

Šemsudin Džopa

Filozofska fakulteta Univerze v Ljubljani

Naturalistični vidiki svobodne volje in moralne odgovornosti

Svobodno voljo in moralno odgovornost lahko obravnavamo na dveh nivojih. Na absolutnem nivoju sta težko združljiva s sodobno znanostjo in mogoče celo nekonsistentna sama na sebi. Na kompatibilističnem nivoju pa ju lahko opredelimo kot evolvirani zmožnosti človeka kot biološkega bitja. Determinizem osnovnega nivoja ne samo, da ne nasprotuje tako pojmovani svobodni volji in moralni odgovornosti, ampak celo pomaga pri evoluciji vedno boljših izogibalcev na višjem nivoju. Z aplikacijo zapornikove dileme na evolucijo je možno utemeljiti moralno obnašanje kot evolucijsko stabilno strategijo, tudi ob odsotnosti močnih pojmov svobodne volje in moralne odgovornosti, ki ju kompatibilizem ni zmožen zagotoviti. Za kompatibilistično svobodo je pomembno, da je človek kot racionalni akter zmožen razumeti in reflektivno ovrednotiti razloge, po katerih želi ravnati, ter zmožen nadzirati svoje delovanje v skladu s tako reflektiranimi razlogi. Sodobne psihološke in nevroznanostvene raziskave kažejo na probleme pri obeh zgornjih zmožnostih ter na človekovo veliko dovzetnost za različne kognitivne disonance, samoprevare in podobno.

Ključne besede: svobodna volja, moralna odgovornost, kompatibilizem, evolucija, racionalni akter

1. Uvod

Pojma svobodne volje in moralne odgovornosti lahko obravnavamo na dveh različnih nivojih, absolutnem in kompatibilističnem (Smilansky, 2000). Absolutna oziroma močna pojava sta težko združljiva z determinizmom in sodobno znanostjo, mogoče pa celo nekonsistentna sama na sebi. Kompatibilistični nivo obravnave pa utemeljuje njuno šibko različico, ki se bistveno razlikuje od njenega močnega, metafizičnega pojmovanja. Ravno iz nerazlikovanja in nepriznavanja teh nivojev, ki bistveno vplivata na samo definicijo in vsebino pojmov, izhajajo pogosti nesporazumi med kompatibilisti in inkompatibilisti.¹

¹ Iz predpostavke dveh različnih nivojev argumentacije sledi, da je inkompatibilistični spor med libertarci in trdimi deterministi utemeljen, spor med inkompatibilisti in kompatibilisti pa ne.

V pričujočem članku bomo absolutni, (meta)fizični nivo pustili ob strani in se posvetili kompatibilističnemu pojmovanju tako, da bomo svobodno voljo in moralno odgovornost utemeljevali kot evolvirani zmožnosti človeka kot biološkega bitja. Pomagali si bomo predvsem z evolucijsko teorijo, s katero bomo najprej poskusili dokazati konsistentnost pojma izogibljivosti v determinističnem vesolju. Determiniranost in izogibljivost bomo obravnavali kot lastnosti dveh različnih nivojev, za razliko od tradicionalnega, inkompatibilističnega enačenja determiniranosti in neizogibnosti. Determinizem osnovnega nivoja se bo izkazal celo kot pozitiven za evolucijo vedno boljših izogibalcev (ang. *avoiders*) na višjem nivoju.

V drugem delu bomo s pomočjo zapornikove dileme iskali pogoje, v katerih sodelovanje, moralno obnašanje in celo altruizem postanejo evolucijsko stabilne strategije, kar nasprotuje razširjeni predpostavki, da darvinizem implicira skrajnje sebično, sovražno in anti-socialno obnašanje.

V tretjem, zadnjem delu članka pa bomo raziskali razum kot človekovo glavno biološko *differentio specifico* in njegov odnos do zmožnosti svobodne volje in moralne odgovornosti. Razum omogoči »ujetje« razlogov delovanja, ki so v celotnem dotedanjem poteku evolucije bili »prosto lebdeči«. Ravno zmožnost racionalne refleksije in zmožnost samonadzora predstavljata najpomembnejši vidik kompatibilistične svobode. Ta vidik je praktično celo veliko bolj pomemben kot metafizična razprava na absolutnem nivoju. Zanimalo nas bo namreč, v kakšni meri so človeku dostopni resnični razlogi njegovega delovanja in v kakšni meri je zmožen reflektiranega samonadzora ter koliko je po drugi strani dovzeten za različne kognitivne disonance, samoprevare, *ad hoc* prilagajanje svojega vrednostnega sistema in racionaliziranje kot opravičevanje svojega početja. Velika dovzetnost za našeto bi namreč predstavljala pomembno oviro človekovi svobodi in ideji racionalnega akterja.

