istraživanje je pokazalo da osobe s urođenom neosetljivošću na bol mogu da prepoznaju i da saosećaju s bolnim iskustvima drugih. Prilikom posmatranja delova tela u situacijama koje su bolne (makaze koje seku prst), kao i prilikom posmatranja facijalne ekspresije koja signalizira bolno iskustvo, kod individua s urođenom neosetljivošću na bol aktiviraju se slična moždana područja kao i kod osoba koje su osetljive na bol – prednji deo središnjeg cingularnog korteksa i prednji deo insularnog korteksa. Ova moždana područja, uz još neka druga, se pouzdano i konzistentno vezuju za sopstveno iskustvo bola, kao i za posmatranje tuđeg bolnog iskustva (Danziger, Faillenot, and Peyron 2009, Decety 2011a). Navedeni autori pretpostavljaju da aktivacija ovih moždanih područja nije u tolikoj meri vezana za zajedničke reprezentacije emocionalnog iskustva bola, koliko za emocionalne reakcije na negativne stimulse uopšte, koje mobilišu organizam da preduzme odgovarajuće postupke kako bi se negativno emocionalno iskustvo izbeglo.

Uprkos tome što je izložena brojnim kritikama teorija zajedničkog kodiranja percepcije i akcije koja počiva na sistemu neurona ogledala je trenutno dominantna naučna paradigma. Kritičari često zaboravljaju da zastupnici ove teorije takođe veruju da je nervni sistem – sistem – i da njegovi podsistemi verovatno ne mogu da funkcionišu potpuno izolovano. Tako ni sistem neurona ogledala ne može da funkcioniše bez pomoći drugih sistema. Iako je empatija kompleksna sposobnost i u njenom ispoljavanju učestvuju brojni neurofiziološki mehanizmi i drugi regioni mozga, kao što su orbitofrontalni i ventromedijalni prefrontalni korteks, tako da ne može da se svede isključivo na aktiviranje neurona ogledala (Decety 2011a, b), ipak možemo da pretpostavimo da je odgovarajuća klasa ovih neurona temeljni činilac najpre za evoluciju mehanizma zajedničkog kodiranja percepcije i akcije, a zatim i za evoluciju i razvoj bazičnih empatičkih kapaciteta. Ovi neuroni omogućavaju da postupci i iskustva drugih budu sagledani iznutra, predstavljeni kao sopstveni posredstvom zajedničkih neuronskih reprezentacija. Sistem neurona ogledala na neki način predstavlja bazični i rudimentarni most između osoba i put do tuđih svesti, a pretpostavlja se da nedostaci u ovom sistemu verovatno vode problematičnim obrascima ponašanja, za šta još nema pouzdanih dokaza (Iacoboni 2009, Rizzolatti and Craighero 2005, Iacoboni and Dapretto 2006).

Neki autori smatraju da rezultati dobijeni proučavanjem osoba s neurološkim oštećenjima za sada ne idu u prilog teoriji zajedničkog kodiranja i ulozi sistema neurona ogledala u doživljavanju empatije. Nedostaci i oštećenja onih moždanih područja koji se vezuju za ovaj sistem ne dovode do disfunkcija u empatiji, saosećanju ili moralnom rasuđivanju, dok su oštećenja orbitofrontalnog i ventromedijalnog prefrontalnog korteksa povezana s ovim disfunkcijama (Decety 2011a). Međutim, ima i drugačijih mišljenja. Među istraživačima postoji konsenzus da je empatija kompleksna sposobnost koja uključuje emocionalnu i kognitivnu komponentu, što znači da su u doživljavanje empatije uključena dva anatomski i funkcionalno različita sistema. Jedan je onaj koji se vezuje za sistem neurona ogledala (donja frontalna ganglija i donji deo temenog režnja) i on je uključen u emocionalnu komponentu empatije. Drugi region, vezan za kognitivnu dimenziju empatije je ventromedijalni prefrontalni korteks. Kao veza između ovih sistema predlaže se tempoparijetalni spoj – mesto spajanja slepoočnog i temenog režnja – koji učestvuje u različitim procesima među kojima su i teorija uma i moralno rasuđivanje. Na osnovu proučavanja osoba s lezijama mozga utvrđeno je da osobe s oštećenim ventromedijalnim prefrontalnim korteksom imaju deficite u kognitivnoj empatiji, dok oni s oštećenjima donje frontalne vijuge pokazuju nedostatke u doživljavanju emocionalne empatije (Shamay-Tsoory 2011, Shamay-Tsoory, Aharon-Peretz, and Perry 2009). Prema modelu empatije koji je zasnovan na mehanizmu zajedničkog kodiranja percepcije i akcije, bihevioralno ispoljavanje empatije opada s godinama i iskustvom, a kao jedan od razloga se navodi inhibicija primarnih reakcija zbog razvića kontrolnih mehanizama prefrontalnog korteksa (Preston and de Waal 2002b).

Danas je poznato i opšte prihvaćeno da osobe s nedostatkom pre svega emocionalne empatije i još nekim emocionalnim deficitima imaju značajnih poteškoća u razumevanju, usvajanju i poštovanju moralnih pravila. O posledicama nedostatka empatije biće reči u narednom poglavlju.

## 4 PSIHOPATIJA

Iako bihevioralno ispoljavanje empatije opada s godinama i iskustvom, usled delovanja različitih kontrolnih mehanizama koji se razvijaju tokom odrastanja, na unutrašnjem planu mogu da se registruju znaci uznemirenosti prilikom opažanja uznemirujućih sadržaja uprkos kognitivnoj kontroli ponašanja ili preusmeravanju pažnje. Prema modelu empatije koji je zasnovan na mehanizmu percepcije i akcije, „ovaj unutrašnji čuvar uznemirenosti može da bude mehanizam zadužen za negativna osećanja kao što su krivica i kajanje“ (Preston and de Waal 2002b, 5). Međutim, postoji posebna klasa individua, prisutna u svim kulturama, koja nikada nije iskusiła (emocionalnu) empatiju, krivicu i kajanje, kao ni niz drugih emocija, i koja je zbog toga potpuno neosetljiva na uznemirenost i patnje drugih što je čini klasom savršenih društvenih predatora. O ovim osobama se obično govori kao o psihopatama ili osobama s psihopatskim karakteristikama. Drugi uobičajeni termini kojim se označavaju ove individue su sociopata ili osoba s antisocijalnim poremećajem ličnosti.

### 4.1 Šta je psihopatija?

Psihopatija je trajni poremećaj ličnosti i jedan je od najkontroverznijih pojmova u psihologiji i psihijatriji. Jedan od razloga za kontroverzu je i taj što sam termin nema samo deskriptivnu, već i vrednosnu komponentu značenja i često se u svakodnevnom govoru koristi kao diskvalifikativ. U *Dijagnostičkom i statističkom priručniku za mentalne poremećaje* Američke psihijatrijske asocijacije, psihopatija se identifikuje kao antisocijalni poremećaj ličnosti (American Psychiatric Association 2013). Mada se većina pobrojanih karakteristika koje u ovom priručniku vezuju za antisocijalni poremećaj ličnosti, odnose i na psihopatiju, postoje značajne razlike. Antisocijalni poremećaj ličnosti se definiše primarno preko nasilničkog ponašanja i kršenja prava drugih, dok se psihopatija pre svega vezuje za određene kognitivne i emocionalne crte ličnosti, tj. njihove nedostatke (Hart, Hare, and Harpur 1992). Psihopate su neretko nasilne, ali nasilništvo nije ono što ih na specifičan način čini različitim od ostatka populacije. Ono što može da ih na specifičan način razlikuje od normalne populacije, jeste način na koji koriste zastrašivanje i nasilje. Dok većina osoba koje nisu psihopate postaje nasilna pod

uticajem snažnih emocija, psihopate nasilje koriste jednostavno, hladnokrvno i instrumentalno (Hare 2001). Filmski antiheroji kao što su Norman Bejts (*Psiho*) ili Kožno lice (*Teksaški masakr motornom testerom*), iako su zastrašujući i nasilni, nisu arhetipovi psihopata.

Psihopatiju je prvi jasno konceptualizovao Hervei Klekli (Hervey Cleckley) u knjizi *Maska normalnosti* koja predstavlja temeljnu studiju u proučavanju ovog poremećaja ličnosti. On navodi 16 karakteristika specifičnih za psihopatiju koje će postati osnova za Herovu (Robert D. Hare) *Kontrolnu listu za psihopatiju* (eng. *Psychopathy Checklist, Psychopathy Checklist – Revised*; skr. PCL, PCL-R), standardni test za utvrđivanje psihopatskih crta ličnosti (Cleckley 1988, Hare 1999). Na Kleklijevoj listi karakteristika se nalaze: površni šarm i dobra inteligencija, odsustvo zabluda ili bilo kog oblika iracionalnog mišljenja, odsustvo nervoze ili psihoneurotskih manifestacija, nepouzdanost, neistinoljubivost i neiskrenost, nedostatak kajanja i stida, neadekvatno motivisano antisocijalno ponašanje, loše rasuđivanje i teškoće da se uči iz iskustva, patološka egocentričnost i nesposobnost da se voli, opšta manjkavost u glavnim afektivnim reakcijama, poseban nedostatak uvida, neosetljivost u interpersonalnim relacijama, ćudljivo i neprivačno ponašanje pod dejstvom alkohola, bez samoubilačkih tendencija, bezličan, trivijalan i slabo integrisan seksualni život i nesposobnost da se sledi bilo koji životni plan (Cleckley 1988, 338-339).

Prema Herovoj listi, na kojoj se nalazi 20 karakteristika (vidi npr. Hare 2001), osobe s psihopatijom su u interpersonalnim odnosima slatkorečive i površno šarmantne, u odnosu prema drugima poseduju grandiozni osećaj sopstvene vrednosti, sklone su patološkom laganju, zapovednički su nastrojene i manipulativne. Odlikuju ih plitke emocionalne reakcije, nedostatak krivice i kajanja, neosetljive su na tuđu patnju i bez empatije, kao i nesposobne da preuzmu odgovornost za sopstvene postupke. Njihov životni stil karakteriše stalna potreba za stimulacijom, parazitizam, nedostatak realističnih dugoročnih životnih ciljeva, impulsivnost i neodgovornost. Psihopate imaju lošu kontrolu ponašanja, rane probleme u ponašanju koji se manifestuju još u detinjstvu<sup>42</sup>, poseduju istoriju maloletničke delinkvencije, a u kriminalnim

---

<sup>42</sup> Iz moralnih razloga deca se ne označavaju kao psihopate, već se o njima govori kao o neosetljivim i neemocionalnim.

aktivnostima su svestrani i nesposobni da se dugo zadrže na slobodi. Odlikuje ih još promiskuitet i veliki broj kratkotrajnih bračnih veza.

*Tabela 7: Kontrolna lista za psihopatiju – revidirana verzija (prilagođeno prema Skeem et al. 2011)*

FAKTOR 1 Interpersonalna i afektivna skala		FAKTOR 2 Antisocijalna skala	
FASETA 1 Interpersonalni odnosi	FASETA 2 Afektivno	FASETA 3 Životni stil	FASETA 4 Antisocijalno ponašanje
Slatkorečivost/površni šarm Preuveličani osećaj sopstvene vrednosti Patološko laganje Zapovednički nastrojen/manipulativan	Nedostatak krivice ili kajanja Plitki afekti Neosetljivost/nedostatak empatije Nesposobnost da se preuzme odgovornost za sopstvene postupke	Potreba za stimulacijom/sklonost ka dosadi Parazitizam Nedostatak realističnih dugoročnih ciljeva Impulsivnost Neodgovornost	Loša kontrola ponašanja Rani problemi u ponašanju Maloletnička delinkvencija Opozivanje uslovnog otpusta Kriminalna svestranost
Ostale karakteristike: <b>promiskuitetno seksualno ponašanje, mnoštvo kratkotrajnih bračnih veza</b>			

Poznavanje taksativno pobrojanih karakteristika ne naoružava nekoga sposobnošću da prepoznaje psihopate u svakodnevnom životu. Utvrđivanje da li je neko psihopata ili ne rezultat je detaljnog kliničkog ispitivanja. Obično se smatra da je osoba psihopata ako na PCL-R skali (0 – 40) ima skor 30 i više (Hart, Hare, and Harpur 1992), mada neki istraživači radije govore o psihopatskim karakteristikama kao o skupu osobina koji se pojavljuje u kontinuitetu (Raine and Glenn 2014). To znači da individua može da poseduje manje ili više psihopatskih karakteristika, a da ne bude u pravom smislu psihopata. Drugim rečima, ova lista omogućava da se napravi razlika između takozvanih uspešnih i neuspešnih psihopata.

## 4.2 Neposredni uzroci psihopatije

Među naučnicima postoji opšta saglasnost da su uzroci psihopatije biološki i sredinski; da biološki činoci imaju između 40 i 60 odsto udela u razvoju psihopatskih crta ličnosti i ispoljavanju antisocijalnog ponašanja, što znači da je udeo bioloških i sredinskih činilaca u razvoju psihopatije približno jednak (Raine and Glenn 2014). Genetički i sredinski činoci utiču na razviće neurofizioloških deficita koji zatim rezultiraju odgovarajućim bihevioralnim obrascima specifičnim za psihopatiju. Zahvaljujući genetičkim činocima psihopatske

karakteristike su stabilne tokom vremena. Budući da je veza između pojedinačnog gena i ponašanja slaba, nalazi većine bihevioralno genetičkih istraživanja o psihopatiji, ukazuju na to da je u razvoj psihopatije uključeno više gena, kao i da je genetički efekat aditivan. Aditivnost znači da se efekti više genskih alela sumiraju kako bi dali određenu fenotipsku karakteristiku. Neke studije ukazuju da isti genetički činioci doprinose različitim fasetama psihopatije.

Uzimajući u obzir visok komorbiditet između bolesti zavisnosti i psihopatije, istraživanja u molekularnoj genetici se orijentišu na korelaciju između gena koji su se ranije povezivali s rizičnim ponašanjem i bolestima zavisnosti kao što su alkoholizam i narkomanija, s jedne strane, i psihopatije, s druge. Osnovna pretpostavka ovih istraživanja je da postoje zajednički geni koji učestvuju u kodiraju neurofiziološkog sistema zaduženog za gratifikaciju i povećanog rizika od razvoja zavisnosti, i koji su, s druge strane, odgovorni za izraženo traženje stimulacije i nagrađivanja kod psihopata. U istom kontekstu impulsivnost specifična za psihopatiju se tumači kao posledica umanjene sposobnosti da se odloži gratifikacija.

Takođe povišen nivo testosterona i snižen nivo kortizola, mogu da budu uzroci umanjenog funkcionisanja amigdale, kao i slabe komunikacije između subkortikalnih struktura mozga i korteksa, tako da emocionalni signali iz amigdale ne dospevaju do korteksa, što ometa donošenje odluka. Uz povišen stepen agresivnosti i hiporeaktivnost na stres, što su neke od karakteristika psihopatije, koje se takođe povezuju s disbalansom testosterona i kortizola, ovo može uticati na emocionalne deficite kao što su neosetljivost i odsustvo empatije, kao i bihevioralne obrasce koji uključuju sklonost ka preuzimanju rizika i instrumentalno korišćenje nasilja.<sup>43</sup>

Deficiti u regionima mozga koji se uglavnom navode kao neurobiološki supstrat empatičkih kapacitete se dovode u vezu s psihopatskim karakteristikama. Među naučnicima postoji opšta saglasnost kada je reč o korelaciji određenih karakteristika psihopatije, pre svega emocionalnoj neosetljivosti, impulsivnosti i agresivnosti, i nedostataka, bilo funkcionalnih bilo anatomskih, u ventromedijalnom prefrontalnom korteksu, orbitofrontalnom korteksu i amigdali (Blair and White 2013, Blair et al. 2006, Hare 1998). Osim toga, postoje deficiti u sistemu koji se

---

<sup>43</sup> Kada je reč o uticaju hormona na ponašanje treba imati u vidu da je ova veza statistička (korelacija), a ne uzročna. Za kritiku uspostavljanje kauzalne veze između nivoa testosterona i agresivnosti vidi (Sapolsky 1998).

vezuje za percepciju i osećaj bola, a verovatno je da se ovi deficiti odnose i na sistem neurona ogledala, ali je ovo područje još nedovoljno istraženo (Fecteau, Pascual-Leone, and Théoret 2008, Blair and White 2013). Konačno, značajan anatomski nedostatak je karakterističan za mozak psihopata. Reč je o redukovanoj sivoj masi i povećanoj količini bele mase u frontalnom korteksu, što je najverovatnije posledica razvojnih propusta u kresanju sinapsi (Grisolia 2001).

Na ovom mestu je važno istaći da postoje značajne neurobiološke razlike između uspešnih i neuspešnih psihopata koje se odražavaju na različite fasete psihopatije (Gao and Raine 2010). Uspešnim psihopatama se smatraju one osobe koje imaju izražene psihopatske karakteristike, čine krivična dela, ali izbegavaju da budu uhvaćeni i osuđeni. Za razliku od populacije neuspešnih psihopata koje na neurobiološkom planu karakteriše umanjen volumen orbitofrontalnog korteksa i amigdale, kao i izvesni nedostaci u hipokampusu, što su regioni mozga zaduženi za više izvršne funkcije, izražavanje emocija, odlučivanje i doživljavanje i reakcije na strah, kod uspešnih psihopata većina ovih funkcija je očuvana, a neke su čak izraženije nego kod pripadnika normalne populacije. Njih karakteriše veoma dobra obrada informacija, bolje donošenje odluka i osetljivost na strah. Osim toga, uspešne psihopate imaju izražene izvršne funkcije, što ih osposobljava za uspešno laganje, obmanjivanje i manipulaciju. Populacija uspešnih psihopata ima izraženu kognitivnu empatiju, ali ih kao i neuspešne psihopate odlikuje nedostatak emocionalne empatije i samokontrole, impulsivnost, traženje uzbuđenja, nesposobnost prepoznavanja facijalne ekspresije straha i umanjena reaktivnost na averzivne stimulse. Očuvane neurobiološke i kognitivne funkcije uspešnih psihopata posebno ih čine sklonim patološkom laganju i obmanjivanju, varanju i interpersonalnoj manipulaciji u čemu mogu da budu izuzetno uspešni.

### **4.3 Emocije i psihopatija**

Odnos psihopata prema emocijama se često poredi s odnosom osobe koja je slepa za boje prema bojama (Cleckley 1988, Hare 1999, Fallon 2013). Kao što daltonisti mogu da nauče da su neke stvari crvene, a neke zelene, i da funkcionišu u svetu u kome ovi sekundarni kvaliteti za većinu individua postoje, na sličan način psihopate mogu da nauče da određene ekspresije znače tugu ili strah i da ih zatim oponašaju, a da pri tom ne iskušavaju ništa slično, da nemaju nikakvo emocionalno iskustvo koje bi bilo u relaciji sa spoljašnjim manifestacijama ovih emocija ili je



njihova emocionalna reakcija znatno slabijeg intenziteta nego kod prosečne osobe. Jednostavno rečeno, emocionalni repertoar ovih osoba je znatno osiromašen. Ali, ove osobe su sposobne da tako dobro kamufliraju svoje emocionalne nedostatke da retko ko može da ih primeti.

Već je rečeno da su averzivni stimuli evoluirali da bi bili prekinuti, da u posmatraču izazivaju uznemirenost i nelagodu, i da je uobičajeno da iniciraju empatičku reakciju i odgovarajući bihevioralni obrazac koji će uznemirenost da prekine (Preston and de Waal 2002b). Osim što izazivaju empatiju, ovi uznemirujući signali su najčešće supresori agresivnog ponašanja, a slično važi i za tužne i uplašene izraze lica, koji u društvenim kontekstima uslovljavaju decu u razvoju da izbegavaju one oblike ponašanja koji izazivaju ove ekspresije (Marsh and Blair 2008). Međutim, to neće biti slučaj ukoliko individue ne mogu da prepoznaju ove signale i nemaju emocionalne kapaciteta da iskuse empatiju na način na koji je to iskustvo dostupno osobi s takvim kapacitetima. Na primer, kada beba plače, uobičajena reakcija je da joj se pomogne, da se nahrani ili da se presvuče. Za psihopatu, s druge strane, to može biti signal da je uhvati za noge i razlupa o zid (Hare 1999).

Psihopatija je u jakom smislu povezana s nesposobnošću da se prepoznaju facijalne ekspresije straha, tuge i iznenađenja, dok je ova veza nešto slabija kada je reč o prepoznavanju sreće, ljutnje i gađenja. Najizraženija nesposobnost za psihopate je nesposobnost prepoznavanja facijalne ekspresije straha (Marsh and Blair 2008). Eksperimentalna evidencija pokazuje da deca koja su neosetljiva i neemocionalna, kao i odrasli s psihopatskim karakteristikama ispoljavaju teškoće u prepoznavanju vizuelnih i vokalnih signala koji se vezuju za strah i tugu (Marsh et al. 2008, Blair and White 2013). Ovakvo prepoznavanje je značajno za sposobnost da se doživi empatija i ovi nalazi su konzistentni s modelom empatije koji je zasnovan na mehanizmu percepcije i akcije. Takođe, Bler i saradnici naglašavaju vezu koja postoji između psihopatije i deficita u prepoznavanju ekspresije straha, kao i umanjenu neurofiziološku reaktivnost (autonomne reakcije) i redukovano održavanje pažnje prilikom izloženosti tužnim izrazima lica. Pored ovih nedostataka, osobe s psihopatskim karakteristikama su načelno manje osetljive na neprijatne i preteće stimulse (elektrošok ili ton visokog intenziteta) i pokazuju sniženu elektrodermalnu reakciju u iščekivanju ovih draži u odnosu na populaciju bez takvih karakteristika (Hare, Frazelle, and Cox 1978, Ogloff and Wong 1990). Ovi rezultati ukazuju na činjenicu da su psihopate manje osetljive na anticipatorni strah od ostatka populacije. Do sličnih

rezultata dolaze i studije koje se bave proučavanjem reakcije trzanja (eng. *startle response*) na neprijatne stimuluse (Patrick 1994, 2001). Reakcija trzanja je defanzivna reakcija koja se javlja prilikom susreta s neprijatnim i odbojnim stimulusima. Zapaženo je da psihopate, za razliku od normalne populacije, pokazuju inhibiciju treptaja prilikom posmatranja fotografija bilo prijatnog bilo neprijatnog sadržaja, odnosno prirodna defanzivna reakcija na preteće stimuluse je kod njih odložena, što ide u prilog tezi da psihopatija uključuje deficite u anksioznosti i strahu.

Psihopate razumeju emocije na konceptualnom (kognitivnom) nivou, ali ne mogu da saosećaju, ne mogu da budu uvučene u emocionalni svet druge osobe. U kontekstu modela empatije koji se oslanja na mehanizam percepcije i akcije, ove osobe imaju teškoća da osećaju strah i da saosećaju sa strahom drugih, a slično važi i za tugu. Osim toga, oni ne doživljavaju strah na isti način kao prosečna osoba; za njih strah ne predstavlja neprijatno iskustvo. Opisujući žrtve silovanja jedan silovatelj izjavljuje da su one preplašene, ali da on ne može da razume zbog čega se osećaju neprijatno, jer strah za njega nije neprijatna emocija (Hare 1999).

Među naučnicima postoji konsenzus da psihopate nemaju sposobnost za emocionalnu dimenziju empatije, ali da je kod ove populacije sposobnost za kognitivnu empatiju koja je shvaćena kao teorija uma očuvana (Blair and White 2013). To znači da osobe s psihopatskim karakteristikama mogu na osnovu dostupnih signala da *znaju* da osoba oseća strah ili tugu, ili bar da je uznemirena, kao što na osnovu posmatranja mogu da saznaju šta izaziva ove emocije, ali ne mogu da *saosećaju* s njenim iskustvom, jer ih osiromašeni emocionalni repertoar čini nesposobnim za to. Ako je emocionalna dimenzija isključena iz doživljavanja empatije, onda ne iznenađuje što su psihopate sklone instrumentalnom izazivanju visokog nivoa uznemirenosti drugih osoba samo da bi posmatrale kako individua pati, da bi doživele uzbuđenje zbog njene patnje ili da bi postigle neki sebičan cilj.

Da bi manipulisale, osobe s psihopatskim karakteristikama koriste emocionalno obojen i metaforičan jezik, ali pri bližem ispitivanju se ispostavlja da reči s emocionalnom dimenzijom značenja, jednako kao i one koje su neutralne, za njih nisu ništa više od rečničkih definicija. Psihopate imaju teškoće da razlikuju neutralne i reči s emocionalnom dimenzijom značenja, kao i da naprave razliku između pozitivnih i negativnih emocionalno obojenih reči. Za razliku od psihopata, u zadacima identifikacije reči s emocionalnim (npr. smrt) i neutralnim značenjem

(npr. papir) osobe koje nisu psihopate reaguju brže na reči s emocionalnim značenjem nego na one s neutralnim (Williamson, Harpur, and Hare 1991). U zadacima *afektivnog primovanja*, psihopate pokazuju slabije rezultate od kontrolne grupe, a njihovo ocenjivanje reči kao pozitivnih ili negativnih po značenju odudara od populacije bez psihopatskih karakteristika, pre svega po tome što reči s neutralnim značenjem ocenjuju kao pozitivne (Blair et al. 2006).<sup>44</sup> Osobe s psihopatijom imaju konceptualno znanje o emocijama i mogu da klasifikuju reči kao pozitivne ili negativne, ali njihove emocionalne reakcije su znatno slabijeg intenziteta, pogotovo kada je reč o negativnim stumulusima, u odnosu na osobe koje nisu psihopate. Takođe, psihopate ocenjuju reči s negativnom konotacijom kao mnogo manje loše u odnosu na kontrolnu grupu. Ovi rezultati ukazuju na to da osobe s psihopatijom ne koriste reči s emocionalnom dimenzijom značenja na adekvatan način. To može biti ilustrovano primerom koji jasno ukazuje na emocionalne deficite psihopata. Na testovima neki od njih procenjuju iskaz “Čovek je crv koji živi na lešu zemlje” kao pozitivan, a iskaz “Ljubav je protivotrov za svetske nedaće” kao krajnje negativan po značenju (Hare 1998). Samo značenje reči „strah“ je za neke psihopate zagonetno. Oni primećuju spoljašnje manifestacije straha kod drugih ljudi, da uplašene osobe drhte, da im srce ubrzano kuca i da ne mogu da progovore, ali im nije jasno zašto je to tako, jer za razliku od prosečne osobe, psihopata ne vezuje strah za ove telesne senzacije, budući da ih sam ne iskušava (Hare 1999). Ima čak i takvih slučajeva gde individua ne može da imenuje ovu emociju, ali ume da da adekvatnu deskripciju: „ljudi tako izgledaju pre nego što ih izbodem“ (ABCTVCatalyst 2012).

#### **4.4 Evolucija psihopatije**

Mada na prvi pogled može izgledati sasvim očigledno da psihopatske karakteristike doprinose povećanju adaptivne vrednosti organizma, iz istih onih razloga zbog kojih u evolucionim modelima sebične jedinke prolaze bolje od altruista, ovakva ocena je preuranjena i nedovoljno utemeljena. Zapravo, mali broj psihopata u populaciji ukazuje na to da oni ne prolaze baš najbolje u evolucionoj utakmici. S druge strane, konstantan broj osoba s psihopatskim

---

<sup>44</sup> Primovanje je metod koji se koristi da bi se utvrdilo da li jedan stimulus utiče na drugi. Uobičajeno je da reč koja se koristi kao primer (npr. sto) utiče na brzinu kojom se reaguje na semantički povezanu reč (npr. stolica). Individue brže reaguju ukoliko su primer i reč koju je potrebno identifikovati semantički povezane. U zadacima afektivnog primovanja kao primer se koriste reči s emocionalnom dimenzijom značenja.

karakteristikama (oko 1% onih koji su 30+ na PCL-R i nešto veći procenat onih koji imaju visok skor na PCL-R, ali nisu u kliničkom smislu psihopate) ukazuje na to da se ove karakteristike na neki način uspešno održavaju u populaciji.

Obično se smatra da psihopate nemaju probleme, već da oni predstavljaju problem drugim osobama s kojima su u kontaktu i društvu kao celini, budući da iako su malobrojni u populaciji, čine najveći deo teških krivičnih dela. Kako nemaju kapaciteta da se suoče sa sopstvenom odgovornošću, psihopate uglavnom misle da je s njima sve u redu, a da su za probleme koje uzrokuju odgovorni ili drugi ili okolnosti. Iz egocentrične perspektive psihopata, drugi su obično zaslužili ono što su dobili, jer su slabići, mekušci ili gubitnici, u svetu koji je podeljen na snažne i slabe, pobednike i gubitnike. Mada većina njih dobar deo svog životnog veka provodi iza rešetaka, u ovoj podeli osobe s psihopatskim karakteristikama sebe doživljavaju kao pobednike.

Zapravo, nije dovoljno jasno da li je ponašanje psihopata po njih u biološkom smislu štetno ili korisno (Murphy and Stich 2000, Glenn, Kurzban, and Raine 2011). Neke od karakteristika psihopatije mogu biti korisne kada je reč o individualnom preživljavanju i reprodukciji, posebno u nepovoljnim životnim uslovima, dok druge mogu biti veoma štetne. Na primer, egoizam, odsustvo empatije, osećaja krivice i kajanja, kao i promiskuitet, mogu da olakšaju nesmetano iskorišćavanje drugih i efikasno obezbeđivanje resursa i seksualnih partnera u uslovima u kojima je takvih resursa malo. S druge strane, agresivno i antisocijalno ponašanje koje se vezuje za psihopatiju, takođe može da vodi akumulaciji resursa, ali može biti povezano s učestalim izlaganjem opasnostima, smrtonosnim povredama ili ekskomunikacijom. Kako u populaciji imamo relativno konstantan broj psihopata i kako psihopatija nije kulturno uslovljena, već je prisutna u svim kulturama, potencijalno evoluciono objašnjenje bi moralo da odgovori na pitanja o tome na koji način psihopatija može biti adaptacija (rezultat prirodne selekcije ili nekog drugog evolucionog mehanizma), tj. na koji način eventualni gubici na jednoj strani (individualno preživljavanje) mogu biti kompenzovani na drugoj (reprodukcija).<sup>45</sup> Može se

---

<sup>45</sup> Ovo može da se dovede u vezu s teorijom životnih istorija. Teorija životnih istorija polazi od pretpostavke da je životni vek organizma ograničen, a resursi oskudni, tako da on mora da čini izvesne ustupke (eng. *trade offs*), prilikom rešavanja različitih adaptivnih problema. Ova ograničenja onemogućavaju podjednako ulaganje napora u rešavanje dva različita adaptivna problema, što jedinku stavlja pred dilemu koja se rešava tako što ona mora da se

dogoditi da ove individue, zbog specifično rizičnog i agresivnog životnog stila imaju kraći životni vek od proseka u populaciji. Takođe, one mogu da budu ekskomunicirane (duga zatvorska kazna), čime mogu da se umanje njihove šanse za reprodukciju. Ali, to možda može da bude kompenzovano seksualnim strategijama koje koriste psihopate – kratkoročne veze i promiskuitet (Lalumière, Mishra, and Harris 2008). U tom smislu i sama psihopatija možda može da se posmatra i kao vrsta seksualne strategije, mada to može biti prilično kontroverzno. Realnije objašnjenje je da je psihopatija vrsta životne strategije.

Psihopatske karakteristike su neposredna posledica delovanja odgovarajućih neurobioloških mehanizama. Mada se o psihopatskim karakteristikama obično govori kao o nedostacima, iz toga ne sledi da su neurobiološki mehanizmi psihopata disfunkcionalni. Naprotiv, moguće je da neurobiološki mehanizmi osoba s psihopatskim karakteristikama funkcionišu upravo onako kako *treba* da funkcionišu, što znači da psihopate imaju karakteristike koje su iz evolucione perspektive normalne (Millikan 1984). Ukoliko postoje neurobiološki mehanizmi koji su adaptacije i koji za posledicu imaju, određene psihopatske karakteristike i antisocijalne bihevioralne obrasce koje su po pretpostavci takođe adaptivni, onda možemo da damo (kauzalno) evoluciono objašnjenje u kome su ove karakteristike i bihevioralni obrasci objašnjeni kao *prave funkcije* (eng. *proper function*) ovih mehanizama. Millikan (Ruth Garrett Millikan) definiše prave funkcije na sledeći način.<sup>46</sup> Da bi *A* imalo funkciju *F* kao pravu funkciju, nužno je (i skoro dovoljno) da bude zadovoljen jedan od ova dva uslova:

1) *A* je nastalo kao reprodukcija (kao kopija ili kopija kopije) neke prethodne stvari koja je, delom *zahvaljujući* tome što je imala svojstvo koje je reprodukovano, izvršavala *F* u prošlosti, i *A* postoji zato što je *F* izvršavano. 2) *A* je nastalo kao proizvod nekog prethodnog mehanizma koji je, pod datim okolnostima, imao *F* kao pravu funkciju i koji, pod tim okolnostima, po pravilu uzrokuje da *F* bude izvršeno tako što proizvodi stvari kao što je *A* (Millikan 1993, 13-14).

---

opredeli da li će da ulaže u sopstveni telesno razviće nasuprot reprodukciji, roditeljstvo nasuprot parenju, kvalitet nasuprot kvantitet potomaka i buduću nasuprot sadašnjoj reprodukciji. U zavisnosti od toga kako su rešene ove adaptivne dileme, razlikuju se brze od sporih životnih strategija organizama. Za psihopate se pretpostavlja da usvajaju brze životne strategije, tj. da uglavnom daju prednost drugoj komponenti navedenih parova (Glenn, Kurzban, and Raine 2011). Osim toga vidi i (Alexander 2009 [1987]).

<sup>46</sup> Za detaljnu analizu pravih funkcija vidi (Millikan 1984).

Jednostavnije rečeno, da bi određena karakteristika imala *pravu funkciju* neophodno je da ima odgovarajuću kauzalnu istoriju, odnosno neophodno je da bude član *reproduktivno ustanovljene porodice*; karakteristika o kojoj je reč je morala da bude *selektovana za* i prenošena iz generacije u generaciju, upravo zato što je doprinosila preživljavanju i reprodukciji organizma koji ju je posedovao, više nego neka alternativna karakteristika (Millikan 1993, 2002). Ako su psihopatske karakteristike u predačkim populacijama i pod nepovoljnim uslovima životne sredine doprinosile povećanju adaptivne vrednosti organizma više nego neke druge (npr. altruističke) karakteristike, onda su one mogle da budu proširene u populaciji, a za individue koje su ih posedovale mogli bismo da kažemo da su bile evoluciono uspešne. U tom smislu psihopatija nije vrsta poremećaja (disfunkcija), već adaptacija, mada iz toga ne sledi da su karakteristike specifične za psihopatiju u sadašnjim uslovima adaptivne (Murphy and Stich 2000). Ako u evolucionom kontekstu psihopatiju ne možemo smatrati poremećajem, onda je reč o tome da su neke osobe biološki programirane da nemilosrdno obmanjuju, budu nasilne, eksploatišu druge individue i sistem društvene saradnje.

Psihopatija je prema jednom broju autora proizvod prirodne selekcije i predstavlja adaptivnu obmanjujuću strategiju u nepovoljnim uslovima životne sredine (Mealey 1995, Murphy and Stich 2000, Lalumière, Mishra, and Harris 2008). Najpoznatiji evolucioni model psihopatije predložila je Linda Mili (Linda Mealey) i on kombinuje dva evolucionarna mehanizma – (1) negativnu selekciju zavisnu od učestalosti gena i (2) evoluciono-razvojni mehanizam uslovne adaptacije (fenotipska plastičnost).

Načelno, selekcija zavisna od učestalosti gena se odnosi na situaciju u kojoj adaptivna vrednost nekog fenotipa (ili genotipa) zavisi od njegove zastupljenosti u populaciji (Tucić 2003). Prema Tucićevim rečima, negativna selekcija zavisna od učestalosti gena (ili balansirajuća selekcija) održava genetičku varijabilnost populacije i deluje u situaciji u kojoj adaptivna vrednost određenog genotipa (fenotipa) predstavlja inverznu funkciju njegove učestalosti u populaciji. Jednostavnije rečeno, negativna selekcija zavisna od učestalosti gena čini da adaptivna vrednost genotipova (fenotipova) opada kako učestalost tog genotipa (fenotipa) raste u populaciji (Ridley 2003). S druge strane, mehanizam uslovne adaptacije sugerise da organski sistemi jedinki, uključujući i nervni sistem, evoluiraju osposobljeni da fleksibilno reaguju na različite uslove životne sredine iz čega sledi da će u zavisnosti od sredinskih uslova jedinke

usvajati različite životne strategije (Giudice, Ellis, and Shirtcliff 2011, Glenn, Kurzban, and Raine 2011).

U evolucionom objašnjenju psihopatije Mili se oslanja na distinkciju između *primarne i sekundarne psihopatije* i svoje nalaze kombinuje u jedinstven evoluciono-razvojni model. *Primarna psihopatija* je rezultat delovanja prirodne selekcije, preciznije selekcije zavisne od učestalosti gena i Mili je opisuje kao genetički fiksiranu obmanjujuću, predatorsku i manipulativnu strategiju čija je zastupljenost oko jedan odsto u opštoj populaciji. Reč je dakle o onim osobama čiji je skor na PCL-R veći od 30. Ova strategija je analogna (ili čak identična) strategiji AILD u *Ponovljenoj dilemi zatvorenika* i moglo bi da se očekuje da se održi u populaciji ukoliko je niskofrekventna. To znači da bi relativno mali broj osoba s psihopatskim karakteristikama mogao da se održava u populaciji u kojoj je dominantna strategija saradnja i ako su šanse za njihovo otkrivanje male (Mealey 1995, Lalumière, Mishra, and Harris 2008). Niska frekventnost umanjuje šanse za njihovo otkrivanje, kao što se šanse za njihovo otkrivanje umanjuju veštím maskiranjem psihopatskih karakteristika, što je konzistentno s nalazima da su psihopate veoma uspešne u prikrivanju emocionalnih deficita.

Jedan od problema s ovom teorijom je taj što prirodna selekcije nastoji da eliminiše, a ne da održava one gene koji imaju i najmanji negativan efekat na adaptivnu vrednost organizma. Na primer, ako su psihopate genetički determinisane da se učestalo izlažu opasnostima, kroz rizično, nasilničko, antisocijalno ponašanje, itd., onda se može očekivati da njihov broj u populaciji opada s obzirom na mogućnosti povređivanja i potencijalno visoku stopu smrtnosti. Drugi problem je što evolucija retko isporučuje jednostavne, genetički fiksirane i nepromenljive strategije. Zapravo organizam koji bi na takav način bio programiran ne bi mogao adekvatno da odgovori na promenljive životne uslove i njegov evolucioní uspeh bi u takvim uslovima mogao lako da bude doveden u pitanje.

Drugačije evoluciono objašnjenje psihopatije se oslanja na koncepciju *fenotipske plastičnosti* i ono se u literaturi obično označava kao *pomeranje kontingenata* ili *uslovna adaptacija* (Buss 2009, Giudice, Ellis, and Shirtcliff 2011, Glenn, Kurzban, and Raine 2011). *Fenotipska plastičnost* se odnosi na sposobnost genoma da u jednim uslovima životne sredine proizvede određeni fenotip, a da u drugačijim uslovima životne sredine, isti genom, proizvede

različiti fenotip. Kako genetički činioci utiču na fenotipsku plastičnost, a evolucija se shvata kao promena u učestalosti gena u populaciji, fenotipska plastičnost ne može da bude isključena iz evolucionog procesa. Prema ovom eksplanatornom modelu „životna sredina nije samo arena u kojoj različiti fenotipovi postižu različiti stepen preživljavanja i reproduktivnog uspeha, već je i činilac razvića, koji značajno utiče na opseg fenotipova koje će dati genotip da proizvede“ (West-Eberhard 1989, 252). Kako je u ovom kontekstu jedinica selekcije fenotip, uticaj životne sredine na razviće se reflektuje i na smer evolucije. Vest-Eberhard (Mary Jane West-Eberhard), kao posebno plastičnu vrstu fenotipa navodi ponašanje koje je podložno bržoj i efikasnijoj promeni u različitim uslovima životne sredine, u odnosu na njegove neurobiološke supstrate. Budući da su organizmi tokom evolucije bili izloženi i povoljnim i nepovoljnim uslovima životne sredine, prirodna selekcija je proizvela organske sisteme sposobne da se prilagode različitim uslovima. Na taj način, u sredinama koje su bogate stresorima, jedinke će biti usmerene da usvoje strategije koje su adaptivne za tu sredinu, dok će u drugačijim uslovima biti usmerene da izaberu drugačije strategije (Giudice, Ellis, and Shirtcliff 2011). Nalazi koji ukazuju na hiporeaktivnost na stres, koja je indikovana sniženim nivoom kortizola, kao i manjak anksioznosti u iščekivanju neprijatnih stimulusa kod osoba s psihopatskim karakteristikama su konzistentni s predviđanjima ove teorije (Glenn, Kurzban, and Raine 2011, Raine and Glenn 2014).

Mili koristi model uslovne adaptacije da bi objasnila evoluciju sekundarne psihopatije. Sekundarna psihopatija se odnosi na prisustvo psihopatskih karakteristika, tj. psihopatskih strategija, koje se usvajaju tokom razvoja, a pod uticajem delovanja nepovoljnih sredinskih uslova, kao što su neadekvatno roditeljsko staranje, razdvajanje roditelja i dece u ranom uzrastu, zanemarivanje, fizičko zlostavljanje, itd. (Mealey 1995, Buss 2009, Glenn, Kurzban, and Raine 2011). Teška socio-ekonomska situacija ili nepovoljne društvene okolnosti (glad, rat, itd.) takođe mogu da uslove ispoljavanje psihopatskih karakteristika. S druge strane, i u povoljnim sredinskim uslovima individua u ranom periodu razvoja, ali i kasnije, može da usvoji životnu strategiju specifičnu za psihopatiju zbog različitih fenotipskih karakteristika koje joj omogućavaju da relativno lako koristi, recimo, zastrašivanje i instrumentalno nasilje. Takve fenotipske karakteristike mogu da budu krupna telesna građa i nasledna emocionalna neosetljivost. S obzirom da je ovaj oblik psihopatije rezultat delovanja mehanizama uslovne



adaptacije, njena zastupljenost u populaciji varira. Ipak, ni u ovom slučaju ne treba zanemariti uticaj koji geni mogu imati na ispoljavanje psihopatskih karakteristika i takvih bihevioralnih obrazaca, s obzirom na to da roditelji koji su, recimo, načelno skloni fizičkom zlostavljanju, deci mogu da prenose gene koji ih čine sklonim sličnim bihevioralnim manifestacijama, tj. antisocijalnom ponašanju (Glenn, Kurzban, and Raine 2011). Psihopate u ovom drugom smislu fleksibilno koriste različite strategije u zavisnosti od okolnosti. Ukoliko bi trebalo napraviti analogiju s nekom strategijom koja je razvijena u teoriji igara, onda se sekundarni psihopata verovatno najčešće koristi strategijom PAVLOV, koja ga čini sposobnim za saradnju, ali i za nemilosrdnu eksploataciju sistema društvene saradnje. Ali, ovo treba uzeti s rezervom.

Alternativno objašnjenje psihopatije je da je ona štetna mutacija, tj. posledica *mutacionog opterećenja* (Buss 2009, Glenn, Kurzban, and Raine 2011). *Mutaciono opterećenje* predstavlja verovatnoću da će prosečna jedinka u populaciji imati nižu stopu preživljavanja i reprodukcije zato što je nosilac nepovoljne mutacije (Tucić 2003). Prilikom replikacije, DNK se obično kopira precizno, tako da je kopija verna originalu, međutim ponekada se dešavaju greške u kopiranju koje se obično ispravljaju delovanjem određenih enzima. Ali, čak i nakon ispravki neke od grešaka ostaju. Ove greške u kopiranoj DNK se nazivaju mutacije. Neke mutacije su neutralne, tj. nemaju nikakvog efekta na adaptivnu vrednost organizma, neke utiču povoljno na preživljavanje i reprodukciju i takve mutacije će biti favorizovane prirodnom selekcijom, ali najveći broj mutacija je štetan i njih prirodna selekcija uglavnom eliminiše iz populacije. Ukoliko je psihopatija štetna mutacija, teorija evolucije predviđa da će ona da bude eliminisana iz populacije. Međutim, ako su nepovoljni efekti mutacije mali, moguće je da oni izmiču delovanju prirodne selekcije. U tom slučaju, psihopatija bi bila posledica kumulativnog delovanja mnoštva mikromutacija koje nepovoljno utiču na razviće i funkcionisanje mozga.

#### **4.5 Šta deca mogu, a psihopate ne mogu?**

Tokom razvoja i vaspitanja roditelji ili staratelji ispostavljaju deci najrazličitiji niz pravila i od njih očekuju da ih (uglavnom) bespogovorno poštuju. Ova pravila su najčešće negativno formulisana u formi zabrana, koje određuju koji postupci nisu ispravni, dopušteni ili opravdani, ali ima i onih koja su formulisana pozitivno u obliku preporuka. U ovom procesu roditelji ili staratelji deci ne skreću posebnu pažnju na to koje zapovesti se odnose na *moralna pravila*, a

koja se odnose na *pravila etikecije*, odnosno *konvencionalna pravila*. Oni jednostavno govore „ne bacaj hranu na pod“, „ne žvači otvorenih usta“, „deli svoje igračke“, „ne udaraj decu“, „ne čupaj devojčice za kosu“, „ne otimaj se“, „ne galami“, „reci dobar dan“, „ne vuci noge po prašini“, „ne laži“, „budi dobar“, „čekaj svoj red“, itd. Jedina uočena razlika se odnosi na sledeće. U određenom periodu razvoja, opomene zbog kršenja moralnih pravila se podupiru apelom na dečje emocije i to obično u obliku pitanja kao što je „Kako bi se ti osećao da te neko udari?“, dok se na kršenje konvencionalnih pravila obično reaguje insistiranjem da se određeno ponašanje, koje se smatra neprihvatljivim, prekine, npr. „Prekini da galamiš!“ (Nucci 2001). Mada roditelji razliku između konvencionalnih i moralnih pravila ne čine eksplicitnom, deca uprkos tome mogu da ih razlikuju.

Kako se razlika između moralnih i konvencionalnih pravila u svakodnevnom kontekstu ne čini eksplicitnom, ovu distinkciju nisu pravili ni rani istraživači u društvenim i humanističkim naukama, već su se moralna i konvencionalna pravila tumačila kao ekvivalentna (Turiel 1980). Da ova razlika zaista postoji, postaje jasno kada se fokusiramo na način na koji se kršenje različitih pravila kao što su „ne bacaj hranu na pod“ i „ne čupaj devojčice za kosu“ prosuđuju (Weston and Turiel 1980). Konvencionalna pravila su određene društvene pravilnosti koje počivaju na zajedničkom znanju i društvenim očekivanjima i kojima se usklađuju postupci individua u socijalnom kontekstu. Ova pravila se tumače kao promenljiva, zavisna od konteksta u kome se pojavljuju i intersubjektivne saglasnosti. S druge strane, moralna pravila se tumače kao univerzalno preskriptivna, nezavisna od društvenog konteksta i intersubjektivne saglasnosti, a njihovo kršenje se sagledava u terminima nanošenja povreda (eng. *harm*), kršenja prava i činjenja štete. Drugim rečima, ona se *smatraju* objektivnim. Pretpostavlja se da je ova razlika rezultat apstrakcije iz iskustava, što je poduprto adekvatnim emocionalnim reakcijama i empatijom (Turiel 1980, 2002). Na primer, bacanje hrane na pod deca prosuđuju kao zabranjeno, ali smatraju da bi roditelji ili staratelji mogli da ukinu ili suspenduju ovu zabranu. Udaranje i čupanje, deca takođe prosuđuju kao zabranjeno, ali smatraju da ova zabrana važi nezavisno od datog konteksta i da ne bi mogla da se ukloni ili suspenduje odlukom autoriteta, jer ovi postupci nanose bol i povređuju druge i značajno su ozbiljniji od bacanja hrane. Slično važi i za krađu ili laganje.

Oslanjajući se na kriterijume objektivnosti, univerzalnosti, nanošenja povreda (štete) i ozbiljnosti, Keli i saradnici (Kelly et al. 2007, 118) sumiraju karakteristike moralnih i konvencionalnih pravila na sledeći način:

1. Moralna pravila imaju *objektivnu*, preskriptivnu snagu; nisu zavisna od autoriteta, bilo koje osobe ili institucije.
2. Moralna pravila imaju *opšte* važenje, tj. ne važe samo lokalno; ona važe nezavisno od bilo kog geografskog ili istorijskog konteksta.
3. Kršenje moralnih pravila obično uključuje *žrtvu* koja je povređena, čija prava su prekršena ili koja je pretrpela nepravdu.
4. Kršenje moralnih pravila je načelno *ozbiljnije* od kršenja konvencionalnih pravila.

Nasuprot moralnim, konvencionalna pravila imaju sledeće karakteristike:

1. Konvencionalna pravila su arbitrarna, to su pravila zavisna od situacije čija je funkcija da podupiru društvenu koordinaciju i organizaciju; ona *nemaju objektivnu* preskriptivnu snagu i mogu da budu suspendovana ili promenjena odlukom odgovarajućeg autoriteta.
2. Konvencionalna pravila su često *lokalna*; ona su zavisna od društvenog, kulturnog i istorijskog konteksta.
3. Kršenje konvencionalnih pravila *ne uključuje žrtvu* koja je povređena, čija su prava prekršena ili koja je pretrpela nepravdu.
4. Kršenje konvencionalnih pravila je *manje ozbiljno* od kršenja moralnih pravila.

Na osnovu eksperimentalnih rezultata, istraživači koji se oslanjaju na distinkciju između moralnih i konvencionalnih pravila tvrde da deca već u najranijem uzrastu mogu da naprave razliku među ovim pravilima. Tako deca uzrasta dve i po godine mogu da naprave razliku na osnovu jednog ili dva kriterijuma, trogodišnjaci prave razliku na osnovu nekoliko kriterijuma, a četvorogodišnjaci ova pravila razlikuju na osnovu svih predloženih kriterijuma. S određenim izuzecima, ovi nalazi su potvrđeni u brojnim kroskulturalnim studijama.<sup>47</sup>

---

<sup>47</sup> Za pregled vidi (Turiel 2006).

Na primer, na osnovu navedenih kriterijuma, adolescenti katoličke veroispovesti prave razliku između moralnih zabrana, kao što su zabrana ubistva, silovanja ili krađe i konvencionalnih, tj. onih koja zavise od autoriteta sveštenika, pape ili boga, kao što su zabrana stupanja u seksualne odnose pre braka ili masturbacija, iako je za kršenje obe zabrane zaprećena drastična kazna (pakao) (Nucci 2001). Takođe, velika većina amiša i ortodoksnih Jevreja različitog uzrasta, kojima se sva pravila predstavljaju kao božje zapovesti, pravi razliku između moralnih i konvencionalnih pravila, tako da moralne prestupe kao što su ubistvo ili krađa, smatra pogrešnim, čak i u protivčinjeničkoj situaciji u kojoj ne bi postojale božje zapovesti koje ove postupke zabranjuju. Konvencionalna pravila se tumače kao zavisna od božanskog autoriteta, dok se zabrana moralnih prestupa tumači kao univerzalno važeća, čak i u slučaju da ih bog dopušta.

Kada je reč o moralnim pravilima, izgleda da je za ljude moralni objektivizam podrazumevana metaetička pozicija (Nichols and Folds-Bennett 2003). Ako se neki postupci prosuđuju kao pogrešni, oni su jednostavno pogrešni, nezavisno od toga ko ih je učinio, kao i od toga da li ih bilo ko odobrava ili ne i to je kraj svake diskusije. Ovu karakteristiku uobičajenog (svakodnevnog, zdravorazumskog) razumevanja morala lepo ilustruje Meki:

Običan korisnik moralnog jezika namerava da kaže nešto o bilo čemu što smatra moralnim, kao što je na primer mogući postupak, kakav je po sebi, kakav bi bio da se realizuje, a ne o [...] svom ili bilo čijem stavu ili odnosu prema njemu. Ali nešto što želi da kaže nije samo deskriptivno, svakako nije inertno, već je nešto što uključuje poziv na delanje ili uzdržavanje od delanja, i nešto što je apsolutno, a ne nešto što je uslovljeno bilo kojom željom ili preferencijom ili politikom ili izborom... Pitanje nije [...] da li on stvarno želi da izvrši određeni postupak, da li će ga to učiniti zadovoljnim ili ne, da li će dugoročno imati odobravajući stav prema njemu, pa čak ni to može li ovaj postupak srećno i iskreno da preporuči u svim slučajevima koji su slični na relevantan način. Niti se pita da li da preporuči takav postupak u svim slučajevima koji su slični na relevantan način. On želi da zna da li bi ovaj kurs delovanja bio pogrešan po sebi (Mackie 1990 [1977], 33-34).

Razlikovanje moralnih i konvencionalnih pravila u najranijem uzrastu navelo je Nikolsa (Shaun Nichols) na zaključak da (afektivno) zauzimanje perspektive, tj. kognitivna empatija, koja se često identifikuje s teorijom uma, nije od posebnog značaja za moralno suđenje (Nichols 2004). Ipak, treba imati na umu da je područje istraživanja teorije uma prilično opskurno i da nije

precizno utvrđeno kada se teorija uma kod dece pojavljuje. Neki eksperimenti, o kojima je već bilo reči, pokazuju da su deca sposobna za afektivno zauzimanje perspektive već s 18 meseci starosti, na osnovu čega može da se zaključi da su neki elementi teorije uma prisutni već u ovom uzrastu, što znači pre nego što počnu da prave razliku između konvencionalnih i moralnih pravila (Vaish, Carpenter, and Tomasello 2009). Nezavisno od vremena kada se teorija uma pojavljuje kod dece, iz proučavanja psihopatije znamo da kognitivna empatija zaista može biti od manjeg značaja za evoluciju morala od emocionalne empatije, što je konzistentno s Nikolsovim predlogom o „sentimentalnim pravilima“ koja imaju normativnu i emocionalnu dimenziju.

Za razliku od dece koja spontano i s lakoćom razlikuju konvencionalna i moralna pravila, osobe s psihopatskim karakteristikama imaju značajne teškoće da naprave ovu razliku (Blair 1995). Bler je pretpostavio da psihopate sva pravila tumače kao konvencionalna. Ali, nasuprot njegovom očekivanju, pokazalo se da oni zapravo sva pravila tumače kao moralna, tj. objektivna i univerzalno važeća. Međutim, u kontekstu opravdanja se nikada ne pozivaju na to da kršenje nekog moralnog pravila nije dopušteno zato što takvi postupci nanose bol i povređuju druge, već na to da takvi postupci nisu društveno prihvaćeni. Kako je istraživanje vršeno u penalnim uslovima, Bler tumači ovaj nalaz pre kao posledicu želje psihopata da pokažu da su se promenili i da navodno prihvataju pravila koja većina populacije prihvata, nego kao posledicu stvarne moralizacije konvencionalnih pravila. Zapravo, zbog nedostatka empatije i drugih emocionalnih deficita koji ih onesposobljavaju da se užive u emocionalni svet druge osobe, psihopate uopšte nisu sposobne da razumeju moralna pravila, niti da razumeju šta moralno pravilo čini moralnim, pa onda ne mogu ni da stvarno moralizuju pravila koja su konvencionalna (Levy 2010).

Ipak, nisu svi uvereni da distinkcija između moralnih i konvencionalnih pravila ima onu važnost za filozofiju morala koja joj se obično pripisuje. Kritičari su razvili niz domišljatih scenarija koji ne uključuju primere iz školskog dvorišta (obdaništa) kakvi se obično koriste u testiranju dece (a i odraslih, uključujući i psihopate), kako bi pokazali da individue bar ponekada tumače moralne prestupe kao zavisne od autoriteta i društveno-istorijskog konteksta (Kelly et al. 2007). U ovim eksperimentima subjekti prosuđuju fizičko kažnjavanje (bičevanje mornara) kao moralno prihvatljivo pre 300 godina, ali ne i danas. Takođe, smatraju da je ropstvo bilo opravdano u grčko-rimskom svetu, ali ne i na američkom jugu, da je kanibalizam prihvatljiv u drugim kulturama, ali ne u SAD, itd. Na osnovu ove kritike Šumejker (David W. Shoemaker)

zaključuje da je pozivanje na distinkciju između moralnih i konvencionalnih pravila, zapravo pozivanje na utvaru, jer takva načelna distinkcija ne postoji, već je pre reč o nizu specifičnijih poddistinkcija (dopustivo/nedopustivo, ozbiljno/manje ozbiljno, zavisno od autoriteta/nezavisno od autoriteta), koje se donekle i ponekada podudaraju (Shoemaker 2011).

Mada je ova kritika zanimljiva, ona ne pruža dovoljne razloge za napuštanje distinkcije između moralnih i konvencionalnih pravila. Situacije predstavljene u scenarijima su složene i dvosmislene. One uključuju više aspekata, od kojih neki mogu da se tumače kao moralni, dok drugi mogu da se prosuđuju kao konvencionalni, što subjekta može da dovede u zabunu. Uzmimo primer bičevanja. Najpre, bičevanje uključuje nanošenje bola i povreda, što može da se tumači kao moralno pogrešno, nedopustivo i neopravdano. Ali, s druge strane, bičevanje kao telesno kažnjavanje može da bude protumačeno kao konvencionalno pravilo koje je nekada bilo prihvaćeno, ali je danas jednostavno napušteno, jer se smatra da je bičevanje ljudi pogrešno.<sup>48</sup>

Predmet narednog poglavlja je *samoobmanjivanje* kao specijes specifičan fenomen značajan za moralno postupanje, usvajanje i održanje moralnih verovanja, kao i verovanja koje se odnosi na objektivnost moralnih svojstava. U kontekstu analize evolucionog porekla samoobmanjivanja i njegovog značaja za evoluciju morala, još jednom ćemo se vratiti na rezultate psiholoških i neurobioloških istraživanja psihopatije. Ova istraživanja će nam poslužiti kao osnova za kritiku Triversovog evolucionog modela samoobmanjivanja.

---

<sup>48</sup> Za sličan odgovor na prigovore vezane za distinkciju između moralnih i konvencionalnih pravila vidi (Prinz 2007).

## 5 SAMOOBMANJIVANJE: KONCEPTUALNI PROBLEMI I EVOLUCIJA

Prirodna selekcija je u kompleksne organizme kakvi su ljudi „ugradila“ kognitivne mehanizme koji ih čine sposobnim da usvajaju verovanja, nezavisno od njihove istinosne vrednosti.<sup>49</sup> Usvajanje verovanja koja su pogrešna čini se gotovo očigledno maladaptivno, ali iza ovog fenomena mora postojati neki skriveni plan. Na primer u situacijama analognim *Ponovljenoj dilemi zatvorenika* strategija kao što je *milo za drago* može biti uspešna i proces reciprocijacije može da otpočne samo ako individua koja počinje interakciju (pogrešno) veruje da je druga strana nesebična i da će na saradnju odgovoriti saradnjom. Ukoliko ona ne neguje ovaj *nerealistični optimizam* proces reciprocijacije je osuđen na neuspeh pre nego što je i otpočeo. Na taj način, ako je usvajanje pogrešnog verovanja u pojedinim situacijama i odgovarajućim uslovima životne sredine vodilo povećanju adaptivne vrednosti organizma, prirodna selekcija je mogla da favorizuje one organizme koji usvajaju ova verovanja, uprkos tome što krše uobičajene epistemičke standarde. U tom smislu, koncept samoobmanjivanja bi mogao da ima važnu eksplanatornu funkciju kada je reč o evoluciji morala.

Filozofske interpretacije fenomena samoobmanjivanja obično se dele u dve struje. Jedna je intencionalistička, dok je druga neintencionalistička. Oni autori koji zastupaju intencionalistički pristup samoobmanjivanju su obično divizionisti i pretpostavljaju sledeće: da bi samoobmanjivanje uopšte bilo moguće i da bi moglo da bude objašnjeno, kao i da bi se izbegle zamke dinamičkog i statičkog paradoksa, u objašnjenju se moramo pozivati na funkcionalnu ili na temporalnu raščlanjenost uma. S druge strane, neintencionalisti su obično i nedivizionisti i pretpostavljaju da objašnjenje fenomena samoobmanjivanja ne mora da uključuje nikakvo deljenje uma i da može da se objasni pozivanjem na kauzalne procese (Talbot 1995).<sup>50</sup> Osim toga, neki autori smatraju da je debata o intencionalnosti ili neintencionalnosti samoobmanjivanja deplasirana, a shodno tome i rasprave o tome da li je um podeljen ili ne, jer

---

<sup>49</sup> Deo ovog poglavlja objavljen je u časopisu *Theoria*.

<sup>50</sup> Izuzetak od ovog pravila je sam Talbot koji je intencionalista, ali nije divizionista, zapravo prihvata jednu vrstu umerenog, benignog divizionizma.

ona sledi iz pogrešnog razumevanja uma koji je prema njihovom mišljenju istovremeno jedinstven, složen, decentralizovan i opterećen brojnim protivrečnostima koje se reflektuju na bihevioralnom planu i u svakodnevnom životu (Sahdra and Thagard 2003). Prema takvoj interpretaciji, to važi ne samo za um koji se samoobmanjuje, već i za svaki drugi.

## 5.1 Interpersonalno obmanjivanje, laganje i dovođenje u zabludu

Interpersonalno obmanjivanje uključuje najmanje dve osobe i predstavlja proces u kome jedna od njih svesno, intencionalno i uspešno navodi drugu da veruje nešto za šta prva zna da nije slučaj. Sam proces obmanjivanja koji uključuje dve individue može formalno da se prikaže na sledeći način. Poći ću od standardne definicije znanja kao istinitog opravdanog verovanja. Pretpostaviću da je  $p$  istinito, a  $ne-p$  lažno, što  $A$  zna, tako da nastoji da navede  $B$  da veruje da  $ne-p$ . Nakon obmanjivanja  $B$  veruje da  $ne-p$ .

- i.  $A$  zna da je  $p$  istinito, a da je  $ne-p$  lažno.
- ii.  $B$  u trenutku  $t$  nema formirano verovanje da  $p$  ili da  $ne-p$ .
- iii.  $B$  u trenutku  $t$  ima formirano verovanje da  $p$ , ali je nesiguran u pogledu istinosne vrednosti  $p$ .
- iv.  $A$  ima razloga da veruje da  $B$  u trenutku  $t$  nema formirano verovanje da  $p$  ili da  $ne-p$  ili da je nesigurna u pogledu istinosne vrednosti da  $p$ , ako ima formirano verovanje da  $p$ .
- v.  $A$  tvrdi da  $ne-p$  ili na neki drugi način nastoji da navede  $B$  da veruje da  $ne-p$ .
- vi.  $B$  u trenutku  $t_1$ , pod uticajem  $A$ , veruje da  $ne-p$ .

Sličnost između obmanjivanja i laganja se čini očiglednom, međutim ipak nije reč o istim činovima. Pre svega, za obmanjivanje je ključan kriterijum uspeha. Da bi individua bila dovedena u stanje obmane, osoba koja obmanjuje bi morala da bude uspešna u obmanjivanju. Ako obmanjivanje ne dovodi do obmane, onda ne bismo mogli uopšte da govorimo o obmanjivanju. S druge strane, za laganje kriterijum uspeha nije presudan.  $A$  može da tvrdi da  $ne-p$  uprkos znanju da  $p$  i da pokuša da uveri  $B$  da  $ne-p$ , ali  $B$  ne mora da formira verovanje u skladu s nastojanjima lažova (McLaughlin 1988). Strogo govoreći, nema neuspešnih obmanjivača, ali ima neuspešnih lažova. Možda bismo mogli da kažemo da obmanjivanje može da uključuje



laganje, ali laganje nije ni nužan ni dovoljan uslov za obmanjivanje, jer jedna od strategija obmanjivanja može da bude i veoma učestalo govorenje istine.

Uzmimo sledeći primer (Baier 1996). Pretpostavimo da  $A$  ima nameru da izvrši samoubistvo i svakodnevno ponavlja ovu svoju nameru ljudima iz svog bliskog okruženja – porodici, rođacima i prijateljima. Budući da  $A$  iz dana u dan ponavlja da će se ubiti, osobe koje su mu bliske, iako najpre uznemirene ovakvim saopštavanjem  $A$ -ovih namera, konačno prestaju da veruju da će  $A$  zaista izvršiti samoubistvo. Budući da ga svakodnevno susreću živog i zdravog, oni počinju da pripisuju ovakve izjave  $A$ -ovoj ekscentričnosti ili nečemu sličnom, da bi  $A$  jednog dana svoje namere sproveo u delo tako što se obesio, pošto se prethodno nagutao tabletama za spavanje. Može se dodati da tek uspešni lažov postaje obmanjivač, iako uspešni obmanjivač može da bude i onaj ko učestalo govori istinu (Hirstein 2004).

Za obmanjivanje i laganje je ključna namera, dok za dovođenje u zabludu to ne mora nužno da bude slučaj. Biti u zabludi jednostavno znači da  $A$  u nedostatku adekvatne evidencije pogrešno veruje da  $p$ , iako bi pod pritiskom empirijskih dokaza verovao da  $ne-p$ . Subjekt može biti doveden u zabludu ako bi mu nekim slučajem, ali ne nužno intencionalnim delovanjem drugih subjekata, relevantna evidencija bila uskraćena ili ako bi mu bila selektivno prikazana tako da bi mogao da pogrešno veruje da  $p$ . Sam čin obmanjivanja i dovođenja u zabludu se razlikuju iako bi sadržaj verovanja subjekta do koga dovode ovi činovi mogao da bude podudaran. Učiniti da neko veruje da  $p$  uprkos evidenciji da  $ne-p$  i *namerno* učiniti da neko veruje da  $p$  uprkos evidenciji da  $ne-p$  nije isto. Kada je reč o dovođenju u zabludu evidencija može jednostavno da ne bude dostupna, ali da to ne znači da je namerno uskraćena, dok, kada je reč o obmanjivanju, evidencija subjektu nije tek jednostavno nedostupna, već mu je ili ciljano pogrešno predstavljena ili mu je namerno uskraćena.<sup>51</sup>

Da bi osoba bila obmanuta ili slagana važna je namera. Važno je i to da se osoba koja laže ili obmanjuje predstavlja kao da zaista veruje ono što tvrdi da je slučaj, a zašta zapravo zna da nije slučaj, kao i da nastoji da svoje namere prikrije (Davidson 1987). Kada je reč o laganju, osim namere važna je i verbalna ekspresija onoga za šta subjekt koji laže zna da nije slučaj.

---

<sup>51</sup> Na primer ljudi su dugo verovali da se Sunce kreće oko Zemlje i da je smer njegovog kretanja od istoka prema zapadu. Oni su bili u zabludi, a da ih niko nije doveo u zabludu.

Strogo govoreći, verbalna ekspresija nije nužna za obmanjivanje, jer osoba može da bude obmanuta i prećutkivanjem ili skretanjem pažnje (Pears 1987).<sup>52</sup> Neko bi mogao da primeti da individua može prećutkivanjem ili skretanjem pažnje da bude dovedena i u zabludu. Ipak, postoji jedan aspekt dovođenja u zabludu koji ovaj čin izdvaja od obmanjivanja i laganja, a on se sastoji u sledećem: osoba *greškom*, dakle potpuno neintencionalno, može da bude dovedena u zabludu, ali ne može greškom da bude obmanuta ili slagana (McLaughlin 1996).

Pretpostavimo da je  $p$  istinito, a  $ne-p$  lažno, ali da  $A$  na osnovu raspoložive evidencije veruje da je  $ne-p$  istinito.  $A$  tvrdi da  $ne-p$ , ali nema nameru da obmane  $B$  da  $ne-p$ .

- i.  $p$  je istinito, a  $ne-p$  lažno.
- ii. Evidencija kojom raspolaže  $A$  je dobijena na osnovu procedure za koju veruje da je nepristrasna ili na osnovu svedočanstva u koje  $A$  ne sumnja.
- iii. Evidencija kojom raspolaže  $A$  ide u prilog tome da je  $ne-p$  istinito.
- iv.  $A$  na osnovu raspoložive evidencije veruje da je  $ne-p$  istinito.
- v.  $A$  nema nameru da obmane  $B$  da  $ne-p$ .
- vi.  $B$  u trenutku  $t$  nema formirano verovanje da  $p$  ili da  $ne-p$ .
- vii.  $A$  iskreno tvrdi da  $ne-p$ .
- viii.  $B$  na osnovu svedočanstva  $A$  veruje da  $ne-p$ .

Onaj ko na osnovu raspoložive evidencije ima razloga da veruje da je  $ne-p$  istinito ne može drugu osobu da obmane tvrdeći  $ne-p$ , ali može da je dovede u zabludu. Takođe, ljudi mogu da budu u zabludi, a da ih niko nije intencionalno doveo u zabludu. Ljudi su vekovima verovali da je Zemlja ravna ploča, da postoji kraj sveta – ivica sa koje se može pasti u bezdan. Oni se nisu samoobmanjivali niti su bili obmanuti, već su bili u zabludi. Na osnovu pomorskih istraživanja i raspoložive evidencije oni su formirali određeno verovanje. Ali, novim pomorskim istraživanjima koja su usledila ustanovljeno je da Zemlja nije ravna ploča i da kretanje u bilo kom pravcu uvek na kraju vodi na isto mesto, da nema nikakvog kraja sveta sa čije ivice bi moglo da se padne u ambis; u svetlu novih dokaza ljudi su postepeno promenili svoja verovanja.

---

<sup>52</sup> Pirs misli da preusmeravanje nečije pažnje s evidencije koja potvrđuje da je nešto slučaj nije dovoljan razlog da bismo govorili o obmanjivanju.

Uzmimo drugi primer. Zamislimo situaciju u kojoj osoba *A* zamoli osobu *B* da prenese nešto osobi *C*, za šta osoba *A* zna da nije slučaj, pri čemu osoba *B* nema razloga da pretpostavlja da osoba *A* ima želju ili nameru da slaže ili obmane osobu *C*, tj. ne sumnja u iskrenost onoga ko (u ovom slučaju *A*) saopštava. Pretpostavimo da osoba *B* prenese osobi *C* ono što joj je *A* rekla, zar onda ne bismo imali razloga da tvrdimo da je osoba *B* slagala ili obmanula osobu *C*. Ma koliko bilo primamljivo da se na ovo pitanje odgovori potvrdno, odgovor je ipak – odričan, jer strogo govoreći osoba *A* je ta koja laže ili obmanjuje, dok je *B* samo komunikacioni kanal između osoba *A* i *C*. Najviše što bismo mogli da kažemo jeste da je osoba *B* (nenamerno) dovela osobu *C* u zabludu.

Takođe, ako pretpostavimo da samoobmanjivanje nije intencionalno, kao što ja pretpostavljam, zanimljivo je to da osoba koja se samoobmanjuje ne može da slaže ili obmane drugu osobu u vezi samoobmanjujućeg verovanja koje je usvojila i koje saopštava drugoj osobi, ali može da je dovede u zabludu. Osoba koja se samoobmanjuje da *p* i iskreno tvrdi da *p*, pokušavajući da drugu osobu uveri da *p*, može samo da je dovede u zabludu u vezi verovanja da *p*, ali ne i da je slaže ili obmane, jer da bi je obmanula ili slagala neophodno je da to čini svesno i namerno, što je u ovom slučaju izgleda isključeno. Naravno, osobi kojoj se saopštava da *p* uvek ostaje otvorena mogućnost da izrazi sumnju u istinitost onoga što joj se saopštava, čak i onda kada se to čini iskreno.

## 5.2 Samoizazvano obmanjivanje i samoobmanjivanje

Samoobmanjivanje se takođe razlikuje od samoizazvanog obmanjivanja (McLaughlin 1988). Jedan od scenarija koji bi mogao da ilustruje samoizazvano obmanjivanje je sledeći. Zamislimo osobu koja je imala dugogodišnju burnu političku karijeru opterećenu aferama, kontroverzama, političkim podlostima, splotkama i grubim makijavelizmom u razračunavanju s političkim protivnicima. Nakon nekoliko meseci političkih previranja i pritisaka, dobro razmislivši odlučuje da je za nju bolje da se povuče s visoke funkcije koju je zauzimala i da se posveti pisanju memoara. Postepeno, na papiru, događaji iz njene političke prošlosti počinju da poprimaju drugačiji izgled. Neki su ulepšani, neki grubo falsifikovani, a neki izmišljeni. Kontroverze, afere, podlosti i splotke postaju nestašluci iz mladosti i posledica političkog neiskustva ili najobičnija podmetanja političkih neistomišljenika. Tokom pisanja ova osoba

postepeno sebe uvlači u obmanjujući zaplet koji je sama osmislila, sećanja se dopunjuju maštarijama, tako da se granica između onoga što se zaista dogodilo i onoga što je izmišljeno i ulepšano postepeno zamagljuje. Po završetku pisanja ona ima formirano verovanje čiji je sadržaj mešavina konfabulacija i zbilje koju smatra istinitom.

Navedeni primer pokazuje da osoba može da obmanjuje sebe na sličan način kao što bi obmanjivala nekog drugog, ali to zapravo ne bi bio pravi ili uobičajeni primer samoobmanjivanja. Drugi i verovatno najčuveniji primer samoizazvanog obmanjivanja je Dekartova metoda sumnja. Dekart je samoizazvano obmanjivanje učinio delom filozofske metodologije (Dekart 1998). Kartezijanski „projekat čistog istraživanja“ (Vilijams) započinje refleksijom o sebi, svetu i njihovom odnosu. Dekart najpre generalizuje svima dobro poznatu činjenicu nepouzdanosti čulnog opažanja u pojedinim situacijama, zatim uočava da se naši opažaji u trenutku kada sanjamo kvalitativno ne razlikuju od onih koje imamo u budnom stanju, tako da bi svet i objekti u njemu uključujući i nas same mogli da predstavljaju samo snoviđenje, da bi konačno došao do pretpostavke o zlom demonu - svemoćnom obmanjivaču – čijim se delovanjem čitava stvarnost transformiše u puku iluziju. Pretpostavka o zlom demonu sumnju radikalizuje na takav način da je za Dekarta svaki iskaz o realnosti ili lažan ili iskaz o pogrešnoj interpretaciji realnosti, jer je ono što se percipira kao realnost u stvari samo predstava indukovana delovanjem zlog demona. Zastrašen mogućnošću da ga zli demon obmanjuje i protivno zdravom razumu on odlučuje da ne veruje ono za šta ima više razloga da veruje. Dekart se dakle smišljeno i sopstvenom odlukom obmanjuje da ga najpre obmanjuju čula, zatim se obmanjuje da sanja iako je budan, i na kraju se obmanjuje da ga zli demon obmanjuje i da je strategija kojom to čini tako podmukla da je sve ono za šta on koji meditira veruje da je istina zapravo laž.

Kao što se laganje, obmanjivanje i dovođenje u zabludu međusobno razlikuju, tako se i samoobmanjivanje razlikuje od laganja, obmanjivanja i dovođenja u zabludu, iako je sličnije (samoizazvanom) dovođenju u zabludu, nego obmanjivanju, laganju ili samoizazvanom obmanjivanju.<sup>53</sup> Grubo govoreći, osoba se samoobmanjuje kada usvaja ili održava već usvojeno

---

<sup>53</sup> Džonston, na primer, smatra da je samoobmanjivanje slično dovođenju u zabludu, dok Dejvidson misli da je ono sličnije laganju i obmanjivanju. Vidi npr. (Davidson 1987, Johnston 1988).

motivisano verovanje da  $p$  koje je pogrešno, iako joj je dostupna evidencija da  $ne-p$ , koju na neki način ignoriše, zanemaruje, pristrasno tumači ili propušta da uvidi služeći se različitim strategijama koje omogućavaju samoobmanjivanje, ali nisu iskorišćene *da bi se* individua samoobmanjivala.

Prvi pokušaj naučnog objašnjenja fenomena samoobmanjivanja nalazimo u radovima Harolda Sackeima i Rubena Gura (Harold A. Sackeim; Ruben C. Gur). Prema eksplanatornom modelu koji su predložili ovi psiholozi, osoba se samoobmanjuje ako istovremeno ima dva protivrečna verovanja na različitim nivoima svesti, od kojih je jedno motivisano (Gur and Sackeim 1979). Oni su, u sada već klasičnom eksperimentu, ispitivali reakcije subjekata na samoprepoznavanje, polazeći od saznanja da ljudi reaguju emocionalno na audio-vizuelne predstave sebe ili drugih, i da se dešava da dolazi do pogrešnih identifikacija u zavisnosti od toga da li osoba ima pozitivnu ili negativnu sliku o sebi (Gur and Sackeim 1979). Na osnovu eksperimentalnih rezultata Gur i Sackeim su došli do zaključka da su bar neki slučajevi pogrešne identifikacije zapravo slučajevi samoobmanjivanja.<sup>54</sup>

Kao psihofiziološki indikator prepoznavanja uzeta je elektrodermalna reakcija. Gur i Sackeim su se usredsredili na pogrešne odgovore i uočili da je u situacijama kada ispitanici daju *lažno odričan* odgovor, tj. pogrešno identifikuju svoj glas kao tuđ, elektrodermalna reakcija povišena, dok je u onim situacijama kada su odgovori *lažno potvrdni*, tj. kada pogrešno identifikuju tuđ glas kao svoj, elektrodermalna reakcija je snižena. U zavisnosti od toga da li su njihovi odgovori u situacijama samoopažanja ispravni ili ne, i u zavisnosti od slike o sebi menjala se i elektrodermalna reakcija. Promena u elektrodermalnoj reakciji je za ove psihologe bila indikator da subjekti imaju protivrečna verovanja, a budući da je slika o sebi imala udela u ovim promenama, kao i da nisu bili svesni napravljenih grešaka, to je ukazivalo da su njihovi odgovori *motivisani*. Na osnovu dobijenih rezultata ovi autori su došli do zaključka da su nužni i dovoljni uslovi za samoobmanjivanje sledeći: (1.) da individua simultano ima dva protivrečna

---

<sup>54</sup> Slično, eksperimenti koje su izveli Kvatron i Tverski pokazuju da promene u toleranciji na hladnu vodu pod uticajem određenih verovanja nisu svesne što ovi autori tumače kao vid samoobmanjivanja (Quattrone and Tversky 1987). Tako, oni subjekti koji veruju da imaju loše srce duže drže ruku u hladnoj vodi ako veruju da je to indikator dobrog zdravlja. Obrnuto, oni koji veruju da je kraće držanje ruke u hladnoj vodi pokazatelj dobrog zdravlja kraće i drže ruku u hladnoj vodi.

verovanja; (2.) da nije svesna jednog od verovanja; i (3.) da je čin koji određuje koje je od verovanja predmet svesti motivisani čin.<sup>55</sup>

### 5.3 Paradoksi samoobmanjivanja

Interpersonalno obmanjivanje predstavlja proces u kome jedna osoba svesno, intencionalno i uspešno navodi drugu da veruje nešto za šta prva zna ili opravdano veruje da nije slučaj. Tumačenje samoobmanjivanja po modelu obmanjivanja na prvi pogled deluje primamljivo i prihvatljivo. Ako se oslonimo na etimologiju, onda samoobmanjivanje nije ništa drugo do obmanjivanje samog sebe, tako da se čini opravdanim da se taj fenomen tumači po modelu interpersonalnog obmanjivanja. Međutim, tumačenje samoobmanjivanja kao analognog interpersonalnom obmanjivanju dovodi do paradoksa koji se odnose (1) na sam čin samoobmanjivanja (dinamički ili strateški paradoks) i (2) na doksatičko stanje do koga dovodi uspešan čin samoobmanjivanja (statički ili doksatički paradoks).

#### 5.3.1 *Dinamički ili strateški paradoks*

Kada je reč o obmanjivanju, onda na umu imamo najmanje dve osobe  $A$  i  $B$  od kojih jedna  $A$  obmanjuje drugu  $B$  u vezi istinitosti verovanja  $p$ , pri čemu se pretpostavlja da osoba  $A$  zna istinosnu vrednost verovanja  $p$ , ali navodi drugu osobu  $B$  da veruje suprotno od onoga u šta sama veruje, dok osoba  $B$  ili uopšte nema formirano verovanje da  $p$ , tako da  $A$  nastoji da kod  $B$  indukuje verovanje da  $p$ , ili ako  $B$  ima formirano verovanje da  $p$ , onda je nesigurna o istinosnoj vrednosti tog verovanja. Nakon obmanjivanja  $B$  veruje ono za šta  $A$  tvrdi da je slučaj. Međutim, u slučaju samoobmanjivanja obmanjivač i obmanuti su jedna ista osoba tako da je nejasno kako je samoobmanjivanje uopšte moguće. Samoobmanjivanje je individualni čin, tako da se čini nemogućim da  $A$  obmanjuje  $B$ , ako je  $A$  u stvari  $B$ , jer u slučaju obmanjivanja, da bi  $A$  uspešno obmanuo  $B$ ,  $B$  ne bi trebalo ništa da zna o namerama, planovima i strategijama  $A$ . Ali, u slučaju samoobmanjivanja gde su obmanjivač i obmanuti identični, čini se da osoba koja se obmanjuje mora da zna sopstvene namere i planove i da bude svesna strategija kojima se služi. Ukratko, ako samoobmanjivanje tumačimo po modelu obmanjivanja, onda bi osoba morala da obmanjuju sebe

---

<sup>55</sup> Ovi autori navode četiri uslova, ali sam drugi, uslov simultanosti protivrečnih verovanja radi jednostavnosti podveo pod prvi.

znajući da se obmanjuje. To znači da bi osoba *A* svesno, namerno i uspešno morala da navodi sebe da pogrešno veruje da *p*.

### 5.3.2 Statički (doksatički) paradoks

Ako samoobmanjivanje tumačimo po analogiji s interpersonalnim obmanjivanjem, onda se statički ili doksatički paradoks pojavljuje kao posledica uspešnog samoobmanjivanja. Pod pretpostavkom da su samoobmanjivanje i interpersonalno obmanjivanje analogni, posledica uspešnog samoobmanjivanja je da individua istovremeno veruje da je *p* istinito i da je *ne-p* istinito, odnosno ima dva protivrečna verovanja, što je neprihvatljivo. To je takođe i razlog zašto se na samu mogućnost samoobmanjivanja gleda s izvesnim stepenom skepticizma.

## 5.4 Intencionalističko i divizionističko objašnjenje samoobmanjivanja

Ako se za trenutak vratimo uslovima za samoobmanjivanje koje su predložili Gur i Sakeim, onda možemo da uočimo da prvi uslov konstituiše paradoks, dok bi drugi trebalo da ga razrešava. Prvi uslov daje kriterijum dvostrukog verovanja. Osoba istovremeno ima verovanje da *p* i verovanje da *ne-p*, što zatim vodi zaključku da ima verovanje (*p* i *ne-p*). Ali, ovakav zaključak bi bio preuranjen. Iako tvrde da individua istovremeno ima verovanje da *p* i verovanje da *ne-p*, ovi autori nigde ne navode da osoba koja se samoobmanjuje u nekom trenutku *t* ima verovanje koje se odnosi na konjunkciju verovanja *p* i *ne-p*. Ako osoba nije svesna jednog od verovanja, onda simultano posedovanje protivrečnih verovanja ne konstituiše paradoksalno stanje svesti.

Da bi se izbegao paradoks bilo je dakle neophodno postulirati nezavisne i izolovane delove svesti koji kauzalno mogu da utiču jedan na drugi i da na neki način utiču na ponašanje. Sakeim i Gur smatraju da je svest *funkcionalno* podeljena na svesno i nesvesno i to tako što je nesvesno shvaćeno kao nezavisan kontrolni sistem koji može da ima kontrolu nad svesnim i da *svrhovito* utiče na ponašanje (Gur and Sackeim 1979). Oni tumače samoobmanjivanje po analogiji s interpersonalnim obmanjivanjem, tako da je samoobmanjivanje intencionalan čin motivisan interesima ega, a svest je raščlanjena tako da jedan njen deo može da utiče na drugi. U ovakvim ekspanatornim modelima individua može simultano da ima protivrečna verovanja ako su ona mentalno ili temporalno izolovana, ili ako ta verovanja pripadaju različitim sferama

života. Na primer, deca koja iskreno veruju u postojanje Deda Mraza, ipak od roditelja očekuju poklone za novogodišnje praznike (Elster 1987).

#### **5.4.1 Mentalno raščlanjivanje**

Stupanj mentalnog raščlanjivanja može da varira od strogog, homunkularističkog pristupa, gde se delovi svesti tumače kao nezavisni subjekti sposobni za odlučivanje (Rorty 1987, 1988), do umerenog, gde su delovi shvaćeni samo kao funkcionalno nezavisni i labavo integrisani podsistemi (Davidson 1982, 1987, Pears 1987). Da bi samoobmanjivanje bilo uopšte moguće, takođe se pretpostavlja da je odnos među različitim delovima ili funkcionalno nezavisnih podsistema, asimetričan na takav način da obmanuti ne može povratno da utiče na deo koji obmanjuje, jer deo koji je obmanut ignoriše postojanje obmanjivača (Elster 1987). Nemoguće je imati dva podsistema koji se uzajamno obmanjuju.

Rorti (Amelie O. Rorty) misli da su uobičajeni psihički procesi i oni koji vode samoobmanjivanju identični, a da se razlika odnosi samo na ono na šta su oni usmereni (Rorty 1987).<sup>56</sup> Dok je u jednom slučaju posledica njihovog delovanja istina ili, preciznije, adekvatna slika realnosti u granicama evoluiranih kognitivnih kapaciteta, u drugim slučajevima posledica njihovog rada je pristrasno izobličavanje stvarnosti. Ukoliko um posmatramo kao jedinstven fenomen, onda se čini da nemamo izbora nego da samoobmanjivanju pristupimo s dozom skepticizma, jer jedinstveno, racionalno, refleksivno, transparentno i strogo integrisano sopstvo, tj. um shvaćen kao „panoptički skener“, nije sposoban za samoobmanjivanje, jer ne može da poseduje protivrečna verovanja koja su izolovana, a ako se protivrečnosti i pojave, one lako mogu da budu otklonjene (Rorty 1988). U ovom eksplanatornom modelu samoobmanjivanje je shvaćeno kao intencionalno, a um je opisan kao raščlanjen na neodređeni broj podsistema, ali tako da racionalnost nije samo jedan u nizu jednakih podsistema, već zauzima centralno i dominantno mesto, što je čini važnim korektivom samoobmanjivanja i drugih oblika iracionalnosti (Rorty 1988). Zbog specifičnog razumevanja intencionalnosti koja se proteže od preintencionalnih perceptivnih razlikovanja do samosvesnih i kritički evaluiranih

---

<sup>56</sup> Taj stav deli i Fingaret, prema čijem mišljenju samoobmanjivanje nastaje kao posledica obrazaca usmeravanja pažnje (selektivna pažnja) koji omogućavaju subjektu da u svakodnevnom kontekstu funkcioniše bez ometanja. Zaključak do koga dolazi je da samoobmanjivanju nije potrebno neko posebno objašnjenje (Fingarette 1998).



propozicionalnih verovanja, samoobmanjivanje može da se tumači kao intencionalno, ali ne i motivisano, jer može da bude posledica navika ili usmeravanja pažnje (Rorty 1987, 1988). Kao što na preintencionalnom nerefleksivnom nivou osoba može da razlikuje toplo od hladnog kada dodiruje neki predmet, ali ne mora da ima propozicionalno znanje da je ovo razlikovanje uzrokovano kretanjem molekula, subjekt može da usvoji ili održava pogrešno verovanje uprkos evidenciji za suprotno, a da nije svestan uzročno-posledičnog odnosa među različitim delovima svesti niti toga da je njegovo verovanje posledica tog delovanja. Iako to zvuči prilično nategnuto, samoobmanjivanje je intencionalno mada nije svesno.

Pirs i Dejvidson predlažu nešto umerenije divizionističke teorije zasnovane na kriterijumu funkcionalne izolovanosti – um je podeljen na funkcionalno autonomne strukture sposobne da jednosmerno deluju jedna na drugu. (Davidson 1987, Pears 1987). Dok u Pirsovom modelu međudelovanje različitih podsistema posreduje altruistička želja (jedan od podsistema čini nešto što je u kratkoročnom interesu drugog podsistema ili individue kao celine), što stanje samoobmane čini na neki način subjektivno poželjnim, za Dejvidsona je pre reč o delimično slepom kauzalnom procesu čiji konačni rezultat ne mora nužno da bude prijatno stanje svesti (Davidson 1987). Pirs razlikuje dva podsistema svesti. Jedan je aktivni, ali efemerni i instantni proizvod želje nezainteresovan za istinitost verovanja, sličan psihoanalitičkom konceptu predsvesnog, dok je drugi (manje više) pasivni sistem orijentisan ka istini (Pears 1987).<sup>57</sup> Ako je istina neprijatna, onda aktivni podsistem usmerava objekt svog delovanja u drugom pravcu (Pears 1987). Ako podsisteme domišljato obeležimo sa *A* i *B*, onda samoobmanjivanje funkcioniše na sledeći način. *A* želi da *B* veruje da *p*. *B* želi da je *p* istinito i da *B* veruje da *p*. S druge strane, *A* želi da *B* veruje da *p* nezavisno od njegove istinosne vrednosti, što samoobmanjivanje čini mogućim. Želja *B* da veruje da *p* isključivo kada je *p* istinito u pojedinim situacijama je neispunjiva i suspendovana (Pears 1987). Iako, prema ovom stanovištu, osoba može simultano da ima protivrečna verovanja pod pretpostavkom funkcionalne raščlanjenosti uma, to ipak nije nužan uslov za samoobmanjivanje. Na sličan način kao što prilikom interpersonalnog obmanjivanja osoba može da bude navedena da veruje da je nešto slučaj, a da prethodno nema formirano verovanje čiji je sadržaj negacija tog verovanja, ona može i da se

---

<sup>57</sup> Predsvesno predstavlja područje uma čijeg sadržaja osoba trenutno nije svesna, ali koga bi uz određen napor mogla da postane svesna.

samoobmanjuje, usvajajući verovanje da  $p$ , a da prethodno nema formirano verovanje da  $ne-p$  (Pears 1987).

Dejvidson samoobmanjivanje objašnjava postojanjem razgraničenih mentalnih struktura koje su sposobne da kauzalno deluju jedna na drugu, tako da jedno mentalno stanje može biti uzrok drugog. Ovo razgraničenje koje uzrok i posledicu na neki način izoluju omogućava koegzistenciju protivrečnih verovanja (Davidson 1982). Tako, iako je moguće verovati da  $p$  i  $ne-p$  istovremeno, ipak nije moguće verovati ( $p$  i  $ne-p$ ), što znači da  $p$  i  $ne-p$  moraju da pripadaju različitim područjima svesti, ali budući da su ova područja podložna međusobnom uticaju, Dejvidson tvrdi da je reč o razdvojenim, a ipak preklapajućim područjima, a sličnog mišljenja je i Pirs (Davidson 1987, Pears 1987). U ovom kontekstu delove uma ne treba razumeti kao stvarne entitete, jer funkcionalna raščlanjenost predstavlja samo pomoćni konceptualni aparat koji omogućava da se fenomen samoobmanjivanja objasni (Davidson 1982). Ipak, ovi modeli na neki način žrtvuju celovitost uma, kako bi sačuvali fenomen samoobmanjivanja.

#### **5.4.2 Vremensko raščlanjivanje**

Drugi način da se izbegnu paradoksi samoobmanjivanja, a da se ipak sačuva sam fenomen, jeste taj da se samoobmanjivanje tumači kao događaj u vremenu. Prema zastupnicima ovog eksplanatornog modela samoobmanjivanja, osoba može intencionalno da se dovede u stanje samoobmane, da ima protivrečna verovanja, ali ona moraju da budu dovoljno vremenski udaljena, tako da ne dolaze u konflikt. S druge strane, s obzirom na protok vremena moguće je proizvesti više privremenih sopstava, tako da samoobmanjivanje može da bude protumačeno kao interpersonalno obmanjivanje samo u kontekstu jedne osobe – privremeno sadašnje sopstvo može da obmanjuje potencijalno buduće sopstvo (Baier 1996). Ako je samoobmanjivanje, jednako kao i interpersonalno obmanjivanje, događaj u vremenu i ne nastupa odjednom kao napad „epistemičke kijavice“ (Rorti), već je reč o dugotrajnom postepenom procesu, onda individua na kraju može potpuno da izgubi kontakt s početnim intencijama (motivima), ali da njen čin ipak ostane intencionalan (Sorensen 1985, Bermúdez 2000).

Bermudez smatra da je samoobmanjivanje intencionalno, ali da intencije ne podležu principu deduktivne zatvorenosti, tako da subjekt može da formira intenciju da usvoji verovanje

da  $p$ , da zna da  $p$  povlači  $q$  koje je pogrešno, ali da ne formira intenciju da usvoji pogrešno verovanje i kao obrazloženje navodi to da bi subjekt u protivčinjeničkoj situaciji u kojoj  $p$  ne povlači  $q$  i dalje nastojao da veruje da  $p$  (Bermúdez 2000). Osim toga, da bi samoobmanjivanje bilo intencionalno i bez paradoksa, dovoljno je da verovanja  $p$  i  $q$  budu dovoljno vremenski udaljena. U skladu s ovim napomenama samoobmanjivanje će biti moguće i neparadoksalno kao događaj u vremenu kada:

- i. postoji trenutak  $t_1$  u kome  $S$  veruje da  $p$ ;
- ii. postoji vremenski interval  $t_1 - t_2$  u kome  $S$  zaboravi da  $p$ ;
- iii. postoji trenutak  $t_2$  u kome  $S$  veruje da  $ne-p$ ;
- iv. ne postoji trenutak  $t_n$  ili vremenski interval  $t_i$  u kome  $S$  veruje da  $(p \text{ i } ne-p)$ .

Kao primer modela vremenskog raščlanjivanja u objašnjenju samoobmanjivanja često se navodi sledeće. Pretpostavimo da neko želi da izbegne neprijatan sastanak s nekom osobom, ali iz nekog razloga to ne može da učini jednostavno odbijajući da se s njom sretne. Tako, da bi taj sastanak izbegao on strateški u svoj rokovnik zapiše pogrešan datum sastanka, svestan da će od trenutka zapisivanja do trenutka njegove potencijalne realizacije da zaboravi pravi datum dogovorenog sastanka i činjenicu da je u rokovnik zapisao pogrešan datum. Međutim, temporalno raščlanjivanje teško da može da objasni uobičajene slučajeve samoobmanjivanja koji ne eksploatišu mehanizme pamćenja i zaboravljanja (Johnston 1988). Osim toga, ako je protok vremena važan činilac samoobmanjivanja, onda ne bismo mogli da objasnimo samoobmanjivanje u kraćim vremenskim intervalima, jer ima tako kratkih vremenskih intervala da je veoma teško zamisliti da je u njima moguće interpersonalno obmanjivanje, kao što ima jednako kratkih vremenskih intervala u kojima je teško zamislivo da bi osoba mogla da se samoobmanjuje služeći se navedenom strategijom (McLaughlin 1988). Konačno, primer sa izbegavanjem sastanka je pre slučaj samoizazvanog obmanjivanja, gde je sopstvo podeljeno na vremenske intervale tako da sadašnje sopstvo na neki volšeban način obmanjuje buduće, bez ugrožavanja njegove mentalne celovitosti (McLaughlin 1988, McLaughlin 1996).

## 5.5 Neintencionalističko i nedivizionističko tumačenje samoobmanjivanja

Da bi se na adekvatan način objasnio fenomen samoobmanjivanja, kao i da bi se izbegle zamke statičkog i dinamičkog paradoksa neophodno je da određeni uslovi, koje u objašnjenju samoobmanjivanja predlažu intencionalisti, budu oslabljeni ili odbačeni. Protiv intencionalističkog razumevanja samoobmanjivanja se tvrdi da postojanje intencije za formiranje verovanja predstavlja prepreku za njegovo formiranje, a intencionalisti ne mogu da objasne zbog čega subjekt nije svestan ove intencije (Lazar 1999). Na prvom mestu, dakle, treba odbaciti uslov intencionalnosti. Na drugom mestu, neintencionalisti odbacuju uslov dvostrukog verovanja, što povlači i to da nije neophodno postulirati nezavisne delove svesti u kojima su simultana protivrečna verovanja pohranjena.

Prema deflacionim modelima objašnjenja fenomena samoobmanjivanja, dovoljno je da osoba koja se samoobmanjuje želi ili da je podstaknuta određenom emocijom da usvoji verovanje koje je pogrešno. Želja (ili anksiozna želja) da  $p$ , motiviše  $S$  da usvoji pogrešno verovanja da  $p$ , uprkos evidenciji za  $ne-p$  (Johnston 1988, Audi 1997, Mele 2001, Barnes 2007). Prema ovim modelima, formiranje (i održanje) verovanja nije čin koji je izolovan od različitih afektivnih stanja i želja koje subjekt ima. Emocije i želje utiču na formiranje i održanje verovanja, što znači da je jedan broj subjektivih verovanja, koji ne mora da bude mali, formiran na epistemički nepouzdan način (Lazar 1999).

### 5.5.1 Samoobmanjivanje i psihička tenzija

Prema jednom broju autora, osnovna funkcija samoobmanjivanja je umanjenje anksioznosti (Barnes 2007, Johnston 1988), dok drugi posebno naglašavaju stanje psihičke tenzije u kojoj se osoba koja se samoobmanjuje nalazi i koja čini stanje samoobmane nestabilnim (Audi 1997, Bach 1981). Prema jednom stanovištu, samoobmanjivanje je posledica delovanja subintencionalnih procesa, mentalnih tropizama, koji nisu ništa drugo do kauzalni odnosi između određenih afektivnih stanja i verovanja (Johnston 1988). O mentalnim tropizmima Džonston još govori kao o svrhovitim mentalnim pravilnostima koje uzrokuju događaje u umu: pojavljivanje određene anksiozne želje rezultira pojavljivanjem određenog verovanja, što dalje vodi umanjenju

anksioznosti. Slično misli i Barns (Annette Barnes). Međutim, objašnjenja koja samoobmanjivanje dovode u vezu s anksioznošću i njegovim umanjnjem ne mogu da objasne one slučajeve u kojima osoba ne usvaja poželjna, već nepoželjna verovanja koja uvećavaju anksioznost (Scott-Kakures 2000, Mele 2001). Osim toga, ima slučajeva kada se individua ne oslanja na samoobmanjivanje iako bi mu ono pomoglo da umanjiti anksioznost.