2. Evolucija znotraj determinističnega vesolja

2.1 Evolucija in igra *Življenje* Johna Conwaya

Za prikaz tega, kako lahko v strogo determiniranem svetu pridemo do pojava »izogibnosti« ali »izogibljivosti« (ang. *avoidance*), uporabi Daniel Dennett primer iz igre *Življenje* (*Life*), ki si jo je zamislil britanski matematik John Conway. Kot že samo ime igre pove, je igra simulacija pojava življenja. Po enostavnih, determinističnih pravilih igre, ki so prispodoba naravnih zakonov, se v dvodimenzionalnem prostoru, glede na začetno postavitev, oblike spreminjajo, izginjajo, rastejo, jejo, razmnožujejo itd. Nekatere oblike, na primer *glider* (jadralec), se izkažejo za zelo obstojne, nekatere druge pa zelo hitro izginejo. Vse pa je seveda zelo odvisno od začetne postavitve, naprej se vse dogaja samo od sebe, torej determinirano s preteklim oziroma začetnim stanjem in tremi enostavnimi »fizikalnimi« zakoni.

Na *osnovnem* nivoju, recimo mu fizikalni, je vse, kar se dogaja, zgolj prižiganje in ugašanje »osnovnih delcev« oziroma kvadratov, ni nobenega gibanja. Na *višjem* nivoju, ko spremljamo, kaj se dogaja z oblikami sestavljenimi iz več kvadratov, pa vidimo, da se nekatere oblike obdržijo oziroma »preživijo« dlje kot druge. Nekatere oblike so bolj trajne, odporne in se lažje »izognejo uničenju« oziroma ugašanju kot druge (Dennett, 2003: 40). Na višjem nivoju spremljamo premikanje nekaterih oblik, kar sicer pomeni, da jih na osnovnem nivoju vsakič predstavljajo drugi kvadrati. Nekatere oblike imajo zmožnosti, ki jih druge nimajo, zahvaljujoč zgolj svoji *obliki* (prav tam: 42).

Čeprav torej na osnovnem nivoju ne moremo govoriti o nobenih zmožnostih, se nam na višjem zdi zelo smiselno primerjati različne oblike glede na njihove »zmožnosti preživetja« oziroma vztrajanja v času. Nekatere bolj kompleksne strukture so se tako zmožne bolj kot druge *izogniti* uničenju, tudi tako, da dobijo informacije iz okolja, »zaznajo in predvidijo« prihajajoče oblike. Temu pojavu Dennett pravi *rojstvo izogibljivosti* (Dennett, 2003: 43).

Kot smo omenili, samo ime igre nakazuje na avtorjevo tezo o nastanku in razvoju življenja.² Zato se prispodoba z evolucija že sama vsiljuje. V igri sicer igralec oziroma ustvarjalec vsakič nastavi začetno postavitev in tako pride do bolj zanimivih oblik in dogajanja. Če pa poskusimo odmisлити ustvarjalca in si zamislimo igro s skoraj neskončnim številom kvadratov, ki se *naključno* znajdejo v začetni postavitvi, brez nekega ustvarjalca (boga), se bodo zanimive oblike, zmožne učenja, predvidevanja, izogibanja in reprodukcije, razvile same od sebe. Vse to dobimo v dvodimenzionalnem prostoru s tremi čisto determinističnimi pravili in zgolj dvema vrednostma vsakega kvadrata. Lahko bi rekli, da je analogija z našim veseljem celo preveč enostavna.