Uzmimo sledeći primer. Pretpostavimo da se osoba  $A$  uputila s porodicom automobilom na more. Negde na pola puta  $A$  počinje da sumnja da kočnice automobila otkazuju, jer kako joj se čini, kada god pritisne kočnicu, papučica se sve više i više približava podu automobila. Iako nije potpuno sigurna da kočnice ne rade valjano,  $A$  formira verovanje da kočnice automobila nisu ispravne što za posledicu ima anksioznost. Usvajanje verovanja da su kočnice ispravne bi umanjilo anksioznost, ali  $A$  radije bira da se na prvoj benzinskoj stanici zaustavi i proveri ispravnost vozila (Talbot 1995). Ukoliko bi usvojila verovanje da su kočnice ispravne, iako nisu,  $A$  bi se samoobmanjivala i umanjila anksioznost, ali bi to moglo da ima pogubne posledice. S druge strane, ukoliko se nakon provere ispravnosti automobila ustanovi da su kočnice zapravo ispravne, to bi značilo da se  $A$  samoobmanjivala, odnosno da su njene sumnje bile neopravdane i da je bila usvojila nepoželjno verovanje koje je pogrešno.

S druge strane, samoobmanjivanje se može sagledati u kontekstu asimetrije između mišljenja kao mentalnog događaja i verovanja kao mentalnih stanja (Bach 1981). U ovom kontekstu samoobmanjivanje je u relaciji s mišljenjem, ali ne i verovanjima. Individua može imati brojna verovanja o kojima ne misli ili o kojima nikada nije mislila, a može imati i različite misli koje nisu propozicionalno izražena verovanja. Pretpostavimo da subjekt zna sve relevantne podatke o džinovskim mravojedima i slonovima. Međutim, dok u mislima ne dovede u korelaciju mravojede i slonove on neće imati nikakvo znanje (ili opravdano verovanje) o tome da su mravojedi, pa i oni džinovski, manji od slonova; verovanje može, ali i ne mora imati bilo kakvu manifestaciju u mišljenju (Bach 1981). Kada se samoobmanjuje, osoba veruje da  $p$ , intenzivno istrajava u tome da ne misli o  $p$  i nikada ne mora da formira verovanje da  $ne-p$ , iako želi da  $ne-p$  bude slučaj (Bach 1981). Za samoobmanjivanje je dovoljno da osoba samo ne misli na ono što veruje, ali to proizvodi neprijatnost. Odnos želje (da  $ne-p$ ) i verovanja (da  $p$ ) koje subjekt ima ga motiviše da izbegava da misli na ono što veruje (Bach 1981). Mada je mehanizam koji vodi samoobmanjivanju intencionalan, samoobmanjivanje je neintencionalno, jer ovaj

mehanizam nije korišćen da bi se subjekt samoobmanjivao (Bach 1981). Kada je reč o ovom modelu samoobmanjivanja, ostaje nejasno zašto osoba nikada ne napravi korak od izbegavanja da misli da  $p$ , ka tome da veruje da  $ne-p$ , jer, kako se čini, najjednostavniji i kognitivno najekonomičniji način da osoba izbegava da misli o  $p$ , jeste da konačno formira verovanje da  $ne-p$ .

S druge strane, samoobmanjivanje se može sagledati i iz perspektive postojanja diskrepancije između onoga što osoba veruje i iskreno saopštava (Audi 1997). Prema ovom modelu, osoba se samoobmanjuje u odnosu na propoziciju  $p$  ako i samo ako (1) nesvesno zna da  $ne-p$ ; (2) iskreno tvrdi da  $p$ ; (3) ima bar jednu želju koja objašnjava zašto nesvesno veruje da  $ne-p$  i zašto tvrdi  $p$  uprkos dokazima protiv  $p$ . Pri tome, treba zapaziti da osoba koja se samoobmanjuje u vezi  $p$ , nema formirano verovanje da  $p$ , iako iskreno saopštava da  $p$ . Time je isključena mogućnost da osoba ima simultana protivrečna verovanja. Jedino verovanje koje ima je verovanje da  $ne-p$  koje je od nje skriveno. Međutim, to je prilično neobično, jer osoba koja tvrdi da  $p$  obično veruje da je  $p$  istinito ili lažno, odnosno podstaknuta je da tvrdi ono što veruje. Konačno, najjednostavniji i kognitivno najisplativiji način da se osoba samoobmanjuje je da konačno usvoji ono što iskreno tvrdi. Izgleda da navedeni autori oklevaju da načine ovaj korak kako se ne bi izgubila psihička tenzija karakteristična za samoobmanjivanje.

### **5.5.2 Samoobmanjivanje i pragmatičko testiranje hipoteza**

Objašnjenja koja na zadovoljavajući način odgovaraju na pitanje kako individue u svakodnevnom kontekstu dolaze bilo do poželjnih, bilo do nepoželjnih (pogrešnih) verovanja nalazimo u oblasti socijalne psihologije. Ova objašnjenja se oslanjaju na proučavanje toga kako motivacija utiče na usvajanje verovanja i koji psihički procesi u ovome učestvuju, odnosno na koji način osoba koja usvaja ili nastoji da održi određeno verovanje procenjuje da li i kada to da učini ili ne. Verovatno najzanimljivija i najpotpunija objašnjenja samoobmanjivanja oslanjaju se upravo na ova istraživanja. Ali, da bismo uopšte mogli da razumemo ove teorijske modele samoobmanjivanja, pažnju najpre moramo da preusmerimo na psihološka istraživanja koja objašnjavaju odnos kognicije i motivacije.

Nezavisno od toga da li donose ishitrene zaključke ili pažljivo rasuđuju, da li verovanja usvajaju u skladu sa željama ili su orijentisane ka istini, sve individue su podložne *konfirmacijskoj pristrasnosti*, tako da proces i rezultati rasuđivanja neće uvek biti i ne moraju nužno da budu epistemički pouzdani (Kunda 1990). *Konfirmacijska pristrasnost* znači da ljudi pokazuju tendenciju da traže, lakše se sećaju i obraćaju više pažnje na evidenciju i informacije koje potvrđuju njihova verovanja, dok potcenjuju, ignorišu, zanemaruju i odbacuju evidenciju i informacije koje ih opovrgavaju. Oni takođe mogu da na pristrasan način biraju pravila zaključivanja, tako da potvrđuju ono što žele, kao i da usvajaju nova verovanja koja su konzistentna sa inicijalnim skupom verovanja koji je možda formiran na epistemički nepouzdan način. Šta više, konfirmacijska pristrasnost može da bude pojačana individualnom orijentacijom ka usvajanju istinitih verovanja. Kako se verovanja usvajaju s obzirom na subjektivnu procenu koristi i štete, pristrasnost će biti veća u slučaju kada potencijalno pogrešno odbacivanje željenog verovanja kao neistinitog može da ima visoku subjektivnu cenu. Tako, ako individua želi da veruje da  $p$  i uz to želi da  $p$  bude istinito, čemu pridaje visoku subjektivnu vrednost, onda će ova kombinacija želje i orijentacije ka istini rezultirati intenzivnom pristrasnošću (Kunda 1990).

Konzistentan s prethodnim razmatranjima je *model pragmatičkog testiranja hipoteza* prema kome je testiranje hipoteza motivisano cenom inferencijalnih grešaka koje se odnose na pogrešno prihvatanje ili odbacivanje hipoteze u odnosu prema trošku za pribavljanje informacija (Trope and Liberman 1996, Trope, Gervy, and Liberman 1997). Budući da željena i neželjena hipoteza obično proizvode asimetriju u tome u kojoj meri se greške nepovoljno odražavaju na subjekta, on će biti prirodno podstaknut da izbegava one greške koje imaju najvišu subjektivnu cenu (Trope, Gervy, and Liberman 1997, Trope and Liberman 1996). Prema ovom modelu koji se fokusira na izbegavanje grešaka, subjekt nije toliko zainteresovan za istinu, koliko je zainteresovan da izbegne greške u rešavanju određenih adaptivnih problema, jer, kako se čini, evolucija favorizuje one jedinke koje najmanje greše, a ne one koje troše vreme i energiju na falsifikaciju (Friedrich 1993). Na primer, u evolucionom smislu za jedinku može biti probitačnije da izbegava da pogreši kada je reč o ulaganju u srodnike i da iz tog razloga ponekada ulaže i u nesrodnike, nego da izbegava da greši kada je reč u ulaganju u nesrodnike i da iz tog razloga propušta da ulaže u neke jedinke s kojima je u srodstvu.

Kada je reč o pragmatičkom testiranju hipoteza, treba imati na umu da motivacija samo inicira postavljanje određene vrste pitanja, ali ne vodi dalje rasuđivanje, već se ono odvija uz pomoć uobičajenih kognitivnih mehanizama (Kunda 1990, Friedrich 1993). Individua želi da  $p$  i da je  $p$  istinito, tako da postavlja pitanje „Da li  $p$ ?” ili „Da li je  $p$  istinito?”, što dalje vodi pristrasnom pretraživanju memorije i prikupljanju podataka koji potvrđuju željenu hipotezu, ali se sam proces testiranja istinitosti hipoteze ne razlikuje od uobičajenog nepristrasnog testiranja (Kunda 1990). Na taj način, osoba koja se samoobmanjuje samo postavlja pitanje koje je motivisano, a zatim se oslanja na uobičajene strategije (metode) za testiranje hipoteza. Ove strategije ne moraju da budu promišljene i svesne već pre automatske i nefleksibilne (Friedrich 1993).

U procesu prihvatanja ili odbacivanja hipoteza, subjekt može na dva načina da napravi ispravan izbor, kao što može da napravi i dve greške. Najpre, on može da usvoji hipotezu koja je ispravna ili da odbaci hipotezu koja je pogrešna. Ali, on može i da pogreši, to jest da prihvati (usvoji) hipotezu kada je pogrešna ili da je odbaci kada je ispravna. Svaka od mogućih grešaka ima pozitivnu ili negativnu subjektivnu vrednost, koja je izražena potencijalnim gubicima koji su posledica pogrešnog prihvatanja, odnosno odbacivanja hipoteze (Trope and Liberman 1996). Na osnovu procene koristi i štete subjekt postavlja dva praga pouzdanosti (eng. *confidence threshold*), jedan za prihvatanje i drugi za odbacivanje hipoteze. U onim situacijama kada pogrešno prihvatanje hipoteze ima visoku subjektivnu cenu, subjekt će prag prihvatljivosti postaviti visoko, što znači da će pre nego što prihvati hipotezu skup evidencije koji uzima u razmatranje biti veći i da će testiranje biti detaljnije. Subjektu će, s druge strane, biti potrebna mala količina opovrgavajućih informacija da bi hipotezu odbacio. Kada je obrnut slučaj, to jest kada je cena pogrešnog odbacivanja hipoteze visoka, subjekt će postaviti prag za odbacivanje hipoteze visoko, što znači da će količina opovrgavajuće evidencije morati da bude znatno veća od količine informacija neophodnih da bi se hipoteza prihvatila (Trope and Liberman 1996, Trope, Gervy, and Liberman 1997).

U odnosu na subjektivnu procenu koristi i štete, individua će nastojati da izbegne najskuplje greške koje mogu negativno da se odraze na njenu adaptivnu vrednost (Friedrich 1993). Da bi se hipoteza prihvatila kada je prag prihvatljivosti postavljen visoko, ona mora da bude veoma verovatna, dok će u istom kontekstu morati da bude samo umereno neverovatna da



bi bila odbačena, odnosno neprihvaćena. To znači da se traže nužni uslovi za njeno prihvatanje, a samo dovoljni za odbacivanje. S druge strane, ako je prag za odbacivanje postavljen visoko, biće neophodno da hipoteza bude veoma neverovatna da bi bila odbačena, ali samo umereno verovatna da bi bila prihvaćena. U ovom slučaju traže se nužni uslovi za odbacivanje hipoteze, a samo dovoljni za njeno prihvatanje. Ovakva vrsta asimetrije vodi pristrasnosti u odnosu na prihvatanje ili odbacivanje hipoteza (Trope and Liberman 1996, Trope, Gervy, and Liberman 1997). Asimetrično postavljanje praga pouzdanosti služi kao kriterijum odlučivanja za prekidanje testiranja hipoteza, koji sa svoje strane može da motiviše pristrasno prikupljanje informacija i ranije prihvatanje evidencije koja naizgled potkrepljuje željene zaključke (Trope and Liberman 1996, Trope, Gervy, and Liberman 1997).

Ljudi radije biraju poželjne nego nepoželjne hipoteze kao polaznu tačku testiranja, jer je o njima prijatno razmišljati i lako su kognitivno dostupne (Kunda 1990). Konačno, poželjne hipoteze se čine *a priori* verovatnijim, jer osobe ulažu individualni napor da ih ostvare, što uz konfirmacijsku pristrasnost vodi njihovom potvrđivanju (Kunda 1990, Trope and Liberman 1996, Trope, Gervy, and Liberman 1997). Kada se evidencija crpi iz prethodnog iskustva, odnosno memorije, onda je ona koja ide u prilog poželjnoj hipotezi istaknutija i dostupnija nego ona koja je opovrgava. Takođe, željenoj hipotezi se obično poklanja visok stepen inicijalnog poverenja, evidencija se ravna s hipotezom i pitanja se postavljaju tako da se dobijaju odgovori koji su konzistentni s testiranom hipotezom, što dalje vodi njenom potvrđivanju (Trope and Liberman 1996).

Osnovna pretpostavka opisanog modela je da ljudi prilikom testiranja hipoteza nisu toliko orijentisani ka istini, već pre ka izbegavanju toga da naprave krupne greške u svakodnevnom kontekstu. Ne radi se o tome da postoji neki poseban kognitivni mehanizam koji vodi samoobmanjivanju, već uobičajeni kognitivni mehanizmi koji procesiraju informacije vode ne samo istinitim verovanjima, već i verovanjima koja su nepouzdana. Ako ove kognitivne mehanizme nazovemo strategijama, onda bismo mogli da kažemo sledeće: da bi samoobmanjivanje bilo intencionalno bilo bi neophodno da je subjekt svestan ne samo strategija koje koristi, već pre svega činjenice da su te strategije namenjene samoobmanjivanju. Tek kada bi subjekt bio svestan ove činjenice, mogli bismo da kažemo da je samoobmanjivanje intencionalno.

Verovatno najuticajnija neintencionalistička teorija samoobmanjivanja je ona koju je predložio Mel (Alfred Mele), a koja se oslanja na prethodne psihološke analize. Individua se samoobmanjuje tako što se služi različitim strategijama koje se odnose na „manipulaciju“ dostupnim podacima koji idu u prilog ili protiv verovanja koje usvaja ili koje je usvojila, a za koje je vezana na neki afektivni način. Ove strategije su strukturirane tako da individua dostupne podatke tumači na motivaciono pristrasan način –  $A$  želi da  $p$  bude istinito, ili pak  $A$  želi da  $p$  bude pogrešno – i u skladu s tim interpretira i prikuplja podatke koji bi potkrepili, odnosno opovrgli određeno verovanje. Strategije o kojima je reč su sledeće:

- 1) *negativna pogrešna interpretacija* – želja da  $p$  navodi individuu da opovrgavajuću evidenciju protiv  $p$  ne priznaje kao opovrgavajuću, iako bi je tumačila kao opovrgavajuću da te želje nema.;
- 2) *pozitivna pogrešna interpretacija* – želja da  $p$  navodi individuu da kao potkrepljujuću evidenciju tumači onu koju bi, da te želje nema, tumačila kao opovrgavajuću;
- 3) *selektivno obraćanje pažnje* – želja da  $p$  dovodi do toga da  $A$  ne usmerava pažnju na evidenciju protiv  $p$  i da se fokusira na onu koje ide u prilog  $p$ ;
- 4) *selektivno prikupljanje podataka* – želja da  $p$  dovodi do toga da  $A$  previđa očiglednu evidenciju za  $ne-p$  i nastoji da pronalazi manje pristupačnu evidenciju za  $p$  (Mele 2001, 25-27).

Mehanizam navedenih strategija obezbeđuje da se osoba samoobmanjuje, ali da se izbegnu paradoksi samoobmanjivanja, simultano postojanje protivrečnih verovanja i divizionizam. Mel predlaže četiri uslova koja su, prema njegovom mišljenju, dovoljna da bi moglo da se kaže da se osoba samoobmanjuje kada usvaja određeno verovanje.

1. Verovanje da  $p$  koje  $A$  usvaja je pogrešno;
2.  $A$  tumači relevantne ili bar naizgled relevantne podatke za istinosnu vrednost da  $p$  na motivaciono pristrasan način;
3. Ovaj pristrasan način je nedevidijantan uzrok  $A$ -ovog prihvatanja verovanja da  $p$ ;
4. Skup podataka koje  $A$  poseduje u vreme kada usvaja verovanje pružaju više osnova da  $ne-p$ , nego da  $p$  (Mele 2001, 51-52).

U svojim ranijim radovima on se uglavnom pridržavao toga da su ovi uslovi dovoljni za samoobmanjivanje (Mele 1987, 2001), ali kako se oni odnose samo na sam proces samoobmanjivanja i ne impliciraju postojanje psihičke tenzije kod osobe koja se samoobmanjuje (tj. ne ukazuju na njeno psihičko stanje), što je za pojedine autore važna karakteristika samoobmanjivanja, Mel uvodi peti uslov:

5. *A* svesno veruje da postoje velike šanse da *ne-p* (Mele 2012, 11).

Može se zapaziti da za razliku od Audiya, Mel ne insistira na tome da je verovanje da *ne-p* deo korpusa nekakvog nesvesnog znanja, koje stoji nasuprot onome što subjekt iskreno tvrdi, već naglašava da je subjekt svestan činjenice da verovanje *ne-p* možda može biti istinito, što ostavlja prostora za psihičku tenziju i ukazuje na nestabilnost stanja samoobmane. Uprkos tome što svesno veruje da postoje velike šanse da *ne-p*, jer evidencija ukazuje na to, subjekt ipak tu evidenciju na neki način zanemaruje ili joj ne pridaje dovoljno važnosti, što mu omogućava da uprkos svemu usvoji verovanje da *p*.

Neki autori pretpostavljaju da je reflektivno, kritičko rasuđivanje od ključnog značaja za proces samoobmanjivanja i da ona bića koja nemaju više kognitivne sposobnosti i mogućnost kritičke evaluacije znanja, kao na primer životinje, ne mogu da se samoobmanjuju, već mogu samo da budu „zahvaćene procesom koji je van njihove kontrole“ (Scott-Kakures 2002, 583). Iako čovek aktivno učestvuje u svom samoobmanjivanju, to ne znači da je aktivnost o kojoj je reč intencionalna. Aktivno učestvovanje u ovom procesu znači da samoobmanjivanje može da bude izbegnuto, ali s obzirom na sklonost ka različitim oblicima pristrasnosti koja je tako duboko ukorenjena u funkcionisanje ljudskog kognitivnog sistema šanse za to su male (Scott-Kakures 2002). U procesu samoobmanjivanja racionalnost se okreće protiv sebe i ometa ispravno rasuđivanje. Zapravo, uzimajući pragmatičko testiranje hipoteza za polazište u razumevanju samoobmanjivanja, možemo da kažemo da oni koji se samoobmanjuju, jednako kao i oni koji to ne čine, testiraju hipoteze oslanjajući se na reflektivno rasuđivanje, s tom razlikom što samoobmanjivači ne znaju šta inicira testiranje, to jest da su pokretači testiranja emocije, želje i lični interesi. Pod uticajem teorije koju je razvio Skot-Kakures, Mel dodaje i poslednji, šesti uslov:

6.  $A$ -ovo usvajanje verovanja da  $p$  je proizvod refleksivnog, kritičkog rasuđivanja, i  $A$  greši kada to rasuđivanje smatra pravilno usmerenim (Mele 2012, 12).

Ovaj eksplanatorni model samoobmanjivanja može na direktan način da objasni na koji način individue usvajaju ne samo poželjna, već i nepoželjna verovanja. Iako se testiranje hipoteza uglavnom započinju poželjnim alternativama, ima i onih koji testiranje započinju negativnim mogućnostima (Trope and Liberman 1996). Budući da motivacija samo inicira testiranje hipoteza, a zatim se proces prepušta uobičajenim kognitivnim mehanizmima, one individue koje otpočinju testiranje negativnim mogućnostima uz podršku konfirmacijske pristrasnosti mogu da dođu do nepoželjnih zaključaka koji ne moraju da budu istiniti. Osoba koja s porodicom automobilom ide na more, može da se samoobmanjuje o (ne)ispravnosti kočnica i da intenzivnije testira negativnu hipotezu, jer je prag za njeno odbacivanje u ovom slučaju postavljen visoko, budući da odbacivanje hipoteze da su kočnice neispravne ima visoku subjektivnu cenu. Osoba koja testira ovu i slične hipoteze može da to čini intenzivnije, da evidenciju tumači na pristrasan način i da se sve više učvršćuje u verovanju da su kočnice neispravne, iako su možda ispravne.

Ostaje još jedan problem koji ću samo pomenuti. Prema modelu pragmatičkog testiranja hipoteza koji usvajaju Mel i Skot-Kakures, zapravo nema jasne distinkcije između samoobmanjivanja i *mišljenja u skladu sa željama*, a ne činjenicama (eng. *wishful thinking*); ne radi se o različitim fenomenima, već je pre reč o istom fenomenu u njegovim različitim stupnjevima.

## 5.6 Evolucija samoobmanjivanja: Triversov model

Kvajn (Willard van Orman Quine) je verovao da ako je opažanje (ispitivanje) svojstava urođeno, onda bi bilo razumno očekivati da prirodna selekcija favorizuje one jedinke kod kojih ova sposobnost vodi do najpouzdanijih indukcija. One jedinke koje greše u induktivnom rasuđivanju pokazuju „žalosnu ali hvale vrednu tendenciju da umru pre nego što se razmnože“ (Quine 1969, 126). Na primer, iako su boje „sekundarni kvaliteti“, opažanje i razlikovanje boja je veoma značajno za prikupljanje hrane i izbegavanje opasnosti, a jedinke koje imaju teškoća u opažanju i razlikovanju boja biće kratkovečne u neprijateljskom okruženju u kome takvo

razlikovanje može da znači razliku između života i smrti. Edvard O. Vilson (Edward O. Wilson) ovo objašnjava ilustrativnim primerom koji predstavlja kratak vodič za opstanak u prirodi (Wilson 2002). Prvo, ako je životinja mala, neobično lepa i uz to ženka - verovatno je otrovna i opasna; i drugo, ako nije samo jarkih boja i ženka, već je i neuznemirena našim prisustvom tako da može lako da bude uhvaćena, onda je verovatno smrtonosna. Ovo nas upućuje na još jedan aspekt važnosti razlikovanja svojstava: informacije do kojih dolazimo induktivnim rasuđivanjem moraju da budu na neki način korisne, a u evolucionom kontekstu korisno je uglavnom ono što ima pozitivnog efekta na adaptivnu vrednost organizma (ili je bar adaptivno neutralno).

Mada se obično smatra da nam perceptivni i kognitivni mehanizmi transgeneracijski pružaju sve bolje kapacitete za precizno opažanje realnosti, ovakva predstava evolucije pomenutih mehanizama nije ispravna, jer danas znamo da nam oni ne pružaju takvu sliku stvarnosti. Nešto izmenjena predstava evolucije perceptivnog i kognitivnog sistema bi bila da, iako nam on transgeneracijski ne daje preciznu ili sve precizniju sliku realnosti, on nam ipak pruža dovoljno adekvatnu predstavu stvarnosti koja omogućava uobičajene životne aktivnosti i snalaženje u svakodnevnim situacijama. Uprošćena predstava evolucije kognitivnih kapaciteta bi izgledala ovako:

1. Perceptivni i kognitivni kapaciteti/mehanizmi organizma su nastali u procesu biološke evolucije delovanjem prirodne selekcije;
2. mehanizmi percepcije i mišljenja *uglavnom* pružaju adekvatnu sliku stvarnosti u granicama evoluiranih neurobioloških kapaciteta;
3. informacije koje jedinka stiče na ovaj način moraju na neki način da budu korisne, odnosno moraju da utiču na povećanje adaptivne vrednosti organizma ili da bar budu adaptivno neutralne; svakako ne bi smele da budu štetne;
4. jedinka mora moći da na efikasan način u kratkom vremenskom intervalu iskoristi ove informacije, kako bi one imale efekta na povećanje adaptivne vrednosti;
5. dakle, prirodna selekcija će favorizovati one jedinke koje *uglavnom adekvatno* opažaju realnost i koje mogu dobijene informacije efikasnije da iskoriste nego druge jedinke.

Adekvatno opažanje realnosti i efikasno iskorišćavanje dobijenih informacija može da poveća adaptivnu vrednost organizma. Ali, adaptivna vrednost jedinke može da bude povećana i na drugačiji način: tako što će adaptivna vrednost drugog organizma da bude na neki način umanjena. Pretpostavimo da se dve jedinke *A* i *B* u određenoj populaciji takmiče za oskudne resurse. Ako jedinka *A* uspe na neki način da manipuliše percepcijom i interpretacijom realnosti jedinke *B* da se, na primer, resursi ne nalaze tamo gde se zapravo ne nalaze, ona time umanjuje adaptivnu vrednost organizma *B*, jer će ovaj trošiti vreme i energiju u potrazi za resursima koji se uopšte ne nalaze tamo gde ih on traži. Za to vreme jedinka *A* može da dođe do resursa i potraži partnere za parenje (Surbey and McNally 1997).

Ako je adekvatno opažanje realnosti od presudne važnosti za preživljavanje, onda je evolucija sposobnosti kognitivnog sistema za sistematsko kršenje epistemičkih standarda i usvajanje pogrešnih verovanja, uprkos evidenciji i bez uplitanja i manipulacije drugih, veoma zagonetna. Dok je adaptivna prednost organizama sposobnih za obmanjivanje gotovo očigledna, to nije slučaj s organizmom koji se samoobmanjuje.

Robert Trivers je u predgovoru za knjigu *Sebični gen* Ričarda Dokinsa skicirao ono što je nazvao naučnom teorijom samoobmanjivanja, koju je kasnije detaljnije razvio, a koja taj fenomen tumači kao rezultat evolucione trke u naoružanju između jedinki koje obmanjuju i sposobnosti onih jedinki koje su potencijalna meta obmanjivača da detektuju obmanjivanje (Trivers 1976/2006, Trivers 2002, Von Hippel and Trivers 2011a).

[...] ako je obmanjivanje osnovna stvar u saobraćanju među životinjama, onda mora postojati i oštra selekcija u korist uočavanja obmanjivanja, što, dalje, treba da dovede do izdvajanja višeg stupnja samoobmanjivanja, odnosno do toga da neke radnje i pobude postanu nesvesne kako se – tananim znacima koji bi poticali iz svesti o njima - ne bi odala obmana koja se primenjuje. Stoga bi ono uobičajeno gledište da prirodno odabiranje daje prednost stvaranju takvog nervnog sistema koji će omogućavati sve tačnije i tačnije predstave sveta moralo biti veoma naivno viđenje mentalne evolucije (Trivers 1976/2006, xx).

Iako je ta teorija najdetaljnije razvijena u (Von Hippel and Trivers 2011a), o njoj ću nadalje govoriti samo kao o *Triversovoj hipotezi* ili *Triversovoj teoriji*, prema kojoj je osnovna ili *prava funkcija* samoobmanjivanja efikasnije obmanjivanje drugih. Da bismo razumeli ovu teoriju, neophodno je da imamo na umu distinkciju između obmanjivanja uopšte i

interpersonalnog obmanjivanja. Ova razlika je konceptualno značajna i znatno olakšava razumevanje. U najgrubljem smislu, za dva organizma  $A$  i  $B$ , reći ćemo da  $A$  obmanjuje  $B$  kada  $B$  pogrešno tumači signale koje  $A$  šalje, ako je *prava funkcija* tih signala da navedu  $B$  da pogrešno interpretira realnost. Kada bi organizmu  $B$  bila poznata prava funkcija ovih signala, on bi ih tumačio na pravi način i ne bi bio obmanut. Ovakvo obmanjivanje nije nužno intencionalno i može biti objašnjeno pozivanjem na kauzalne relacije, jer signali mogu da budu hemijski ili da se odnose na neke vidljive fenotipske odlike, kao u slučaju mimikrije i kamuflaže. S druge strane, interpersonalno obmanjivanje je uvek intencionalno. To je aktivnost u kojoj učestvuju najmanje dve individue, gde jedna svesno, intencionalno i uspešno navodi drugu da usvoji pogrešno verovanje. U ovom kontekstu, za dve individue  $A$  i  $B$  ćemo reći da  $A$  obmanjuje  $B$  o  $p$ , kada  $A$  svesno, namerno i uspešno navodi  $B$  da usvoji ili zadrži verovanje da  $p$ , za koje  $A$  u vreme obmanjivanja zna ili iskreno veruje da je pogrešno.

Trivers je u pravu kada obmanjivanje prepoznaje na svim nivoima života, od interakcija gena, preko odnosa virusa i bakterija i njihovih domaćina, insekata i biljaka, predatora i plena, pa sve do socijalnih interakcija sisara i naravno čoveka (Trivers 2010). Međutim, ovde je neophodan oprez, jer je interpersonalno obmanjivanje (kao i samoobmanjivanje) *specijes specifična* karakteristika čoveka i ne treba je mešati s obmanjivanjem koje se pojavljuje van ljudskog socijalnog konteksta.

Trivers tumači samoobmanjivanje po analogiji s interpersonalnim obmanjivanjem i usvaja intencionalistički i divizionistički model dvostrukog verovanja (Trivers 2002, Von Hippel and Trivers 2011b). U podeli uma na svesno i nesvesno, svesni deo uma predstavlja neku vrstu „socijalnog fronta“ koji individue koriste kako bi zavarale druge i izvukle za sebe neku korist plasiranjem pristrasnih informacija, dok istinite informacije ostaju zatvorene u sferi nesvesnog. Reč je o tome da individua prilikom obmanjivanja, preteće i po nju nepovoljne informacije prikriva od drugih, a najbolji i najefikasniji način da to neprimetno izvede jeste da ih prikriva i od sebe.

Da bi samoobmanjivanje evoluiralo, ono mora na neki način da povećava adaptivnu vrednost individue koja se samoobmanjuje, a ono to čini ako pospešuje efikasnost obmanjivanja. Ako su najuspešniji obmanjivači zapravo samoobmanjivači, onda će oni imati prilike da dođu do

više resursa, više partnera za parenje i da ostvare veći broj potencijalnih začeća na račun drugih jedinki. Shodno tome, samoobmanjivanje će moći da evoluiru, jer će one jedinke koje se samoobmanjuju biti uspješnije u obmanjivanju i moći će da povećaju svoju adaptivnu vrednost na račun obmanutih jedinki.

Ako su ljudi sposobni da prepoznaju obmanjivanje, onda će po pretpostavci postojati evolucionu pritisak da se obmanjivanje prikriva, a rezultat će biti samoobmanjivanje kao najefikasniji način za prikrivanje obmanjivanja. Ako je prema ovom modelu najbolja strategija za prikrivanje obmanjivanja ta da se iz bihevioralnog repertoara uklone one manifestacije koje ga odaju, kao što su podrhtavanje glasa, znojenje dlanova ili skretanje pogleda, onda će najefikasniji način da se one uklone biti da prirodna selekcija organizme učini sposobnim za samoobmanjivanje koje će ove pojave da neutrališe, tj. koje će da neutrališe *kognitivno opterećenje* – kognitivne troškove svesno posredovanog obmanjivanja (Von Hippel and Trivers 2011a). Time se umanjuju šanse za otkrivanje obmanjivanja i retribuciju ako je obmanjivanje otkriveno. Prema tome, samoobmanjivanje je ofanzivna strategija nastala u evolucionoj trci u naoružanju koja individui omogućava da obmanjuje druge (Von Hippel and Trivers 2011a). Ako je funkcija samoobmanjivanja interpersonalno obmanjivanje i ako obmanjivanje drugih u određenim situacijama povećava adaptivnu vrednost individue koja obmanjuje na račun obmanute individue, onda samoobmanjivanje može da ima povoljne biološke posledice.

Ali, da bi ova teorija bila opravdana, potrebno je nekoliko dodatnih pretpostavki ili uslova koji moraju da budu zadovoljeni. Evolucija samoobmanjivanja zavisi od sposobnosti jedinki koje su meta obmanjivanja da prepoznaju obmanjivanje. Potkrepljenje za stav da su ljudi vešti u otkrivanju obmanjivača dolazi iz oblasti evolucione psihologije (Cosmides and Tooby 1992). Na osnovu rezultata *Vejsonovog zadatka izbora* (eng. *Wason selection task*) istraživači su došli do zaključka da ljudi poseduju kognitivni mehanizam za koji se pretpostavlja da je rezultat delovanja prirodne selekcije i koji ih osposobljava za otkrivanje varalica. Naime, kada je *Vejsonov zadatak izbora* predstavljen u apstraktnim terminima, individue načelno krše logička pravila pokušavajući da dođu do rešenja, ali kada je zadatak s identičnom formalnom strukturom „uokviren“ drugačije, tako da reflektuje potencijalno obmanjivanje u socijalnom kontekstu, ljudi relativno lako dolaze do rešenja, kršeći u mnogo manjoj meri pravila formalne logike.



Međutim, to ne znači da u socijalnom kontekstu ne dolazi do kršenja epistemičkih standarda, logičkih procedura i donošenja pogrešnih zaključaka. Greška koja se pravi je jednaka onoj do koje dolazi i u apstraktnom kontekstu (afirmacija konsekvensa), ali se ona u socijalnom kontekstu ne tumači kao izraz iracionalnosti, već kao *heurističko sredstvo za brzu detekciju varalica i izbegavanje skupih grešaka*. U evolucionom kontekstu je mnogo važnije izbegavati sebične jedinke sklone eksploataciji nego precizno detektovati altruiste; važnije je izbegavati skupocene greške, nego biti orijentisan ka istini (Friedrich 1993). Na taj način, individua koja izbegava sebične jedinke, može da propusti da sarađuje s nekim altruistom kog je pogrešno identifikovala kao sebičnog, ali njoj se više isplati da ponekada propusti priliku za saradnju nego da bude eksploatisana.

Trivers pretpostavlja da su ljudi dobri u detektovanju obmanjivanja, a da bi bili uspešni u prepoznavanju obmanjivanja moraju da budu zadovoljeni sledeći uslovi: 1) obmanjivanje mora da ima štetne posledice po onoga ko je obmanut; 2) potencijalna žrtva obmanjivanja mora da ima priliku da propituje potencijalnog obmanjivača; 3) neophodno je da se obmanjivač i žrtva poznaju; i 4) da interakcije subjekata budu učestale (Von Hippel and Trivers 2011a, 4). Međutim, kako se čini, ove uslove je u uobičajenim situacijama interpersonalnog obmanjivanja teško zadovoljiti. Čak i kada su zadovoljeni, ispostavlja se da oni ne utiču značajnije na razotkrivanje obmanjivača. Kao potkrepljenje ovom stavu mogu se navesti podaci dobijeni proučavanjem laganja (Vrij 2008, 2011). Na prvom mestu, većina laži (i obmana) u svakodnevnom kontekstu je trivijalna, zatim, ljudi su po prirodi lakoverni, dok s druge strane nastoje da budu uverljivi kada lažu ili obmanjuju, i konačno, kada se nalaze u emotivnim vezama, nastoje da laganje i obmanjivanje specifikuju i uobliče za posebnu relaciju o kojoj je reč.

Za adaptacionističko objašnjenje samoobmanjivanja važna je i odgovarajuća mera (Von Hippel and Trivers 2011a). Takođe, važna je i selektivnost.<sup>58</sup> Ako se jedinka neumereno samoobmanjuje u svim prilikama, to svakako neće imati pozitivnog efekta na njenu adaptivnu vrednost. Jedinke će morati da bude tako programirane da se samo ponekad, u specifičnim situacijama i o određenim aspektima realnosti samoobmanjuje, dok će u drugim sferama

---

<sup>58</sup> O selektivnosti kao prepreci za objašnjenje samoobmanjivanja vidi (Bermúdez 2000, Talbott 1995).

delovanja morati da zadrži adekvatnu predstavu stvarnosti. Osim toga, uzimajući u obzir količinu energije koju mozak troši, mogli bismo verovatno da očekujemo da će prirodna selekcija da favorizuje kognitivne mehanizme za samoobmanjivanje koji nisu intencionalni, čime bi se takođe umanjili kognitivni troškovi (Surbey 2004). To znači da bi prirodna selekcija radila u prilog deflacionog objašnjenja fenomena samoobmanjivanja, uprkos Triversovom insistiranju na intencionalističkom i divizionističkom modelu.

## 5.7 Šta nije u redu s Triversovom teorijom?

Iako na prvi pogled deluje veoma prihvatljivo, Triversova teorija je opterećena jednim brojem značajnih problema. Najpre, Trivers ne pravi razliku između laganja i interpersonalnog obmanjivanja, iako ta razlika može biti konceptualno značajna. Takođe, nije svako govorenje neistine laganje, kao što nije svako laganje obmanjivanje; za laganje i obmanjivanje su važni motivi, a ne samo posledice potencijalnog govorenja neistine. Čini se da je laganje kao i obmanjivanje za samoobmanjivače konceptualno nemoguće, tako da je Triversova teorija već na tom nivou neodrživa. Osoba koja se samoobmanjuje u vezi verovanja da  $p$  i iskreno saopštava da  $p$ , može drugu osobu samo da dovede u zabludu u vezi verovanja da  $p$ , ali ne i da je slaže ili obmane, jer da bi je obmanula ili slagala neophodno je da to čini ciljano, svesno, namerno, i u slučaju obmanjivanja uspešno. Čini se da je za osobe koje se samoobmanjuju to isključeno, osim ukoliko laganje i obmanjivanje trivijalno ne definišemo kao prosto govorenje neistine. Individue koje se samoobmanjuju nisu sposobne za obmanjivanje i laganje kada je reč o verovanju u vezi koga se samoobmanjuju. Ako osobe koje se samoobmanjuju nisu sposobne za obmanjivanje i laganje, onda prava funkcija samoobmanjivanja ne može da bude efikasno obmanjivanje drugih.

Na drugom mestu, jednostavnim misaonim eksperimentom se može pokazati da je stanovište prema kome je funkcija samoobmanjivanja efikasno obmanjivanje drugih gotovo očigledno pogrešno (Ramachandran 1996, Van Leeuwen 2007). Preciznije rečeno, samoobmanjivanje čija bi *prava funkcija* bila efikasno obmanjivanje bi bilo maladaptivno, tako da bi jedinke koje se samoobmanjuju verovatno bile istisnute iz populacije delovanjem prirodne selekcije. Pretpostavimo da jedinka  $A$ , u uslovima oskudice zna gde se nalazi hrana bogata mastima i ugljenim hidratima koja je važna za opstanak. Pretpostavimo da jedinka  $A$  nije dominantna i pretpostavimo da druga dominantna jedinka  $B$ , koja ne zna ništa o tome gde se

nalaze oskudni resursi, želi da sazna od jedinke *A* gde se hrana nalazi. Jedinka *A* ima dve mogućnosti na raspolaganju. Jedna je da kaže istinu i ostane bez resursa (ili da se nada da će *B* s njom da ih podeli) i drugi, da pokuša da obmane jedinku *B*. Ako uspešno obmane jedinku *B*, može samostalno da uživa u hrani, osvoji više partnera za parenje i da se izbori za bolju poziciju u društvenoj hijerarhiji. Ako pokuša da obmane *B* i pri tome potencijalna obmana bude otkrivena jedinka može da trpi dvostruku štetu: može da bude izložena nekoj vrsti retribucije zbog obmanjivanja i da ujedno ostane bez resursa. Prema Triversovoj hipotezi, da bi se obmanjivanje vešto prikriilo i bilo uspešno biće najbolje da se jedinka samoobmanjuje. Međutim, ako i sama jedinka koja obmanjuje ima verovanje koje je pogrešno kada je reč o mestu gde se resursi nalaze, onda ni ona sama neće moći do njih da dođe, što će svakako da umanja njenu adaptivnu vrednost. Izgleda da samoobmanjivanje ne može da evoluirala na takav način.

## 5.8 Alternativna evolucionarna teorija

Alternativna evolucionarna teorija se oslanja na specijalizaciju moždanih hemisfera i istraživanja *anosognozije*, kliničkog stanja koje se manifestuje poricanjem hemiplegije (Ramachandran 1996). Specijalizacija moždanih hemisfera se obično odnosi na sledeće: leva hemisfera mozga specijalizovana za verbalne funkcije, dok je desna zadužena za vizuelno-prostorne odnose. Pri tome treba imati na umu da zbog specifične organizacije nervnog sistema leva hemisfera inervira desnu stranu tela, dok desna inervira levu stranu tela. Anosognozija predstavlja specifično kliničko stanje koje se pojavljuje kod osoba koje su doživele moždani udar koji pogađa desnu hemisferu mozga, tako da im je paralisana leva strana tela (Ramachandran 1996). Ove osobe, mada ne sve i ne sve u istoj meri, uprkos evidenciji poriču da im je leva strana tela paralisana. Kada se od njih zahteva da na neki način upotrebe levu ruku upuštaju se u niz zanimljivih, a ponekada čak i komičnih konfabulacija čija funkcija je opravdanje postojećeg deficita (Ramachandran 1996).<sup>59</sup> Prema ovom modelu, samoobmanjivanje je evoluiralo ne da bi olakšalo interpersonalno obmanjivanje, iako i to može da bude jedna od njegovih funkcija, već

---

<sup>59</sup> Neki od njih tvrde da je pomeraju iako je ne pomeraju, dok drugi tvrde ne samo da je pomeraju, već i da vide da je pomeraju, itd. Osim toga, neki izričito odbijaju da je pomere, ali insistiraju da je ruka inače pokretljiva, mada je oni trenutno ne pokreću zbog reumatskih tegoba (a ne zato što je paralisana), dok drugi saopštavaju da je ne pomeraju jer nisu navikli da primaju naređenja (Ramachandran 1996).

da bi se održala konzistentnost i koherentnost sistema verovanja, kao i kontinuitet ponašanja individue.

Za održanje konzistenosti je prevashodno odgovorna leva hemisfera. Individua nastoji da ima konzistentan i koherentan skup verovanja, kao i da u dužem vremenskom intervalu zadrži kontinuitet određenog bihevioralnog obrasca koji je prethodno usvojila. Kada se skup verovanja koji osoba ima suoči s novim informacijama koje ugrožavaju koherentnost sistema, osoba ne odbacuje sva prethodna verovanja i ne počinje iz početka, već pre pokušava da popravi brod na uzburkanom moru, tj. traži načine kako da nove informacije učini konzistentnim s već postojećim sistemom verovanja. Osoba koja je suočena s neprijatnim ili nepoželjnim informacijama nastoji da mehanizmima kao što su *konfabulacija* i *poricanje* sačuva celovitost postojećeg sistema verovanja za šta je zadužena leva hemisfera mozga (Ramachandran 1996). Međutim, postoji određena, mada ne jasno postavljena, granica do koje konfabulacije i poricanje mogu uspešno da osposobljavaju individuu da zadrži postojeći sistem verovanja. Kada je ta gornja granica dostignuta ona mora da ima neki kontrolni mehanizam koji je osposobljava da prepravi čitav model. Taj mehanizam se nalazi u desnoj hemisferi mozga (Ramachandran 1996).

Osnovna zamisao je da se leva i desna hemisfera na različit način nose s nedostacima, odnosno da imaju različite strategije za njihovo prevazilaženje. Leva hemisfera je „konzervativna“ i nastoji da održi *status quo*, po cenu poricanja i konfabulacija, dok je desna hemisfera, koju Ramačandran naziva „detektor anomalija“ revolucionarna, tako da kada informacije koje ugrožavaju koherentnost sistema verovanja dostignu određeni prag, desna hemisfera odlučuje da je vreme da se leva hemisfera prisili na kunovsku promenu paradigmi i krene iz početka (Ramachandran 1996). Međutim, osobama s anosognozijom nedostaje ovaj kontrolni mehanizam, jer im je desna hemisfera mozga oštećena, što onemogućava da promene verovanja koja su jednom usvojili, a koje leva hemisfera uporno „forsira“ uprkos evidenciji, tj. novim informacijama o određenom telesnom nedostatku. Individue s anosognozijom su u svakom drugom pogledu uglavnom svesne svog okruženja i realnosti. Njihovo poricanje i konfabulacije se odnose pretežno na specifični telesni nedostatak, čime bi bar delimično mogla da se objasni selektivnost samoobmanjivanja i kod zdravih osoba.

Veza anosognozije sa samoobmanjivanjem je gotovo očigledna, osoba *A* usvaja verovanje da *p* koje je pogrešno uprkos evidenciji za *ne-p*, i to usvajanje opravdava određenim razlogom *r* ili skupom razloga. S druge strane, treba imati na umu da se prema ovom modelu osoba sa anosognozijom samoobmanjuje tako što zapravo održava skup verovanja koja su u jednom dugom vremenskom intervalu za tu osobu bila istinita (odsustvo hemiplegije), a onda su naglo postala lažna, čega osoba nije svesna jer su neuronski putevi prekinuti. Osoba sadašnje nepriznavanje prisustva hemiplegije potkrepljuje konstruisanim razlozima. Ali, ti razlozi iz pozicije prethodnog (zdravog) stanja, koje leva hemisfera sada nastoji da održi po svaku cenu uprkos evidenciji za postojanje deficita, imaju smisla i zapravo nisu nikakvi neosnovani konstrukti. Problem je u tome što je osnova na kojoj su stajali iznenada nestala s čime nervni sistem kao da ne može tako lako da se pomiri. To je konzistentno sa stanovištem da osoba ne može da usvoji ili da dugotrajno održava verovanje nezavisno od brojnih pozadinskih verovanja koja već ima (Kunda 1990).

Jedna od prednosti ovog modela samoobmanjivanja u odnosu na Triversov jeste u tome što se objašnjenje evolucije samoobmanjivanja ne oslanja na interpersonalne relacije i konverzacioni kontekst. Robinzon Kruso može da postane „žrtva“ samoobmanjivanja. Jedna od teškoća je to što se ovo objašnjenje ipak oslanja na model dvostrukog verovanja. Ramačandran opisuje slučaj pacijentkinje koja tvrdi da je uspešno vezala pertle i to *s obe ruke* (Ramachandran 1996). Isticanje da je to učinila s obe ruke nije specifično za osobe koje zaista mogu da vežu pertle i ova vrsta akcentovanja, kao i nalazi eksperimenata s vestibularnom stimulacijom ukazuju da osoba na neki način „zna“, ima do određene mere pristupačno verovanje ili „snažno izraženu sumnju“ da joj nešto nedostaje (Levy 2007, 269-272). To znači da divizionistički modeli objašnjenja samoobmanjivanja ne mogu potpuno da budu odbačeni, ali ni potpuno rehabilitovani, već se teret dokazivanja prebacuje na pristalice deflacionog objašnjenja. Uzimajući u obzir Melove uslove za samoobmanjivanje i pojedine aspekte Audijeve teorije, možda je ipak moguće ovu teoriju učiniti kompatibilnom s deflacionizmom.

## 5.9 Triversova hipoteza i psihopatija

Najbolji način da se empirijski testira Triversova hipoteza da je osnovna funkcija samoobmanjivanja efikasno obmanjivanje drugih, jeste da se ustanovi da li postoji klasa

individua koja je vešta u interpersonalnom obmanjivanju, ali koja se ne samoobmanjuje. Moja pretpostavka je da psihopate čine takvu klasu individua. Psihopatama nedostaje moralna sposobnost, ne iskušavaju empatiju, krivicu i kajanje, imaju snižen nivo anksioznosti i straha u susretu i u iščekivanju neprijatnih stimulusa, veoma su vešte u laganju i interpersonalnom obmanjivanju i ne osećaju neprijatnost u slučaju da budu otkriveni (Hare, Forth, and Hart 1989). Ako je samoobmanjivanje oblik iracionalnog mišljenja, i uzimajući u obzir Kleklijevu tvrdnju da je jedna od karakteristika psihopatije „odsustvo zabluda ili drugih znakova iracionalnog mišljenja“ (Cleckley 1988, 339), moramo da se zapitamo da li su psihopate sposobne za samoobmanjivanje i ako jesu u kojoj meri u odnosu na normalnu populaciju.<sup>60</sup>

Mada postoji obimna literatura koja se bavi problemom samoobmanjivanja, radovi koji dovode u vezu samoobmanjivanje i psihopatiju ne postoje. Prema rečima Džeјmsa Blera (lična komunikacija) osobe s psihopatskim karakteristikama često govore neobične stvari o svojoj prošlosti koje su očigledno neistinite, ali zapravo nije poznato da li zaista veruju u ono što govore. U osnovnoj studiji koja se na eksperimentalan način bavila problemom samoobmanjivanja promena u elektrodermalnoj reakciji subjekata je ukazivala da su neka od njihovih verovanja proizvod procesa samoobmanjivanja (Gur and Sackeim 1979). U skladu s tim, elektrodermalna hiporeaktivnost u susretu s pretećim stimulusima koja je specifična za psihopatiju mogla bi da ide u prilog pretpostavci da psihopate nisu sposobne za samoobmanjivanje, mada su veoma vešte u obmanjivanju drugih. Ipak ovu tvrdnju bi trebalo uzeti s rezervom, s obzirom na to da prema nekim autorima psihopate jednako kao i normalne osobe prolaze loše na poligrafskom testiranju (Hare, Forth, and Hart 1989).

Ako su psihopate potpuno svesne svojih obmanjujućih, manipulativnih i egoističkih smicalica, kao što verujem da jesu, tj. ako se ne samoobmanjuju, onda samoobmanjivanje nije evoluiralo kao funkcija efikasnog obmanjivanja, već iz nekog drugog razloga. Von Hipel (lična komunikacija), koji je sarađivao s Triversom, tvrdi da je ova linija argumentacije plauzibilna s

---

<sup>60</sup> Treba imati na umu da samoobmanjivanje može bar ponekada da bude instrumentalno racionalno. Na primer, kada potencijalni samoubica iz određenih razloga, koji ovde nisu od presudnog značaja, na životne teškoće koje oku nepristrasnog posmatrača zaista mogu delovati poražavajuće, počne da gleda kroz ružičaste naočare što ga zatim odvraća od toga da počinu samoubistvo, on se samoobmanjuje, ali je njegovo samoobmanjivanje instrumentalno racionalno, pod pretpostavkom da život tumačimo kao cilj vredan ostvarenja.

obzirom na emocionalne deficite psihopata, koji ne osećaju krivicu i kajanje, tako da imaju manje razloga da se samoobmanjuju u svakodnevnim instancama interpersonalnog obmanjivanja, jer im to nije neophodno da bi prikrili svoje obmanjujuće namere, planove, itd. S druge strane, on misli da se psihopate jednako kao i drugi samoobmanjuju kada je reč slici o sebi i preteranom samopouzdanju, ali da eksperimentalna evidencija koja bi potvrdila ili opovrgla jednu ili drugu hipotezu ne postoji.

Suviše precizne predstave stvarnosti i svog mesta u svetu mogu da uključuju i one njene aspekte koji su neprijatni i nepoželjni, a to ponekada može da bude parališuće i poražavajuće (Surbey 2004). Poznato je da se za razliku od normalne populacije koja svet uglavnom posmatra kroz ružičaste naočare, depresivne individue manje samoobmanjuju i imaju manje više precizne reprezentacije stvarnosti, što je poznato kao *depresivni realizam* (Pham 2007), ali za razliku od psihopata one nisu emocionalno „oštećene“ i ove precizne predstave stvarnosti čine da se osećaju loše, a ne da obmanjuju druge. Iz tog razloga one nisu dobri kandidati za testiranje Triversove hipoteze. S druge strane, psihopate svoje precizne reprezentacije stvarnosti (npr. većina osoba je lakoverna) koriste da bi obmanjivali druge i eksploatisali ih radi zadovoljenja sebičnih ciljeva. Oni to čine hladnokrvno, bez osećaja krivice, kajanja ili anksioznosti.

U pionirskom radu o psihopatiji Klekli zapaža da psihopate ne pokazuju nikakve znake psihoze ili bilo kog drugog psihičkog oboljenja, kao da ih odlikuje izvrsno logičko rasuđivanje i adekvatno opažanje realnosti, a takođe mogu ispravno da predvide posledice svojih nepromišljenih i antisocijalnih postupaka (Cleckley 1988). Međutim, ako racionalnost razumemo instrumentalno, kao ispravno razumevanje odnosa ciljeva i raspoloživih sredstava, nepristrasnom posmatraču njihovi postupci mogu da izgledaju veoma iracionalno (Cleckley 1988).

Prema jednom broju autora, psihopatija je poremećaj u donošenju odluka (Koenigs and Newman 2013). Osobe s psihopatskim karakteristikama zbog specifičnih emocionalnih i kognitivnih deficita propuštaju da rade ono što je u njihovom najboljem interesu. Međutim, da li su ove osobe racionalni donosioci odluka ili ne stvar je otvorene debate. Mišljenja o tome su podeljena i nema načelne saglasnosti o ovom problemu koji u mnogome zavisi od toga kako se racionalnost konceptualizuje. U literaturi koja problematizuje odnos psihopatije i racionalnosti mogu se pronaći najmanje tri koncepta: 1) *logička racionalnost* koja uključuje rasuđivanje na

osnovu normativnih pravila formalne logike, ekonomsku racionalnost onako kako je formulisana u teoriji odlučivanja i sposobnost ispravnog probabilističkog rasuđivanja (uzimanje u obzir osnovne stope); 2) *materijalna (instrumentalna) racionalnost* koja je uglavnom shvaćena kao razumevanje relacije između postavljenih ciljeva i sredstava za njihovo ostvarenje; i 3) *ekološka racionalnost* koja se tumači kao sposobnost donošenja odgovarajućih odluka pod pritiskom različitih prirodnih ili kulturnih uslova životne sredine (Pham 2007).

Osobe s oštećenim ventromedijalnim prefrontalnim korteksom, čije se ponašanje i rasuđivanje donekle podudara s bihejvioralnim obrascima psihopata, prolaze loše na kockarskom zadatku s četiri špila karata (eng. *Iowa Gambling Task*) što implicira nemogućnost da se koriste biološke reakcije (bioregulatorni mehanizmi koji vode ponašanje i odlučivanje, tj. *somatički markeri*) koje bi vodile ponašanje ka boljim ishodima (Bechara et al. 1997, Hughes et al. 2014). Osobe s ovom vrstom oštećenja istrajno biraju „loš“ špil karata, uprkos konceptualnom znanju da je reč o lošem špilju i ne iskušavaju anticipatorni strah (ili anksioznost) kada padnu u iskušenje da prave nepovoljan izbor (Bechara et al. 1997). Takođe, odrasli i dečaci s psihopatskim tendencijama učestalije prave rizičan izbor i prolaze loše u zadatku s četiri špila (Blair, Colledge, and Mitchell 2001, Blair, Mitchell, and Blair 2005). Zapaženo je da je antisocijalna komponenta psihopatije uključena u bolje odlučivanje u ovom zadatku, što je konzistentno s evolucionim argumentom prema kome je psihopatiju možda bolje tumačiti kao posebnu strategiju koja je evoluirala u nepovoljnim uslovima životne sredine, nego kao patološku pojavu (Hughes et al. 2014). Takođe, psihopate prolaze loše i u *Vejsonovom zadatku izbora*, koji je namenjen testiranju sposobnosti valjanog logičkog zaključivanja (Ermer and Kiehl 2010). Za razliku od pripadnika normalne populacije koji poslovično greše kada je *Vejsonov zadatak izbora* predstavljen apstraktno, ali imaju mnogo više uspeha kada je predstavljen u socijalnom kontekstu, psihopate zbog deficita koji pogađaju uobičajene socijalne interakcije prolaze jednako loše u oba zadatka.

S druge strane, čini se da rezultati dobijeni u ekonomskim eksperimentima u kojima su korišćeni standardni modeli *Dileme zatvorenika*, *Ponovljene dileme zatvorenika*, *Igre ultimatum* i *Diktatorske igre* govore u prilog racionalnosti osoba s psihopatskim karakteristikama. U jednoj *Igri ultimatum* osobe s izraženim psihopatskim karakteristikama prihvataju nepravičnu ponudu učestalije od osoba koje nemaju izražene karakteristike psihopatije, što je sa stanovišta ekonomske teorije racionalan izbor (Osumi and Ohira 2010). Navedeni autori ne



tumače ovakav izbor kao posledicu impulsivnosti koja je specifična za psihopate, već kao preferenciju da se uzme bilo šta u odnosu na ništa, što je konzistentno s elementarnom pretpostavkom ekonomske racionalnosti da individue uvek preferiraju više u odnosu na manje. Osobe bez psihopatskih karakteristika, kao i one s manje psihopatskih karakteristika pokazuju hiperreaktivnost kada je reč o elektrodermalnoj reakciji prilikom nepravične ponude u *Igri ultimatum*, dok individue s izraženim psihopatskim karakteristikama reaguju identično i na pravične i na nepravične ponude, što ukazuje da one nisu osetljive na kršenje norme pravičnosti.

Međutim, prema jednom broju autora, korelacija između nedostataka u ventromedijalnom prefrontalnom korteksu i psihopatije se reflektuje na *loše* postignuće u *Igri ultimatum* i *Igri diktatora*, što znači da se psihopate ne ponašaju kao racionalni donosioci odluka (Koenigs, Kruepke, and Newman 2010, Koenigs and Newman 2013). Prema eksperimentalnim rezultatima ovih istraživača svaka od osoba s psihopatskim karakteristikama bar jednom odbija nepravičnu ponudu u *Igri ultimatum*, što se kosi s pretpostavkama ekonomske teorije, ali s druge strane, u *Diktatorskoj igri* po pravilu nude manje, što je konzistentno s predviđanjima ekonomske teorije. Odbijanje da se prihvati nepravična ponuda u *Igri ultimatum* može da se tumači kao čin agresije ili osvete koja ne može da se kontroliše zbog frustriranosti ili iritacije (zdrave osobe to mogu da kontrolišu), dok manja ponuda u *Igri diktatora* može da bude posledica nedostatka empatije i drugih socijalnih emocija.

U *Dilemi zatvorenika* manipulativne i egoističke individue manje saraduju i nisu spremne da učestvuju u recipročnim relacijama (Curry, Chesters, and Viding 2011). Odsustvo anksioznosti u očekivanju negativnih posledica, kao i otpornost na stres pozitivno se odražavaju na nesaradnju, što je dominantna racionalna strategija u ovoj igri, kao i na odbijanje reciprocijacije. Ono što je važno, zahvaljujući tendenciji za neodložno traženje gratifikacije, psihopate ne samo da manje saraduju u *Dilemi zatvorenika*, već prolaze bolje od onih koji saraduju, budući da ih eksploatišu (Mokros et al. 2008). Nestalnost psihopata ide u prilog tezi da oni frekventnije stupaju u jednokratne ili kratkoročne interakcije, tako da oni koji s njima stupaju u kontakt nemaju vremena da u strateškim interakcijama analognim *Dilemi zatvorenika* pređu sa saradnje na nesaradnju, tj. ne mogu da ih kazne. Aktivacija orbitofrontalnog korteksa kada osoba odlučuje da saraduje u *Dilemi zatvorenika* i kada je rezultat igre obostrana kooperacija, reflektuje ulogu emocija u donošenju odluka i efekat koji saradnja ima – doživljavanje zadovoljstva

(Rilling et al. 2002, Rilling et al. 2007). Odluka da se saraduje je posredovana emocijama za koje su odgovorni regioni mozga za koje se pretpostavlja da su kod psihopata anatomske ili funkcionalno deficitarne. U *Dilemi zatvorenika*, u slučajevima jednostrane saradnje, aktivacija amigdale ukazuje na to da se nesaradnja doživljava kao averzivni stimulus, kao društvena pretnja, a budući da psihopate manje saraduju u *Dilemi zatvorenika*, moguće je da jednostrana saradnja koja je averzivni ishod u *Ponovljenoj dilemi zatvorenika* za njih nije u tolikoj meri odbojan, na šta ukazuje slaba aktivacija amigdale u situacijama jednostrane saradnje kod psihopata. S druge strane, osobe koje imaju izražene psihopatske karakteristike osećaju manje intenzivan unutrašnji konflikt kada ne saraduju i potrebno im je manje kognitivnog napora da tako postupaju.

Iako možda ponekada greše i u različitim eksperimentalnim uslovima ne uspevaju da pritisnu pravo dugme, poznato je da psihopate u stvarnom životu i realnim međuljudskim interakcijama umeju da pritisnu pravo dugme i to veoma uspešno i pravovremeno (Hare, Forth, and Hart 1989). Emocionalni i lingvistički deficiti specifični za psihopatiju omogućavaju ovoj populaciji da reči koristi kao najobičnije oruđe za manipulaciju bez suvišnog emocionalnog uplitanja. Moguće je da psihopate lažu i obmanjuju čak i kada za to ne postoji nekakva posebna potreba, što bi verovatno moglo da bude protumačeno kao posledica njihove potrebe za stimulacijom. Sposobnost psihopata da obmanjuju je toliko da čak ni oni koji imaju dugogodišnje iskustvo u radu s ovom populacijom u penalnim uslovima ne ostaju imuni na njihove obmane (Hare, Forth, and Hart 1989).

Ako psihopatiju razumemo kao adaptaciju, tj. kao specifičnu životnu strategiju proizvedenu delovanjem prirodne selekcije u nepovoljnim uslovima životne sredine, onda naizgled iracionalno ponašanje osoba s psihopatskim karakteristikama deluje sasvim uobičajeno sredstvo za ostvarivanje odgovarajućih bioloških ciljeva (Mealey 1995). Dakle, ako psihopate razumemo kao nemilosrdne društvene grabljivce, to nam možda omogućava da naizgled besmislene bihevioralne obrasce učinimo smislenim i instrumentalno i ekološki racionalnim (Hare 1998). Osim toga, moguće je da psihopate lažu i obmanjuju bez ikakve naznake samoobmanjivanja, kao što je moguće da se samoobmanjuju, ali u manjoj meri u odnosu na prosečnog pripadnika opšte populacije. Ako se psihopate samoobmanjuju u određenoj meri i ako psihopatske karakteristike razumemo kao deo kontinuuma, kao i s obzirom neurobiološke razlike

između tzv. uspešnih i neuspešnih psihopata koje se reflektuju na različite fasete *Herove kontrolne liste za psihopatiju* (PCL-R), racionalno je očekivati da se i nivo samoobmanjivanja razlikuje u ove dve kategorije. Da bi se odnos psihopatije i samoobmanjivanja precizno utvrdio neophodna su empirijska istraživanja, s obzirom na to da eksperimentalna evidencija o ovom problemu za sada ne postoji.

Pod pretpostavkom da veza samoobmanjivanja i psihopatije zavisi od toga na koji način konceptualizujemo i šta uzimamo kao osnovnu funkciju samoobmanjivanja, u svetlu prethodnih razmatranja možemo da kažemo sledeće. Ako kao jednu od potencijalnih funkcija samoobmanjivanja uzmemo umanjeње anksioznosti (Barnes 2007, Johnston 1988), što je konzistentno s *Triversovom teorijom*, prema kojoj je uklanjanje spoljašnjih manifestacija svesno posredovanog obmanjivanja značajno za uspešno obmanjivanje, suočićemo se s problemom ako u razmatranje uzmemo populaciju psihopata, koja je vešta u obmanjivanju, ali koja poslovično nije anksiozna i ne pokazuje uobičajene emocionalne reakcije u stresnim uslovima. Takođe, ako samoobmanjivanje definišemo kao posledicu uobičajenih kognitivnih procesa u svakodnevnom testiranju hipoteza i težnje da se izbegnu skupe greške (Mele 2001, Scott-Kakures 2002), što je takođe konzistentno s Triversovim evolucionim argumentima, i u ovom slučaju ćemo se suočiti s problemima, ako u razmatranje uzmemo psihopatsku populaciju, budući da je ova populacija sklona traženju senzacija, preuzimanju rizika i manje više je indiferentna prema pravljenju grešaka. Konačno, ako kao pravu funkciju samoobmanjivanja uzmemo održanje koherentnosti sistema verovanja (i ponašanja) (Ramachandran 1996, Levy 2007), i tada ćemo se suočiti s problemom s obzirom na to da su psihopate nezainteresovane za nekonzistentnosti u onome što saopštavaju, da ih protivrečnosti ne uznemiravaju i da su indiferentni prema nekoherentnim detaljima sopstvenih saopštenja (Williamson 1991). To se može ilustrovati sledećim primerom. Na pitanje kako je započeo kriminalnu karijeru jedan psihopata je odgovorio:

To mora da ima veze s mojom majkom, najdivnijom osobom na svetu. Bila je snažna i vredno je radila da bi odgajala četvero dece. Divna osoba. Počeo sam da kradem njen nakit kada sam bio u petom razredu. Znae, zapravo nikada nisam poznao kučku – krenuli smo različitim putevima... Morao sam ponekada da kradem da bih otišao iz grada, da, ali ja *nisam... kriminalac* (Hare 1999).

Iz svega navedenog sledi da nezavisno od toga na koji način konceptualizujemo samoobmanjivanje i na koji način protumačimo njegovu funkciju, suočićemo se s teškoćama da

samoobmanjivanje pripišemo osobama s psihopatskim karakteristikama. Ako su ove osobe veoma vešte u obmanjivanju, a pokazuju odsustvo ili manjak samoobamnjivanja u odnosu na prosečnu populaciju, onda prava funkcija samoobamnjivanja nije efikasno obmanjivanje drugih.

Ranije je rečeno da je u evolucionom kontekstu adekvatno opažanje i interpretacija realnosti od vitalnog značaja. Ali, ako u pojedinim situacijama važnim za samoodržanje, racionalna, refleksivna i precizne predstave stvarnosti vode smanjenju adaptivne vrednosti, onda bi prirodna selekcija mogla da favorizuje druge, upravo suprotne mentalne mehanizme. Ono je dakle od vitalnog značaja, ali ne uvek i ne nužno. Samoobmanjivanje u takvim situacijama može da pomogne individui da izađe na kraj s pretećim okruženjem. Drugačiji način da se izađe na kraj s pretećim okruženjem je nemilosrdno obmanjivanje drugih, eksploatacija sistema društvene saradnje i pribavljanje resursa za sebe. Ovaj drugi način je specifičan za osobe s psihopatskim karakteristikama. Osim toga, poznato je da samoobmanjivanje u situacijama konflikta interesa može povoljno da utiče na evoluciju saradnje, održanje romantičnih veza, veće roditeljsko ulaganje, itd. (Surbey 2004), a to su stvari koje ne prepoznajemo kod psihopata. Nedostatak samoobmanjivanja, uz druge emocionalne deficite, može da se reflektuje na nesposobnost psihopata da funkcionišu uobičajenom socijalnom i moralnom kontekstu, što delimično može da objasni i njihov relativno mali broj u populaciji.

Ako samoobamnjivanje nije prisutno u populaciji individua s psihopatskim karakteristikama za koju se vezuje odsustvo emocionalne empatije i bilo kakvih moralnih obzira, a važno je za evoluciju i jačanje socijalnih veza, onda je njegova uloga u evoluciji morala nedvosmislena. Takođe, ako usvajanje moralnih verovanja ne zavisi od istinosne vrednosti tih verovanja, već od toga što se njihovo usvajanje pozitivno odražava na adaptivnu vrednost, prirodna selekcija je mogla da favorizuje one organizme koji usvajaju ova verovanja, uprkos tome što krše uobičajene epistemičke standarde. Prema stanovištu koje zastupam, samoobmanjivanje je ako ne nužan, onda jedan od najvažnijih činilaca za evoluciju morala i njegovu objektivaciju. Konačno, filozofski verovatno najvažnija implikacija ovakvog pristupa je da tumačenje odnosa etike i biologije preko samoobmanjivanja povlači za sobom specifičnu metaetičku poziciju koja je poznata kao *evolucioni antirealizam*: evoluciono objašnjenje uzroka koji stoje u pozadini naših moralnih verovanja podriva svaki razlog da ta verovanja smatramo epistemički opravdanim. O tome će biti reči u sledećem poglavlju.

## 6 EVOLUCIJA I ETIKA

### 6.1 Normativne implikacije teorije evolucije

Od same pojave Darwinove knjige *Postanak vrsta*, i među laicima i među stručnjacima, pojavile su se spekulacije o potencijalnim implikacijama teorije evolucije za društveni i moralni život čoveka. Razlog za to je bio jednostavan. Prema teoriji evolucije, čovek je proizvod prirodnog procesa, prirodne selekcije, a ne božanskog čina stvaranja. On je deo prirode jednako kao što su to ševe, kornjače ili bubašvabe i ne zauzima nikakav poseban položaj u velikom lancu bića. Zapravo, takav lanac bića uopšte ne postoji. Evolucionistička zamisao o zajedničkom poreklu isključila je postojanje supstancijalnih razlika između ljudi i drugih životinja: svi se nalaze na različitim granama istog evolucionog stabla, a uspeh se meri preživljavanjem i reprodukcijom, a ne poezijom, filozofijom ili astronomijom.

Ako su perceptivni i kognitivni kapaciteti organizma nastali delovanjem prirodne selekcije i ako nema supstancijalne razlike između čoveka i drugih životinja (jer ma koliko da je velika razlika između čoveka i ostalih sisara kada je reč o kognitivnim sposobnostima, ta razlika se odnosi na stepen, a ne na vrstu sposobnosti), onda takvo objašnjenje može navesti na pomisao da je davanje oduška svojoj prirodi, koja je životinjska, dopustivo i moralno opravdano, što bi ugrozilo konvencionalnu moralnost (up. Darwin 1981). Istini za volju, sam Darwin nije izvodio normativne implikacije iz svoje teorije, niti je smatrao da je evolucionni proces upravljen ka nekom cilju ili da nužno vodi moralnom napretku. Takvu interpretaciju teorije evolucije su predložili socijaldarvinisti, koji su osnovni mehanizam evolucione promene, prirodnu selekciju ili u Spenserovoj (Herbert Spencer) terminologiji *preživljavanje najsnažnijih*, pokušali da primene u razmatranjima koja se tiču morala i drugih društvenih, političkih i ekonomskih procesa.

Ilustracije radi, uzmimo nekoliko primera. U Spenserovom progresivističkom tumačenju evolucije, evolucionni napredak se odnosi na kretanje materije od neodređenog, nekoherentnog i homogenog ka određenom, koherentnom i heterogenom; od nižeg ka višem nivou kompleksnosti čiji je najviši oblik i krajnji cilj moralno ponašanje (Spencer 1879). On identifikuje dobro

ponašanje kao ono što je dostiglo viši stupanj evolucije, a loše kao ono čiji je stupanj evolucije niži, na nekakvoj zamišljenoj evolucionoj lestvici, iz čega zatim sledi da moralne karakteristike osoba i različitih etničkih grupa zavise od inherentnih prirodnih svojstava koje na ovoj evolucionoj skali mogu da budu na višem ili nižem mestu. Na vrhuncu evolucionog procesa nalazi se čovek, kao moralno i društveno biće, čiji postupci podležu moralnim pravilima ponašanja, koja su usmerena na saradnju i uzajamnu pomoć (Spencer 1978 [1879]).

S druge strane, pozivajući se na teoriju evolucije Vilijam Grejem Samner je, kao uostalom i Spenser, opravdavao *laissez-faire* ekonomiju i mušku dominaciju; bio je ubeđen da priroda nagrađuje najsnažnije, a zatim kada su prirodne nedaće jednom ukroćene, one s najviše kapitala (Sumner 1914). Svako uplitanje u ovaj prirodni poredak stvari može biti pogubno. Izbor se čini između slobode, nejednakosti i opstanka najsposobnijih, s jedne strane, i neslobode, jednakosti i opstanka najnesposobnijih, s druge strane. U progresivističkom ključu koji se vezuje za rane istraživače u društvenim naukama koji su se oslonili na evolucionu paradigmu Samner tvrdi da je prva alternativa motor društvenog i individualnog napretka, dok druga povlači sve nadole i favorizuje najgore.

U pre-evolucionoj fazi slično je pisao i Spenser protiveći se bilo kakvom državnom intervencionizmu.

Slepi za činjenicu da pod prirodnim poretom stvari društvo stalno isključuje svoje nezdrave, maloumne, spore, kolebljive, i nesavesne članove, ovi nepromišljeni, mada dobronamerni ljudi se zalažu za uplitanje koje ne samo da zaustavlja proces pročišćenja, već i pospešuje izopačenje – potpuno ohrabrujući umnožavanje nerazboritih i nesposobnih nudeći im neiscrpne povlastice, a obeshrabrujući umnožavanje sposobnih i promišljenih tako što im otežava očuvanje porodice (Spencer 1851, 324).

Socijaldarvinizam dobija krajnje mračne tonove u delu Fridriha fon Bernhardija (Friedrich von Bernhardi) koji u Darwinovoj teoriji pronalazi sredstvo za opravdanje rata i osvajačkih aspiracija (von Bernhardi 1912). U ovoj teoriji zakon jačeg je predstavljen kao univerzalni prirodni zakon, rat i osvajanje se tumači kao biološka nužnost, a borba za opstanak kao živototvorni princip i osnova svakog zdravog razvoja. Rat je shvaćen ka zamajac progresa, jer eliminiše inferiorne koji bi u suprotnom nadvladali iz čega bi sledilo opšte propadanje;

ratničke vrline donose opšti napredak. Iako rat počiva na biološkim zakonima, on nije samo prirodna nužnost, već i moralna obaveza.

Normativne implikacije teorije evolucije za *svet ljudskih stvari* Darwinu su se činile neprimerene, jer je smatrao da je teorija evolucije deskriptivna i on s podsmehom gleda na one koji je tumače kao da iz nje sledi identitet moći i prava, opravdanje osvajačkih pohoda ili tržišne utakmice u kojoj ne važe nikakva pravila.<sup>61</sup> Takođe, on sa skepticizmom pristupa distinkciji *više/nije*. Njegov sledbenik, Tomas Henri Haksli je takođe osporavao da se iz teorije evolucije mogu izvesti bilo kakvi normativni standardi ili moralni principi. Prema Hakslijevom mišljenju u prirodi ne možemo tražiti nikakve moralne uzore, jer priroda nije „škola vrline, već štab neprijatelja“, a moralni progres čovečanstva ne leži u oponašanju prirode, niti u bežanju od nje, već u tome da joj se suprotstavimo (Huxley 2009, 27).

Na osnovu dosadašnjih razmatranja o biološkim uticajima na rasuđivanje uopšte i moralno suđenje posebno, čitaocu je verovatno postalo jasno da bežanje ili suprotstavljanje prirodi nije ni lako, a ponekada nije ni celishodno. Zato je možda bolje sagledati kakve su biološke činjenice i evoluciona istorija čoveka, a zatim na osnovu toga pokušati pružiti neka filozofska objašnjenja specifično ljudskih fenomena, kakav je moral. Ako se fokusiramo na otkrivanje bioloških (neurofizioloških) procesa koji uopšte omogućavaju mišljenje u moralnim terminima i moralno suđenje, onda bismo možda mogli da kažemo nešto o metafizičkom statusu moralnih svojstava i epistemičkom statusu moralnih verovanja, ali teško da bismo isključivo na osnovu toga mogli da kažemo šta treba da činimo ili ne, šta je ispravno, a šta pogrešno.

Savremeni filozofi koji se oslanjaju na evolucionu paradigmu, a čije pozicije će biti razmotrene u ovom poglavlju, prihvataju metodološki naturalizam, tj. polaze od naučnih saznanja o evolucionom procesu i njegovim učincima, a zatim pokušavaju da sagledaju kakve mogu biti filozofske implikacije poznavanja evolucione istorije. Međutim, iako prihvataju metodološki naturalizam, oni ne prihvataju etički naturalizam, kao specifičnu varijantu moralnog realizma, jer smatraju da evoluciono objašnjenje funkcioniše kao *pobijač* ove metaetičke pozicije. Ovi autori, dakle, nisu zainteresovani za pružanje epistemičkog opravdanja moralnih

---

<sup>61</sup> Darwin Correspondence Database, <http://www.darwinproject.ac.uk/letter/entry-2782>, posećeno 27. 1. 2014.

verovanja na osnovu evolucione paradigme, niti veruju da je takvo opravdanje uopšte moguće dati.

Pre nego što se usredsredim na filozofske argumente ovih autora, najpre ću izložiti neke klasične probleme vezane za etički naturalizam. Za odbacivanje naturalizma u etici postoje dva formalna razloga. Jedan od njih je *Hjumov zakon* koji zabranjuje premošćavanje jaza između činjenica i vrednosti, dok je drugi formulisao Džordž Edvard Mur i on je poznat kao *naturalistička greška*. *Naturalistička greška* je obrazložena *argumentom otvorenog pitanja*.

## 6.2 Hjumov zakon

Hjum je ustanovio da nije moguće formulisati logički valjan argument čije bi sve premise bile iskazi o činjenicama, a zaključak vrednosni sud. Drugim rečima, nemoguće je deduktivno izvesti *treba iz jeste*.

Ne mogu da se uzdržim a da ovom razmatranju ne dodam jedno zapažanje koje će se, možda, pokazati od izvesnog značaja. U svakom sistemu morala s kojim sam se dosada susreo uvek sam zapažao da se autor jedno vreme kreće uobičajenom linijom rasuđivanja i uspostavlja postojanje boga, ili pravi zapažanja koja se odnose na ljudske stvari; kada odjednom, na moje iznenađenje, otkrijem da umesto uobičajenih veznika iskaza, *jeste* i *nije*, ne nalazim nijedan iskaz koji nije spojen sa jednim *treba* ili *ne treba*. Ova promena je neprimetna; ali je, ipak, od najveće važnosti. Jer, kako ovo *treba* i *ne treba* izražavaju neku novu relaciju ili tvrđenje, onda je nužno da ono bude zapaženo i objašnjeno, a istovremeno treba pružiti razlog za ono što izgleda potpuno neprihvatljivo, kako ova nova relacija može da bude dedukovana iz drugih, koje su od nje potpuno različite (Hume 1960, 469 – 470).

Hjumov argument možemo da tumačimo na dva načina. Prvi je da je nemoguće *deduktivno izvesti treba iz jeste*, dok je drugi da je *načelno* nemoguće izvesti moralne (vrednosne) sudove iz činjeničkih premisa, tj. da ne postoji nikakav valjani argument kojim bi mogao da bude premošćen jaz koji deli činjenice i vrednosti.

Jedan od najčuvenijih pokušaja da se ovaj jaz deduktivno premosti učinio je Serl (John R. Searle). Njegov argument se oslanja na teoriju govornih činova i koncepciju institucionalnih činjenica koje nastaju primenom konstitutivnih pravila na domen sirovih činjenica (Searle 1964).



Najjednostavnija varijanta tog argumenta, bez pomoćnih premisa koje uključuju empirijske tvrdnje, tautologije i pravila o upotrebi jezika, kojima je izvođenje upotpunjeno, izgleda ovako:

1. Džons je izrekao reči: „Ovim obećavam da ću ti, Smite, platiti pet dolara”.
2. Džons je obećao da će Smitu platiti pet dolara.
3. Džons je sebe stavio u obavezu (preuzeo) da Smitu plati pet dolara.
4. Džons je u obavezi da Smitu plati pet dolara.
5. Džons treba da plati Smitu pet dolara.

U osnovnom Serlovom argumentu polazeći od izgovaranja određenih reči, preko toga da izgovaranje tih reči u određenom institucionalnom kontekstu znači preuzimanje obaveze, dolazi se do zaključka koji je moralni sud. Kako su sve premise argumenta iskazi o činjenicama, u prvih se činilo da uvođenje institucionalnih činjenica može da premosti jaz između činjeničkih iskaza i moralnih sudova. Ali, odmah po objavljivanju Serlovog članka pojavio se veliki broj kritika koje su osporavale njegovu valjanost. One se mogu svrstati u dve grupe, u jednu koja se odnosi na deskriptivno i normativno značenje moralnih termina, i drugu, koja se fokusira na jednu od pomoćnih premisa, tj. uslov *ceteris paribus* u kome kritičari traže skrivenu vrednosnu premisu u obliku inferencijalnog pravila.

Detaljno izlaganje Serlovog argumenta i kritika koje su njemu upućene prevazilazi domašaje ovog rada i iz tog razloga ću ih preskočiti.<sup>62</sup> Ovde ću samo ukratko predstaviti Mekijevu (John Mackie) kritiku koja donekle pokriva obe ove grupe argumenata (Mackie 1990 [1977]).

Prema Mekijevom mišljenju, moguća su dva načina na koji opisujemo institucije, ili institucionalne činjenice u slučaju Serlovog deduktivnog argumenta. Najpre, (a) institucije možemo da opišemo kao kada bismo nekome strancu opisivali neku društvenu praksu s kojom se prvi put susreće. Ovaj način opisivanja možemo da nazovemo *spoljašnjim*. Na drugom mestu, (b) institucije možemo opisati iz konteksta same institucije, kao što to čine učesnici u njoj. Ovaj drugi način možemo da nazovemo *unutrašnjim*.

---

<sup>62</sup> Detaljna analiza i pregled kritika Serlovog argumenta na srpskom jeziku može se naći u (Babić 1977).

Polazeći od (a), da bismo nekoga upoznali s takvom institucijom kao što je obećanje imajući na umu da nikada ranije nije imao priliku da se upozna s tom institucijom, morali bismo da ga upoznamo s pravilima koja važe za tu instituciju. Institucija obećanja počiva na nekim pravilima, kao što šah počiva na određenim pravilima. Pravila obećanja su u Serlovom argumentu specifikovana na način kao što se nekome ko ne ume da igra šah objašnjavaju pravila kretanja svake od šahovskih figura, da bi razumeo samu igru. Privilegija skakača da može da se kreće po šahovskoj tabli u obliku slova *L* i preskače druge figure bi bilo jedno od takvih pravila, a takva pravila bi mogla da se navedu za sve ostale figure. Na sličan način možemo da predstavimo pravila koja važe za obećanje. Ako neko u kontekstu određene institucije, u ovom slučaju institucije obećanja, izgovara reči „Ovim obećavam da ću ti, Smitu, platiti pet dolara”, onda je on u kontekstu te institucije, obećao da će Smitu da plati pet dolara. Dalje, na osnovu pravila koja važe za instituciju obećanja, onaj ko je nešto obećao u navedenom kontekstu je na sebe preuzeo obavezu da ispuni dato obećanje, itd. Ova pravila vode tome da onaj ko ih razume može lako, polazeći od prve premise Serlovog argumenta, da deduktivno izvede zaključak koji je vrednosni sud. Taj argument onda funkcioniše na osnovu same logike.

Međutim, Meki se pita da li je na ovaj način zaista premošćen jaz između činjenica i vrednosti ili je samo nekome ko ne poznaje instituciju obećanja objašnjeno na koji način, odnosno po kojim pravilima ta institucija funkcioniše za nekoga ko tu instituciju prihvata. U tom slučaju zaključak da „Džons treba da plati Smitu pet dolara“ je samo *deskripcija* onoga kako neko treba da se ponaša u kontekstu određene institucije koja je zasnovana na određenim pravilima. Ako je tako, onda Serl nije izveo *treba iz jeste, već jeste iz jeste*, a institucionalne činjenice nemaju ništa drugačiji status od bilo kojih drugih sirovih činjenica.

S druge strane, ako pođemo od (b), onda iz konteksta same institucije možemo da kažemo: „Ne smeš da pomeriš topa jer ćeš time dovesti kralja u šah poziciju“ ili „Džons treba da plati Smitu pet dolara, zato što je to obećao“, čime pretpostavljamo da već prihvatamo određena pravila igranja šaha ili davanja obećanja. Neko ko igra šah zaključuje na osnovu prihvatanja i poznavanja pravila igre da pomeranje topa u određenoj konfiguraciji figura znači šah, kao što neko ko izgovara reči „Ovim obećavam da ću...“, pod određenim prihvatljivim uslovima zaključuje da zaista *obećava* (daje obećanje). Meki tvrdi da je u tom slučaju, argument vođen posebnom logikom koja važi samo za instituciju o kojoj je reč i na osnovu prihvatanja pravila za

tu instituciju. Ako smo kao eksplicitne ili prećutne premise argumenta prihvatili i određena pravila o upotrebi jezika ili inferencijalna pravila, onda Serl nije deduktivno izveo *treba iz jeste*, nego *treba iz treba*, budući da pravila načelno mogu da budu formulisana kao *treba* iskazi. Uzimajući u obzir (a) i (b), možemo da zaključimo zajedno s Mekijem da Serl nije *nedvosmisleno* dokazao da je moguće deduktivno izvesti *treba* iz *jeste*.

Deduktivno izvođenje *treba iz jeste* opterećeno je teškoćama, ali dedukcija nije jedina vrsta zaključivanja koje nam je na raspolaganju, iako se smatra najpouzdanijom. Da bismo logički izveli *treba iz jeste* možemo da iskoristimo druga pravila koja su manje pouzdana od dedukcije, ali koja su takođe legitimna. Jedan od primera koje je ponudio Prajor (A. N. Prior) zasnovan je na logičkom pravilu za uvođenje disjunkcije (Prior 1960).

1. Pijenje čaja je uobičajeno u Engleskoj.
2. Ili je pijenje čaja uobičajeno u engleskoj ili svi Novozelanci *treba* da budu streljani.

U logičkom smislu u ovom zaključku nema ničeg neuobičajenog, on je valjano izveden, a premisa koja je uzeta kao polazna tačka je svakako iskaz o činjenicama, dok je zaključak moralni sud. Do sličnog zaključka možemo da dođemo ako se poslužimo disjunktivnim silogizmom:

1. Ili su morski ježevi sisari ili  $A$  treba da učini  $X$ .
2. Morski ježevi nisu sisari.
3. Dakle,  $A$  treba da učini  $X$ .

Da bismo na nešto drugačiji način izveli *treba iz jeste* možemo da se poslužimo logičkim pravilom koje kaže da iz protivrečnosti sledi bilo šta:

1. Morski ježevi su sisari i morski ježevi nisu sisari.
2. Morski ježevi su sisari.
3. Morski ježevi nisu sisari.
4. Morski ježevi su sisari ili  $A$  treba da učini  $X$ .

## 5. Dakle, $A$ treba da učini $X$ .<sup>63</sup>

Dakle, ako Hjumov argument razumemo tako da se njime tvrdi da je potpuno nemoguće izvesti *treba* iz *jeste*, onda on nije tačan. Međutim, treba imati na umu da se problem sastoji u tome što se upravo prikazano izvođenje smatra trivijalnim i praznim, jer se iskaz „ $A$  treba da učini  $X$ “ može zameniti bilo kojim vanmoralnim iskazom, a da se pri tome ne promeni istinsna vrednost zaključka (Pigden 1989).

Osim ovih načina za izvođenje *treba* iz *jeste*, možemo da se oslonimo na induktivno zaključivanje, koje se ne smatra pouzdanim, ali koje u mnogome olakšava svakodnevno snalaženje u uobičajenim situacijama. Zapravo, bez sposobnosti induktivnog zaključivanja ili s umanjenom sposobnošću za donošenje zaključaka na osnovu zapažanja učestalosti različitih događaja, uobičajene životne aktivnosti bi bile veoma otežane. Takođe, valjano induktivno zaključivanje može individui da pruži određene adaptivne prednosti. Dok u deduktivnom zaključivanju istinitost zaključka *nužno* sledi iz istinitih premisa, u induktivnom zaključivanju iz istinitih premisa *verovatno* sledi istinit zaključak. U induktivnom zaključivanju obično pretpostavljamo da će budućnost ličiti na prošlost, da će se događaji koji su se s određenom učestalošću ponavljali u prošlosti, s istom učestalošću ponavljati i u budućnosti. Ako su do sada opaženi objekti ( $O_1, O_2, O_3, \dots, O_n$ ) iz klase  $K$  imali svojstvo  $F$ , onda će i naredni opaženi objekat  $O_{n+1}$  iz klase  $K$  imati svojstvo  $F$ . Iz toga zaključujemo da svi  $O$  iz klase  $K$  imaju svojstvo  $F$ . Ako su sve do sada opažene vrane bile crne, zaključujemo da će i naredna opažena vrana biti crna. Dakle, sve vrane su crne. Da bi se na jednostavan način pokazalo da je ovakvo zaključivanje nepouzdanost, dovoljno je vrane zameniti labudovima, a crno belim. Ali, uprkos nepouzdanosti, na osnovu prethodnih zapažanja, čini se da imamo dobar *razlog* da zaključimo da će svaka naredna

---

<sup>63</sup> Ako s  $P$  obeležimo iskaz „Morski ježevi su sisari“, a „treba iskaz“ sa  $Q$ , onda to možemo da izrazimo jezikom iskaznog računa:

1.  $P \wedge \neg P$
2.  $P$  (simplifikacijom iz 1.)
3.  $\neg P$  (simplifikacijom iz 1.)
4.  $P \vee Q$  (primenom pravila za uvođenje disjunkcije na 2.)
5.  $Q$  (disjunktivni silogizam iz 3. i 4.)

opažena vrana biti crna, kao što na osnovu uobičajene prakse pozajmljivanja i vraćanja novca imamo dobar razlog da verujemo da *treba* da vratimo novac ako smo ga pozajmili.

*Jeste/treba* problemu možemo da pristupimo i na nešto drugačiji način, posmatrajući ga u eksplanatornoj ravni, gde možemo da razlikujemo dva nivoa, onaj u kome se na činjeničke iskaze pozivamo u kontekstu opravdanja i onaj u kome se moralni iskazi (treba iskazi) pojavljuju kao bolja objašnjenja za datu evidenciju od alternativnih objašnjenja za istu evidenciju. Najpre, u kontekstu opravdanja moralnih sudova možemo da se pozivamo na iskaze o činjenicama. Zapravo, moguće je celokupno opravdanje zasnovati isključivo na činjeničkim iskazima. Uzmimo kao primer slučaj iz kratke priče Edgara Alana Poa (Edgar Allan Poe) *Pad kuće Ašera* u kojoj glavni junak Roderik Ašer (Roderick Usher), zbog specifičnog zdravstvenog stanja, živi u osami, tami i tišini. Pogledajmo zatim sledeći moralni sud:

1. Svako ko ulazi u dom Roderika Ašera *treba* da se uzdrži od pravljenja buke i podizanja zastora.<sup>64</sup>

Potencijalno objašnjenje za ovaj sud može da bude sledeće:

2. Roderik Ašer boluje od alodinije.
3. Oboleli od alodinije doživljavaju bol kada su izloženi stimulusima koji obično ne izazivaju bol.
4. Roderik Ašer doživljava bol kada je izložen glasnim zvucima i svetlosti.
5. Roderik Ašer izbegava da bude izložen stimulusima koji su za njega averzivni.

Ali, činjenica da Roderik Ašer izbegava da bude izložen nadražajima koji su pod uobičajenim okolnostima manje-više neutralni, ali koje on doživljava kao neprijatne i bolne nije dovoljna da bi se bilo ko od onih koji dolaze u njegov dom uzdržao od pravljenja buke i podizanja zastora. Zbog toga je neophodno pružiti još neke dodatne razloge. Ovi razlozi su zasnovani na razmatranju problema empatije i uticaju emocija na rasuđivanje.

---

<sup>64</sup> Mada se može činiti da ovde nije reč o moralnom sudu, već o najobičnijem konvencionalnom pravilu, kakva su pravila o ponašanju za stolom ili pravila odevanja u kontekstu neke institucije, kao što će se pokazati, to ipak nije slučaj. Reč je moralnom sudu zasnovanom na principu nepovređivanja.

6. Vokalna i facijalna ekspresija bola su averzivni stimulusi.
7. Ovi stimulusi su evoluirali da bi bili prekinuti.
8. Ljudi su evoluirali tako da izbegavaju da budu izloženi averzivnim stimulusima.
9. Pod uobičajenim okolnostima ljudi izbegavaju da kod drugih osoba izazivaju visok nivo uznemirenosti (vokalnu i facijalnu ekspresiju bola).
10. Osobe koje dolaze u dom Roderika Ašera izbegavaju da izazivaju averzivne stimuluse.

Neko bi mogao da prigovori da navedeni razlozi i dalje nisu dovoljni za opravdanje suda da svako ko ulazi u Ašerov dom *treba* da se uzdrži od pravljenja buke i podizanja zastora. I može se dogoditi da je on sasvim u pravu. Ali, pitanje koje se nameće je da li je uopšte moguće dati dovoljno iscrpnu listu razloga koja bi bila imuna na ovaj prigovor. U ovom kontekstu dovoljno znači samo *dovoljno*, dovoljno ne znači *sve*. Na sličan način kao što porotnici koji donose osuđujuću (ili oslobađajuću) presudu zaključuju uzimajući u razmatranje evidenciju koja izvan *osnovane sumnje* upućuje na krivicu ili nevinost optuženog, ali ne i evidenciju koja van *svake sumnje* upućuje na njegovu krivicu ili nevinost, možemo da zaključimo da je ova lista razloga zaista dovoljna, iako nije imuna na potencijalne prigovore. Još jedan od takvih prigovora može biti da je argument podlozan optužbi za neku vrstu naturalističke greške, u kojoj se *treba* definiše preko onoga što osoba želi, o kojoj će biti više reči kasnije, i taj prigovor je verovatno sasvim na mestu.

Filozofi naklonjeni naturalističkom pristupu etici insistiraju na tome da jaz između činjenica i vrednosti nije toliko veliki, a da svakako nije nepremostiv. Prema stanovištima ovih filozofa, ovaj problem bismo mogli da sagledamo u evolucionoj ravni i da onda premošćavanje jaza između činjenica i vrednosti tumačimo u kontekstu rešavanja nekog adaptivnog problema. Za potrebe ovog rada njegova precizna identifikacija nije od posebnog značaja. Tako, Patriša Čerčland *jeste/treba* problem izmešta iz strogo logičko-filozofskog konteksta i sagledava ga u kontekstu svakodnevnog snalaženja u svetu (Churchland 2011). Ukoliko bismo se oslanjali isključivo na dedukciju u rešavanju nekog svakodnevnog problema, kao što je prelaženje ulice, odlazak kod lekara ili nešto slično, suočili bismo se s nepremostivim teškoćama i parališućom neodlučnošću, što bi s evolucione tačke gledišta bilo poražavajuće.

Čerčlandova se oslanja na Tagardov (Paul Thagard) model koherencije koji on definiše u terminima problema zadovoljenja ograničenja (Thagard and Verbeurgt 1998). Prema tom modelu, u kontekstu morala, neophodno je postizanje deduktivne, eksplanatorne, deliberativne i analoške koherencije (Thagard 2000). Deduktivna koherencija se odnosi na usaglašavanje principa i sudova, eksplanatorna na usaglašavanje principa i sudova s empirijskim hipotezama, deliberativna na nastojanje da se usaglase sudovi i ciljevi, i konačno analoška koherencija se odnosi na usaglašavanje sudova sa drugim sudovima u sličnim slučajevima. Tagard naglašava da ostvarenje nekih ciljeva ima prioritet u odnosu na neke druge, što je sa svoje strane povezano sa eksplanatornom koherencijom, odnosno s evaluacijom empirijskih hipoteza o prirodi čoveka i društva. Tako, ciljevi koji vode povećanju adaptivne vrednosti jedinke imaju prioritet u odnosu na neke druge ciljeve. Individua koja rasuđuje na ovaj način zaključuje na najkoherentniji plan za rešenje određenog adaptivnog problema. U svakodnevnim situacijama se oslanjamo na sposobnost rešavanja problema koja nalikuje na rešavanje problema zadovoljenja ograničenja na koji većina svakodnevnih problema može da se redukuje. U skladu s evolucionom paradigmom, rešavanje problema zadovoljenja ograničenja ispostavlja adekvatna, ali ne i optimalna rešenja adaptivnih problema (Churchland 2011).

Neki autori smatraju da moralni iskazi mogu da se pojavljuju kao bolja objašnjenja za datu evidenciju od alternativnih objašnjenja za istu evidenciju na sličan način kao što na osnovu činjenice da smo opazili veliki broj crnih vrana i nijednu koja nije crna, pa zatim iz toga zaključujemo da su sve vrane crne, zato što je to najbolje objašnjenje zbog čega smo opazili veliki broj crnih vrana i nijednu koja nije crna. To bi onda bio slučaj *abdukcije* ili zaključivanja na najbolje objašnjenje, koje ima karakteristike koje su slične indukciji, na primer da zaključak prevazilazi ono što se nalazi u premisama, ali za razliku od induktivnog zaključivanja u zaključivanju na najbolje objašnjenje se implicitno ili eksplicitno pozivamo na eksplanatornu funkciju zaključka (Douven 2011).

Harman (Gilbert Harman) ovu vrstu zaključivanja tumači kao poseban oblik indukcije i predlaže da se u rečenicama koje imaju oblik „*X* uzrokuje *Y*“ kauzalne relacije izražene terminom *uzrokuje* zamene eksplanatornim operaterima kao što su *jer*, *zato*, *zato što* ili jednostavno – *objašnjava* (Harman 1973). Induktivni zaključci su zaključci na najbolje objašnjenje, a kauzalna objašnjenja su posebna vrsta ovih zaključaka. U kontekstu moralne

epistemologije, prilikom zaključivanja na najbolje objašnjenje, zaključak (moralni sud) je izveden iz premisa koje su činjenički iskazi zasnovani na opservacijama, sećanju ili introspekciji i dodatne premise kojom se tvrdi da je ovaj zaključak bolje objašnjenje od bilo kog alternativnog objašnjenja (Zimmerman 2010). Jednostavan abduktivni argument izgleda ovako:

1. Ulice su mokre.
2. Da je padala kiša je najbolje objašnjenje zašto su ulice mokre.
3. Padala je kiša.

Jedan od uobičajenih kriterijuma za izbor među alternativnim objašnjenjima istog skupa podataka je jednostavnost, koja predstavlja indikator relativnih šansi da je teorija istinita, tako da kompleksna objašnjenja moraju da ustupe mesto jednostavnijim eksplanatornim modelima, a najjednostavnija objašnjenja su ona koja su najbolja. Kako je sam koncept jednostavnosti višesmislen, Harman predlaže da se jedna hipoteza smatra jednostavnijom od druge ako ju je lakše koristiti za dobijanje odgovora na postavljena pitanja. To je ono što on naziva pragmatička teorija jednostavnosti (Harman 1999). Da je padala kiša je jednostavnije objašnjenje prisustva vode na ulicama od alternativnih objašnjenja, a jednostavna objašnjenja su bolja od drugih eksplanatornih modela, zato je zaključak da je padala kiša najbolje objašnjenje otkuda voda na ulicama.<sup>65</sup> Međutim, Harman smatra da je pozivanje na moralne činjenice za objašnjenje opservacione evidencije redundantno i irelevantno, o čemu će više reći biti kasnije. Ali, Sturdžen misli da Harman nije u pravu i predlaže nekoliko primera u kojima se moralne činjenice, prema njegovom mišljenju, pojavljuju kao najbolja objašnjenja za dostupnu evidenciju (Sturgeon 2006 [1984]). Uzmimo samo jedan od tih primera kao ilustraciju da za moralnog realistu (naturalistu) nije sve izgubljeno:

1. Hitler je počinio brojne zločine.
2. Hitlerova moralna izopačenost je najbolje objašnjenje zašto je Hitler počinio brojne zločine.
3. Hitler je bio moralno izopačen.

---

<sup>65</sup> Ukoliko bi se ovaj argument ispisao u obliku silogizma lako bi se uvidelo da nije reč o deduktivnom argumentu budući da transformacija u silogizam donosi grešku koja je poznata kao *afirmacija konsekvensa*. Takođe, stiče se utisak da je argument cirkularan, ali se obično smatra da je to benigna vrsta cirkularnosti (Lipton 2001).



Ostavljajući za sada po strani Harmanovu i Sturdženovu raspravu o eksplanatornoj moći moralnih činjenica, prema Sturdženovom stanovištu, pozivanje na Hitlerov moralni karakter je najbolje, najjednostavnije i najprirodnije objašnjenje za počinjene zločine. Otvoreno je pitanje da li je na ovaj način zaista izvedeno *treba iz jeste*, ali za potrebe ovog rada je dovoljno da se samo ukaže na to da etički naturalista može da se osloni na različite izlazne strategije kako bi izbegao *Hjumovu giljotinu*.

U nastavku ću razmotriti drugi problem koji se postavlja pred etičkog naturalistu – problem naturalističke greške.

### 6.3 Naturalistička greška i test otvorenog pitanja

Sam termin naturalistička greška u godinama nakon pojavljivanja Murove knjige *Principi etike* postao je neka vrsta magične reči kojom su se svi pokušaji naturalističkog pristupa moralu jednostavno diskvalifikovali ili se teret dokazivanja prebacivao na onu stranu koja je navodno počinila ovu kobnu grešku. Na osnovu Murovih argumenata, u prvi mah je izgledalo da je svaki pokušaj formulisanja naturalističke teorije morala osuđen na neuspeh. Naime, svaka takva teorija proglašavana je pogrešnom ili lišenom eksplanatorne snage. Uprkos tome, Murova kritika naturalizma nije imala onu snagu koju joj je on pripisivao; ona nije zadala odlučujući udarac naturalizmu iz nekoliko razloga koji će ovde biti kritički razmotreni.

Mada se često poistovećuje s *jeste/treba* problemom, naturalistička greška se ipak odnosi na sasvim drugačiju teškoću. U Murovoj knjizi nema ni jedne jedine rečenice koja se odnosi na ovaj problem, a pojam dedukcije se koristi u kontekstima koji ne ukazuju na to da je Murova briga uopšte bila mogućnost ili nemogućnost deduktivnog (ili bilo kakvog drugog) izvođenja vrednosnih (moralnih) sudova iz premisa koje su iskazi o činjenicama. Dok se *Hjumov zakon* odnosi na navodnu nemogućnost logičkog *izvođenja* moralnih sudova iz činjeničkih premisa, naturalistička greška se odnosi na nemogućnost *definisanja* pojedinih moralnih (ali ne samo moralnih) pojmova, kao što je pojam *dobrog*, kao i na nemogućnost *redukcije* tog pojma na neko drugo (ne nužno prirodno) svojstvo.

Kada je reč o problemu definisanja pojma *dobro*, oštrica Murove kritike nije okrenuta ka nominalnim definicijama koje se tiču upotrebe određenog pojma, što najčešće zavisi od

intersubjektivne saglasnosti. Ovakvu vrstu definisanja Mur smatra trivijalnom. Njega pre svega preokupira pojam *dobrog* koji se pojavljuje u realnim definicijama, u kojima se tvrdi *suštinski* identitet definienduma i definiensa. Prema njegovim rečima, on je pre svega zainteresovan šta dobro u metafizičkom smislu jeste, za „prirodu tog predmeta ili te ideje“, a ne šta reč dobro znači i kako se upotrebljava u nekom jeziku (Moore 1963, 11).

Mur polazi od pretpostavke da su sve definicije analitički iskazi čija istinitost zavisi od značenja upotrebljenih termina, tako da kada pokušamo da definišemo dobro nikada nećemo moći da dobijemo analitički iskaz ma koji definiens da koristimo. Razlog za to je što su svi iskazi u kojima se pojavljuje *dobro* sintetički, a ne analitički. Za razliku od pojedinačnog dobra, samo *dobro* se jednostavno ne može definisati, jer nije identično ni sa jednim prirodnim (ili metafizičkim) svojstvom, a Mur pod prirodnim podrazumeva sve što je predmet prirodnih nauka i psihologije. „Može se reći da to uključuje sve što je postojalo, što postoji i što će postojati u vremenu (Moore 1963, 38-39).“

Vrednosni status dobra, tj. činjenica da je ono različito od bilo kog prirodnog svojstva, nije odlučujuća karakteristika koja onemogućava njegovo definisanje. Stav da dobro nije moguće definisati je logički nezavisan od stava da dobro nije nešto prirodno, jer određeno svojstvo može biti prirodno, a da ga ipak nije moguće definisati, kao što je slučaj sa žutim ili sa zadovoljstvom koje je takođe prirodno svojstvo i koje se takođe ne može definisati (Frankena 1967, Moore 1963). S druge strane ispravno takođe nije prirodno, a ipak ga je moguće definisati. To znači da pojmovni par *prirodno/neprirodno* ne korespondira pojmovnom paru *definljivo/nedefinljivo*. Iz stava da se i žuto i zadovoljstvo ne mogu definisati, sledi da se naturalistička greška ne pripisuje isključivo onima koji neko moralno svojstvo, kao što je dobro, definišu nekim prirodnim ili natprirodnim svojstvom, već na samu mogućnost definisanja određenih pojmova, bilo da su oni moralni ili ne.<sup>66</sup> Razlog za nemogućnost definisanja *dobrog* i njemu sličnih pojmova leži u tome što je dobro prost pojam kao što su osnovne boje prosti pojmovi. Mur zaključuje da su dobro i žuto pojmovi iz kojih se sastoje definicije, a koji sami izmiču mogućnosti daljeg definisanja.

---

<sup>66</sup> Budući da je za Murov argument irelevantno da li se dobro definiše prirodnim ili neprirodnim svojstvom, Frankena upozorava da sam naziv naturalistička greška nije adekvatan i da bi se pre moglo govoriti o definicionističkoj grešci (Frankena 1967).

Predmet definisanja može biti samo nešto što je složeno, a kako dobro i žuto nisu složeni, oni ne mogu da budu definisani. Na primer, možemo da definišemo konja. Konj je krupna životinja, kopitar iz klase sisara, biljojed kratkog krzna, bujne grive i dugog repa, ima između 36 i 44 zuba, pripitomljen služi za jahanje ili kao zaprežna životinja. Konja možemo da definišemo jer se sastoji iz mnogo svojstava i osobina od kojih sve mogu da se nabroje, ali ne i dobro ili žuto.

Ako sam upitan „Šta je dobro?“ moj odgovor je da je dobro dobro, i to je kraj te stvari. Ili, ako sam upitan „Kako da se dobro definiše?“, moj odgovor je da se ono ne može definisati i to je sve što imam da kažem o tome (Moore 1963, 11).

Kao što je nedefinljivo, dobro je takođe ireducibilno; ono ne može biti svedeno na neki prirodni (ili neprirodni) predmet različit od njega samog ili skup takvih predmeta. Sve što je do sada rečeno ne znači da *dobro* ne može da se pojavljuje u iskazima oblika „*A* je dobro“, naprotiv, sve dok se veznik „je“ u ovim iskazima ne tumači kao relacija identiteta, već samo u svojstvu pripisivanja karakteristika, opravdano je koristiti ovakve iskaze bez opasnosti da se počini greška. Greška nastaje kada se iskazi pripisivanja u kojima se pojavljuje pojam dobro identifikuju s iskazima identiteta.

Da bi dokazao da se dobro ne može definisati ili redukovati na neko prirodno ili neprirodno svojstvo, Mur se služi argumentom koji je poznat kao *argument otvorenog pitanja*. *Argument otvorenog pitanja* treba razumeti kao instrument kojim možemo da testiramo ispravnost različitih definicija. Argumentacija ima sledeći tok. Najpre, *dobro* nije moguće analizirati i definisati jer je reč o prostom svojstvu, a test otvorenog pitanja funkcioniše na sledeći način. Pretpostavimo da neko tvrdi da je *N*, gde je *N* neko prirodno svojstvo, dobro.

1. Ako je neko prirodno svojstvo *N* identično s dobrim (dobrotom), onda je pitanje „Da li je istinito da je *N* dobro?“ besmisleno i kompetentan govornik koji razume sve termine iskorišćene u rečenici ne može smisljeno da ga postavi.
2. Ali, pitanje „Da li je istinito da je *N* dobro?“ nije besmisleno. Reč je, dakle, o otvorenom pitanju koje kompetentan govornik može da postavi.
3. Dakle, *N* nije dobro.

U Murovom originalnom argumentu  $N$  je zadovoljstvo. Kako se ovaj rad bavi biološkim osnovama morala, a Mur je mislio da je biologija, tačnije teorija evolucije, potpuno irelevantna za etiku, pretpostavimo da je prirodno svojstvo  $N$  o kome je reč *ono (bilo šta) što povećava adaptivnu vrednost organizma* i da je ono dobro, a zatim formulišimo argument u obliku testa otvorenog pitanja.

1. Ako je ono što povećava adaptivnu vrednost organizma identično s dobrim, onda je pitanje „Da li je istinito da je ono što povećava adaptivnu vrednost organizma dobro?“ besmisleno i kompetentan govornik koji razume sve termine iskorišćene u rečenici ne može smisljeno da ga postavi.
2. Ali, pitanje „Da li je istinito da je ono što povećava adaptivnu vrednost organizma dobro?“ nije besmisleno. To je otvoreno pitanje.
3. Dakle, dobro nije ono što povećava adaptivnu vrednost organizma.

Ako evolucionari etičari definišu dobro kao ono što povećava adaptivnu vrednost organizma ili kao ono što doprinosi očuvanju biodiverziteta ili kao ono što vodi očuvanju genskog fonda čoveka, onda je njihov poduhvat osuđen na neuspeh. Socijaldarvinističke teorije, koje su bile jedna od meta Murove kritike, bile su osetljive na ove prigovore jer su se bavile problemom opravdanja (eng. *vindication*) i nastojale su da definišu pojam dobrog kao ono što je „više evoluiralo“, za razliku od onoga što je loše ili zlo i što je sledstveno „manje evoluiralo“. Međutim, predmet savremene evolucione etike uglavnom nije semantički status moralnih sudova i opravdanje.

Drugi Murov argument se odnosi na definiciju dobrog kao onoga što želimo da želimo i nešto je kompleksniji od prethodnog. „Kada mislimo da je  $A$  dobro, mi mislimo da je  $A$  jedna od stvari koju mi želimo da želimo“ (Moore 1963, 19). Dakle, za bilo koje  $A$ ,  $A$  je dobro, ako i samo ako je  $A$  ono što želimo da želimo. U tom slučaju možemo da postavimo sledeći niz pitanja:

1. Da li želimo da želimo  $A$ ?
2. Da li je  $A$  dobro?

Međutim, to nije sve što možemo da pitamo, već možemo da postavimo pitanje o tome da li je to što želimo da želimo  $A$  dobro i da zatim jednostavnom zamenom pojma *dobro* u tom

pitanju *onim što želimo da želimo*, dobijemo pitanje čija složenost prevazilazi ono što nam je uobičajeno kognitivno dostupno kada postavljamo određeno pitanje.

3. Da li je dobro da želimo da želimo *A*?
4. Da li je želja da želimo *A* ono što želimo da želimo?

Mur zaključuje da nama kognitivno nikada nije dostupan tako kompleksan predmet kao što je pitanje: „Da li želimo da želimo da želimo da želimo *A*?“. Prema tome, *dobro* i *ono što želimo da želimo* nisu identični i kada razmišljamo o *dobrom* i *onome što želimo da želimo*, zapravo razmišljamo o dva posebna pojma. Prema Murovom mišljenju, složenost pitanja s kojim se suočavamo je dovoljan argument za opovrgavanje naturalizma koji *dobro* definiše kao *ono što želimo da želimo*. Međutim, moguće je da je složenost koja nastaje kada se pitanje pod rednim brojem tri preformuliše u pitanje pod rednim brojem četiri posledica pažljivog izbora primera, kao što se može dogoditi da kada definišemo neki pojam koji je moguće definisati prema Murovim kriterijumima, takođe nemamo na umu svu složenost na koju taj pojam referira (Joyce 2006). Ako postavimo pitanje „Da li je ovo konj?“, onda svakako u mišljenju nemamo nešto tako složeno kao što je pitanje „Da li je ovo krupna životinja, kopitar iz klase sisara, biljojed kratkog krzna, bujne grive i dugog repa, koji ima između 36 i 44 zuba, ...?“.

Međutim, nije tako teško dati primere u kojima test otvorenog pitanja ne daje očekivane rezultate i navodi nas na stranputicu. Na primer, definišimo trougao kao geometrijski oblik koji ima tri ugla. Ako smo trougao tako definisali, onda pretpostavljam da je svima jasno da je u pojmu trougla već sadržano da on ima tri ugla, tako da neko ko razume taj pojam, ne može smisljeno da postavi pitanje kao što je „Ovo je trougao, ali da li ima tri ugla?“. Budući da je u pojmu trougla analitički sadržano da on ima tri ugla, to pitanje je zatvoreno. Izgleda da nam u slučaju trouglova test otvorenog pitanja pruža zadovoljavajući rezultat. Ali, trougao možemo da definišemo i drugačije. Na primer, definišimo trougao kao mnogougao sa najmanjim brojem uglova. Pitanje „Da li je trougao mnogougao sa najmanjim brojem uglova?“ se čini otvorenim. Ono ne zvuči tako besmisleno kao pitanje o trouglastim trouglovima, jer kako se čini, neko bi sasvim smisljeno mogao da postavi to pitanje. Zapravo ovo pitanje izgleda kao da je otvoreno, a u stvari je zatvoreno. Izgleda da nas argument otvorenog pitanja u ovom slučaju navodi na pogrešan trag.

Polazeći od pretpostavke da su sve definicije analitički iskazi čija istinitost zavisi od značenja upotrebljenih termina, Mur je zanemario čitavi klasu iskaza identiteta koja ne počivaju na analitičkoj identifikaciji, kao u slučaju trougao i geometrijsko telo koje ima tri ugla ili momak i neoženjeni muškarac, već na sintetičkoj, kao što je to slučaj s pojmom voda i hemijsko jedinjenje koje se sastoji od dva atoma vodonika i jednog atoma kiseonika (Kirchin 2012). Ono što test otvorenog pitanja eventualno može da pobije je definicionistički naturalizam, ali ne i metafizički. Podsetimo se da je Mur zainteresovan za to šta stvari jesu, a ne za to šta reči znače i kako se upotrebljavaju u jeziku; zanima ga ono na šta pojam dobro referira, a to je, prema njegovom mišljenju, neprirodno svojstvo dobrote. U tom smislu, istinitost definicije ne bi trebalo da zavisi od značenja upotrebljenih termina, već od stanja stvari u svetu (Fisher 2011).

Frege (Gottlob Frege) je ustanovio da je pravilna veza između znaka (imena), smisla i referencije takva da znaku odgovara određeni smisao, smislu određena referencija, dok nije nužno da referenciji ili jednom predmetu pripada samo jedan znak (Frege 1995). To znači da dva pojma mogu da referiraju na isti predmet, tj. da imaju isti objekat referencije, ali da imaju različit smisao, što se može ilustrovati uobičajenim primerom sa Zornjačom i Večernjačom. Zornjaču i Večernjaču možemo da definišemo na sledeći način.

1. Zornjača je nebesko telo koje se vidi na istoku pre izlaska Sunca.
2. Večernjača je nebesko telo koje se vidi na zapadu nakon zalaska Sunca.

U ovom slučaju Zornjača i Večernjača imaju različit smisao, ali imaju istu referenciju – planetu Veneru. Možemo da razumemo smisao termina Večernjača i Zornjača, ali da ne znamo njihov objekat referencije i da sasvim smisljeno postavimo pitanje „Da li je nebesko telo koje se vidi na istoku pre izlaska Sunca i koje imenujemo Zornjača zapravo planeta Venera?“ (Miller 2003). Za one osobe koje ne znaju da Zornjača, Večernjača i Venera kao objekat referencije imaju jedno isto nebesko telo, to pitanje bi bilo otvoreno, iako je ono zapravo zatvoreno.

Uzmimo još jedan primer. Definišimo vodu kao hemijsko jedinjenje koje se sastoji od dva atoma vodonika i jednog atoma kiseonika. Zamislimo da se sada neko ko opaža supstanciju bez ukusa, mirisa i boje, i koju inače označava imenom voda, ali ne zna ništa o njenom hemijskom sastavu, pita „Da li je ova supstancija bez ukusa, mirisa i boje koju nazivam voda,

zapravo H<sub>2</sub>O?“. Čini se da je tu reč o otvorenom pitanju. Formuliramo sada to u obliku testa otvorenog pitanja.

1. Ako je voda H<sub>2</sub>O, onda je pitanje „Da li je istinito da je voda H<sub>2</sub>O?“ besmisleno, i kompetentan govornik koji razume sve termine iskorišćene u rečenici ne može smisljeno da ga postavi.
2. Ali, pitanje „Da li je istinito da je voda H<sub>2</sub>O?“ nije besmisleno.
3. Dakle, voda nije H<sub>2</sub>O.

Ako argument otvorenog pitanja treba da nam ispostavi test istinitosti određenih definicija, onda on bar u ovom slučaju u tome ne uspeva. Jer, iako postoji razlika u značenju reči voda i H<sub>2</sub>O, supstancija bez ukusa, mirisa i boje koju označavamo imenom voda *jeste* identična hemijskom jedinjenju koje se sastoji od dva atoma vodonika i jednog atoma kiseonika. Voda *jeste* H<sub>2</sub>O iako istinitost iskaza „Voda je H<sub>2</sub>O“ nije utvrđena pojmovnom analizom, već empirijskim istraživanjem. Prema zastupnicima kauzalne teorije referencije, značenje određenih reči kauzalno je uslovljeno na takav način da je referencija fiksirana prvobitnim aktom imenovanja (Kripke 2011 [1971], Putnam 1973). Ne samo da voda *jeste* H<sub>2</sub>O, nego je i *nužno* da je voda H<sub>2</sub>O, jer je termin voda rigidni designator koji označava određeno hemijsko jedinjenje i samo to hemijsko jedinjenje u svim mogućim svetovima. Načelno, ne postoji razlog zbog koga relacija identiteta između nekog prirodnog svojstva *N* i dobrog ne bi bila kao u slučaju vode i hemijskog jedinjenja koje označavamo s H<sub>2</sub>O. Iako za sada ne znamo koje je to prirodno svojstvo, to ne znači da u budućnosti ono neće biti otkriveno empirijskim istraživanjem. Međutim, kao što ćemo kasnije videti, ovo stanovište je prilično problematično.

Moguće je da je Frankena u pravu kada naturalističku grešku smatra domišljatim trikom, a sam Murov argument tumači kao *petitio principii* (Frankena 1967). Dobro je teško definisati, ali na osnovu prethodnih razmatranja izgleda da ne možemo da isključimo načelnu mogućnost njegovog definisanja. U tom slučaju, test otvorenog pitanja ne treba razumeti kao onaj koji zadaje odlučujući udarac etičkom naturalizmu, već kao izazov na koji naturalista treba da pruži adekvatan odgovor i navede razloge zašto definicije dobrog izgledaju otvoreno (Fisher 2011).

Pre nego što pređem na razmatranje argumenata protiv moralnog realizma koji su zasnovani na evolucionoj teoriji i koji, kako se čini, poseduju potencijal da zadaju odlučujući udarac etičkom naturalizmu, neophodno je upoznati se s nekim argumentima na kojima je evoluciono antirealizam zasnovan. Dva argumenta, *teorija greške* Džona Mekija i argument eksplanatorne irelevancije moralnih činjenica, koji je poznatiji kao *Harmanov izazov*, imala su odlučujuću ulogu u raščišćavanju prostora za evoluciono podrivanje moralnog realizma (naturalizma). Najpre ću ukratko izložiti temeljni Mekijev argument na kome je zasnovana teorija greške, prema kojoj su svi moralni sudovi pogrešni, a koji je važan za Rusov evoluciono argument protiv objektivnosti morala. Taj argument Meki naziva *argument na osnovu neobičnosti*. Zatim ću predstaviti *Harmanov izazov* koji čini osnovu Džojsovog evolucionog argumenta protiv moralnog realizma.

#### **6.4 Teorija greške: argument na osnovu neobičnosti**

*Argument na osnovu neobičnosti* ima dva komplementarna dela: metafizički i epistemološki. Metafizički argument se odnosi na mogućnost vanmentalnog postojanja i prirodu objektivnih moralnih svojstava, ako takva svojstva postoje, kao i na njihov status u odnosu prema prirodnim svojstvima. Pod pretpostavkom da takva svojstva postoje, epistemološki argument problematizuje mogućnost njihovog saznanja i postavlja pitanje o saznajnim procedurama koje bi vodile saznanju objektivnih moralnih svojstava. Meki nije na strani logičkih pozitivista koji su smatrali da su svi moralni sudovi jednostavno besmisleni, kao što se ne slaže s emotivistima da su moralni sudovi samo izrazi emocija ili subjektivistima da su moralni sudovi izrazi odobravanja ili neodobravanja. Prema njegovom mišljenju moralni sudovi imaju kognitivnu dimenziju značenja, tj. oni izražavaju verovanja koja imaju istinosnu vrednost, ali je istinosna vrednost jednaka za sve moralne sudove. Naime, kako Meki smatra da nezavisna realnost na koju referiraju moralni sudovi ne postoji, *svi moralni sudovi su pogrešni*.

*Argument na osnovu neobičnosti* je veoma jednostavan:

Ako bi postojale objektivne vrednosti, onda bi one bile entiteti ili kvaliteti ili relacije veoma čudne vrste, sasvim drugačije od bilo čega drugog u univerzumu. U skladu s tim, ako bismo ih bili svesni, to bi bilo pomoću neke posebne sposobnosti moralne percepcije ili intuicije, svakako različite od naših uobičajenih načina saznanja svega drugog (Mackie 1990 [1977], 38).



Ako bi postojala mentalno nezavisna moralna svojstva ili moralne činjenice, one bi bile prilično neobična vrsta entiteta, drugačijih od svih drugih poznatih entiteta, za čije saznanje bi nam bila potrebna neka specifična saznajna sposobnost koja odstupa od uobičajenih saznajnih sposobnosti i metoda kojima inače stičemo saznanja o svetu.

Pretpostavimo da objektivna moralna svojstva postoje. Problem je kako uspostaviti vezu između objektivnih moralnih svojstava s motivacionim potencijalom i prirodnih karakteristika. Na primer, uzmimo sledeći iskaz „Postupak *X* izaziva bol“, za kojim sledi moralni sud „Postupak *X* je pogrešan“.<sup>67</sup> Pitanje je onda u kakvom odnosu stoji moralna činjenica (svojstvo) *pogrešnosti* postupka i prirodna činjenica da je postupak bolan. Odgovor bi bio „Postupak *X* je pogrešan *zato što* izaziva bol“, ali Meki se odmah pita šta je u svetu označeno ovim *zato što* i kako saznajemo relaciju koju označava. *Pogrešnost* postupka *X* je na neki način ili posledica *činjenice* da izaziva bol ili je supervenijentna u odnosu na nju iz čega, prema Mekiju, sledi epistemološki problem: koja bi to saznajna sposobnost ujedno obuhvatala prirodnu činjenicu nanošenja bola, moralnu činjenicu pogrešnosti i vezu (recimo supervenijencije) između njih. Takva saznajna sposobnost bi bila neka vrsta moralne percepcije ili intuicije, a pozivanje na intuiciju je, kao zapaža Kičer, epistemogija očaja (Kitcher 2006b).

Mekijeva filozofska pozicija je bezmalo neobična koliko i postojanje moralnih činjenica kao onog na šta referiraju moralni sudovi, koje argument na osnovu neobičnosti osporava. Prema logičkom zakonu isključenja trećeg, jedan iskaz može biti istinit ili lažan. Uzmimo sledeći primer. Iskaz „*X* je pogrešno“ može biti istinit ili lažan. Ako je iskaz „*X* je pogrešno“, lažan, onda je njegova negacija „*X* je ispravno (nije pogrešno)“, istinit. Ali, prema teoriji greške, u moralu ne važi princip bivalencije, tako da su oba ova iskaza *lažna*, jer nezavisna realnost na koju referiraju moralni sudovi ne postoji, ništa nije ni ispravno ni pogrešno, pa su tako svi iskazi koji sadrže ove pojmove lažni. Moralnim svojstvima dobro/loše, ispravno/pogrešno, itd. jednostavno nedostaje objekt referencije, tako da je jezik morala sličan govoru u kome se pojavljuju patuljci, vile, veštice i jednorozni.<sup>68</sup> Međutim, ovakav manevar ima nepovoljne

---

<sup>67</sup> Ovim ne sugerišem da su svi postupci koji su bolni pogrešni. Postoje brojne situacije kada je nanošenje bola ispravno.

<sup>68</sup> Mekdauel (*John McDowell*) se ne slaže s Mekijevom analizom metafizičkog statusa vrednosti (McDowell 1985). On smatra da se ova analiza nekritički oslanja na Lokovu (*John Locke*) analizu primarnih i sekundarnih kvaliteta.

posledice na normativnom planu, jer sugeriše da su svi postupci, nezavisno od njihovog sadržaja i posledica, dopušteni.

Izlaz iz ove situacije moguće je naći u tvrđenju da je teorija greške metaetička teorija koja se bavi *pitanjima drugog reda*, a ne normativna teorija koja se bavi *pitanjima prvog reda*, tako da ako pažljivo razgraničimo ova dva nivoa, onda se metaetička razmatranja ne reflektuju na probleme normativne etike. Ovaj odnos možemo da uporedimo s igrom kakva je recimo fudbal. Jedno od regulativnih pravila fudbala je da se ne sme dodavati lopta igraču koji se nalazi između golmana i poslednjeg odbrambenog igrača protivničkog tima. Za igrača koji se nalazi u ovoj poziciji se kaže da se nalazi u ofsajdu (eng. Offside). Za nas ovo pravilo važi ili drugačije rečeno, iskaz „Igrač *A* se nalazi u ofsajd poziciji“ možemo da smatramo istinitim ili lažnim samo ako se nalazimo „u igri“ i samo ako se igrač nalazi, odnosno ne nalazi u nedozvoljenoj poziciji. Ali, ako se pozicioniramo izvan igre, onda ovo pravilo ne važi, jer izvan same aktivnosti kakva je fudbal nema nikakvih fudbalskih pravila, pa shodno tome sva pravila koja se pojavlju u igri mogu da se sagledaju kao lažna. Ako se nalazimo na nekoj poljani i neobavezno šutiramo loptu s grupom prijatelja niko se nikada neće naći u ofsajdu ma gde da se lopta nalazi i ma kome da se dodaje. Na normativnom planu, tj. kada se nalazimo „u igri“ kakva je moral, iskazi kao što su „*X* je pogrešno“ ili „*X* je ispravno“, mogu biti istiniti ili lažni i iz toga da je lažno da je *X* pogrešno, sledi da je istinito da *X* nije pogrešno. Ali, kada se nalazimo „izvan igre“, u ovom slučaju izvan

---

Primarni kvaliteti (fizička svojstva predmeta) su za razliku od sekundarnih kvaliteta objektivni, što znači da njihov metafizički status ne zavisi od subjektivnih aspekata opažanja. Metafizički status sekundarnih kvaliteta zavisi od primarnih kvaliteta i subjektivnih aspekata opažanja, a prefilozofska svest greši kada sekundarnim kvalitetima pripisuje sposobnost uzročnog delovanja. Na sličan način prefilozofska svest greši kada o vrednostima pripisuje postojanje nezavisno od subjektivne sveti. Uprkos tome što sekundarni kvaliteti ne mogu da objasne objektivnu realnost, Mekdauel smatra da razlika između primarnih i sekundarnih kvaliteta nije tako oštra. Zapravo, on tvrdi da nije nužno da subjekt greši kada neki objekt označava kao crven i veruje da je on zaista crven. Kada je reč o vrednostima, one takođe ne objašnjavaju objektivnu realnost i nema objektivne moralne realnosti koja u nama uzrokuje moralna verovanja, ali su vrednosti stvarne u smislu koji se donekle razlikuje od onog u kome su boje stvarne. Kada nekoj osobi, postupku ili situaciji pripisujemo neku vrednost, onda se ne radi o pukom zauzimanju stava (*attitude*), već o tome da ona (osoba, postupak, situacija) to i *zaslužuje*. Adekvatan odgovor na određeni postupak je da se on vrednuje, jer zaslužuje da bude vrednovan, kao što neki objekat *zaslužuje* da se prosuđuje kao zastrašujući, ako je adekvatna reakcija na njega strah.

morala i preispitujemo njen metafizički i epistemološki status, onda iz toga da je lažno da je  $X$  pogrešno, *ne sledi* da je istinito da je  $X$  ispravno (nije pogrešno), jer izvan morala ništa nije ispravno ili pogrešno. U tom smislu, rešavanje metaetičkih problema (pitanja drugog reda), kojima se Meki prevashodno bavi, predstavlja slobodnu igru loptom s grupom prijatelja (drugih filozofa) i nikako se ne reflektuju na svakodnevni život i normativnu etiku. O ovome će još biti reči kasnije, kada se upoznamo s Rusovim i Džojsovim evolucionim argumentima protiv moralnog realizma.

Ovde se može dodati još nešto što će biti važno za kasnija razmatranja. Kao što je već rečeno, uprkos tome što je objektivno postojanje morala iluzija, jer ne postoji nekakva posebna realnost čiji entiteti predstavljaju moralne činjenice na koje referiraju moralni termini, objektivizam je uobičajena, zdravorazumska, metaetička pozicija. Meki smatra da za usvajanje ovog pogrešnog verovanja postoje dobri motivi.

Moral nam je potreban za regulisanje međuljudskih odnosa, za kontrolisanje pojedinih načina na koje se ljudi ponašaju jedni prema drugima, često u suprotnosti prema suprotstavljenim sklonostima. Mi zato želimo da naši moralni sudovi budu autoritativni za druge delatnike kao i za nas same: objektivno važenje bi im dalo neophodni autoritet (Mackie 1990 [1977], 43).

Za razliku od Mekija koji smatra da postoje dobri *motivi* za objektivizaciju morala, filozofi koji se oslanjaju na evolucionu paradigmu smatraju da je objektivizacija vođena biološkim *uzrocima*. Ovi filozofi usvajanje pogrešnog verovanja o objektivnosti morala stavljaju u kontekst evolucione efikasnosti: usvajanje ovog pogrešnog verovanja je u predačkim populacijama vodilo povećanju adaptivne vrednosti organizma. Stav koji se zastupa u ovom radu jeste da je samoobmanjivanje bilo *neposredni uzrok* ove objektivizacije.

## 6.5 Harmanov izazov: argument eksplanatorne irelevancije

Nasuprot Mekiju, čiji argument ima metafizičku osnovu, Harman smatra da je moguće dati isključivo epistemološko objašnjenje kojim se podriva moralni realizam i eksplanatornu moć moralnih činjenica (Harman 1977, 1986). Naime, Harman je mišljenja da su moralne činjenice eksplanatorno irelevantne i da uopšte ne doprinose objašnjenju moralnih sudova (sudjenja). Da bi ovo pokazao on pravi paralelu između uloge koju opservacije o prirodnim

činjenicama imaju u naučnim objašnjenjima i uloge koje opservacije o moralnim činjenicama imaju u etici. Dok prirodne činjenice doprinose naučnim objašnjenjima tako što mogu da potvrde ili opovrgnu teoriju, moralne činjenice su *eksplanatorno suvišne* i mogu da budu eliminisane iz objašnjenja jednim potezom *Okamove oštrice*.

Da bismo bolje razumeli Harmanov moralni skepticizam neophodno je ukratko se upoznati s njegovim opštijim epistemološkim gledištima. Harman prihvata radikalnu verziju empirističke epistemologije koju je nasledio od Kvajna i koja može da se sažme u tri tačke: 1) čulno opažanje je osnovni izvor naučnog znanja o svetu; 2) svako čulno opažanje je inferencijalno zavisno od prethodno prihvaćenog teorijskog okvira i 3) nema relevantne razlike između iskaza koji su istiniti zahvaljujući upotrebljenim terminima i pravila za upotrebu jezika (analitički iskazi) i rečenica koje su istinite zahvaljujući činjenicama (sintetički iskazi).

Načelno, svako opažanje je zasnovano na teoriji. Da bi osoba mogla da tvrdi da se pred njom nalazi šolja s kafom ona mora da ima nekakva znanja o svetu, tj. neki prihvaćeni teorijski okvir: da postoji spoljašnji svet, da u njemu postoje trodimenzionalni objekti, da su neki trodimenzionalni objekti predmeti za upotrebu i da su neki predmeti za upotrebu šolje, da postoje topli napici i da je neki od njih kafa, itd. Samo čulno opažanje nije dovoljno za formiranje verovanja koje pretenduje da ima dignitet znanja, već je neophodno da svako verovanje bude formirano *inferencijalnim* procesom čije premise čine određene pozadinske teorije i informacije o tome kako nam stvari izgledaju (Harman 1973). Harmanov moralni skepticizam je zasnovan na ovim epistemološkim osnovama.

Iako moralni skeptik, Harman odbacuje skepticizam koji dovodi u pitanje perceptualna verovanja. Ovo je važno za predmet ovog rada, jer nijedan od evolucionih antirealista ne prihvata skepticizam u pogledu ovih verovanja. Harmanovo odbacivanje opšteg skepticizma je zasnovano na *epistemičkom konzervativizmu*: verovanja su nevinna dok se ne dokaže da su kriva, a teret dokazivanja se prebacuje na onog ko ih dovodi u sumnju i insistira na njihovoj reviziji (Harman 1999). Za odbacivanje već usvojenih verovanja potrebni su posebni razlozi, koje skeptik treba da ponudi. S druge strane, da bi osoba nastavila da veruje u ono šta veruje navođenje takvih posebnih razloga (posebnog opravdanja) nije neophodno. Ako je jedno od verovanja podrazumevana pozicija da spoljašnji svet postoji, a drugo skeptička alternativa da smo mozgovi

u posudi koje zli neuronaučnik stimuliše proizvedeći u njima određena verovanja, čime se dovodi u pitanje opravdanost verovanja o postojanju spoljašnjeg sveta, onda prvo verovanje ima *prima facie* opravdanost, koja skeptičkoj alternativni nedostaje. Razlog za opravdanost je postojanje asimetrije između verovanja koja je osoba već usvojila i skeptičkih alternativa, što prema Harmanu skepticizam u vezi perceptualnih verovanja čini neprihvatljivim. Prihvatanje skeptičkih alternativa nije jednako razumno kao zadržavanje prethodno usvojenih verovanja, budući da su prethodno usvojena verovanja *već usvojena* (Harman 1973). Ovaj temporalni primat već usvojenih verovanja je slaba odbrana od opšteg skepticizma, jer nije teško pronaći protivprimere u kojima je prethodno usvojeno verovanje bilo neopravdano. Izlaz iz ove situacije jeste da se u drugom koraku preispita sam proces usvajanja verovanja. Mada su uobičajena perceptualna verovanja *prima facie* opravdana, za *potpuno* opravdanje potrebna je refleksija o procesu koji je doveo do njihovog usvajanja (Harman 2010). Prema Harmanovom mišljenju, iz činjenice da osoba ne zna ništa o procesu usvajanja verovanja, ne sledi da treba da odbaci to verovanje, jer (prosečna) individua uglavnom ne zna ništa o načinu na koji je došla do ogromnog broja verovanja. Ali, ako ustanovi da je proces kojim je došla do verovanja epistemički nepouzdan, onda je time konstituisan razlog za njegovo napuštanje ili reviziju ukoliko je to moguće.

Da bi objasnio razliku između opservacija u nauci i opservacija u moralu, odnosno eksplanatornu relevantnost prirodnih naspram moralnih činjenica, Harman nas suočava s dva primera. U prvom osoba opaža da dečaci spaljuju mačku. U drugom naučnik opaža trag u maglenoj komori. U prvom slučaju osoba donosi sud da je to što dečaci čine pogrešno. U drugom naučnik donosi sud da je kroz maglenu komoru prošao proton. Pogledajmo Harmanove primere:

### **Primer 1: zapaljena mačka**

Zamislite da dok prolazite ulicom vidite grupu dečaka kako polivaju mačku benzinom, da bi je zatim zapalili. Istog trenutka *vidite* da je njihov postupak *pogrešan*, nema potrebe da *zaključujete* da je *pogrešan*. Rezultat je moralni sud „Spaljivanje mačke je pogrešno“.

### **Primer 2: trag u maglenoj komori**

Zamislite da fizičar opazi trag isparenja u maglenoj komori. Nakon ovog opažaja, on *odmah* donese opservacioni sud da je proton prošao kroz maglenu komoru. Rezultat je iskaz „Proton je prošao kroz maglenu komoru“.

Harman se pita da li je uzrok moralnog suda „Spaljivanje mačke je pogrešno“, *pogrešnost* spaljivanja mačke, kao što je pojavljivanje protona uzrok naučnikovog iskaza ili je samo odraz određenog psihološkog stanja. Ako u skladu s Harmanovim epistemološkim stanovištem pretpostavimo da su ova dva verovanja *prima facie* opravdana, onda je teško uočiti da među njima postoji relevantna razlika. Iako su ova dva primera naizgled slična, Harman smatra da se oni razlikuju u relevantnom smislu, a razlika postaje evidentna ako postavimo pitanje *zašto* je osoba usvojila verovanje da je spaljivanje mačke pogrešno, odnosno da je proton prošao kroz maglenu komoru. Dakle, razlika u moralnim i naučnim verovanjima postaje jasnija ako se fokusiramo na *objašnjenje procesa* formiranja verovanja i toga kakvu ulogu u objašnjenju imaju moralne, odnosno prirodne činjenice, pogrešnost postupka i pojavljivanje protona. Harman smatra da je postuliranje fizičke činjenice u drugom primeru *ekspalanatorno relevantno* na način na koji postuliranje moralnih činjenica nije.

[...] opservacije imaju ulogu u nauci kakvu izgleda da nemaju u etici. Razlika je u tome što morate da imate neke pretpostavke o fizičkim činjenicama da biste objasnili pojavljivanje opservacije koja potkrepljuje fizičku teoriju, ali ne morate da imate bilo koju pretpostavku o moralnim činjenicama da biste objasnili pojavljivanje moralnih opservacija. U slučaju morala dovoljne su samo pretpostavke o psihologiji i moralnoj osetljivosti osobe koja čini moralnu opservaciju (Harman 1977, 6).

U slučaju traga u maglenoj komori, moramo da upotrebimo opaženu činjenicu u objašnjenju verovanja da je proton prošao kroz maglenu komoru i osim toga, u objašnjenju procesa formiranja verovanja *zašto* je naučnik doneo sud da je proton prošao kroz maglenu komoru. Odgovor na ovo pitanje je *zato što* je proton prošao kroz maglenu komoru. (Naučnik je doneo sud da je proton prošao kroz maglenu komoru *zato što* je proton *stvarno* prošao kroz maglenu komoru.) U slučaju spaljivanja mačke, moralni sud da je spaljivanje mačke pogrešno se objašnjava činjenicom da osoba opaža dečake koji spaljuju mačku i da ima određeni moralni senzibilitet ili implicitnu moralnu teoriju da je spaljivanje mačke pogrešno. Time se ne implicira da je spaljivanje mačke *stvarno* pogrešno. U objašnjenju procesa formiranja verovanja da je

spaljivanje mačke pogrešno ne moramo da se pozivamo na moralnu činjenicu *pogrešnosti* spaljivanja mačke, dovoljno nam je poznavanje empirijskih okolnosti i činjenica moralne psihologije. Tako, u odgovoru na pitanje *zašto* osoba donosi sud da je spaljivanje mačke pogrešno, izostaje činjenica *pogrešnosti* tog postupka i na njeno mesto dolaze činjenice moralne psihologije. Drugim rečima, moralne činjenice nikada nisu premise zaključivanja na najbolje objašnjenje čiji je zaključak moralni sud. Kako su moralne činjenice eksplanatorno irelevantne, time je ugrožena opravdanost moralnih verovanja.

Da bi moralne činjenice bile eksplanatorno relevantne morale bi da budu reducibilne na prirodne, tj. moralo bi biti moguće da moralni sudovi budu redukovani na iskaze o prirodnim svojstvima (Harman 1977, 1986). Uzimajući u obzir Harmanovu epistemološku poziciju, to znači da bi moralo biti moguće pružiti definicije moralnih svojstava koje počivaju na sintetičkoj identifikaciji s nekim prirodnim svojstvom  $N$ , kao u slučaju vode i  $H_2O$ . Ako je ova redukcija nemoguća, onda nam na raspolaganju ostaje jedino nekognitivizam. Kako primećuje Džojls, oslanjajući se na Stivenzona, redukcija moralnih na prirodna svojstva je prilično jednostavna: nešto je moralno dobro ako i samo ako je ružičasto s žutim prugama (Stevenson 1937, 14, Joyce 2006). Međutim, nije dovoljno da *ne-analitička* redukcija bude načelno moguća, već je neophodno da bude smisljena i prihvatljiva, a Harman misli da još nikom nije pošlo za rukom da na neproblematičan način redukuje moralna na prirodna svojstva (Harman 1986).

Pogledajmo jedan pokušaj takve redukcije. Oslanjajući se na Patmanovu i Kripkevu teoriju značenja, Bojd (Richard Boyd) je predložio sledeći model za *ne-analitičku redukciju* moralnih na prirodna svojstva.

Grubo govoreći, i za uobičajene slučajeve, termin  $t$  referira na vrstu (svojstvo, relaciju, itd)  $k$  samo u slučaju da postoji kauzalni mehanizam čija tendencija je da dovede, vremenom, do toga da će ono na šta se odnosi (predicated) termin  $t$  biti približno istinito  $k$ ... Takav mehanizam će tipično uključivati postojanje procedura za prepoznavanje članova vrste ili slučajeva  $k$  (bar u lakim slučajevima) koje su aproksimativno precizne i koji na relevantan način upravljaju upotrebom  $t$ , društvenim prenošenjem određenih na relevantan način aproksimativno istinitih verovanja o  $k$ , koja su formulisana kao tvrdnje o  $t$ ... Kada se relacije ove vrste pojave, možemo misliti o svojstvima  $k$  kao o onima koja regulišu upotrebu  $t$  (putem kauzalne relacije), i možemo misliti o onome što je rečeno koristeći  $t$  kao o onome što nam obezbeđuje društveno koordinisani epistemički pristup do  $k$ ;  $t$  referira na  $k$  (u uobičajenim slučajevima) samo u

slučaju da društveno koordinisana upotreba  $t$  obezbeđuje značajan epistemički pristup do  $k$ , a ne drugim vrstama (svojstvima, itd.) (Boyd 1997).

Međutim, Bojdov model je problematičan, jer se pokazuje da semantika moralnih termina nije analogna semantici termina koji se koriste u prirodnim naukama, budući da u drugom slučaju, kauzalno delovanja objekata (činjenica) i jezika ima presudnu funkciju u konstituisanju značenja na način na koji to nije slučaj u moralu. Da bi to pokazali Horgan i Timons (Terence Horgan, Mark Timmons) su, na osnovu Patmanove originalne zamisli, konstruisali misaoni eksperiment koji je poznat kao *Moralna Zemlja bliznakinja* (Horgan and Timmons 1991). Zapažimo, najpre, da je u Bojdovom argumentu kauzalna relacija određena preko načina na koji se moralni termini upotrebljavaju, a zatim pogledajmo kako izgleda Horganov i Timonsov eksperiment. Pod pretpostavkom da je čitalac upoznat s osnovnim Patmanovim misaonim eksperimentom, sažeta verzija Horganovog i Timonsovog eksperimenta izgleda ovako:

**Na Zemlji:**  $X$  je  $M$  (neko moralno svojstvo) akko je *uzrokovano* prirodnim svojstvom  $N$  i  $M$  se *upotrebljava* kao da je artikulirano normativnom teorijom  $K =$  [konsekvencijalistički princip maksimacije sreće];  $M$  na normativnom planu referira na ovo *prirodno svojstvo*.

**Na Zemlji bliznakinji:**  $X$  je  $M$  (neko moralno svojstvo) akko je *uzrokovano* prirodnim svojstvom  $N$  i  $M$  se *upotrebljava* kao da je artikulirano normativnom teorijom  $D =$  [postupanje iz poštovanja prema moralnom zakonu];  $M$  na normativnom planu referira na ovo *deontološko svojstvo*.

U osnovnom Patmanovom misaonom eksperimentu inicijalna pomisao stanovnika Zemlje i stanovnika Zemlje bliznakinje prilikom njihovog eventualnog susreta bi bila da reč „voda“ ima isto značenje na obe planete. Ova inicijalna pomisao se koriguje kada se utvrdi da je voda na Zemlji  $H_2O$ , a na Zemlji bliznakinji XYZ. Stanovnici Zemlje i Zemlje bliznakinje se inicijalno razumeju, na obe planete voda *znači* tečnost bez ukusa, mirisa i boje, mada se razilaze oko *prave prirode* te tečnosti koja pokriva 70 procenata njihove planete. Ali, u Horganovom i Timonsovom misaonom eksperimentu se čini da se stanovnici Zemlje i stanovnici Moralne Zemlje bliznakinje uopšte ne razumeju kada koriste termine koji se odnose na neko moralno svojstvo, jer referiraju na različite normativne teorije, mada je njihova referencija fiksirana istim uzročnim (prirodnim) procesom koji je identičan na obe planete. Ako su moralna svojstva identična s prirodnim



svojstvima, onda Horganov i Timonsov misaoni eksperiment osporava *ne-analitičku* verziju naturalizma.

Vratimo se sada Harmanu. Da proton nije prošao kroz maglenu komoru, naučnik ne bi formirao verovanje da je proton prošao kroz maglenu komoru. Činjenica prolaska protona ima kauzalnu ulogu u formiranju verovanja; ona je kauzalno relevantna za naučnikovo verovanje. Zar onda ne bi bilo slučaj i sledeće: da Hitler nije bio moralno izopačen, ne bismo verovali da je Hitler bio moralno izopačen (Sturgeon 2006 [1984]). Da li svojstvo moralne izopačenosti ima jednaku kauzalnu ulogu u formiranju verovanja da je Hitler bio moralno izopačen, kao što u slučaju traga u maglenoj komori prolazak protona ima kauzalnu ulogu u formiranju naučnikovog verovanja. Harman misli da nema. On odbacuje stanovište da je bilo kakva redukcija moralnih na prirodna svojstva moguća, ali prihvata epifenomenalizam i tezu o supervenijenciji moralnih svojstava i tvrdi da:

Moralni epifenomenalista može da prihvati relevantan protivčinjenički sud, a da ne mora da pretpostavi da moralna svojstva ikada objašnjavaju bilo koje nemoralne činjenice. Moralni epifenomenalista smatra moralna svojstva epifenomenima koji su supervenijentni na prirodna svojstva u smislu da se posedovanje moralnih svojstava objašnjava posedovanjem relevantnih prirodnih svojstava i ništa nije objašnjeno posedovanjem moralnih svojstava (Harman 1986, 63).

Jednostavno rečeno, relacija supervenijencije je zapravo jednosmerna relacija zavisnosti i svodi se na sledeće: za bilo koje skupove svojstava  $P$  i  $Q$ , reći ćemo da je  $P$  supervenijentno u odnosu na  $Q$  (koje je subvenijentno) *akko* dva objekta  $x$  i  $y$  ne mogu da se razlikuju s obzirom na svojstvo  $P$ , a da se ne razlikuju s obzirom na svojstvo  $Q$  (Horgan 1993). Kada se relacija supervenijencije kombinuje s epifenomenalizmom dobija se da je  $P$  supervenijentno u odnosu na  $Q$  (koje je subvenijentno) *akko* je  $Q$  kauzalno relevantno za  $P$  i nije slučaj da je  $P$  kauzalno relevantno za  $Q$ . To znači da za bilo koja dva objekta  $x$  i  $y$ , ako  $x$  ima svojstvo  $P$  i  $y$  ima svojsvo  $Q$ , onda je  $y$  uzrok  $x$  i nije slučaj da je  $x$  ikada uzrok  $y$ . Uzimajući u obzir asimetričnost ovih relacija Harman zaključuje da su moralna svojstva epifenomeni prirodnih svojstava, neurofizioloških i psiholoških procesa, bez ikakve kauzalne funkcije, tj. moralne činjenice nikada ne mogu da se pojavljuju kao premise zaključka na najbolje objašnjenje, budući da su lišene kauzalne i eksplanatorne snage.

## 6.6 Prizivanje duhova, Napoleoneove pilule i plovidba na Bermude

U literaturi koja se bavi evolucionim antirealizmom u etici moguće je pronaći dve grupe podrivajućih (eng. *debunking*) argumenata. Jedna je prevashodno metafizička i oslanja se na evolucionu koncepciju filogenetske kontingencije koju nalazimo još kod Darvina. Takvu vrstu argumenta predlažu Rus i Vilson. Mada Rusov i Vilsonov *argument filogenetske kontingencije* kombinuje metafizičke i epistemološke elemente, on je prevashodno metafizički. Koncepcija filogenetske kontingencije se odnosi na činjenicu da evolucioni proces nije usmeren ka određenom fiksiranom cilju i da se moglo dogoditi da pod određenim drugačijim uslovima u evolucionoj istoriji, ljudi evoluiraju na drugačiji način, tako da njihova moralna sposobnosti, moralni pojmovi i moralna verovanja budu potpuno drugačija od aktualnih. Druga grupa argumenata je prevashodno epistemološka i oslanja se na Nozиковu (Robert Nozick) epistemološku koncepciju praćenja istine i princip osetljivosti, kao i na teoriju pouzdanosti procesa. Ovi argumenti tvrde, otprilike sledeće: ako su kognitivne i moralne sposobnosti proizvod prirodne selekcije ili nekog drugog evolucionog procesa i ako postoje moralne činjenice, onda ove kognitivne i moralne sposobnosti nisu evoluirale da prate moralne istine i to je razlog da se mogućnost moralnog znanja dovede u sumnju. Radi jednostavnosti ovaj argument ću zvati *argument na osnovu neosetljivosti*. Strogo govoreći, Rus i Vilson, kao i Strit su antirealisti, tj. ovi autori tvrde da su svi moralni sudovi pogrešni ili da su svi moralni sudovi pogrešni ako uključuju tvrdnje o objektivnosti, dok je Džojls zagovornik metaetičkog skepticizma, što znači da njegov argument ne isključuje mogućnost da moralne činjenice i istine postoje, ali smatra da moralnim verovanjima nedostaje opravdanje. Iako to nije do kraja precizno, radi jednostavnosti ću, kao i do sada, o svim ovim argumentima govoriti kao o *evolucionom antirealizmu u etici*.

Obe ove grupe argumenata su *genealoške*: podrivajući zaključci se izvode iz poznavanja kauzalne istorije verovanja.<sup>69</sup> Ali, kao što je već rečeno prilikom razmatranja Harmanove epistemologije, poznavanje ili nepoznavanje načina na koji je osoba došla do određenog verovanja ne pruža *prima facie* razlog za odbacivanje verovanja, jer kada je reč o većini

---

<sup>69</sup> Ovi argumenti nisu novi. Takve argumente su koristili Marks, Niče i Frojd (Karl Marx, Friedrich Wilhelm Nietzsche, Sigmund Freud).

verovanja osoba ne zna kako je do njih došla. Ali, ako osoba sazna da je način na koji je usvojila verovanje epistemički nepouzdan, onda je time konstituisan razlog za sumnju u njegovu istinitost. Osim toga, ukoliko kauzalna istorija pruža potpuno objašnjenje moralnih verovanja u terminima evolucione teorije, onda je pozivanje na moralne činjenice eksplanatorno irelevantno i redundantno. Argumenti koji se pozivaju na kauzalnu istoriju, takođe, mogu da nam pruže uvid u metafizički status sadržaja tih verovanja. Zaključci ovih argumenata su da je ono što moralni realisti (naturalisti) nazivaju moralnim činjenicama ili istinama, zapravo (korisna) *fikcija*.

Konačno, ovi argumenti se gotovo isključivo fokusiraju na *krajnje uzroke* (evolucionu istoriju) i gotovo ništa ne govore o *neposrednim uzrocima* koji dovode do razvoja moralne sposobnosti i usvajanja moralnih verovanja, kao i verovanja koje se odnosi na objektivnost morala. Konceptija samoobmanjivanja se pojavljuje na marginama ili se uopšte ne pojavljuje u ovim argumentima, iako je, ukoliko sam u pravu, samoobmanjivanje neposredni uzrok koji vodi usvajanje moralnih verovanja i verovanja o objektivnosti morala. U narednim odeljcima izložiću najpoznatije evolucione argumente protiv moralnog realizma i pokušaću da pokažem na koji način konceptija samoobmanjivanja u njima može da ima važnu ulogu.

### **6.6.1 Argument filogenetske kontingencije**

*Argument filogenetske kontingencije* prvi je formulisao Darwin, a Rus i Vilson su doprineli njegovoj naučno-filozofskoj artikulaciji (Darwin 1981, Ruse and Wilson 1986). Osim osporavanja moralnog realizma, primarni cilj argumenta filogenetske kontingencije jeste osporavanje zdravorazumske metaetičke pozicije kakva je objektivizam. Osnovna pretpostavka ovog, kao i drugih evolucionih argumenata koji osporavaju moralni realizam, jeste da je moral proizvod prirodne selekcije i da je sadržaj moralnih verovanja određen onim što doprinosi preživljavanju i reprodukciji. Da ne bude zabune, time se ne tvrdi da je neko moralno svojstvo  $M$  identično s bilo kojim prirodnim svojstvom  $N$ , pa ni sa svojstvom kao što je „ono što doprinosi povećanju adaptivne vrednosti“. Za Rusa i Vilsona *naturalistička greška* je zaista greška i jaz između činjenica i vrednosti je zaista nepremostiv.<sup>70</sup> To da određeni bihevioralni obrasci i

---

<sup>70</sup> Ovde je neophodna mala opreznost u interpretaciji. Zapravo, Rus insistira na tome da je naturalistička greška stvrano greška i da nema deduktivnog užeta koje bi nas prebacilo s jedne na drugu stranu *Hjumove provalije* (Ruse 2002). Međutim, Vilson nije toliko osetljiv na ove filozofske distinkcije. Njegov samostalni rad, kada se dotakne

verovanja doprinose povećanju adaptivne vrednosti i da ih mi ujedno nazivamo moralnim, stvar je čiste slučajnosti. Pod drugačijim okolnostima, moralna verovanja bi bila drugačija, otuda *kontingencija*. Ovi, kao i drugi autori, koji se oslanjaju na evolucionu paradigmu pokušavaju da uvide kakve posledice evolucionu prošlost čoveka ima na sposobnost moralnog suđenja i sadržaj moralnih verovanja. Oni prihvataju *metodološki naturalizam*, ali odbacuju *etički*.

Ako je moral proizvod prirodne selekcije, odnosno bioloških i neurofizioloških procesa i ako imamo na umu da je evolucionu prošlost čoveka mogla da bude drugačija, onda nema razloga da verujemo u objektivnost morala (u vanmentalno postojanje moralnih činjenica), jer bismo u protivčinjeničkoj situaciji, u kojoj drugačiji bihejvioralni obrasci doprinose povećanju adaptivne vrednosti, imali potpuno drugačije shvatanje onoga što je dobro, ispravno i pravedno. Darwin o tome piše sledeće:

Možda bi bilo dobro da najpre pomenem da ne želim da tvrdim da bi ijedna potpuno društvena životinja, kada bi njene intelektualne sposobnosti postale tako delotvorne i visoko razvijene kao kod čoveka, usvojila potpuno isti moralni osećaj kao što je naš. Na isti način kao što različite životinje imaju neki osećaj za lepo, mada se dive veoma različitim stvarima, mogle bi da imaju osećaj za ispravno i pogrešno, a da vođene njime prate potpuno različite smerove ponašanja. Da su, na primer, da uzmemo drastičan slučaj, ljudi odgojeni pod potpuno istim uslovima kao pčele, gotovo da nema sumnje da bi naše neudate žene, kao pčele radilice, mislile da je sveta dužnost ubijati svoju braću i da bi majke težile da ubijaju svoje plodne kćeri; i nikome ne bi palo na pamet da se umeša (Darwin 1981, 73).

Paralelizam između Darwinovog argumenta i osavremenjene verzije tog argumenta koji predlažu Rus i Vilson je očigledna:

Lako je zamisliti vanzemaljsku inteligentnu vrstu koja ima takva evoluirana pravila ponašanja koja smatra visoko moralnim, ali koje nama deluju odvratno, kao što su kanibalizam, incest, ljubav prema tami i raspadanju, ubistvo roditelja i

---

filozofskih problema, predstavlja pravu konceptualnu zbrku (Vidi npr. Wilson 1999, ch. 11). On tvrdi da *treba* nije ništa drugo nego *jeste*, iako među njima postoji značajna razlika, zatim identifikuje *naturalističku grešku* i *Hjumov zakon*, iako je reč o različitoj vrsti problema; jedan se odnosi na definisanje, a drugi na mogućnost deduktivnog izvođenja. Slično važi i za Fransa de Vala (de Waal 2014). Njegovo insistiranje da ponašanje životinja nije izuzeto od nekih normativnih standarda, i da jaz između *jeste* i *treba* nije tako izražen koliko pretpostavljaju filozofi, zapravo promašuje metu, jer se *jeste/treba* problem odnosi na mogućnost deduktivnog izvođenja moralnih iskaza iz iskaza o činjenicama, a ne na ponašanje.

uzajamno jedenje izmeta. Mnoge životinjske vrste čine neke ili sve od ovih stvari u slast i da bi preživele. Da su ljudska bića evoluirala od vrste koja je drugačija od dvonožnih mesoždera, stanovnika savane od kojih smo evoluirali, i mi bismo mogli da činimo isto, osećajući se sigurnim iznutra da je takvo ponašanje prirodno i ispravno. Ukratko, etičke premise su osoben proizvod genetičke istorije, i mogu biti shvaćeni isključivo kao mehanizmi koji su adaptivni za vrstu koja ih poseduje. *Sledi da etički kod jedne vrste ne može da se primeni (ne može da se prevedu) na drugu.* Nikakvi apstraktni moralni principi ne postoje izvan posebne prirode individualne vrste (Ruse and Wilson 1986, 186, moj kurziv).

Kao što je sličnost između navedenog i Darwinovog argumenta očigledna, tako postoje sličnosti između ovog i argumenta *Moralne Zemlje blizakinje*. Vilson i Rus temelje svoje stanovište na činjenici da je kurs evolucije mogao da bude drugačiji. Pod takvim izmenjenim okolnostima moralna pravila bi bila potpuno drugačija, čime se dovodi u pitanje postojanje nepromenljivih moralnih istina.<sup>71</sup>

Ukoliko bismo se sreli s nekom inteligentnom vanzemaljskom vrstom čija je evolucionarna istorija u mnogome drugačija od naše i čija moralna psihologija je, shodno tome, veoma različita od naše, takva vrsta bi imala potpuno drugačija moralna verovanja u odnosu na nas. Razlika u moralnim verovanjima bi bila posledica uzročnog delovanja evolucionih mehanizama i epigenetičkih pravila mentalnog razvoja, a ne delovanja bilo kojih specifičnih moralnih svojstava (bilo da su reducibilna ili ne). Moral (moralna sposobnost i moralna verovanja) je proizvod neurofizioloških mehanizama i ništa više od toga: različiti smerovi evolucije mogu da uzrokuju različite neurofiziološke procese i na kraju različitu moralnu sposobnost i različita moralna verovanja.

Moralna sposobnost može da se redukuje na neurofiziološke procese koji su rezultat prirodne selekcije, i koji se zatim manifestuju na kognitivnom ili emocionalnom planu. Ova činjenica mora da bude iskorišćena da bi se objasnio moral, moralno rasuđivanje i poreklo moralnih verovanja. Ova redukcija jeste naturalistička u metodološkom smislu, ali ovde nije reč o etičkom naturalizmu, budući da se pod etičkim naturalizmom podrazumeva učenje prema kome moralna verovanja mogu da budu istinita ili lažna, da su neka zaista istinita jer referiraju na *moralne činjenice* koje su reducibilne na prirodne (Pigden 1991). Međutim, treba imati na umu da Rus i Vilson *načelno* odbacuju mogućnost postojanja moralnih činjenica: „[...] ne može

---

<sup>71</sup> Ovaj argument može da ospori postojanje moralnih činjenica samo ako se one shvate kao *nužno istinite*.

biti stvarno objektivnih spoljašnjih etičkih premisa. Sve što znamo o evolucionom procesu upućuje na to da takva vantelesna uputstva ne postoje“ (Ruse and Wilson 1986, 186). Kada je jednom pruženo potpuno evoluciono objašnjenje morala, moralne činjenice postaju *eksplanatorno irelevantne i redundantne* i mogu da budu eliminisane iz objašnjenja delovanjem *Okamove oštrice*.

Vratimo se na citirani paragraf u kome se tvrdi da „*etički kod jedne vrste ne može da se primeni (ne može da se prevede) na drugu*“ koji sada može da deluje razumljivije. Izgleda da bi za dve različite (inteligentne) vrste s potpuno različitim evolucionim poreklom moralni relativizam bio sudbina, kao što bi to verovatno bio i konceptualni relativizam. Uzmimo primer koji je najbliži ljudima. Danas je uglavnom poznato da različite vrste primata različito opažaju boje (Jacobs 1996). Mada je to samo nagađanje, planetarno udaljene inteligentne vrste bi verovatno na potpuno drugačiji način interpretirale svet. Ali kada je reč o ljudskoj vrsti, njenim različitim varijetetima i kulturnim praksama, možemo da kažemo da konceptualni, pa onda ni moralni relativizam nije opcija. Razlog za to je nekoherentnost konceptualnog relativizma: on pretpostavlja da postoje potpuno neprevodive konceptualne sheme koje uobličavaju stvarnost na potpuno drugačiji način, što je u suprotnosti s mogućnošću prevođenja (Davidson 1991). To ne znači da Rus i Vilson potpuno odbacuju (normativni) moralni relativizam. Oni usvajaju umerenu verziju tog stanovišta, s obzirom na to da smatraju da su biološke i kulturne sličnosti među ljudima mnogo veće nego razlike (Ruse and Wilson 1986). Oni jednostavno priznaju da postoje različita moralna verovanja u različitim kulturama, ali dodaju i to da su podudarnosti značajne.

Uzmimo primer koji smo već koristili: postojanje roditeljske brige je biološki fenomen koji kod altricijalnih vrsta, kakva je ljudska, doprinosi povećanju adaptivne vrednosti. Roditeljska briga se takođe podudara s duboko izraženim osećajem obaveze da se vodi računa o sopstvenom potomstvu, što je dobro dokumentovan fenomen u različitim ljudskim kulturama. Mada uzrast do koga treba brinuti o potomstvu varira od kulture do kulture, nema kulture u kojoj ne postoji obaveza brige za potomstvo ili u kojoj postoji neobično moralno pravilo koje bi nalagalo da roditelji ne treba da brinu o potomstvu. U tom smislu briga o potomstvu je univerzalna. Međutim, zamislivo je da je evolucija mogla da ima drugačiji tok i da evolucija ljudske vrste bude bliža evoluciji reptila. U tom slučaju obaveza da se vodi računa o potomstvu verovatno ne bi postojala.

S druge strane, metaetički relativizam je stanovište prema kome istinitost ili neistinitost moralnih verovanja zavisi od kulturnog, društvenog i istorijskog konteksta (Wong 1991). Ali, ovi autori nisu metaetički relativisti. U svojim samostalnim radovima Rus se eksplicitno poziva na Mekija i svoje stanovište temelji na *teoriji greške* koja je upotpunjena evolucionom pretpostavkom (Ruse 1995a, 2002, 2006a, b). Na taj način, uzimajući u obzir sve ono što znamo o evolucionom procesu Rus misli da su sva moralna verovanja podjednako pogrešna, u ma kojoj kulturi da se pojavljuju. Ovakav pristup predstavlja neku vrstu metodološkog trika: na normativnom planu moralna verovanja se u različitim kulturama razlikuju, na metaetičkom planu sva moralna verovanja su podjednako pogrešna u ma kojoj kulturi da se pojavljuju. Takođe, Rus izbegava da koristi reč „pogrešno“ kada govori o moralnim iskazima koji su prema njegovom mišljenju neutemeljeni, jer smatra da su iskazi kao što je „silovanje je pogrešno“ ili „ubistvo je pogrešno“ istiniti, mada isključivo u kontekstu svakodnevnog moralnog postupanja.<sup>72</sup> To je svakako neobična pozicija.

[Ne] samo da verujemo u supstantivan moral, već verujemo da supstantivan moral ima objektivnu osnovu. Kao što je Meki primetio pre mene, dobar deo moralnog iskustva se odnosi na objektivaciju naše supstantivne etike. Zapravo nema temelja, ali mi verujemo da u nekom smislu ima temelja. Postoje dobri biološki razlozi zašto to činimo.... Supstancijalna moralnost ostaje na mestu kao efektivna iluzija, zato što mislimo da nije iluzija nego prava stvar.... [E]pistemološka osnova evolucione etike je moralni ne-realizam, ali je važan deo evolucione etike da mi mislimo da je to neka vrsta moralnog realizma (Ruse 2006b, 21-22).

Dakle, moral je *univerzalna* i moćna kolektivna iluzija, vešta obmana naših gena u funkciji promovisanja biološkog altruizma i saradnje (Ruse 2006b, Ruse and Wilson 1986). Međutim, Rus ne govori ništa o specifičnim, neposrednim, mehanizmima koji ovu iluziju čine mogućom, a takvi mehanizmi su neophodni, jer strogo govoreći, geni, iako mogu da obmanjuju, nisu obmanjivači koji mogu da plasiraju lažna verovanja. Osim toga, biološka evolucija operiše genima (ili fenotipovima ili možda grupama), a ne verovanjima. Na osnovu onoga što znamo o evoluciji i fenotipskoj plastičnosti, geni ne usađuju u ljudske mozgove nikakve gotove i nepromenljive algoritme, niti bilo koja opštija verovanja, a još manje tako specifično verovanje kao što je ono o objektivnosti morala. Da bi samoobmanjivanje moglo da bude neposredan

---

<sup>72</sup> Rus koristi sličnu analogiju s igrom i pravilima igre kakvu sam koristio objašnjavajući Mekijevu metaetičku poziciju.

mehanizam koji omogućava ovu iluziju, naslućuje se iz Rusove analogije sa spiritualizmom (Ruse 2006b). Usvajanje moralnih verovanja i verovanja o objektivnosti morala je slično prizivanju duhova: osoba veruje da komunicira s mrtvima, iako samo osluškuje glas sopstvene imaginacije. Prema Rusu, u slučaju morala taj glas bi činile odgovarajuće emocije.

Prema stanovištu koje se zastupa u ovom radu, neposredni mehanizam koji ovu iluziju čini mogućom je *samoobmanjivanje*: (specijes) specifična kognitivna karakteristika ljudi da usvajaju različita verovanja pod uticajem bioloških podsticaja i različitih emocija koji ih čine pristrasnim, što proces usvajanja verovanja čini epistemički nepouzdanim. Na taj način, ako je usvajanje pogrešnog verovanja u pojedinim situacijama i odgovarajućim uslovima životne sredine, a pod uticajem navedenih bioloških i emocionalnih činilaca, imalo važnu ulogu u preživljavanju i reprodukciji organizma, onda je prirodna selekcija mogla da favorizuje one jedinke koje usvajaju ova verovanja, iako krše uobičajene epistemičke standarde. Individue koje su se samoobmanjivale, pogrešno verujući da *treba* da sarađuju u situacijama koje imaju strukturu sličnu dilemi zatvorenika, da *ne treba* da varaju (iako bi od toga mogli da imaju koristi), da *treba* da pomažu susedima (mada bi zbog toga mogli da trpe štetu), itd., su zbog toga, a s obzirom na ono što znamo o delovanju direktnog i indirektnog reciprociteta, mogle da imaju značajne adaptivne prednosti u odnosu na druge jedinke.

U kontekstu teze o filogenetskoj kontingenciji, Rus i Vilson tvrde da je kod neke zamišljene vrste moralna sposobnost mogla da evoluirala na potpuno drugačiji način, tako da bi svaki oblik pristrasnosti bio eliminisan, ali da bi čak i u tom slučaju najmanja odstupanja usmerila jedinke te vrste u pravcu motivacione pristrasnosti (Ruse and Wilson 1986). Kada je reč o ljudima, a s obzirom na njihovo biološko ustrojstvo, mogućnost nepristrasnosti je, iako zamisliva, za ogromnu većinu ljudske populacije praktično isključena.<sup>73</sup> U kontekstu ljudskih

---

<sup>73</sup> Činjenica je da se načelno mogu formulisati različita moralna pravila koja potpuno odstupaju od bioloških imperativa preživljavanja i reprodukcije, kao što je na primer Šopenhauerovo (Artur Schopenhauer) zalaganje za univerzalno uzdržavanje od ostavljanja potomaka ili Godvinovo (William Godwin) zagovaranje krajnje nepristrasnosti. Ali, uprkos tome, uveren sam da se može pokazati da nepisani kodeksi ponašanja u kontekstu različitih kultura, kao i etički sistemi i moralna pravila definisana u kontekstu religije ili filozofije morala, retko kada gube vezu s biološkim imperativima preživljavanja i reprodukcije, pod pretpostavkom da pretenduju na praktičnu primenljivost i univerzalno važenje.



zajednica samo jedan poseban varijetet ljudske vrste se prema svim osobama ophodi na nepristrasan način. Reč je o psihopatama. Ali, psihopatama nedostaje emocionalna empatija, oni imaju brojne emocionalne i kognitivne deficite, ne mogu da se samoobmanjuju, zbog čega nisu sposobni da razumeju moralna pravila, niti da razumeju šta moralno pravilo čini moralnim, tako da su van igre koju zovemo moral (Up. Ruse and Wilson 1986, 188). Oni nemaju nikakve iluzije (osim možda iluziju o sopstvenoj grandioznosti i povlašćenom položaju), pa ni iluziju o tome da druge osobe ne treba isključivo instrumentalizovati (normativna etika), kao ni iluziju o objektivnosti morala (metaetika) (Up. Ruse 2012, 183). Da je ljudska evolucija *načelno* pratila evolucionu putanju psihopata, danas uopšte ne bismo raspravljali o moralnim problemima, jer nikakvih moralnih problema ne bi ni bilo.

Na osnovu psiholoških istraživanja znamo da je formiranje verovanja, a pogotovo moralnih verovanja, pod snažnim uticajem bioloških i emocionalnih činilaca. Ljudi nemaju apsolutnu slobodu da veruju šta god žele, već to mogu da čine samo pod izvesnim ograničenjima koje im postavljaju brojna verovanja koja su već usvojili (Kunda 1990). Pod uticajem bioloških činilaca ljudi u moralnom i socijalnom kontekstu usvajaju verovanja koja doprinose povećanju njihove adaptivne vrednosti, nezavisno od njihove istinitosti. U svetlu ovih inicijalnih moralnih verovanja mogu da formiraju nova i potrudu se da za njih pruže koherentno racionalno *opravdanje*, koje je uglavnom *post hoc* racionalizacija. Koherentnost u ovom slučaju nije *kriterijum istinitosti*, jer su inicijalna verovanja usvojena s obzirom na njihove biološke posledice, a ne zato što su istinita. Ako je samoobmanjivanje evoluiralo radi održanja koherentnosti sistema verovanja i očuvanja celovitosti uma (Ramachandran 1996), onda usvajanje koherentnog sistema moralnih verovanja, koja su sva bez razlike pogrešna, ima smisla, jer u tom slučaju „moralna konzistencija hrani mentalnu koherenciju“ (Ruse and Wilson 1986, 188). Ovde je važno podsetiti se da iz proučavanja psihopatije znamo da osobe s psihopatskim karakteristikama, kojima nedostaje moralna sposobnost, nisu zainteresovane ni za koherentnost. Izgleda da je pozivanje na fenomen samoobmanjivanja neizbežno u objašnjenju evolucije morala.

### 6.6.2 Argument na osnovu neosetljivosti

Na pokvarenom satu Martovskog kunića uvek je šest sati popodne, tako da je uvek vreme za čaj. Zamislimo da Alisa koja ne zna da sat Martovskog kunića uvek pokazuje isto vreme, bilo da je vreme za čaj ili ne, pogledavši na njegov časovnik izjavi „Vreme je za čaj“ i da se ispostavi da je *zaista* šest sati i vreme za čaj. Da li je Alisa *znala* da je šest sati i vreme za čaj? Odgovor na ovo pitanje je odričan, jer njeno verovanje nije *osetljivo* na to koliko zaista ima sati, tj. da li je stvarno šest sati i vreme za čaj ili ne. Čak i da nije bilo šest sati i vreme za čaj, Alisa bi zajedno s društvom na *Ludoj čajanci* verovala da je kucnuo čas za šoljicu čaja. Alisino verovanje da je vreme za čaj je *slučajno istinito*, a slučajno istinita verovanja nismo skloni da smatramo znanjem (Lazović 1994, 2012). Da bi Alisina izjava da je vreme za čaj bila tačna, bio bi joj neophodan neki „test realnosti“ koji bi mogao da pokaže da su njena verovanja na adekvatan način kauzalno povezana s iskustvom i spoljašnjim svetom, što bi je onda usmerilo ka istini kao korelaciji verovanja, iskustva i sveta.<sup>74</sup> Jednostavnije rečeno, Alisino verovanje da je vreme za čaj ne *prati istinu*.

Ova linija rasuđivanja se oslanja na Nozikove kondicionalne kriterijume za znanje:

- 1)  $p$  je istinito;
- 2)  $S$  veruje da  $p$ ;
- 3) Da  $p$  nije istinito,  $S$  ne bi verovao da  $p$ ;
- 4) Da je  $p$  istinito,  $S$  bi verovao da  $p$  (Nozick 1981).

Prva dva od ovih kriterijuma su preuzeta iz klasične definicije znanja kao istinitog opravdanog verovanja, dok se uslovi 3) i 4) odnose na uslov praćenja istine. Za verovanja koja zadovoljavaju uslove 3) i 4) se kaže da su osetljiva na istinu. Ovi uslovi tvrde sledeće: za bilo koju osobu  $S$  koja u aktualnom svetu veruje da  $p$  koje je istinito, važi da  $S$  u nekoj *protivčinjeničkoj situaciji* u kojoj  $p$  nije istinito, ne bi verovala da  $p$ , i da bi  $S$  koja veruje da je  $p$  istinito u aktualnom svetu, u *bliskim* svetovima u kojima je  $p$  istinito takođe verovala da  $p$ . Da bi verovanje bilo znanje moralo bi da može da prati istinu u svim *bliskim* mogućim svetovima.

---

<sup>74</sup> Zapravo, kauzalna veza je prisutna, ali nije odgovarajuće vrste, jer u datim okolnostima njeno verovanje nije osetljivo na istinu (Lazović, lična komunikacija).

Uzmimo Harmanov primer s protonom. Uslov 3) tvrdi da naučnik veruje da je proton prošao kroz maglenu komoru, i da u protivčinjeničkoj situaciji u kojoj proton nije prošao kroz maglenu komoru, naučnik ne bi verovao da je prošao. Dok uslov 4) tvrdi ne samo da je proton prošao i da naučnik to veruje, već i da kad bi proton prošao kroz maglenu komoru, naučnik bi verovao da je prošao. U situaciji u kojoj bi naučnik verovao da je proton prošao kroz maglenu komoru i da on nije prošao, njegovo verovanje ne bi bilo osetljivo na istinu.

Majkl Rus je prvi formulisao argument na osnovu neosetljivost koji je usmeren protiv moralnog realizma. Kasnije su Ričard Džojns i Šeron Strit (Sharon Street) predložili svoje sofisticiranije verzije tog argumenta. Rus u saradnji s Vilsonom tvrdi da „evoluciono objašnjenje čini objektivnu moralnost redundantnom, jer čak i da eksterne etičke premise nisu postojale, mi bismo nastavili da mislimo o ispravnom i pogrešnom na način na koji sada činimo“ (Ruse and Wilson 1986, 187). A na drugom mestu nas poziva da zamislimo dva sveta koja su potpuno identična po fizičkim svojstvima, ali se razlikuju po moralnim, tako što jedan sadrži objektivne moralne činjenice, dok ih drugi ne sadrži, a zatim tvrdi da bi ljudi u oba sveta evoluirali tako da imaju ista moralna verovanja koja imaju u aktualnom svetu i da donose iste moralne sudove koje inače donose, na primer da treba brinuti za bolesne (Ruse 1995a).

Džojns predlaže sličan argument, ali se on u njegovim kasnijim radovima ne pojavljuje:

Pretpostavite da stvarni svet sadrži stvarne kategoričke zahteve – takve koji bi moralni diskurs nužno učinili istinitim. U takvom svetu ljudi bi bili skloni da donose moralne sudove (uopštenije, da veruju da kategorički zahtevi postoje), jer bi ih prirodna selekcija učinila takvima. Sada zamislite da stvarni svet uopšte ne sadrži takve zahteve – ništa što bi moralni diskurs učinilo istinitim. U takvom svetu ljudi bi *i dalje* bili skloni da donose ove sudove (uopštenije, da veruju da kategorički zahtevi postoje), kao što su to činili u prvom svetu, jer ih je prirodna selekcija učinila takvima (Joyce 2001, 163).

Kako su ova dva argumenta prilično slična, sada ću se fokusirati samo na Rusov argument, a jedan drugačiji Džojsov argument ću izložiti kasnije. Ono što Rus tvrdi je sledeće. Ako s moralnim realitom pretpostavimo da moralne činjenice koje su istinite postoje i da imamo određena moralna verovanja, onda bismo imali *prima facie* opravdanje tih verovanja, ali ako potpuno evoluciono objašnjenje morala (moralnih verovanja) ne iziskuje pozivanje na njihovu istinitost, što je Rusovo stanovište, onda moralna verovanja nisu osetljiva na istinu, jer u slučaju

da takvih činjenica nema, mi bismo i dalje imali ista moralna verovanja. On iz toga da moralna verovanja nisu osjetljiva, zaključuje da moralnih činjenica nema. Međutim, ovaj argument nije validan, jer iz činjenice da bismo imali ista moralna verovanja koja imamo, nezavisno od toga da li moralne činjenice postoje ili ne, ne sledi da moralne činjenice ne postoje (Braddock 2012).

Osim toga, Vilenberg kritikuje Rusov epistemološki argument s obzirom na to da je u suprotnosti s tezom o supervenijenciji (Wielenberg 2010). Podsetimo se, tezom o supervenijenciji se tvrdi da za bilo koje skupove svojstava  $P$  i  $Q$ , važi da je  $P$  supervenijentno u odnosu na  $Q$  (koje je subvenijentno) *akko* dva objekta  $x$  i  $y$  ne mogu da se razlikuju s obzirom na svojstvo  $P$ , a da se ne razlikuju s obzirom na svojstvo  $Q$  (Horgan 1993). Ako prihvatimo tezu o supervenijenciji, onda dva sveta koja su identična po prirodnim svojstvima moraju da budu identična i po moralnim svojstvima. Za moralna i prirodna svojstva  $M$  i  $N$ ,  $M$  je supervenijentno u odnosu na  $N$  *akko* nije moguće da postoje dva sveta koja su identična s obzirom na prirodna svojstva  $N$ , a da se razlikuju s obzirom na moralna svojstva  $M$ . Dakle, Rusov svet koji je identičan po prirodnim svojstvima, ali se razlikuje po moralnim, i u kome imamo ista moralna verovanja kao i u aktualnom svetu, *nije mogući svet*.<sup>75</sup>

Vilenbergov argument je još suptilniji. U njemu se zapravo tvrdi da su moralna verovanja *nužno istinita*, tj. istinita u svim mogućim svetovima. Kada je sadržaj verovanja takav da uključuje nužno istinite (ili nužno neistinite) tvrdnje, onda je antecedens protivčinjeničkog kondicionala *nemoguć*, a sam kondicional *prazno istinit*. Na primer, ako je nužno istinito da je ubistvo pogrešno, onda je kondicional „da nije slučaj da je ubistvo pogrešno, mi bismo i dalje verovali da je pogrešno“ prazno istinit, budući da svet u kome nije slučaj da je ubistvo pogrešno, *nije mogući svet*. Uzimajući u obzir tezu o supervenijenciji, Vilenberg tvrdi da je teško uvideti zbog čega bi činjenica da imamo pogrešna verovanja u nemogućem svetu podrivala moralno znanje u aktualnom (Wielenberg 2010). Kako su svi protivčinjenički kondicionali s antecedensom koji nije moguć istiniti, što je neprihvatljivo, Džojls predlaže da se rešenje ovog problema može naći ako se u ocenu istinosne vrednosti protivčinjeničkog kondicionala uključe *nemogući svetovi* (tj. ako se semantika ovih kondicionala dopuni ovom opcijom) (Joyce 2014a).

---

<sup>75</sup> Argument filogenetske kontingencije koji može da se sumira tako da glasi „Da je ljudska evolucionarna istorija bila drugačija, moralna pravila bi takođe bila drugačija“ nije u suprotnosti s tezom o supervenijenciji, s obzirom na to da je svet u kome je evolucionarna istorija drugačija i u kome su pravila drugačija mogući svet.

Tada bi kondicional „da nije slučaj da je ubistvo pogrešno, mi bismo i dalje verovali da je pogrešno“ bio istinit ako i samo ako bi postojao mogući ili nemogući svet u kome ubistvo nije pogrešno i u kome bismo verovali da je pogrešno.<sup>76</sup>

Umesto da se upuštam u tehničke detalje ove analize koja prevazilazi domašaje ovog rada i odvlači pažnju od osnovnih ciljeva koje ovde želim da postignem, usredsrediću se (zajedno s Džojsem) na sam proces formiranja verovanja, bez upuštanje u ocenu istinitosti protivčinjeničkih kondicionala.

Zamislimo da sat Martovskog kunića radi savršeno; da podmazivanje satnog mehanizma najkvalitetnijim maslacem daje rezultate. Zamislimo sada da Alisa veruje da je šest sati i vreme za čaj i izjavi „Vreme je za čaj“, ali da je njeno verovanje uzrokovano time što je pre nego što se pridružila društvu na *Ludoj čajanci* pojela pečurku, od čijeg dejstva umesto da raste i smanjuje se, kako je ranije bio slučaj, sada ima određena verovanja, bilo da su istinita ili lažna. Njeno verovanje nije formirano tako što je pogledala na sat Martovskog kunića. Ali, baš kad se Alisa pridružila *Ludoj čajanci* bilo je *zaista* šest sati na časovniku Martovskog kunića i vreme za čaj. Alisa je, dakle, imala istinito verovanje da je vreme za čaj, jer je pojela komadić pečurke. Pitanje je onda, da li je Alisa *znala* da je vreme za čaj? Odgovor je opet odričan, a razlog je taj što *proces formiranja verovanja* nije epistemički pouzdan i ne prati istinu. Alisino verovanje nije opravdano. Ovo nam ukazuje na to da je način na koji dolazimo do verovanja važan, jer se reflektuje na uspeh u praćenju istine (Goldman 1979, Nozick 1981).

Važna premisa evolucionih genealoških argumenata protiv moralnog realizma je da uzročna istorija moralnih verovanja može da generiše podrivajuće zaključke o njihovoj istinitosti (Rus) ili opravdanosti (Džojse). Jedna od osnova za kritiku ovih argumenata jeste da oni počivaju na *genetičkoj grešci*. Grubo govoreći, genetička greška se odnosi na tumačenje *uzroka* kao

---

<sup>76</sup> Jedan način da se izađe na kraj s ovim problemom je da se tvrdi da Rus misli na nešto drugo. Moguće je da Rus samo hoće dodatno da ukaže na to da moralne činjenice nemaju nikakvu kauzalnu ulogu u formiranju verovanja, a da se zatim pomalo neobazrivo služi koncepcijom mogućih svetova. Ako u formiranju verovanja kauzalnu ulogu imaju samo prirodne činjenice, onda bismo u oba sveta imali ista verovanja, a moralne činjenice bi bile eksplanatorno irelevantne, bilo da ih ima ili da ih nema, jer ne dodaju niti oduzimaju bilo šta u objašnjenju porekla moralnih verovanja. To bi onda bilo samo ponavljanje argumenta koji je nezavisan od teze o supervenijenciji i analize protivčinjeničkih kondicionala.

*razloga* za ili protiv verovanja. Oslanjajući se na Sobera, Džojls smatra da jednostavno nije tačno da se način na koji dolazimo do verovanja nikako ne reflektuje na šanse za njegovu istinitost ili lažnost (istinosnu vrednost) (Joyce 2001). Možda bi bolje bilo ovu grešku shvatiti kao upozorenje na koji način evolucionari argumenti mogu da budu neuspešni nego kao stvarnu prepreku da se argument uopšte razvije (Railton 2000). Sober, naime, tvrdi da postoje genetički argumenti koji nisu pogrešni, ali koji nemaju deduktivnu, već probabilističku formu (Sober 2000). Soberova argumentacija se zasniva na sledećim primerima:

- 1) Osoba *A* izvlačeći ceduljicu s brojem 78 iz kutije u kojoj se nalaze cedulje s brojevima, izjavljuje da veruje da u prostoriji ima 78 osoba.
- 2) Osoba *B* je pažljivo prebrojala broj prisutnih i utvrdila da ih ima ukupno 104, na osnovu toga veruje da ih ima 104.

Uzimajući u obzir način na koji su verovanja formirana, na osnovu 1) možemo da zaključimo da, s obzirom na to da je verovanje dobijeno nepouzdanim procesom verovatno netačno, odnosno da u prostoriji verovatno nema 78 osoba, dok na osnovu 2) možemo da zaključimo je verovanje da u prostoriji ima 104 osobe verovatno tačno, jer je formirano na osnovu pouzdanog procesa prebrojavanja prisutnih.

Mada je teško odrediti šta jedan proces čini pouzdanim, uglavnom se tvrdi da je to tendencija procesa da učestalije vodi formiranju istinitih, nego neistinitih verovanja (Goldman 1979). Iako nije jednostavno utvrditi da li je sazajni proces pouzdan, čini se da je lakše uvideti šta ga čini nepouzdanim. Pod pretpostavkom da je moralna sposobnost koja generiše moralna verovanja proizvod evolucionog procesa, koji je načelno „nezainteresovan“ za istinosnu vrednost ili opravdanost verovanja, Džojls tvrdi da pokazati da ova sposobnost može sistematski da proizvodi neopravdana verovanja, znači pokazati da je sposobnost jednostavno nepouzdana (Joyce 2001).

Pre nego što se posvetim Džojsovom argumentu želeo bih da skrenem pažnju na jednu važnu stvar. Ako je moralna sposobnost proizvod evolucionog procesa, ali ne prirodne selekcije, nego je sporedni proizvod procesa selekcije bez adaptivne vrednosti (eng. *spandrel*), onda utoliko gore po tu sposobnost kada je reč o pouzdanosti i praćenju istine (Kahane 2011). Ako

moralna sposobnost jednim delom počiva na samoobmanjivanju, koje je po definiciji proces koji ne prati istinu, onda su izgledi da će ova sposobnost generisati istinita ili opravdana verovanja gotovo nepostojeći.

Pogledajmo sada Džojsov argument koji je zasnovan na misaonom eksperimentu koji je poznat kao *Napoleonove pilule* (Joyce 2006, 179-181).

### **6.6.3 *Napoleonove pilule***

Pretpostavimo da postoje pilule koje uzrokuju da onaj ko ih popije ima određena verovanja. Pretpostavimo da postoji crvena pilula koja uzrokuje verovanje čiji je sadržaj „Napoleon je pobedio u bici kod Vaterloa“ i druga plava pilula koja uzrokuje verovanje čiji je sadržaj „Napoleon je izgubio u bici kod Vaterloa“; i pretpostavimo da za svaku od ovih pilula postoji odgovarajući protivotrov koji neutrališe njihovo dejstvo. Pretpostavimo i to da je neko nekome u određenom periodu života podmetnuo plavu pilulu. Osoba kojoj je podmetnuta pilula sada veruje da je Napoleon izgubio bitku kod Vaterloa. Džojse se pita da li saznanje da je *istinito* verovanje posledica uzimanja pilule podriiva poverenje u to verovanje i odgovara potvrdno: iako je verovanje istinito, trebalo bi ga staviti na listu sumnjivih budućih da je dobijeno nepouzdanim saznavnim procesom.

Naravno, ovo nije kraj Džojsovog argumenta. Njegova nativistička koncepcija evolucije morala je nešto drugačija i sofisticiranija od Rusove. Podsetimo se da Rus smatra da su moralna verovanja, uključujući i verovanje o objektivnosti morala proizvod genetičkog algoritma u funkciji biološkog altruizma i saradnje. Osim što evolucionni proces ne profiliše moralna verovanja, jer verovanja nisu ono na šta prirodna selekcija deluje i uzimajući u obzir njihovu raznolikost u različitim kulturama, malo je verovatno da su specifična verovanja s određenim specifičnim sadržajem proizvod evolucionog procesa (Vidi npr. Prinz 2007, Prinz 2009). Iz tog razloga Džojse smatra da iako moralna verovanja verovatno nisu proizvod evolucionog procesa, moralni pojmovi kao što su *obaveza, vrlina, svojina, zasluga ili pravičnost* nisu toliko raznoliki i mogli bi da budu kandidati koji se dobro uklapaju u nativističku hipotezu. U tom slučaju verovanja u kojima se pojavljuju ovi pojmovi, postaju kandidati za osporavanje na osnovu genealogije (Joyce 2006, Fraser 2014).

Džojns nastavlja tako što od nas zahteva da pretpostavimo da pilule s početka ne uzrokuju neka specifična verovanja, već da je posledica njihovog delovanja sklonost da osoba koja ih je uzela formira verovanje koje uključuje određeni pojam i da se pojam o kome je reč drugačije ne bi pojavio u verovanju (Joyce 2006). Pilula sada ne generiše nikakva posebna verovanja o Napoleonu, kao što je ono da je Napoleon izgubio kod Vaterloa, već je samo reč o *Napoleonovoj piluli* koja deluje tako da kod osobe čini da ima neka načelna verovanja o Napoleonu i bez koje nijedno verovanje o Napoleonu nije moguće formirati.

Pretpostavimo da je nekoj osobi u nekom životnom periodu podmetnuta ovakva *Napoleonova pilula*. Džojns se sada pita da li ovo saznanje podriva sva verovanja o Napoleonu koje osoba ima, i odgovara potvrdno: „Verovanje je podriveno kada je jedan od pojmova koji se u njemu pojavljuju podriven“ (Joyce 2006, 181). Njegova preporuka je uzimanje protivotrova koji bi neutralisao sva pozitivna verovanja o Napoleonu koja osoba ima, nakon čega bi joj bio otvoren put za primenu nekog pouzdanog saznavnog procesa kojim bi bila u stanju da utvrdi koja su od prethodno usvojenih verovanja bila istinita. Ono što važi za *Napoleonove pilule*, važi i za moral. Budući da evolucija nije proces koji prati istinu, potpuno evoluciono objašnjenje porekla moralnih sudova dovodi u pitanje njihovu opravdanost. U nedostatku odgovarajuće evidencije za ili protiv moralnih verovanja, Džojns preporučuje agnosticizam. Oslanjajući se na Harmana, on je uveren da je jedina evidencija koja može da se pruži, zapravo priča o evolucionom poreklu ovih verovanja koja je podrivajuća. Kako se objašnjenje moralnih verovanja iscrpljuje u njihovom evolucionom objašnjenju, drugačija objašnjenja – naturalističko ili supernaturalističko – otpadaju zahvaljujući tome što su *eksplanatorno suvišna*.

Jedan od osnovnih prigovora ovom i sličnim argumentima protiv moralnog realizma je da su samoporažavajući, jer ako bi bili dosledno sprovedeni, vodili bi ne samo u moralni, već i u opšti skepticizam (Railton 2000, Shafer-Landau 2012). Razlog za to je što nije samo moralna sposobnost rezultat delovanja evolucionog procesa, već su načelno sve kognitivne sposobnosti proizvod biološke evolucije tako da bismo bili prinuđeni da odbacimo i perceptualna i matematička verovanja kao neopravdana. Međutim, među zastupnicima evolucionog antirealizma nema onih koji bi išli tako daleko da osporavaju i to da imamo neka perceptualna i matematička znanja. Rus, Džojns, kao i Strit misle da iako moralna sposobnost nije evoluirala da bi pratila istinu, druge perceptualne i kognitivne sposobnosti pružaju adaptivne prednosti



organizmima koji ih poseduju samo ako pružaju adekvatnu predstavu stvarnosti, tj. samo ako su pouzdane i osetljive na istinu.

Na primer, kada je reč o perceptualnim verovanjima, za jedinke u predačkim populacijama adekvatno opažanje stvarnosti bilo je značajno za pribavljanje korisnih informacija o okruženju i potencijalnim opasnostima. One jedinke čije perceptualne sposobnosti nisu pratile istinu imale su manje šanse da prežive i razmnože se, tako da su njihovi geni završavali u evolucionom ćorsokaku. Nije na odmet podsetiti se Kvajnovih reči da one jedinke koje greše u induktivnom rasuđivanju pokazuju „žalosnu ali hvale vrednu tendenciju da umru pre nego što se razmnože“ (Quine 1969, 126). Precizno opažati sabljozubog tigra kada se *zaista* nalazi u vidokrugu i verovati da je sabljozubi tigar u blizini je izuzetno korisno i pruža značajne adaptivne prednosti, jer sabljozubi tigrovi ubijaju, jednako kao što danas vozovi ubijaju (Ruse 2006a).

Takođe, Džojls pretpostavlja da bi jednostavna istinita aritmetička verovanja mogla da budu urođena, kao na primer da je  $1+1=2$ , s obzirom na to da su korisna i da bi jedinke u predačkim populacijama od toga mogle da imaju značajne adaptivne prednosti (Joyce 2006). Sličan argument izneo je Nozik, ali je izveo drugačije zaključke (Nozick 1981). Dok Nozik smatra da je evoluirana moralna sposobnost analogna matematičkoj kada je reč o praćenju istine, Džojls osporava ovaj zaključak.

Pogledajmo slikoviti, ali veoma simplifikovani primer na koji način matematička sposobnost koja prati jednostavne aritmetičke istine može da evoluirati (James 2011). Zamislimo dečake Freda i Barnija koji se u igraju u neposrednoj blizini pećine sabljozubog tigra. Zamislimo da dečaci opaze da u pećinu uđu dva tigra, a da nakon izvesnog vremena izađe jedan. Dečaci u tom trenutku odluče da uđu u pećinu. Ali, Fred okleva, jer mu se čini da nešto nije u redu; njegova matematička „intuicija“ mu kaže da ne ulazi. Međutim, Barni je uporan i odlučuje da sam uđe u pećinu, nakon što je Freda proglasio kukavicom. Na taj način, Barnijeva neobazrivost i matematička nekompetencija završavaju na smetlištu evolucione istorije. Barni je pokazao jadnu, ali hvale vrednu tendenciju da umre pre nego što se njegovi geni prošire populacijom.

Džojls misli da su perceptualna i matematička verovanja korisna samo ako su istinita, ali da ovo ne važi za moralna verovanja. Ona kao i neka druga verovanja dobijena procesom koji je analogan samoobmanjivanju mogu da budu korisna, samo ako su pogrešna (Joyce 2014a).<sup>77</sup> Da se poslužimo Rusovom analogijom sa spiritualizmom. Kao što verovanja koja se generišu tokom spiritualističke seanse mogu da budu *korisna* samo ako su neistinita, tj. samo na taj način mogu da pruže utehu i olakšanje rođacima preminulih, na sličan način moralna verovanja koja su proizvod sposobnosti koja je sama proizvod procesa koji ne prati istinu, mogu da ispunjavaju svoju funkciju (altruizam i saradnja) i da budu korisna, nezavisno od njihove istinosne vrednosti.

Da bi se izbegle skeptičke implikacije, realista (naturalista) bi morao da može precizno da identifikuje ona moralna verovanja koja su imuna na evolucione uticaje i uz to pouzdana (Shafer-Landau 2012). Dakle, drugi prigovor se odnosi na to da nisu sva moralna verovanja koja sadrže navedene ili neke druge moralne pojmove proizvod prirodne selekcije, već su bar neka od njih proizvod racionalne refleksije (Huemer 2008, Brosnan 2011, Fraser 2014, FitzPatrick 2015). Problem je što se kao kandidati za takva verovanja navode ona za koje je uz određeni napor moguće pronaći evoluciono objašnjenje – nepristrasna benevolencija, saosećanje s povređenim strancima, dobrota prema malim životinjama, zabrinutost za udaljene ljude i buduće generacije, itd. (Shafer-Landau 2012) – ili se navode verovanja koja su lišena bilo kakvog konkretnog sadržaja, na primer, „Ako je  $X$  bolje od  $Y$ ,  $Y$  bolje od  $Z$ , onda je  $X$  bolje od  $Z$ “ ili „Ako je dopustivo da se učini  $X$  i ako je dopustivo da se učini  $Y$ , ako  $S$  učini  $X$ , onda je dopustivo da se učine  $X$  i  $Y$ “ (Huemer 2008).

Čak i ako prihvatimo da su moralna verovanja čiji je sadržaj dobrota prema malim životinjama ili formalna tranzitivnost relacije „bolje od“, itd., prava moralna verovanja koja su rezultat racionalne refleksije i ako uz to prihvatimo da je moguće formulisati neograničen broj takvih verovanja, ipak se čini da su ona, kada ih sagledamo iz darvinističke perspektive, irelevantna. Trivijalno je tačno da je moguće formulisati neograničeni broj moralnih sudova s najrazličitijim sadržajem, ali je otvoreno pitanje da li su svi oni relevantni. Radi poređenja, jedan takav sud možemo da pozajmimo od Rusa: „Budi dobar prema kupusu petkom“ (Ruse 1995a,

---

<sup>77</sup> Džojls ovde ne govori o samoobmanjivanju, ali koristi argumente iz psiholoških teorija koje čine osnovu za deflacioni pristup problemu samoobmanjivanja.

235). Moralno verovanje da treba biti dobar prema kupusu određenog dana u nedelji je potpuno irelevantno, naspram moralnog verovanja da ne treba mučiti decu iz zabave. Broj relevantnih moralnih sudova, tj. onih koji upravljaju svakodnevnim aktivnostima ljudi (a samo ti sudovi su značajni za evoluciono razumevanje morala) je prilično ograničen. Bernard Gert (Bernard Gert), filozof koji se ne oslanja na evolucionu paradigmu, identifikuje svega deset moralnih sudova (zabrana) oko kojih se odvija moralni život i oni mogu biti svrstani u dve kategorije 1) ne povređuj druge i 2) ne narušavaj poverenje (Gert 2005). Za sve ove zdravorazumske sudove koje navodi Gert moguće je, uz manje ili više napora, dati evoluciono objašnjenje. Na početku ovog rada je istaknuto da su istraživači koji su zainteresovani za evoluciju utvrdili da je moralni život ljudi fokusiran oko sledećih pet kategorija: 1) *briga za dobrobit/povređivanje*; 2) *pravičnost/reciprocitet*; 3) *unutargrupna kohezija/lojalnost*; 4) *autoritet/poštovanje*; i 5) *čistota/svetost* (Haidt and Joseph 2007, Haidt and Bjorklund 2008). Uobičajeni moralni život ljudi je prilično banalan i nije tako kompleksan kako se to tvrdi u udžbenicima filozofije.

Zapravo u filozofskim tekstovima se susrećemo s moralnim problemima i dilemama s kakvim se obični smrtnici uglavnom ne suočavaju, a koji su našim precima u pleistocenu verovatno bili nezamislivi. Klasičan primer moralnih dilema u filozofiji je ona o pomahnitalom tramvaju (eng. Trolley problem) (Foot 2013 [1967], Thomson 1976). Međutim, ljudi se u svakodnevnom kontekstu gotovo nikada ne bave *trolologijom* (eng. trolleyology), a upravo je ovaj kontekst važan kada se bavimo evolucijom morala i efektima evolucionog procesa na moralno suđenje. Uspešnost moralnih verovanja se meri njihovim praktičnim efektom u svakodnevnom međuljudskim odnosima, zatim njihovim pozitivnim ili negativnim efektima na adaptivnu vrednost njihovih donosilaca, a ne njihovim epistemičkim statusom.

Konačno, Džojls najpre eksplicitno odbacuje stanovište prema kome je samoobmanjivanje, koje on razume kao *samoizazvano obmanjivanje*, uključeno u proces usvajanja i održanja moralnih verovanja i kao razlog navodi da za razliku od *moralnog fiktionaliste*, osoba koja se samoobmanjuje ne može da zauzme kritičku distancu prema usvojenim verovanjima, tj. ne može da ih podvrgne kritičkoj refleksiji (Joyce 2001). Međutim, on odmah zatim uočava sličnost sopstvene filozofske pozicije i Bahove neintencionalističke koncepcije samoobmanjivanja, prema kojoj se samoobmanjivanje suštinski ne odnosi na verovanja koja su mentalna stanja, već na mišljenje koje je mentalni događaj, a

samoobmanjivanje je moguće zahvaljujući asimetriji između verovanja i mišljenja (Bach 1981). Ako prihvatimo da samoobmanjivanje ima ili da je u evolucionoj prošlosti imalo značajnog uticaja na formiranje moralnih verovanja koja sada imamo, onda bi takođe trebalo da uzmemo u obzir i to da racionalna refleksija nije prepreka za samoobmanjivanje, tako da ni ona verovanja koja su rezultat refleksije ne moraju da budu pouzdana (Scott-Kakures 2002, Mele 2012). Podsetimo se, jedan od uslova za samoobmanjivanje jeste da je usvajanje pogrešnog verovanja da *p* proizvod refleksivnog, kritičkog rasuđivanja i osoba greši kada to rasuđivanje smatra pravilno usmerenim (Mele 2012, 12). Ukoliko bi to rasuđivanje bilo pravilno usmereno, ono bi moglo da predstavlja epistemički korektiv i da nam omogući da uđemo u trag istini. Međutim, ako je ljudska moralna sposobnost rezultat delovanja bioloških činilaca i ako je starija od ljudske vrste i racionalne refleksije, kako to misli jedan broj autora (de Waal 1996, de Waal et al. 2006, Flack and De Waal 2000, Warneken and Tomasello 2009a), onda su šanse za korekciju veoma male (Živanović 2012).

#### **6.6.4 Darvinistička dilema**

Argument koji iznosi Šeron Strit u svom uticajnom članku ima oblik dileme, a dolazi do sličnih rezultata do kojih dolaze Rus i Džojns (Street 2006).<sup>78</sup> Strit, takođe, smatra da je evolucija imala značajan uticaj na oblikovanje naših kognitivnih i moralnih sposobnosti, pa i na sadržaj moralnih verovanja. U skladu s tim, prva pretpostavka njenog argumenta je evolucionarna i Strit je tumači kao dobro potkrepljenu naučnu činjenicu: evolucionarni činioci su na presudan način uticali na oblikovanje moralnih verovanja usmeravajući nas ka usvajanju onih koja se pozitivno odražavaju na preživljavanje i reprodukciju. S druge strane, realisti tvrde da postoje nezavisne moralne činjenice, da moralna verovanja referiraju na ove činjenice i da su moralni sudovi iskazi koji mogu biti istiniti ili lažni i da su bar neki od njih istiniti. Zadatak koji Strit stavlja pred moralnog realistu je da objasni *relaciju* između nadmoćnih (eng. tremendous) evolucionarnih pritisaka koji su delovali na oblikovanje moralnih verovanja, s jedne strane, i nezavisnih moralnih istina kao elementarne pretpostavke realizma s druge. Dilema koju konstruiše Strit je

---

<sup>78</sup> Mada Strit govori o evaluativnim činjenicama i istinama i uprkos tome što je opseg ovih činjenica širi od opsega moralnih činjenica i istina, u daljem izlaganju ću se služiti terminologijom kojom sam se služio i do sada, tj govoriću o moralnim činjenicama, istinama, verovanjima, itd., a ne o evaluativnim činjenicama, istinama, verovanjima, itd.

takva da realista ne može da na nju odgovori, a da svoju filozofsku poziciju ne dovede u nepovoljan položaj.

Najpre, ona tvrdi da realista može da poriče da relacija uopšte postoji. Ako relacija ne postoji, onda ovakav stav vodi neprihvatljivom skepticizmu: većina naših moralnih verovanja ne prati istinu zbog *ometajućeg* delovanja evolucionih pritisaka na formiranje moralnih verovanja i donošenje moralnih sudova. Delovanje evolucionih pritisaka je ometajuće, jer je evolucija „zainteresovana“ za preživljavanje i reprodukciju, tj. za ono što pod određenim uslovima životne sredine doprinosi povećanju adaptivne vrednosti, a ne za istinitost moralnih verovanja; evolucija nam pruža podsticaje da usvajamo i održavamo ona verovanja koja omogućavaju naše individualno preživljavanje i reprodukciju ili preživljavanje i reprodukciju naših srodnika ili preživljavanje i reprodukciju osoba iz neposrednog okruženja s kojima smo u uključeni u recipročne odnose na obostranu korist. Dakle, evolucija nas usmerava tako da usvajamo (i održavamo) moralna verovanja na motivaciono pristrasan način.<sup>79</sup>

Strit kao i Džojls ne poriče da moralne istine možda postoje, ali tvrdi da je malo verovatno da ih moralna sposobnost prati, s obzirom na to da nas evolucionari pritisci usmeravaju ka usvajanju verovanja koja doprinose povećanju adaptivne vrednosti, nezavisno od njihove istinosne vrednosti. Da bi to ilustrovala, Strit se služi analogijom u kojoj praćenje moralnih istina poredi s plovidbom na Bermude brodom kojim niko ne upravlja, već je pod isključivim uticajem vetra i morskih struja. Na sličan način kao što prirodni činioци koji usmeravaju ovaj brod nemaju nikakve veze s tim u kom pravcu želimo da se on kreće, ni evolucionari činioци koji usmeravaju usvajanje moralnih verovanja na motivaciono pristrasan način, nemaju ništa s njihovom istinosnom vrednošću. Sadržaj moralnih verovanja je nezavisan od toga šta je istina, a šta ne. I kao što se može dogoditi da zahvaljujući *pukoj slučajnosti* struje i vetrovi brod nanesu do Bermuda, na sličan način se može dogoditi da se zahvaljujući *pukoj slučajnosti* usvojena moralna verovanja podudaraju s onim što realisti zovu nezavisnim moralnim istinama, ali Strit smatra da

---

<sup>79</sup> Ukoliko pogledamo uobičajene primere usvajanja verovanja na motivaciono pristrasan način (tj. uobičajene primere samoobmanjivanja) koji se obično koriste u literaturi, možemo da zapazimo da se oni uvek odnose ili na lično postignuće (poboljšana slika o sebi) ili na neposredne potomke i srodnike (majka veruje da joj sin nije narkoman, iako jeste) ili na seksualne i bračne partnere (osoba veruje da je bračni partner ne vara, iako je vara); što su primeri koji su konzistentni s predviđanjima evolucionarne teorije (Uporedi npr. Mele 2001).

je verovatnoća da se to dogodi izuzetno mala. Ona zaključuje da je u tom slučaju jedini izbor skepticizam, jer po svemu sudeći većina moralnih verovanja ne prati istinu.

Jedan od mogućih odgovora je da čak i ako teško možemo da utvrdimo gde će brod bez kormilara prepušten vetru i strujama da završi, ipak kada jednom pristane na određeno mesto, mi relativno lako možemo da utvrdimo gde se nalazimo (White 2010). Na sličan način možemo da prepoznamo moralna verovanja koja su istinita, kada na njih naiđemo. Međutim, izgleda da Vajt (Roger White) nije u pravu, jer moralna verovanja mogu da imaju motivacioni potencijal, da budu korisna i da obavljaju funkciju koju im pripisuju evolucionisti, čak i ako samo izgledaju istinito, mada to nisu (Fraser 2014).

Na drugom mestu, potencijalna izlazna strategija je da se pozovemo na racionalnu refleksiju kao epistemički korektiv pristrasnosti koju proizvode evolucionisti uticaji (Huemer 2008, Shafer-Landau 2012). Tako možemo da pretpostavimo da nisu sva, već da je samo dve trećine moralnih verovanja usvojeno pod ometajućim delovanjem evolucionističkih činilaca, dok jednoj trećini možemo da poklonimo poverenje. Zatim, kao kriterijum istinitosti možemo da uzmemo njihovu koherentnost i da verovanja koja su koherentna smatramo aproksimativno istinitim i *prima facie* opravdanim (Huemer 2008). Međutim, ovaj argument se suočava sa sledećim problemom. Verovanja domišljatog, sistematičnog i logički briljantnog paranoika bismo takođe mogli da smatramo racionalnim i epistemički opravdanim, ali svakako ne možemo da se otmemo utisku da takvom skupu verovanja nedostaje veza sa onim što obično nazivamo stvarnost (Sosa 2003). Na sličan način kao što verovanja paranoika ma koliko bila koherentna ne smatramo istinitim i epistemički opravdanim, tako i moralna verovanja koja su „kontaminirana nelegitimnim uticajima“ ne možemo da smatramo epistemički opravdanim i prihvatljivim (Street 2006, 122).

Osim toga, mogli bismo da se oslonimo na koherentistički model opravdanja kakav je *opsežni refleksivni ekvilibrijum* (Rawls 1999b, Daniels 1996), ali se onda suočavamo s dva problema. Pod pretpostavkom da je evolucionarna hipoteza tačna, u inicijalnom skupu moralnih verovanja kojim započinjemo refleksiju će se nalaziti većina onih koja su pod evolucionim uticajem. S obzirom na nadmoćno delovanje motivacione pristrasnosti na izbor moralnih verovanja, kao osnovu bismo verovatno izabrali skup verovanja koji doprinosi povećanju

adaptivne vrednosti, a ne neki drugi, što zatim vodi *Garbage In, Garbage Out* (skr. GIGO) problemu.<sup>80</sup> Ako ovaj problem ostavimo po strani i ustanovimo da skup *pozadinskih teorija* u opsežnom refleksivnom ekvilibrijumu ima funkciju testa realnosti, onda se suočavamo s tim da moralna verovanja koja smatramo racionalnim kao referentni okvir imaju različite moralne teorije, a ne nezavisne moralne činjenice.

Konačno, ovde se može dodati i ono što je već rečeno u analizi Džojsovog argumenta: racionalna refleksija nije prepreka samoobmanjivanju, pogotovo ne prilikom usvajanja onih verovanja koja mogu da se pozitivno odraze na adaptivnu vrednost jedinke. Za prosečnu individuu koherentnost sistema verovanja je od izuzetnog značaja, između ostalog i zbog toga što um ne trpi protivrečnosti, a kada se pojave nastoji da ih eliminiše, tako što teži da ponovo uspostavi narušenu koherentnost (Festinger 1962). Ali, kao što je već rečeno, ovo može da se postigne i na načine koji nisu epistemički pouzdani.

Vratimo se osnovnom argumentu. Rečeno je da moralni realista može da poriče da postoji relacija između evolucionih činilaca i moralnih istina, ali može da odgovori i potvrdno. Ako odgovori potvrdno, onda može da tvrdi da je prirodna selekcija delovala tako da moralnu sposobnost učini osetljivom na istinu i da je u tom slučaju relacija između evolucionih uticaja i moralnih istina takva da evolucionini činiooci ne deluju ometajuće i da moralna sposobnost prati istinu. Međutim, Strit ističe da je stanovište prema kome je moralna sposobnost *selektovana za praćenje moralnih istina* u suprotnosti sa svim onim što znamo o tome kako evolucionini mehanizmi deluju. Ako se argument sigurnog praćenja istine razume kao naučno objašnjenje o načinu delovanja prirodne selekcije, onda je ono manjkavo u odnosu na objašnjenje zasnovano na onome što Strit naziva *adaptivnom vezom*.

Pomenuto objašnjenje je utemeljeno na stanovištu da postoji veza između usvajanja i održavanja određenog verovanja i povećanja adaptivne vrednosti jedinke pod određenim uslovima životne sredine. U tom smislu, usvajanje moralnih verovanja i donošenje moralnih sudova je analogno *refleksnim reakcijama*. Jednako kao što se u slučaju refleksnih reakcija uspostavlja veza između određenog stimulusa i odgovarajuće neurofiziološke reakcije, tako je

---

<sup>80</sup> Ovaj pojam je preuzet iz računarstva i obično se ne prevodi, a referira na činjenicu da rezultat procesa obrade informacija zavisi od kvaliteta unetih podataka. Ukoliko input čine nepozdani podaci i autput će biti nepouzdan.

usvajanje moralnih verovanja zasnovano na uspostavljanju veze između određene situacije i odgovarajućih reakcija na tu situaciju. Na primer, kada je osoba izložena snažnom zvuku, njena refleksna reakcija je trzanje. Kada je reč o domenu morala, nešto najpribližnije refleksnim reakcijama je *impulsivno pomaganje*: kada dete zakorači pred automobil, refleksna reakcija je da ga povučemo nazad. Da bismo to učinili nije nam neophodno da proniknemo u moralnu činjenicu ispravnosti takvog postupka, dovoljno je samo da postupamo na odgovarajući način.

Čini se da je ovo u duhu Harmanovog objašnjenja moralnog suđenja koje je prema njegovom mišljenju potpuno spontano, pod pretpostavkom da imamo implicitne ili eksplicitne teorije o svetu koji nas okružuje: kada vidimo da dečaci spaljuju mačku mi jednostavno „vidimo“ da je to pogrešno. Objašnjenje na osnovu adaptivne veze je jednostavnije, elegantnije i ima veću eksplanatornu moć od objašnjenja zasnovanom na sigurnom praćenju istine. Jedan od uobičajenih kriterijuma za izbor među alternativnim objašnjenjima istog skupa podataka je jednostavnost kao indikator relativnih šansi da je teorija istinita, tako da složeni eksplanatorni modeli moraju da ustupe mesto jednostavnijim. Objašnjenje zasnovano na koncepciji sigurnog praćenja istine je kompleksnije od objašnjenja utemeljenog na adaptivnoj vezi, jer u razmatranje uvodi dodatne entitete kakvi su moralne činjenice i istine. Iz tog razloga je objašnjenje zasnovano na adaptivnoj vezi najbolje objašnjenje.

Strit kao i drugi autori čija su stanovišta ovde razmatrana ne misli da je moralna sposobnost selektovana za praćenje moralnih istina. Ako se oslonimo na Soberovu distinkciju između *selekcije za nešto* i *selekcije nečega*, prethodnom argumentu može da se prigovori da čak i ako nije postojala *selekcija za praćenje*, to ipak ne znači da moralna sposobnost koja prati moralne istine nije *selektovana* (Brosnan 2011). Ali, ako nije postojala selekcija za praćenje, onda moralna sposobnost koja prati moralne istine nije proizvod prirodne selekcije (nije adaptacija), nego je njen nusproizvod (*spandrel*), ili je proizvod nekog drugog sasvim slučajnog evolucionog procesa kakav je genetički drift. Već je rečeno da ukoliko je to slučaj, utoliko gore po moralnu sposobnost kada je reč o praćenju istine. Ako je u proces formiranja moralnih verovanja uključeno samoobmanjivanje, proces koji po definiciji ne prati istinu, onda su šanse da moralna sposobnost prati moralne istine, zanemarljive.



## ZAKLJUČAK

Osnovni cilj ovog rada je bio da se pokaže na koji način su evolucionari činili mogli da utiču na moralnu sposobnost individue i koje implikacije teorije evolucije su relevantne za etiku. U ovom, zaključnom delu rada sumiraću najvažnije rezultate do kojih sam došao baveći se ovim problemima, ali pre toga želeo bih, radi preglednosti, da ukratko sumiram osnovnu argumentaciju koja se pojavljuje u radu. Radi jednostavnosti, brojne filozofske i empirijske pretpostavke koje se inače pojavljuju u disertaciji su ovde izostavljene.

1. Evolucionari procesi deluju tako što usmeravaju individue ka biološkoj i psihološkoj pristrasnosti.
2. Individue usvajaju verovanja koja su relevantna za preživljavanje (koja se reflektuju na adaptivnu vrednost) i učestalo rasuđuju pod uticajem emocionalnih činilaca, tj. usvajaju verovanja, a zatim ih i održavaju na motivaciono pristrasan način.
3. Samoobmanjivanje je motivaciono pristrasan način usvajanja i održanja verovanje.
4. Usvajanje i održavanje moralnih verovanja može imati dalekosežne posledice po adaptivnu vrednost organizma.
5. Individue usvajaju moralna verovanja na motivaciono pristrasan način.
6. Evolucionari procesi su preko samoobmanjivanja na značajan način uticali na formiranje moralne sposobnosti.
7. Ako su evolucionari procesi preko samoobmanjivanja na značajan način uticali na formiranje moralne sposobnosti, onda se individue koje nemaju moralnu sposobnost ili koje imaju teškoća u moralnom rasuđivanju i funkcionisanju u moralnom kontekstu ne samoobmanjuju ili se samoobmanjuju u daleko manjoj meri od individua koje imaju moralnu sposobnost i nemaju teškoća u moralnom rasuđivanju.
8. Individue s psihopatskim karakteristikama nemaju moralnu sposobnost i imaju teškoća u moralnom rasuđivanju i funkcionisanju u moralnom kontekstu.
9. Individue s psihopatskim karakteristikama se ne samoobmanjuju ili se samoobmanjuju u daleko manjoj meri od individua koje imaju moralnu sposobnost i nemaju teškoća u moralnom rasuđivanju.

10. Usvajaje i održanje verovanja koja su relevantna za preživljavanje (koja se reflektuju na adaptivnu vrednost organizma) pod uticajem emocionalnih činilaca, tj. usvajanje i održavanje verovanja na motivaciono pristrasan način nije epistemički pouzdano.
11. Samoobmanjivanje nije epistemički pouzdan način usvajanja verovanja.
12. Ako usvajanje i održavanje moralnih verovanja može imati dalekosežne posledice po adaptivnu vrednost organizma i ako su evolucionim procesi preko samoobmanjivanja na značajan način uticali na formiranje moralne sposobnosti, onda moralna sposobnost nije epistemički pouzdana.
13. Moralna sposobnost nije epistemički pouzdana.
14. Ako je moralna sposobnost epistemički nepouzdana, onda je istinitost (ili lažnost) verovanja koja generiše ta sposobnost u najmanju ruku podložna sumnji.
15. Istinitost (ili lažnost) verovanja koja generiše moralna sposobnost je u najmanju ruku podložna sumnji.

Sve ono što znamo o evolucionim procesima upućuje na to da oni deluju tako što usmeravaju individue ka biološkoj i psihološkoj pristrasnosti, bilo da je reč o srodnicima, shodno predviđanjima teorije inkluzivne adaptivne vrednosti, osobama koje su pouzdani partneri u recipročnim (ili uzajamnim) relacijama, shodno predviđanjima teorije recipročnog altruizma (direktnog reciprociteta), ili članovima grupe kojoj individua pripada, što bi bilo predviđanje teorije grupne selekcije, o kojoj je, s obzirom na teškoće s kojima se suočava ova teorija, u ovom radu bilo malo reči.

Obično se tvrdi da bi sva moralna pravila trebalo da suspenduju egoizam (sebičnost) i da nam pruže nepristrasni kriterijum za donošenje važnih moralnih odluka (Williams 1976). Ali, evoluciono poreklo čoveka, njegovo biološko ustrojstvo i briga za individualno preživljavanje i reprodukciju, što je biološki imperativ, usmeravaju jedinke da u društvenom kontekstu usvajaju verovanja i rasuđuju pod uticajem snažnih emocionalnih činilaca i da daju preferencijalan status sebi, sopstvenim interesima, svojim potomcima i srodnicima, zatim osobama kojima su afektivno privrženi (seksualni partneri, prijatelji, itd.), kao i onima koji na adekvatan način odgovaraju na zahtev reciprocijacije. Ako je tako, onda jednostavno rečeno, sebičnost (egoizam) nije moguće isključiti iz objašnjenja porekla morala, niti očekivati da individue rasuđuju nepristrasno. Nije

reč o tome da osobe načelno ne mogu da rasuđuju nepristrasno. Međutim, tamo gde su njihovi biološki ili bilo koji drugi interesi u igri, to rasuđivanje je znatno otežano. Individue u uobičajenom moralnom rasuđivanju učestalo uzimaju u obzir ove interese čak i onda kada je predmet moralnog ocenjivanja neko drugi, a ne oni sami.

Nije teško osmisliti moralna pravila koja u potpunosti isključuju egoizam (sebičnost) čak i u njegovoj umerenoj varijanti. Ako bi se osoba pridržavala nekog takvog pravila, bila bi prinuđena da se prilikom pružanja eventualne pomoći ne obazire na sebe, svoje potomke, srodnike, itd. U situacijama u kojima postoji konflikt interesa ili bilo kojoj drugoj situaciji čiji aspekti mogu na ovaj ili onaj način da imaju efekata na adaptivnu vrednost, takva osoba bi se ponašala u suprotnosti s biološkim imperativom preživljavanja i reprodukcije. Njeno ponašanje bi bilo u skladu s moralnim pravilom ili pravilima, ali teško da bismo mogli da se otmemo utisku da je takvo ponašanje u najmanju ruku neobično. Na primer, za osobu koja veliki deo svojih primanja daje u humanitarne svrhe, pomažući nezbrinutoj i gladnoj deci, dok njena porodica živi u oskudici, filozofi bi verovatno rekli da je iracionalna, dok bi neko drugi upotrebio neki drugačiji, „sočniji“ i primereniji izraz. Na sreću, ljudi se uglavnom ne ponašaju na takav način; oni u većini realnih situacija jednostavno ne prekoračuju granice koje nameće biološki imperativ preživljavanja i reprodukcije. Šta više, uobičajena moralna pravila (zabrana) oko kojih se odvija svakodnevni moralni život kompatibilna su s predviđanjima evolucione teorije i retko (ili nikada) pred individue ne postavljaju zahteve za koje je malo verovatno da bi ih se iko *dosledno* pridržavao. Takođe, dileme s kojima se suočavamo u svakodnevnom životu prilično su udaljene od primera koji se pojavljuju u filozofskim tekstovima.

Budući da usvajanje i održavanje moralnih verovanja može imati dalekosežne posledice po adaptivnu vrednost organizma, kao i da postoje saznanja da je usvajanje određenih verovanja koja mogu da imaju efekata na preživljavanje i reprodukciju organizma pod snažnim uticajem emocionalnih činilaca, jedna od centralnih pretpostavki koja se pojavljuje u kontekstu ove disertacije jeste ta da je u proces usvajanja i održanja ovih verovanja uključeno samoobmanjivanje, odnosno da su evolucionarni procesi preko samoobmanjivanja na značajan način oblikovali moralnu sposobnost ljudi.

Podsticaj da se samoobmanjivanje na neki način priključi moralnoj sposobnosti pronašao sam u istraživanjima koja pokazuju da ovaj fenomen u specifičnim okolnostima koje mogu da imaju efekata na adaptivnu vrednost može pozitivno da se odražava na psihičko stanje jedinke, zatim da u situacijama konflikta interesa vodi potiskivanju misli o tuđim i sopstvenim sebičnim motivima, podupire poverenje, umanjuje konflikt između roditelja i dece, podstiče bliskost u porodičnim i partnerskim odnosima, pozitivno se odražava na deljenje resursa (hrana), itd (Surbey 2004, Surbey and McNally 1997). Jednom rečju, samoobmanjivanje u određenim socijalnim kontekstima jača socijalne veze. Svi ovi aspekti imaju značajnog udela u evoluciji saradnje, a u evolucionim teorijama se moralu pripisuje važna funkcija kada je reč o saradnji. Naime, smatra se da je osnovna funkcija morala obezbeđivanje efikasne saradnje među individuama sa suprotstavljenim interesima.

Međutim, najznačajniji podsticaj za uključivanje samoobmanjivanja u raspravu o evoluciji morala nalazi se u Triversovoj *naučnoj teoriji samoobmanjivanja* koja ovaj fenomen sagledava kao posledicu evolucione trke u naoružanju između obmanjivača i mete obmanjivanja. Prema ovoj teoriji, *prava funkcija* samoobmanjivanja je efikasnije interpersonalno obmanjivanje. Međutim, hipoteza o samoobmanjivanju u funkciji efikasnog obmanjivanja mi se činila neodrživom. Razlozi za to su bili delom konceptualni, a delom empirijski. Ako je samoobmanjivanje evoluiralo kao potpora efikasnom obmanjivanju, onda bi trebalo da su osobe koje su najveštije u obmanjivanju, istovremeno i one koje se najviše samoobmanjuju. Da bi se opovrgla *Triversova hipoteza* bilo je neophodno pronaći klasu individua koja je vešta u obmanjivanju, svakako veštija od prosečnih pripadnika populacije, ali kod koje samoobmanjivanje izostaje kada je reč o obmanjujućim smicalicama ili kod koje se samoobmanjivanje pojavljuje u znatno manjoj meri. Takva klasa individua je identifikovana među osobama s psihopatskim karakteristikama, za koje je već poznato da poseduju niz emocionalnih deficita, da ne funkcionišu u moralnom kontekstu i da nemaju odgovarajuće razumevanje moralnih pravila. Drugačije rečeno, psihopate nemaju moralnu sposobnost.

Ako psihopate nemaju moralnu sposobnost i ako se uz to ne samoobmanjuju, onda osnovna pretpostavka ove disertacije, da je samoobmanjivanje na značajan način uključeno u evoluciju, a zatim i realizaciju moralne sposobnosti, ima izgleda za uspeh. Pošavši od Kleklijeve inicijalne pretpostavke o racionalnosti psihopata, kroz analizu empirijskih podataka o

rasuđivanju osoba s psihopatskim karakteristikama, došao sam do zaključka da bi se zaista moglo očekivati da se ova klasa individua ili uopšte ne samoobmanjuje ili da se samoobmanjuje u značajno manjoj meri nego ostatak populacije. Ovo je verovatno najvažniji i istovremeno najproblematičniji zaključak ovog rada. Najpre, on je značajan zato što opovrgava *Triversovu hipotezu* koja je najistaknutija evolucionarna teorija samoobmanjivanja. Na drugom mestu, ovaj zaključak je značajan jer se do njega došlo uobičajenim filozofskim pristupom *iz naslonjače*. Na trećem mestu, ovo može biti važan doprinos proučavanju psihopatije.

Treba imati na umu da su i samoobmanjivanje i psihopatija merljive karakteristike, tj. da postoje psihološki instrumenti koji mogu da registruju prisustvo obe ove karakteristike kod ljudskih bića. Međutim, empirijski podaci koji koreliraju ove fenomene ne postoje. Budući da autor ovih redova nije obučan za sprovođenje psiholoških testiranja i statističku obradu podataka, dalje empirijsko proučavanje odnosa samoobmanjivanja i psihopatije mora da bude prepušteno psiholozima.

Uključivanje epistemički nepouzdanog procesa kakav je samoobmanjivanje u evoluciju i oblikovanje moralne sposobnosti, ima značajne metaetičke implikacije. Pod pretpostavkom da usvajanje i održavanje moralnih verovanja može imati dalekosežne posledice po adaptivnu vrednost organizma i ako su evolucionarni procesi preko samoobmanjivanja na značajan način uticali na formiranje moralne sposobnosti, onda moralna sposobnost nije epistemički pouzdana. U poslednjem poglavlju rada, u kome su razmotreni najpoznatiji argumenti koji se vezuju za *evolucionarni antirealizam*, pokazao sam na koji način koncepcija samoobmanjivanja može da bude u njih uključena i kako može da doprinese njihovoj uverljivijoj artikulaciji. Rezultat ovih analiza je ili etički antirealizam ili skepticizam.

## Navedena literatura

- ABCTVCatalyst. 2012. Dr Essi Viding Extended Interview.
- Agoramoorthy, Govindasamy, and Rasanayagam Rudran. 1992. "Adoption in Free-Ranging Red Howler Monkeys, *Alouatta Seniculus* of Venezuela." *Primates* 33 (4):551-555.
- Alexander, Richard D. 1985. "A Biological Interpretation of Moral Systems." *Zygon*® 20 (1):3-20.
- Alexander, Richard D. 2009 [1987]. *The Biology of Moral Systems*. New Brunswick [u.a.: Aldine Transactions.
- Andreoni, James, and John H Miller. 1993. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence." *The economic journal*:570-585.
- Association, American Psychiatric. 2013. *Diagnostic and Statistical Manual of Mental Disorders : Dsm-5*. Washington, DC: American Psychiatric Association.
- Audi, Robert. 1997. *Moral Knowledge and Ethical Character*. New York, N.Y.: Oxford University Press.
- Axelrod, R., and W. D. Hamilton. 1981. "The Evolution of Cooperation." *Science* 211 (4489):1390-6.
- Axelrod, Robert M. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Babić, Jovan. 1977. "Serlova teorija govornih akata i jeste-treba pitanje." *Filozofske studije* IX:151-205.
- Bach, Kent. 1981. "An Analysis of Self-Deception." *Philosophy and Phenomenological Research* 41 (March):351-370.
- Baier, Annette C. 1996. "The Vital but Dangerous Art of Ignoring: Selective Attention and Self-Deception." In *Self and Deception: A Cross-Cultural Philosophical Enquiry*, edited by Roger T. Ames and Wimal Dissanayake. Albany: SUNY Press.
- Barnard, C.J. 2004. *Animal Behaviour: Mechanism, Development, Function and Evolution*. Harlow, England: Pearson Education.
- Barnes, Annette. 2007. *Seeing through Self-Deception*. Cambridge: Cambridge University Press.
- Baron-Cohen, Simon, Alan M Leslie, and Uta Frith. 1985. "Does the Autistic Child Have a "Theory of Mind"?" *Cognition* 21 (1):37-46.

- Batson, C. Daniel. 2009. "These Things Called Empathy: Eight Related but Distinct Phenomena." In *The Social Neuroscience of Empathy*, edited by J. Decety W. Ickes, 3-15. Cambridge, MA, US: MIT Press.
- Batson, C. Daniel. 2011. *Altruism in Humans*. Oxford; New York: Oxford University Press.
- Bechara, Antoine, Hanna Damasio, Daniel Tranel, and Antonio R Damasio. 1997. "Deciding Advantageously before Knowing the Advantageous Strategy." *Science* 275 (5304):1293-1295.
- Bermúdez, José Luis. 2000. "Self-Deception, Intentions, and Contradictory Beliefs." *Analysis* 60 (268):309-319.
- Bittles, Alan H. 2004. "Genetic Aspects of Inbreeding and Incest." In *Inbreeding, Incest, and the Incest Taboo: The State of Knowledge at the Turn of the Century*, edited by Arthur P Wolf and William H Durham. Stanford University Press.
- Blair, James, Derek Robert Mitchell, and Karina Blair. 2005. *The Psychopath : Emotion and the Brain*. Malden, Mass. [u.a.]: Blackwell.
- Blair, James, and Stuart F. White. 2013. "Social Cognition in Individuals with Psychopathic Tendencies." In *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, edited by Simon Baron-Cohen, Michael Lombardo and Helen Tager-Flusberg, 364. Oxford: Oxford University Press.
- Blair, KS, RA Richell, DGV Mitchell, A Leonard, J Morton, and RJR Blair. 2006. "They Know the Words, but Not the Music: Affective and Semantic Priming in Individuals with Psychopathy." *Biological psychology* 73 (2):114-123.
- Blair, R James R. 1995. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57 (1):1-29.
- Blair, RJR, E Colledge, and DGV Mitchell. 2001. "Somatic Markers and Response Reversal: Is There Orbitofrontal Cortex Dysfunction in Boys with Psychopathic Tendencies?" *Journal of abnormal child psychology* 29 (6):499-511.
- Boesch, Christophe, Camille Bolé, Nadin Eckhardt, and Hedwige Boesch. 2010. "Altruism in Forest Chimpanzees: The Case of Adoption." *PLoS ONE* 5 (1).
- Boyd, R., and P. J. Richerson. 1988. "The Evolution of Reciprocity in Sizable Groups." *J Theor Biol* 132 (3):337-56.

- Boyd, Richard. 1997. "How to Be a Moral Realist." In *Moral Discourse and Practice : Some Philosophical Approaches*, edited by Stephen L. Darwall, Allan Gibbard and Peter Albert Railton, 105-137. New York: Oxford University Press.
- Boyd, Robert, and Peter J Richerson. 1989. "The Evolution of Indirect Reciprocity." *Social Networks* 11 (3):213-236.
- Bračanović, Tomislav. 2007. *Evolucijska teorija i priroda morala*. Zagreb: Hrvatski studiji.
- Braddock, Matthew C. 2012. "Debunking Challenges to Moral Realism." Duke University.
- Brandon, Robert. 1982. "The Levels of Selection." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982:315 - 323.
- Brosnan, Kevin. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology & Philosophy* 26 (1):51-64.
- Brown, Kevin M. 1998. "Proximate and Ultimate Causes of Adoption in Ring-Billed Gulls." *Animal Behaviour* 56 (6):1529-1543. doi: <http://dx.doi.org/10.1006/anbe.1998.0913>.
- Bshary, Redouan, and Alexandra S Grutter. 2006. "Image Scoring and Cooperation in a Cleaner Fish Mutualism." *Nature* 441 (7096):975-978.
- Buck, Ross. 2014. *Emotion: A Biosocial Synthesis*. New York: Cambridge University Press.
- Buller, David J. 2005. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, Mass.: MIT Press.
- Buss, David M. 2009. "How Can Evolutionary Psychology Successfully Explain Personality and Individual Differences?" *Perspectives on Psychological Science* 4 (4):359-366.
- Carr, Laurie, Marco Iacoboni, Marie-Charlotte Dubeau, John C Mazziotta, and Gian Luigi Lenzi. 2003. "Neural Mechanisms of Empathy in Humans: A Relay from Neural Systems for Imitation to Limbic Areas." *Proceedings of the national Academy of Sciences* 100 (9):5497-5502. doi: 10.1073/pnas.0935845100.
- Carter, Gerald G, and Gerald S Wilkinson. 2013. "Food Sharing in Vampire Bats: Reciprocal Help Predicts Donations More Than Relatedness or Harassment." *Proceedings of the Royal Society B: Biological Sciences* 280 (1753):20122573.
- Cartwright, J. 2000. *Evolution and Human Behavior: Darwinian Perspectives on Human Nature*. Cambridge, Mass.: MIT Press.



- Cäsar, Cristiane, and RobertJohn Young. 2008. "A Case of Adoption in a Wild Group of Black-Fronted Titi Monkeys (*Callicebus Nigrifrons*)." *Primates* 49 (2):146-148. doi: 10.1007/s10329-007-0066-x.
- Churchland, P.S. 2011. *Braintrust: What Neuroscience Tells Us About Morality*. Princeton and Oxford: Princeton University Press.
- Cleckley, Hervey Milton. 1988. *The Mask of Sanity : An Attempt to Clarify Some Issues About the So-Called Psychopathic Personality*. Augusta, Géorgie: E. S. Cleckley.
- Coleman, James S. 1990. "The Emergence of Norms." In *Social Institutions: Their Emergence, Maintenance and Effects* edited by Michael Hechter, Karl-Dieter Opp and Reinhard Wippler, 35-59. New York: Transaction Publishers.
- Cosmides, Leda, and John Tooby. 1992. "Cognitive Adaptations for Social Exchange." In *The adapted mind : evolutionary psychology and the generation of culture* edited by Jerome H. Barkow, Leda Cosmides, John Tooby. New York: Oxford University Press
- Curry, Oliver, Matthew Jones Chesters, and Essi Viding. 2011. "The Psychopath's Dilemma: The Effects of Psychopathic Personality Traits in One-Shot Games." *Personality and individual differences* 50 (6):804-809.
- Cushman, Fiery, Kurt Gray, Allison Gaffey, and Wendy Berry Mendes. 2012. "Simulating Murder: The Aversion to Harmful Action." *Emotion* 12 (1):2.
- Cushman, Fiery, Liane Young, and Marc Hauser. 2006. "The Role of Conscious Reasoning and Intuition in Moral Judgment Testing Three Principles of Harm." *Psychological science* 17 (12):1082-1089.
- Daly, Martin, and Margo Wilson. 2000. *The Truth About Cinderella: A Darwinian View of Parental Love*. New Haven [u.a.]: Yale Univ. Press.
- Daly, Martin, and Margo Wilson. 2001. "An Assessment of Some Proposed Exceptions to the Phenomenon of Nepotistic Discrimination against Stepchildren." *Annales Zoologici Fennici*.
- Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Avon Books.
- Damuth, John, and I Lorraine Heisler. 1988. "Alternative Formulations of Multilevel Selection." *Biology and Philosophy* 3 (4):407-430.

- Daniels, Norman. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge [England]; New York: Cambridge University Press.
- Danziger, Nicolas, Isabelle Faillenot, and Roland Peyron. 2009. "Can We Share a Pain We Never Felt? Neural Correlates of Empathy in Patients with Congenital Insensitivity to Pain." *Neuron* 61 (2):203-212.
- Darwin, Charles. 1981. *The Descent of Man, and Selection in Relation to Sex*. Princeton, N.J.: Princeton University Press.
- Darwin, Charles. 2009a. *The Expression of the Emotions in Man and Animals*. New York: Cambridge University Press.
- Darwin, Charles. 2009b. *The Origin of Species: By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Cambridge: Cambridge University Press.
- Davidson, Donald. 1982. "Paradoxes of Irrationality." In *Philosophical Essays on Freud*, edited by Richard Wollheim and James Hopkins. New York: Cambridge University Press.
- Davidson, Donald. 1987. "Deception and Division." In *The Multiple Self*, edited by Jon Elster, 79. Cambridge [Cambridgeshire]; New York : Cambridge University Press.
- Davidson, Donald. 1991. "On the Very Idea of a Conceptual Scheme." In *Inquiries into Truth and Interpretation*, edited by Donald Davidson, 183-189. Oxford [Oxfordshire]; New York: Clarendon Press ; Oxford University Press.
- Dawkins, Richard. 1999 [1982]. *The Extended Phenotype : The Long Reach of the Gene*. Oxford: Oxford University Press.
- Dawkins, Richard. 2006 [1976]. *The Selfish Gene: 30th Anniversary Edition*. Oxford: Oxford University Press
- de Quervain, Dominique JF J, Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. "The Neural Basis of Altruistic Punishment." *Science (New York, NY)* 305 (5688):1254-1258.
- de Waal, Frans B. M. 1991. "The Chimpanzee's Sense of Social Regularity and Its Relation to the Human Sense of Justice." *American Behavioral Scientist* 34 (3):335-349. doi: 10.1177/0002764291034003005.
- de Waal, Frans B. M. 1996. *Good Natured*. Harvard: Harvard University Press.

- de Waal, Frans B. M. 2008. "Putting the Altruism Back into Altruism: The Evolution of Empathy." *Annual Review of Psychology* 59 (1):279-300.
- de Waal, Frans B. M. 2012. "The Antiquity of Empathy." *Science* 336 (6083):874-876.
- de Waal, Frans B. M. 2014. "Natural Normativity: The Is and Ought of Animal Behavior." *BEH Behaviour* 151 (2-3):185-204.
- de Waal, Frans B. M., S. Macedo, J. Ober, and R. Wright. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.
- Decety, Jean. 2011a. "Dissecting the Neural Mechanisms Mediating Empathy." *Emotion Review* 3 (1):92-108.
- Decety, Jean. 2011b. "The Neuroevolution of Empathy." *Annals of the New York Academy of Sciences* 1231 (1):35-45. doi: 10.1111/j.1749-6632.2011.06027.x.
- Decety, Jean, and Meghan Meyer. 2008. "From Emotion Resonance to Empathic Understanding: A Social Developmental Neuroscience Account." *Development and psychopathology* 20 (04):1053-1080.
- Dekart, Rene. 1998. *Meditacije o prvoj filozofiji u kojima se dokazuje božja egzistencija i razlika između ljudske duše i tijela : [(prilog kritici građanskog prava)]*. Vol. 11, *Biblioteka Episteme*. Beograd: Plato.
- Dennett, Daniel C. 1996. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
- Dennett, Daniel C. 2013. "The Evolution of Evaluators." In *The Evolution of Economic Diversity*, edited by Antonio Nicita and Ugo Pagano, 66-83. London: Routledge.
- Douven, Igor. 2011. "Abduction." Accessed June 25, 2015. <http://plato.stanford.edu/archives/spr2011/entries/abduction/>.
- Dugatkin, Lee Alan. 1998. "Game Theory and Cooperation." In *Game Theory and Animal Behavior*, edited by Lee Alan Dugatkin, and Hudson Kern Reeve, 38-63. Oxford: Oxford University Press.
- Dugatkin, Lee Alan. 2000. *The Imitation Factor: Evolution Beyond the Gene*. New York: Free Press.
- Dugatkin, Lee Alan. 2007. "Inclusive Fitness Theory from Darwin to Hamilton." *Genetics* 176 (3):1375-1380.

- Dunbar, Robin. 2009. "Kinship in Biological Perspective." In *Early Human Kinship: From Sex to Social Reproduction*, edited by N.J. Allen, H. Callan, R. Dunbar and W. James, 131-150. Chichester, U.K.: Wiley.
- Eisenberger, Naomi I, Matthew D Lieberman, and Kipling D Williams. 2003. "Does Rejection Hurt? An Fmri Study of Social Exclusion." *Science* 302 (5643):290-292.
- Elster, Jon. 1987. *The Multiple Self*. Cambridge [Cambridgeshire]; New York: Cambridge University Press.
- Ermer, Elsa, and Kent A Kiehl. 2010. "Psychopaths Are Impaired in Social Exchange and Precautionary Reasoning." *Psychological science*. 21 (10): 1399-1405.
- Fallon, James. 2013. *The Psychopath Inside: A Neuroscientist's Personal Journey into the Dark Side of the Brain*. New York: Penguin Publishing Group.
- Fecteau, Shirley, Alvaro Pascual-Leone, and Hugo Théoret. 2008. "Psychopathy and the Mirror Neuron System: Preliminary Findings from a Non-Psychiatric Sample." *Psychiatry research* 160 (2):137-144.
- Fehr, E., H. Bernhard, and B. Rockenbach. 2008. "Egalitarianism in Young Children." *Nature* 454 (7208):1079-83. doi: 10.1038/nature07155.
- Fehr, E., and S. Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868):137-40. doi: 10.1038/415137a.
- Festinger, Leon. 1962. "Cognitive Dissonance." *Scientific American* 207 (4):93-106.
- Fingarette, Herbert. 1998. "Self-Deception Needs No Explaining." *Philosophical Quarterly* 48 (192):289-301.
- Fisher, Andrew. 2011. *Metaethics: An Introduction*. Durham: Acumen Pub.
- Fisher, Ronald Aylmer. 1930. *The Genetical Theory of Natural Selection*. Oxford: The Clarendon Press.
- FitzPatrick, William J. 2015. "Debunking Evolutionary Debunking of Ethical Realism." *Philosophical Studies* 172 (4):883-904.
- Flack, Jessica C, and Frans BM De Waal. 2000. "'Any Animal Whatever'. Darwinian Building Blocks of Morality in Monkeys and Apes." *Journal of Consciousness Studies* 7 (1-2):1-29.
- Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review*:305-316.

- Foot, Philippa. 2013 [1967]. "The Problem of Abortion and the Doctrine of the Double Effect." In *Ethical Theory: An Anthology*, edited by Russ Shafer-Landau. Chichester, West Sussex: Wiley-Blackwell.
- Frankena, William K. 1967. "The Naturalistic Fallacy." In *Theories of Ethics*, edited by Philippa Foot. London: Oxford University Press.
- Fraser, Benjamin James. 2014. "Evolutionary Debunking Arguments and the Reliability of Moral Cognition." *Philosophical Studies* 168 (2):457-473.
- Frege, Gottlob. 1995. *Osnove aritmetike i drugi spisi*. Zagreb: Kruzak.
- Friedrich, James. 1993. "Primary Error Detection and Minimization (Pedin) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena." *Psychological review* 100 (2):298.
- Gallese, Vittorio, and Alvin Goldman. 1998. "Mirror Neurons and the Simulation Theory of Mind-Reading." *Trends in cognitive sciences* 2 (12):493-501.
- Gao, Yu, and Adrian Raine. 2010. "Successful and Unsuccessful Psychopaths: A Neurobiological Model." *Behavioral sciences & the law* 28 (2):194-210.
- Gauthier, David P. 1986. *Morals by Agreement*. Oxford [Oxfordshire]; New York: Clarendon Press ; Oxford University Press.
- Gert, Bernard. 2005. *Morality: Its Nature and Justification*. Oxford: Oxford University Press.
- Giudice, Marco Del, Bruce J. Ellis, and Elizabeth A. Shirtcliff. 2011. "The Adaptive Calibration Model of Stress Responsivity." *Neuroscience and biobehavioral reviews* 35 (7):1562-1592. doi: 10.1016/j.neubiorev.2010.11.007.
- Glenn, Andrea L, Robert Kurzban, and Adrian Raine. 2011. "Evolutionary Theory and Psychopathy." *Aggression and Violent Behavior* 16 (5):371-380.
- Godfrey-Smith, Peter. 2000. "The Replicator in Retrospect." *Biology and Philosophy* 15 (3):403-423.
- Goldman, Alvin I. 1979. "What Is Justified Belief?" In *Justification and Knowledge*, edited by Alvin I. Goldman and George Sotiros Pappas, 1-23. Springer Netherlands: Springer.
- Gould, S. J., and R. C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London. Series B, Biological sciences* 205 (1161):581-98.

- Gould, Stephen Jay. 1992. *The Panda's Thumb: More Reflections in Natural History*. New York: WW Norton & company.
- Gould, Stephen, and Elisabeth Vrba. 1982. "Exaptation—a Missing Term in the Science of Form." *Paleobiology* 8 (1):4-15. doi: 10.2307/2400563.
- Greene, Joshua D. 2007. "The Secret Joke of Kant's Soul." In *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong. Cambridge, Mass.: MIT Press.
- Greene, Joshua D, Fiery A Cushman, Lisa E Stewart, Kelly Lowenberg, Leigh E Nystrom, and Jonathan D Cohen. 2009. "Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment." *Cognition* 111 (3):364-371.
- Grisolia, James Santiago. 2001. "Neurobiology of the Psychopath." In *Violence and Psychopathy*, edited by Adrian Raine and José Sanmartín, 79-87. Springer.
- Gur, Ruben C, and Harold A Sackeim. 1979. "Self-Deception: A Concept in Search of a Phenomenon." *Journal of Personality and Social Psychology* 37 (2):147.
- Haidt, Jonathan, and Fredrik Bjorklund. 2008. "Social Intuitionists Answer Six Questions About Morality." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, Mass: MIT Press.
- Haidt, Jonathan, and Craig Joseph. 2007. "The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules." In *The Innate Mind* edited by Peter Carruthers, Stephen Laurence and Stephen P. Stich. Oxford: Oxford University Press.
- Haldane, J. B. S. 1932. *The Causes of Evolution*. London; New York: Longmans, Green and Co.
- Haldane, John BS. 1955. "Population Genetics." *New Biology* 18:34-51.
- Halpin, Zuleyma T. 1991. "Kin Recognition Cues of Vertebrates." In *Kin Recognition*, edited by Peter G Hepper, 220-258. Cambridge: Cambridge University Press.
- Hamilton, William D. 1963. "The Evolution of Altruistic Behavior." *American naturalist*:354-356.
- Hamilton, William D. 1964a. "The Genetical Evolution of Social Behaviour. I." *Journal of Theoretical Biology* 7 (1):1-16. doi: [http://dx.doi.org/10.1016/0022-5193\(64\)90038-4](http://dx.doi.org/10.1016/0022-5193(64)90038-4).
- Hamilton, William D. 1964b. "The Genetical Evolution of Social Behaviour. II." *Journal of Theoretical Biology* 7 (1):17-52. doi: [http://dx.doi.org/10.1016/0022-5193\(64\)90039-6](http://dx.doi.org/10.1016/0022-5193(64)90039-6).

- Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge [Cambridgeshire]; New York: Cambridge University Press.
- Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162 (3859):1243-1248. doi: 10.2307/1724745.
- Hare, Robert D. 1998. "Psychopathy, Affect and Behavior." In *Psychopathy: Theory, Research and Implications for Society*, edited by David J. Cooke, Adelle E. Forth and Robert D. Hare, 105-137. Springer.
- Hare, Robert D. 2001. "Psychopaths and Their Nature." In *Violence and Psychopathy*, edited by Adrian Raine and José Sanmartín, 5-34. Springer.
- Hare, Robert D, Janice Frazelle, and David N Cox. 1978. "Psychopathy and Physiological Responses to Threat of an Aversive Stimulus." *Psychophysiology* 15 (2):165-172.
- Hare, Robert D. 1999. *Without Conscience : The Disturbing World of the Psychopaths among Us*. New York: Guilford Press.
- Hare, RobertD, AdelleE Forth, and StephenD Hart. 1989. "The Psychopath as Prototype for Pathological Lying and Deception." In *Credibility Assessment*, edited by JohnC Yuille, 25-49. Springer Netherlands.
- Harman, Gilbert. 1973. *Thought*. Princeton, N.J.: Princeton University Press.
- Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
- Harman, Gilbert. 1986. "Moral Explanations of Natural Facts—Can Moral Claims Be Tested against Moral Reality?" *The Southern Journal of Philosophy* 24 (S1):57-68.
- Harman, Gilbert. 1999. *Reasoning, Meaning, and Mind*. Oxford; New York: Clarendon Press; Oxford University Press.
- Harman, Gilbert. 2010. "Gilbert Harman." In *A Companion to Epistemology*, edited by Jonathan Dancy, Ernest Sosa and Matthias Steup. Malden, MA: Wiley-Blackwell; John Wiley & Sons.
- Harsanyi, John C. 1958. "Ethics in Terms of Hypothetical Imperatives." *Mind* 67 (267):305-316.
- Hart, Stephen D, Robert D Hare, and Timothy J Harpur. 1992. "The Psychopathy Checklist—Revised (Pcl-R)." In *Advances in Psychological Assessment*, 103-130. Springer.
- Hatfield, E., J.T. Cacioppo, and R.L. Rapson. 1994. *Emotional Contagion*. Cambridge [England]; New York: Cambridge University Press.

- Hickok, Gregory. 2009. "Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans." *Journal of cognitive neuroscience* 21 (7):1229-1243.
- Hirstein, William. 2004. *Brain Fiction: Self-Deception and the Riddle of Confabulation*. Cambridge, Mass: MIT Press.
- Hoffman, Martin L. 1981. "Perspectives on the Difference between Understanding People and Understanding Things: The Role of Affect." In *Social Cognitive Development*, edited by John H. Flavell and Lee Ross, 67-81. Cambridge: Cambridge University Press.
- Hoffman, Martin L. 2000. *Empathy and Moral Development : Implications for Caring and Justice*. Cambridge, U.K.; New York: Cambridge University Press.
- Horgan, Terence. 1993. "From Supervenience to Superdupervenience: Meeting the Demands of a Material World." *Mind* 102 (408):555-586. doi: 10.2307/2254046.
- Horgan, Terence, and Mark Timmons. 1991. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16:447-465.
- Huemer, Michael. 2008. "Revisionary Intuitionism." *Social Philosophy and Policy* 25 (01):368-392.
- Hughes, Melissa A, Mairead C Dolan, Jennifer S Trueblood, and Julie C Stout. 2014. "Psychopathic Personality Traits and Iowa Gambling Task Performance in Incarcerated Offenders." *Psychiatry, Psychology and Law* (ahead-of-print):1-11.
- Hull, David L. 1980. "Individuality and Selection." *Annual Review of Ecology and Systematics* 11:311-332. doi: 10.2307/2096911.
- Hull, David L. 2001. *Science and Selection : Essays on Biological Evolution and the Philosophy of Science*. Cambridge: Cambridge University Press.
- Hume, David Selby-Bigge L. A. 1960. *A Treatise of Human Nature*. Oxford: Clarendon Press.
- Huxley, Thomas H. 2009. *Evolution and Ethics: Delivered in the Sheldonian Theatre, May 18, 1893*. Cambridge: Cambridge University Press.
- Iacoboni, Marco. 2009. "Imitation, Empathy, and Mirror Neurons." *Annual review of psychology* 60:653-670.
- Iacoboni, Marco, and Mirella Dapretto. 2006. "The Mirror Neuron System and the Consequences of Its Dysfunction." *Nature Reviews Neuroscience* 7 (12):942-951.
- Izar, Patricia, Michele P. Verderane, Elisabetta Visalberghi, Eduardo B. Ottoni, Marino Gomes De Oliveira, Jeanne Shirley, and Dorothy Fragaszy. 2006. "Cross-Genus Adoption of a



- Marmoset (*Callithrix jacchus*) by Wild Capuchin Monkeys (*Cebus libidinosus*): Case Report." *American Journal of Primatology* 68 (7):692-700. doi: 10.1002/ajp.20259.
- Jacobs, Gerald H. 1996. "Primate Photopigments and Primate Color Vision." *Proceedings of the National Academy of Sciences* 93 (2):577-581.
- James, Scott M. 2011. *An Introduction to Evolutionary Ethics*. Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.
- Johannsen, W. 1909. *Elemente Der Exakten Erblchkeitslehre: Deutsche Wesentlich Erweiterte Ausgabe in Fünfundzwanzig Vorlesungen*: G. Fischer.
- Johnston, Mark. 1988. "Self-Deception and the Nature of Mind." In *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amélie Rorty, 63-91. Berkeley: University of California Press.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge ; New York: Cambridge University Press.
- Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge, Mass.: MIT Press.
- Joyce, Richard. 2014a. "Evolution, Truth-Tracking, and Moral Skepticism." *Problems of Goodness: New Essays on Metaethics*.
- Joyce, Richard. 2014b. "The Origins of Moral Judgment." *BEH Behaviour* 151 (2-3):261-278.
- Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Noûs* 45 (1):103-125.
- Kant, Immanuel. 1979. *Kritika praktičkog uma*. Beograd: BIGZ.
- Kant, Immanuel. 2008. *Zasnivanje metafizike morala*. Beograd: Dereta.
- Kavka, G.S. 1986. *Hobbesian Moral and Political Theory*. Princeton, N.J.: Princeton University Press.
- Keller, Evelyn Fox. 2000. *The Century of the Gene*. Cambridge, Mass: Harvard University Press.
- Kelly, Daniel, Stephen Stich, Kevin J Haley, Serena J Eng, and Daniel MT Fessler. 2007. "Harm, Affect, and the Moral/Conventional Distinction." *Mind & Language* 22 (2):117-131.
- Kirchin, Simon. 2012. *Metaethics*. New York: Palgrave Macmillan.
- Kitcher, Philip. 1994. "Four Ways of "Biologizing" Ethics." In *Conceptual Issues in Evolutionary Biology*, edited by E. Sober, 439--50. Cambridge, Mass.: The MIT Press. Bradford Books.

- Kitcher, Philip. 2006a. "Between Fragile Altruism and Morality: Evolution and the Emergence of Normative Guidance." In *Evolutionary Ethics and Contemporary Biology*, edited by Giovanni Boniolo and Gabriele De Anna. Cambridge: Cambridge University Press.
- Kitcher, Philip. 2006b. "Biology and Ethics." In *The Oxford Handbook of Ethical Theory*, edited by David Copp. New York: Oxford University Press.
- Kleinginna, Paul R., and Anne M. Kleinginna. 1981. "A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition." *Motivation and emotion* 5 (4):345-379.
- Koenigs, Michael, Michael Kruepke, and Joseph P Newman. 2010. "Economic Decision-Making in Psychopathy: A Comparison with Ventromedial Prefrontal Lesion Patients." *Neuropsychologia* 48 (7):2198-2204.
- Koenigs, Michael, and Joseph P Newman. 2013. "The Decision-Making Impairment in Psychopathy: Psychological and Neurobiological Mechanisms." *Handbook on Psychopathy and Law*:93.
- Korchmaros, Josephine D, and David A Kenny. 2001. "Emotional Closeness as a Mediator of the Effect of Genetic Relatedness on Altruism." *Psychological science* 12 (3):262-265.
- Korchmaros, Josephine D, and David A Kenny. 2006. "An Evolutionary and Close-Relationship Model of Helping." *Journal of social and Personal Relationships* 23 (1):21-43.
- Korsgaard, Christine M. 1986. "The Right to Lie: Kant on Dealing with Evil." *Philosophy & Public Affairs* 15 (4):325-349.
- Krebs, Dennis L. 2011. *The Origins of Morality: An Evolutionary Account*. Oxford University Press, USA.
- Kripke, Saul A. 2011 [1971]. "Identity and Necessity." In *Philosophical Troubles: Collected Papers* edited by Saul A. Kripke, 1-26. New York: Oxford University Press.
- Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3):480-98.
- Lalumière, Martin L., Sandeep Mishra, and Grant T. Harris. 2008. "In Cold Blood: The Evolution of Psychopathy." In *Evolutionary Forensic Psychology: Darwinian Foundations of Crime and Law*, edited by J. D. Duntley and T. K. Shackelford, 176-197. New York, NY, US: Oxford University Press.
- Lazar, Ariela. 1999. "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs under the Influence." *Mind* 108 (430):265-290.

- Lazović, Živan. 1994. *O prirodi epistemičkog opravdanja*. Beograd: Filozofsko društvo Srbije.
- Lazović, Živan. 2012. *Problem filozofskog skepticizma*. Beograd: Institut za filozofiju, Filozofski fakultet Univerziteta u Beogradu.
- Levy, Neil. 2007. *Neuroethics: Challenges for the 21st Century*. Cambridge: Cambridge University Press.
- Levy, Neil. 2010. "Psychopathy, Responsibility and the Moral/Conventional Distinction." In *Responsibility and Psychopathy : Interfacing Law, Psychiatry, and Philosophy*, edited by Luca Malatesti and John McMillan. Oxford; New York: Oxford University Press.
- Lewontin, R C. 1970. "The Units of Selection." *Annual Review of Ecology and Systematics* 1 (1):1-18. doi: doi:10.1146/annurev.es.01.110170.000245.
- Lipton, Peter. 2001. "Inference to the Best Explanation." In *Companion to the Philosophy of Science.*, edited by W. H. Newton-Smith, 184-193. Oxford: Blackwell Publishers Ltd.
- Lloyd, Elisabeth A. 2007. "Units and Levels of Selection." In *The Cambridge Companion to the Philosophy of Biology*, edited by D.L. Hull and M. Ruse. Cambridge University Press.
- Mackie, John. 1990 [1977]. *Ethics: Inventing Right and Wrong*. Harmondsworth ; New York: Penguin Books Limited.
- Madsen, Elaine A, Richard J Tunney, George Fieldman, Henry C Plotkin, Robin IM Dunbar, Jean-Marie Richardson, and David McFarland. 2007. "Kinship and Altruism: A Cross-Cultural Experimental Study." *British Journal of Psychology* 98 (2):339-359.
- Marsh, Abigail A, and RJR Blair. 2008. "Deficits in Facial Affect Recognition among Antisocial Populations: A Meta-Analysis." *Neuroscience & Biobehavioral Reviews* 32 (3):454-465.
- Marsh, Abigail A, Elizabeth C Finger, Derek GV Mitchell, Marguerite E Reid, Courtney Sims, David S Kosson, Kenneth E Towbin, Ellen Leibenluft, Daniel S Pine, and RJR Blair. 2008. "Reduced Amygdala Response to Fearful Expressions in Children and Adolescents with Callous-Unemotional Traits and Disruptive Behavior Disorders."
- Maynard Smith, J. 1987. "Reply to Sober." In *The Latest on the Best: Essays on Evolution and Optimality*, edited by John Dupre. Cambridge, Mass: MIT Press.
- Maynard Smith, John. 1964. "Group Selection and Kin Selection." *Nature* 201 (4924):1145-1147.
- Maynard Smith, John. 1976. "Group Selection." *The Quarterly Review of Biology* 51 (2):277-283. doi: 10.2307/2823631.

- Maynard Smith, John. 1998. "The Origin of Altruism." *Nature* 393:639-640.
- Mayr, E. 1961. "Cause and Effect in Biology." *Science* 134 (3489):1501-6.
- Mayr, E. 1998. *This Is Biology: The Science of the Living World*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- McDowell, John. 1985. "Values and Secondary Qualities." In *Objectivity and Morality*, edited by Ted Honderich, 110-129. London: Routledge.
- McKenzie Alexander, J. 2008. "Cooperation." In *A Companion to the Philosophy of Biology*, edited by Sahotra Sarkar, and Anya Plutynski, 415-430. Malden, Mass.; Oxford: John Wiley & Sons.
- McLaughlin, Brian P. 1988. "Exploring the Possibility of Self-Deception in Belief." In *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amélie Rorty. Berkeley: University of California Press.
- McLaughlin, Brian P. 1996. "On the Very Possibility of Self-Deception." In *Self and Deception: A Cross-Cultural Philosophical Enquiry*. Albany: SUNY Press.
- Mealey, Linda. 1995. "The Sociobiology of Sociopathy: An Integrated Evolutionary Model." *Behavioral and Brain sciences* 18 (03):523-541.
- Mele, Alfred R. 1987. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press.
- Mele, Alfred R. 2001. *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Mele, Alfred R. 2012. "When Are We Self-Deceived?" *Humana Mente Journal of Philosophical Studies* (20): 1-15.
- Miller, Alexander. 2003. *An Introduction to Contemporary Metaethics*. Cambridge, UK; Malden, MA: Polity Press; Blackwell Publishers.
- Millikan, Ruth G. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, Mass.; London: MIT Press
- Millikan, Ruth G. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge: MIT Press.
- Millikan, Ruth G. 2002. "Biofunctions: Two Paradigms." In *Functions*, edited by Andre Ariew, Robert Cummins and Mark Perlman, 113-143. Oxford, New York: Oxford University Press.

- Mokros, Andreas, Birgit Menner, Hedwig Eisenbarth, Georg W Alpers, Klaus W Lange, and Michael Osterheider. 2008. "Diminished Cooperativeness of Psychopaths in a Prisoner's Dilemma Game Yields Higher Rewards." *Journal of abnormal psychology* 117 (2):406.
- Moore, G. E. 1963. *Principi etike*. Beograd: Nolit.
- Muller, Martin N, and John C Mitani. 2005. "Conflict and Cooperation in Wild Chimpanzees." *Advances in the Study of Behavior* 35:275-331.
- Murphy, Dominic, and Stephen Stich. 2000. "Darwin in the Madhouse: Evolutionary Psychology and the Classification of Mental Disorders." In *Evolution and the Human Mind*, edited by Peter Carruthers and A. Chamberlain, 62-92. Cambridge University Press.
- Nichols, Shaun. 2004. *Sentimental Rules : On the Natural Foundations of Moral Judgment*. Oxford; New York: Oxford University Press.
- Nichols, Shaun, and Trisha Folds-Bennett. 2003. "Are Children Moral Objectivists? Children's Judgments About Moral and Response-Dependent Properties." *Cognition* 90 (2):B23-B32.
- Nowak, M. A., R. M. May, and K. Sigmund. 1995. "The Arithmetics of Mutual Help." *Scientific American* 272 (6):76-81.
- Nowak, Martin A, and Karl Sigmund. 1992. "Tit for Tat in Heterogeneous Populations." *Nature* 355 (6357):250-253.
- Nowak, Martin A., and Karl Sigmund. 2005. "Evolution of Indirect Reciprocity." *Nature* 437 (7063):1291-1298.
- Nowak, Martin, and Karl Sigmund. 1993. "A Strategy of Win-Stay, Lose-Shift That Outperforms Tit-for-Tat in the Prisoner's Dilemma Game." *Nature* 364 (6432):56-58.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.
- Nucci, Larry P. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- Ogloff, James R. P., and Stephen Wong. 1990. "Electrodermal and Cardiovascular Evidence of a Coping Response in Psychopaths." *Criminal Justice and Behavior* 17 (2):231-245. doi: 10.1177/0093854890017002006.
- Okasha, Samir. 2001. "Why Won't the Group Selection Controversy Go Away?" *The British journal for the philosophy of science* 52 (1):25-50.
- Okasha, Samir. 2006. *Evolution and the Levels of Selection*. Oxford; Oxford; New York: Clarendon Press ; Oxford University Press.

- Okasha, Samir. 2008. "The Units and Levels of Selection." In *A Companion to the Philosophy of Biology*, edited by Sahotra Sarkar, and Anya Plutynski, 138-156. Malden, Mass.; Oxford: John Wiley & Sons.
- Osumi, Takahiro, and Hideki Ohira. 2010. "The Positive Side of Psychopathy: Emotional Detachment in Psychopathy and Rational Decision-Making in the Ultimatum Game." *Personality and individual differences* 49 (5):451-456.
- Patrick, Christopher J. 1994. "Emotion and Psychopathy: Startling New Insights." *Psychophysiology* 31 (4):319-330.
- Patrick, Christopher J. 2001. "Emotional Processes in Psychopathy." In *Violence and Psychopathy*, edited by Adrian Raine and José Sanmartín, 57-77. Springer.
- Paulus, Frieder M., Laura Müller-Pinzler, Stefan Westermann, and Sören Krach. 2013. "On the Distinction of Empathic and Vicarious Emotions." *Frontiers in Human Neuroscience* 7:196. doi: 10.3389/fnhum.2013.00196.
- Pears, David. 1987. "The Goals and Strategies of Self-Deception." In *The Multiple Self*, edited by Jon Elster, 59. Cambridge [Cambridgeshire]; New York : Cambridge University Press.
- Pedersen, Eric J, Robert Kurzban, and Michael E McCullough. 2013. "Do Humans Really Punish Altruistically? A Closer Look." *Proceedings of the Royal Society B: Biological Sciences* 280 (1758):20122723.
- Pfennig, D. W. 2002. Kin Recognition. In *Encyclopedia of Evolution*, edited by Mark Pagel. Oxford: Oxford University Press.
- Pham, Michel Tuan. 2007. "Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence." *Review of general psychology* 11 (2):155.
- Pigden, Charles. 1991. "Naturalism." In *A Companion to Ethics*, edited by Peter Singer, 421-431. Oxford, UK; Cambridge, Mass., USA: Blackwell.
- Pigden, Charles R. 1989. "Logic and the Autonomy of Ethics." *Australasian Journal of Philosophy* 67 (2):127-151.
- Preston, Stephanie D., and Frans de Waal. 2002a. "The Communication of Emotions and the Possibility of Empathy in Animals." In *Altruism and Altruistic Love*, edited by Stephen G. Post, Lynn G. Underwood, Jeffrey P. Schloss and William B. Hurlbut, 284-308. New York: Oxford University Press.

- Preston, Stephanie D., and Frans de Waal. 2002b. "Empathy: Its Ultimate and Proximate Bases." *Behavioral and brain sciences* 25 (01):1-20.
- Prinz, Jesse. 2007. *The Emotional Construction of Morals*. Oxford; New York: Oxford University Press.
- Prinz, Jesse J. 2009. "Against Moral Nativism." In *Stich and His Critics*, edited by Dominic Murphy and Michael A. Bishop, 167-189. Chichester, U.K.: Wiley-Blackwell.
- Prinz, Jesse J. 2011. "Is Empathy Necessary for Morality?" In *Empathy*, edited by Amy Coplan and Peter Goldie. Oxford: Oxford University Press.
- Prinz, Jesse J. , and Shaun Nichols. 2010. "Moral Emotions." In *The Moral Psychology Handbook*, edited by John M. Doris and Fiery Cushman. Oxford: Oxford University Press.
- Prinz, Wolfgang. 1997. "Perception and Action Planning." *European journal of cognitive psychology* 9 (2):129-154.
- Prior, Arthur N. 1960. "The Autonomy of Ethics." *Australasian Journal of Philosophy* 38 (3):199-206.
- Putnam, Hilary. 1973. "Meaning and Reference." *The Journal of Philosophy* 70 (19):699-711. doi: 10.2307/2025079.
- Quattrone, George A., and Amos Tversky. 1987. "Self-Deception and the Voter' S Illusion." In *The Multiple Self*, edited by Jon Elster. Cambridge [Cambridgeshire]; New York : Cambridge University Press.
- Quine, Willard V. 1969. "Natural Kinds." In *Ontological Relativity and Other Essays*, 114-38. New York: Columbia University Press.
- Railton, Peter. 1997. "Moral Realism." In *Moral Discourse and Practice : Some Philosophical Approaches*, edited by Stephen L. Darwall, Allan Gibbard and Peter Albert Railton. New York: Oxford University Press.
- Railton, Peter. 2000. "Darwinian Building Blocks." *Journal of Consciousness Studies* 7 (1-2):55-60.
- Raine, A., and A.L. Glenn. 2014. *Psychopathy: An Introduction to Biological Findings and Their Implications*. New York: New York University Press.

- Ramachandran, V. S. 1996. "The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues from Anosognosia." *Medical Hypotheses* 47 (5):347-362. doi: 10.1016/s0306-9877(96)90215-7.
- Ramachandran, V.S. 2011. *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human*: W. W. Norton.
- Rawls, John. 1999a. "The Independence of Moral Theory." In *Collected Papers*, edited by Samuel Freeman. Harvard University Press.
- Rawls, John. 1999b. *A Theory of Justice*. Cambridge, Mass.: The Belknap Press of Harvard University Press.
- Redondo, Tomas, Francisco S. Tortosa, and Luis Arias de Reyna. 1995. "Nest Switching and Alloparental Care in Colonial White Storks." *Animal Behaviour* 49 (4):1097-1110. doi: <http://dx.doi.org/10.1006/anbe.1995.0138>.
- Richerson, P.J., and R. Boyd. 2006. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Ridley, Mark. 2003. *Evolution*. Malden, Oxford: Wiley Blackwell.
- Riedl, Katrin, Keith Jensen, Josep Call, and Michael Tomasello. 2012. "No Third-Party Punishment in Chimpanzees." *Proceedings of the National Academy of Sciences* 109 (37):14824-14829.
- Riedman, Marianne L., and Burney J Le Boeuf. 1982. "Mother-Pup Separation and Adoption in Northern Elephant Seals." *Behavioral Ecology and Sociobiology* 11 (3):203-215. doi: 10.1007/BF00300063.
- Rilling, J., D. Gutman, T. Zeh, G. Pagnoni, G. Berns, and C. Kilts. 2002. "A Neural Basis for Social Cooperation." *Neuron* 35 (2):395-405.
- Rilling, James K, Andrea L Glenn, Meeta R Jairam, Giuseppe Pagnoni, David R Goldsmith, Hanie A Elfenbein, and Scott O Lilienfeld. 2007. "Neural Correlates of Social Cooperation and Non-Cooperation as a Function of Psychopathy." *Biological psychiatry* 61 (11):1260-1271.
- Rizzolatti, Giacomo, and Laila Craighero. 2005. "Mirror Neuron: A Neurological Approach to Empathy." In *Neurobiology of Human Values*, 107-123. Springer.
- Rizzolatti, Giacomo, and Corrado Sinigaglia. forthcoming. "A Curious Book on the Mirror Neurons and Their Myth." *American Journal of Psychology*.



- Rorty, Amélie Oksenberg. 1987. "Self-Deception, Akrasia and Irrationality." In *The Multiple Self*, edited by Jon Elster, 115-131. Cambridge [Cambridgeshire]; New York : Cambridge University Press.
- Rorty, Amélie Oksenberg. 1988. "The Deceptive Self: Liars, Layers, and Lairs." In *Perspectives on Self-Deception*, edited by Brian P McLaughlin and Amélie Rorty, 11-28. Berkeley: University of California Press.
- Rose, Steven Peter Russell, Richard Charles Lewontin, and Leon J. Kamin. 1990. *Not in Our Genes: Biology, Ideology and Human Nature*. London: Penguin books.
- Rosenberg, Alexander. 1992. "Altruism: Theoretical Contexts." In *Keywords in Evolutionary Biology*, edited by Evelyn Fox; Lloyd Keller, Elisabeth Anne. Cambridge, MA: Harvard University Press.
- Ruse, Michael. 1980. "Charles Darwin and Group Selection." *Annals of Science* 37 (6):615-630. doi: 10.1080/00033798000200421.
- Ruse, Michael. 1995a. "Evolutionary Ethics: A Phoenix Arisen." In *Issues in Evolutionary Ethics / Edited by Paul Thompson*, edited by Paul Thompson. Albany: State University of New York Press.
- Ruse, Michael. 2002. "A Darwinian Naturalist's Perspective on Altruism." In *Altruism & Altruistic Love: Science, Philosophy, & Religion in Dialogue*, edited by Stephen Garrard Post. Oxford: Oxford University Press.
- Ruse, Michael. 2006a. *Evolutionary Naturalism: Selected Essays*. London; New York: Routledge.
- Ruse, Michael. 2006b. "Is Darwinian Metaethics Possible (and If It Is, Is It Well Taken)?" In *Evolutionary Ethics and Contemporary Biology*. Cambridge University Press.
- Ruse, Michael. 2012. *The Philosophy of Human Evolution*. Cambridge; New York: Cambridge University Press.
- Ruse, Michael, and Edward O. Wilson. 1986. "Moral Philosophy as Applied Science." *Philosophy* 61 (236 ):173-192.
- Rutte, Claudia, and Michael Taborsky. 2007. "Generalized Reciprocity in Rats." *PLoS biology* 5 (7):e196.
- Sahdra, Baljinder, and Paul Thagard. 2003. "Self-Deception and Emotional Coherence." *Minds and Machines* 13 (2):213-231.

- Salmon, Catherine. 2005. "Parental Investment and Parent-Offspring Conflict." In *The Handbook of Evolutionary Psychology*, edited by David M. Buss, 506-527. Hoboken, NJ, US: John Wiley & Sons Inc.
- Sapolsky, Robert M. 1998. *The Trouble with Testosterone: And Other Essays on the Biology of the Human Predicament*. New York: Scribner.
- Sayre-McCord, Geoffrey. 1986. "The Many Moral Realisms<sup>1</sup>." *The Southern Journal of Philosophy* 24 (S1):1-22.
- Scott-Kakures, Dion. 2000. "Motivated Believing: Wishful and Unwelcome." *Noûs* 34 (3):348–375.
- Scott-Kakures, Dion. 2002. "At "Permanent Risk".<sup>1</sup>" *Philosophy and Phenomenological Research* 65 (3):576 - 603.
- Searle, John R. 1964. "How to Derive" Ought" from" Is".<sup>1</sup>" *The Philosophical Review* 73 (1):43-58.
- Sesardic, Neven. 1995. "Recent Work on Human Altruism and Evolution." *Ethics* 106 (1):128-157.
- Sesardic, Neven. 2005. "From Genes to Incest Taboos: The Crucial Step." In *Inbreeding, Incest, and the Incest Taboo: The State of Knowledge at the Turn of the Century*, edited by Arthur P Wolf and William H Durham. Stanford, Calif.: Stanford University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford; Oxford; New York: Clarendon; Oxford University Press.
- Shafer-Landau, Russ. 2012. "Evolutionary Debunking, Moral Realism and Moral Knowledge." *J. Ethics & Soc. Phil.* 7:i.
- Shamay-Tsoory, S. G., J. Aharon-Peretz, and D. Perry. 2009. "Two Systems for Empathy: A Double Dissociation between Emotional and Cognitive Empathy in Inferior Frontal Gyrus Versus Ventromedial Prefrontal Lesions." *Brain* 132 (Pt 3):617-27. doi: 10.1093/brain/awn279.
- Shamay-Tsoory, Simone G. 2011. "The Neural Bases for Empathy." *The Neuroscientist* 17 (1):18-24.
- Shoemaker, David W. 2011. "Psychopathy, Responsibility, and the Moral/Conventional Distinction." *The Southern Journal of Philosophy* 49 (s1):99-124.

- Sigmund, Karl. 2013. "The Basis of Morality: Richard Alexander on Indirect Reciprocity." In *Human Social Evolution - the Foundational Works of Richard D. Alexander*, edited by Kyle Summers and Bernard Crespi, 199-208. Oxford: Oxford University Press.
- Silk, Joan B, Sarah F Brosnan, Jennifer Vonk, Joseph Henrich, Daniel J Povinelli, Amanda S Richardson, Susan P Lambeth, Jenny Mascaro, and Steven J Schapiro. 2005. "Chimpanzees Are Indifferent to the Welfare of Unrelated Group Members." *Nature* 437 (7063):1357-1359.
- Silk, Joan B. 1987. "Adoption and Fosterage in Human Societies: Adaptations or Enigmas?" *Cultural Anthropology* 2 (1):39-49. doi: 10.1525/can.1987.2.1.02a00050.
- Singer, P. 2011. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton: Princeton University Press.
- Sinigaglia, Corrado, and Giacomo Rizzolatti. 2011. "Through the Looking Glass: Self and Others." *Consciousness and cognition* 20 (1):64-74.
- Skeem, Jennifer L., Devon L. L. Polaschek, Christopher J. Patrick, and Scott O. Lilienfeld. 2011. "Psychopathic Personality: Bridging the Gap between Scientific Evidence and Public Policy." *Psychological Science in the Public Interest* 12 (3):95-162.
- Smith, Michael. 1991. "Realism." In *A Companion to Ethics*, edited by Peter Singer. Oxford, UK; Cambridge, Mass., USA: Blackwell Reference.
- Sober, E. 1993. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*: University of Chicago Press.
- Sober, E., and D.S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*: Harvard University Press.
- Sober, Elliott. 1987. "Comments on Maynard Smith's "How to Model Evolution"." In *The Latest on the Best: Essays on Evolution and Optimality*, edited by John Dupre, 133-145. Cambridge, Mass: MIT Press.
- Sober, Elliott. 1988. "What Is Evolutionary Altruism?" *Canadian Journal of Philosophy* 18 (sup1):75-99. doi: 10.1080/00455091.1988.10715945.
- Sober, Elliott. 1994. *From a Biological Point of View: Essays in Evolutionary Philosophy*. Cambridge: Cambridge University Press.
- Sober, Elliott. 2000. *Philosophy of Biology*. Boulder, Colo.: Westview Press.

- Sober, Elliott, and David Sloan Wilson. 1994. "A Critical Review of Philosophical Work on the Units of Selection Problem." *Philosophy of Science* 61 (4):534-555.
- Sorensen, Roy A. 1985. "Self-Deception and Scattered Events." *Mind* 94 (373):64-69.
- Sosa, Ernest. 2003. "Beyond Internal Foundations to External Virtues." In *Epistemic Justification: Internalism Vs. Externalism; Foundations Vs. Virtues*, edited by Laurence Bonjour and Ernest Sosa. Oxford: Blackwell.
- Spencer, H. 1879. *The Data of Ethics*. London: Williams and Norgate.
- Spencer, Herbert. 1851. *Social Statics, or, the Conditions Essential to Human Happiness Specified, and the First of Them Developed*. London: John Chapman.
- Spencer, Herbert. 1978 [1879]. *The Principles of Ethics*. Indianapolis: Liberty Fund.
- Sterelny, Kim, and Philip Kitcher. 1988. "The Return of the Gene." *The Journal of Philosophy* 85 (7):339-361. doi: 10.2307/2026953.
- Stevens, Jeffrey R, Fiery A. Cushman, and Marc D. Hauser. 2005. "Evolving the Psychological Mechanisms for Cooperation." *Annual Review of Ecology, Evolution, and Systematics*:499-518.
- Stevenson, Charles Leslie. 1937. "The Emotive Meaning of Ethical Terms." *Mind* 46 (181):14-31.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1):109-166.
- Sturgeon, Nicholas L. 2006 [1984]. "Moral Explanations." In *Ethics: History, Theory, and Contemporary Issues*, edited by Steven M. Cahn and Peter J. Markie. New York: Oxford University Press.
- Sumner, W. G. 1914. *The Challenge of Facts and Other Essays*. New Haven: Yale University Press.
- Surbey, Michele K. 2004. "Self-Deception: Helping and Hindering Personal and Public Decision Making." In *Evolutionary psychology, public policy, and personal decisions*, edited by Charles Crawford and Catherine Salmon, 117-144. Mahwah, N.J.: Lawrence Erlbaum Associates
- Surbey, Michele K, and Jeffrey J McNally. 1997. "Self-Deception as a Mediator of Cooperation and Defection in Varying Social Contexts Described in the Iterated Prisoner's Dilemma." *Evolution and Human Behavior* 18 (6):417-435.

- Talbott, W. J. 1995. "Intentional Self-Deception in a Single Coherent Self." *Philosophy and Phenomenological Research* 55 (1):27-74.
- Thagard, Paul. 2000. *Coherence in Thought and Action*. Cambridge, Mass.: MIT Press.
- Thagard, Paul, and Karsten Verbeurgt. 1998. "Coherence as Constraint Satisfaction." *Cognitive Science* 22 (1):1-24.
- Thierry, Bernard. 2000. "Building Elements of Morality Are Not Elements of Morality." *Journal of Consciousness Studies* 7 (1-2):60-62.
- Thomson, Judith Jarvis. 1976. "Killing, Letting Die, and the Trolley Problem." *The Monist* 59 (2):204-217.
- Tomasello, Michael. 2009. *Why We Cooperate*. Cambridge (Massachusetts); London (England): MIT Press.
- Tooley, Greg A., Mari Karakis, Mark Stokes, and Joan Ozanne-Smith. 2004. "Generalising the Cinderella Effect to Unintentional Childhood Fatalities." *Evolution and Human Behavior* 27 (3):224-230. doi: 10.1016/j.evolhumbehav.2005.10.001.
- Trivers, Robert. 1971. "The Evolution of Reciprocal Altruism." *Quarterly review of biology*:35-57.
- Trivers, Robert. 2002. *Natural Selection and Social Theory: Selected Papers of Robert Trivers*. Oxford: Oxford University Press.
- Trivers, Robert. 2006. "Reciprocal Altruism: 30 Years Later." In *Cooperation in Primates and Humans: mechanisms and evolution*, edited by Peter M.Kappeler and Carel P. van Schaik, 67-83. Berlin: Springer.
- Trivers, Robert. 2010. "Deceit and Self-Deception." In *Mind the Gap*, edited by Peter M. Kappeler and Joan Silk, 373-393. Berlin; Heidelberg: Springer.
- Trivers, Robert 1976/2006. "Foreword." In *The Selfish Gene*, by Richard Dawkins, 19–20. Oxford: Oxford University Press.
- Trope, Yaacov, Benjamin Gervy, and Nira Liberman. 1997. "Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective." In *The Mythomanias: The Nature of Deception and Self-Deception*, edited by Michael S. Myslobodsky, 105-31. Mahwah, N.J.: L. Erlbaum Associates.

- Trope, Yaacov, and Akiva Liberman. 1996. "Social Hypothesis Testing: Cognitive and Motivational Mechanisms." In *Social psychology: Handbook of basic principles*, edited by Tory E. Higgins and Arie W. Kruglanski. New York: Guilford Press.
- Tucić, Nikola. 2003. *Evoluciona biologija*. Beograd: NNK-International.
- Turiel, Elliot. 1980. "The Development of Moral Concepts." In *Morality as a Biological Phenomenon : The Presuppositions of Sociobiological Research*, edited by Gunther S. Stent. Berkeley, California: University of California Press.
- Turiel, Elliot. 2002. *The Culture of Morality: Social Development, Context, and Conflict*. Cambridge: Cambridge University Press.
- Turiel, Elliot. 2006. "The Development of Morality." In *Handbook of Child Psychology. Volume 3*, edited by Nancy Eisenberg, William Damon and Richard M. Lerner. Hoboken, N. J.: John Wiley & Sons.
- Vaish, Amrisha, Malinda Carpenter, and Michael Tomasello. 2009. "Sympathy through Affective Perspective Taking and Its Relation to Prosocial Behavior in Toddlers." *Developmental psychology* 45 (2):534.
- Van Der Wel, RPRD, Natalie Sebanz, and Günther Knoblich. 2013. "Action Perception from a Common Coding Perspective." In *People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception*, edited by Kerri L. Johnson and Maggie Shiffrar, 101-120. New York: Oxford University Press.
- Van Leeuwen, D. S. Neil. 2007. "The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon." *Philosophical Psychology* 20 (3):329-348.
- von Bernhardt, F. 1912. *Germany and the Next War*: Longmans, Green.
- Von Hippel, William, and Robert Trivers. 2011a. "The Evolution and Psychology of Self-Deception." *Behav Brain Sci* 34 (1):1-16; discussion 16-56.
- Von Hippel, William, and Robert Trivers. 2011b. "Reflections on Self-Deception." *Behavioral and Brain Sciences* 34 (1):41-56.
- Vrij, Aldert. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*: John Wiley & Sons.
- Vrij, Aldert. 2011. "Self-Deception, Lying, and the Ability to Deceive." *Behavioral and Brain Sciences* 34 (01):40-41.

- Warneken, Felix, and Alicia Perez Melis. 2012. "The Ontogeny and Phylogeny of Cooperation." In *The Oxford Handbook of Comparative Evolutionary Psychology*, edited by Todd K. Shackelford and Jennifer Vonk, 381-398. Oxford ; New York: Oxford University Press.
- Warneken, Felix, and Michael Tomasello. 2006. "Altruistic Helping in Human Infants and Young Chimpanzees." *Science* 311 (5765):1301-1303.
- Warneken, Felix, and Michael Tomasello. 2007. "Helping and Cooperation at 14 Months of Age." *Infancy* 11 (3):271-294.
- Warneken, Felix, and Michael Tomasello. 2009a. "The Roots of Human Altruism." *British Journal of Psychology* 100 (3):455-471.
- Warneken, Felix, and Michael Tomasello. 2009b. "Varieties of Altruism in Children and Chimpanzees." *Trends in cognitive sciences* 13 (9):397-402.
- Wedekind, C., T. Seebeck, F. Bettens, and A. J. Paepke. 1995. "Mhc-Dependent Mate Preferences in Humans." *Proc Biol Sci* 260 (1359):245-9. doi: 10.1098/rspb.1995.0087.
- West-Eberhard, MJ. 1989. "Phenotypic Plasticity and the Origins of Diversity." *Annual Review of Ecology and Systematics* 20 (1):249-278.
- West, Stuart A, Claire El Mouden, and Andy Gardner. 2011. "Sixteen Common Misconceptions About the Evolution of Cooperation in Humans." *Evolution and Human Behavior* 32 (4):231-262.
- Weston, Donna R, and Elliot Turiel. 1980. "Act-Rule Relations: Children's Concepts of Social Rules." *Developmental Psychology* 16 (5):417.
- White, Roger. 2010. "You Just Believe That Because...." *Philosophical Perspectives* 24 (1):573-615.
- Wielenberg, Erik J. 2010. "On the Evolutionary Debunking of Morality." *Ethics* 120 (3):441-464.
- Wilkinson, Gerald S. 1988. "Reciprocal Altruism in Bats and Other Mammals." *Ethology and Sociobiology* 9 (2):85-100.
- Wilkinson, Gerald S. 1990. "Food Sharing in Vampire Bats." *Scientific American* 262 (2):76-82.
- Williams, B. 1976. *Problems of the Self: Philosophical Papers 1956-1972*: Cambridge University Press.
- Williams, Bernard. 1995. "Evolution, Ethics, and the Representation Problem." In *Making Sense of Humanity*. Cambridge University Press.

- Williams, George C. 1993. "Mother Nature Is a Wicked Old Witch." In *Evolutionary Ethics*, edited by Matthew Nitecki and Doris Nitecki, 2--17. Suny Press.
- Williams, George C. 1996 [1966]. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, N. J.: Princeton University Press.
- Williamson, Sherrie. 1991. "Cohesion and Coherence in the Speech of Psychopathic Criminals." Doctor of Philosophy - PhD Electronic Thesis or Dissertation, University of British Columbia.
- Williamson, Sherrie, Timothy J Harpur, and Robert D Hare. 1991. "Abnormal Processing of Affective Words by Psychopaths." *Psychophysiology* 28 (3):260-273.
- Wilson, David Sloan. 1992. "On the Relationship between Evolutionary and Psychological Definitions of Altruism and Selfishness." *Biology and Philosophy* 7 (1):61-68.
- Wilson, David Sloan. 2015. *Does Altruism Exist?: Culture, Genes, and the Welfare of Others*. New Haven: Yale University Press, Templeton Press.
- Wilson, David Sloan, and Elliott Sober. 1994. "Reintroducing Group Selection to the Human Behavioral Sciences." *Behavioral and Brain Sciences* 17 (04):585-608.
- Wilson, E.O. 1980. *Sociobiology: The Abridged Edition*. Cambridge, MA [etc.]: Belknap Press of Harvard University Press.
- Wilson, E.O. 1999. *Consilience: The Unity of Knowledge*. New York: Vintage Books.
- Wilson, E.O. 2002. *The Future of Life*. New York: Alfred A. Knopf.
- Wolf, Arthur P. 2005. "Explaining the Westermarck Effect: Or, What Did Natural Selection Select For?" In *Inbreeding, Incest, and the Incest Taboo: The State of Knowledge at the Turn of the Century*, edited by Arthur P Wolf and William H Durham, 76-92. Stanford, Calif.: Stanford University Press.
- Wong, David. 1991. "Relativism." In *A Companion to Ethics*, edited by Peter Singer. Oxford, UK: Blackwell Reference.
- Wu, Jianzhong, and Robert Axelrod. 1995. "How to Cope with Noise in the Iterated Prisoner's Dilemma." *Journal of Conflict resolution*:183-189.
- Wynne-Edwards, Vero Copner. 1962. *Animal Dispersion in Relation to Social Behaviour*. New York: Hafner Pub. Co.
- Wynne-Edwards, Vero Copner. 1963. "Intergroup Selection in the Evolution of Social Systems." *Nature* 200 (4907):623-626.



- Yamamoto, Shinya, Tatyana Humle, and Masayuki Tanaka. 2009. "Chimpanzees Help Each Other Upon Request." *PLoS ONE* 4 (10):e7416. doi: 10.1371/journal.pone.0007416.
- Yamamoto, Shinya, Tatyana Humle, and Masayuki Tanaka. 2012. "Chimpanzees' Flexible Targeted Helping Based on an Understanding of Conspecifics' Goals." *Proceedings of the National Academy of Sciences* 109 (9):3588-3592. doi: 10.1073/pnas.1108517109.
- Zahavi, Amotz. 1995. "Altruism as a Handicap: The Limitations of Kin Selection and Reciprocity." *Journal of Avian Biology* 26 (1):1-3. doi: 10.2307/3677205.
- Zimmerman, Aaron. 2010. *Moral Epistemology*. New York: Routledge.
- Živanović, Igor. 2012. "Refleksivni ekvilibrijum i problem opravdanja u etici." *Filozofski godišnjak* 25 (25):73-98.

## Biografija

Igor Živanović rođen je 1974. godine u Požarevcu gde je završio osnovnu i srednju školu. Filozofiju je diplomirao na Filozofskom fakultetu u Beogradu 2007. godine s prosečnom ocenom 9,11 tokom studija i ocenom 10 na diplomskom ispitu odbranjenom kod Prof. dr Milorada Stupara. Diplomski rad „Hobsova koncepcija prirodnog stanja i problem kooperacije“ bio je u najužem izboru za nagradu „Dr Zoran Đinđić“ za najbolji diplomski ili magistarski rad iz filozofskih i socioloških nauka. Doktorske studije upisao je 2009. godine i položio sve ispite s prosečnom ocenom 9,83. Trenutno je angažovan na projektu „Dinamički sistemi u prirodi i društvu: filozofski i empirijski aspekti“ Instituta za filozofiju Filozofskog fakulteta u Beogradu, koji finansira Ministarstvo nauke Republike Srbije. Proteklih godina Igor Živanović je bio angažovan u nastavi na Odeljenju za filozofiju Filozofskog fakulteta u Beogradu na predmetima Uvod u probleme politike, Filozofija politike, Savremene teorije pravde i Evolucionarna etika. Učestvovao je na naučnom skupu Instituta za filozofiju i društvenu teoriju povodom izdavanja knjige Milorada Stupara *Filozofija politike*.

Objavio je sledeće naučne radove:

„Hobsova koncepcija prirodnog stanja i problem kooperacije“, *Filozofski godišnjak*. God. 20 (2007), str. 76-120.

“Egoizam, altruizam i samoobmanjivanje“, *Theoria*, God. 54, br. 3 (2011), str. 61 – 85.

„Platon, Makijaveli i makijavelizam“, *Filozofija i društvo*, God. 22, br. 3, (2011), str. 45–67.

„Refleksivni ekvilibrijum i problem opravdanja u etici“ *Filozofski godišnjak*. God. 25, br. 25 (2012), str. 73 – 98.

“Da li je život u prirodnom stanju usamljениčki, siromašan, opasan, skotski i kratak”, *Theoria* 2013, vol. 56, br. 3, str. 41 – 78.

“Samoobmanjivanje”, *Theoria*, br. 3, (2015), prihvaćen za objavljivanje

(Bez naslova), prikaz knjige WILSON, D. S. (2015). *Does altruism exist?: culture, genes, and the welfare of others*. New Haven, Yale University Press, Templeton Press. *Filozofija i društvo*, br. 3, (2015)

Prilog 1.

## Izjava o autorstvu

Potpisani-a \_\_\_\_\_

broj upisa \_\_\_\_\_

### Izjavljujem

da je doktorska disertacija pod naslovom

\_\_\_\_\_  
\_\_\_\_\_

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija u celini ni u delovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio intelektualnu svojinu drugih lica.

### Potpis doktoranda

U Beogradu, \_\_\_\_\_

\_\_\_\_\_

Prilog 2.

## Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora \_\_\_\_\_

Broj upisa \_\_\_\_\_

Studijski program \_\_\_\_\_

Naslov rada \_\_\_\_\_

Mentor \_\_\_\_\_

Potpisani \_\_\_\_\_

izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao/la za objavljivanje na portalu **Digitalnog repozitorijuma Univerziteta u Beogradu**.

Dozvoljavam da se objave moji lični podaci vezani za dobijanje akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.

Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

**Potpis doktoranda**

U Beogradu, \_\_\_\_\_

\_\_\_\_\_

Prilog 3.

## Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

---

koja je moje autorsko delo.

Disertaciju sa svim priložima predao/la sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni repozitorijum Univerziteta u Beogradu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio/la.

1. Autorstvo
2. Autorstvo - nekomercijalno
3. Autorstvo – nekomercijalno – bez prerade
4. Autorstvo – nekomercijalno – deliti pod istim uslovima
5. Autorstvo – bez prerade
6. Autorstvo – deliti pod istim uslovima

(Molimo da zaokružite samo jednu od šest ponuđenih licenci, kratak opis licenci dat je na poledini lista).

**Potpis doktoranda**

U Beogradu, \_\_\_\_\_

\_\_\_\_\_