Če si predstavljamo skoraj neskončno število atomov oziroma osnovnih delcev v veselju, ki se po strogo determinističnih zakonih povezujejo v molekule in tako naprej, ali pa razpadejo nazaj v osnovne delce, nam ni potrebno predpostaviti nobenega »inteligentnega načrta« za tako neverjetne stvaritve, kot jih najdemo v naravi. To je bil Darwinov nenavadni obrat v razmišljanju. Po besedah njegovega kritika: *absolutno modrost* je zamenjala *absolutna nevednost* (MacKenzie, v Dennett, 2003: 47). Pred pojavom teorije evolucije je teleološki dokaz za obstoj boga kot stvarnika veljal za enega najmočnejših dokazov za obstoj boga. Ljudje so si težko predstavljali, da bi tako popolna in okolju prilagojena bitja živalskega in človeškega sveta z vsemi neverjetnimi lastnostmi lahko nastala po naključju. Ko obrnemo način razmišljanja, pravi Dennett, bomo videli, kako se inteligenca raz-

² Igra je bila ustvarjena z namenom testiranja miselnega eksperimenta o samoreproducirajočih se avtomatih Johna von Neumanna. In jim je tudi uspelo »ustvariti« takšne avtomate pri številu kvadratov 10^{13} , kar pomeni, da če bi vsak kvadrat bil velikosti enega milimetra, bi celoten »svet« meril približno 3 km. Dennett komentira, da je najbrž (ni še dokazano) potrebna takšna razlika v redu velikosti od osnovnih delcev do strukture, ki se je zmožna reproducirati, kot je od atomov do organizma (Dennett, 2003, str. 49.)

vije iz »intelligence«, kako »slepi urar« ustvari čut vida, kako se izbira razvije iz »izbire« in kako je lahko celota bolj *svobodna* kot njeni deli (Dennett, 2003: 48).

2.2 Evolucija izogibalcev, od generacij do mikrosekund

Česar še nikomur ni uspelo simulirati v Conwayevi igri *Življenje*, je *mutacija*, kar je najbrž posledica tega, da »svetovi« niso dovolj veliki oziroma so premalo »bogati«. Dennettu pa je najbolj bistveno, da je v tem miniaturnem determinističnem svetu dokazal obstoj vseh nujnih gradnikov za evolucijo *izogibalcev* (Dennett, 2003: 51), kar po njegovem zadostuje za zavrnitev kognitivne iluzije, ki determinizem povezuje z neizogibnostjo.

Če se zdaj vrnemo k iskanju začetka izogibalcev v našem svetu, vidimo, da je na začetku evolucije maksimalna hitrost izogibanja določena s hitrostjo menjave generacij. Torej prva možna reakcija na problem oziroma nevarnost, ki se pojavi v času ene generacije, se zgodi v času naslednje generacije. Šele naslednje generacije lahko razvijejo nek odgovor na problem. Tista serija, vrsta oziroma rod potomcev, ki se bolj uspešno ubrani nevarnosti, preživi, ostali rodovi izumrejo. Seveda bi bilo to narobe razumeti v smislu, da so te osnovne življenjske oblike karoli *počele*. Tiste *srečne*, ki so preživele, so preživele zaradi uporabnih mutacij:

Te srečne življenjske oblike so potem imele potomce, katerih srečni potomci so spet imeli potomce in tako naprej, dokler ne pridemo do nas. Mi pa smo tisti srečneži, sestavljeni iz tako uporabnih delov, posebej narejenih zato, da se dobro obnesejo pri izogibanju, ampak tokrat v veliko krajšem času. (Dennett, 2003: 52)

To je bistvo evolucije oziroma naravne selekcije, tukaj ni potrebna nikakršna daljnovidnost niti predvidevanje. Naravna selekcija deluje s preizkusi skozi različne značilnosti mutiranih vrst in rodov, od katerih jih veliko ne preživi. Lahko bi rekli, da so ti preizkusi precej »potratni«, oziroma dolgotrajni.

Vendar pa ni potreben noben čudež, da bi kratkoviden proces ustvaril daljnovidnega. Slepi ali vsaj kratkovidni proces naravne selekcije si, kot pravi Dennett, pomaga z »žerjavi« in ne z »nebesnimi kljukami«, kjer »raziskovanje v času ene generacije (generacije) lahko vpliva na osnovni proces genske evolucije /.../« (Dennett, 2003: 53). Danes so človeška bitja zmožna zaznati stvari pri ogromni razdalji, prirojenim organom pomagajo protetične naprave, kar omogoča zaznavo in rešitev problemov s hitrostjo, ki se bliža maksimalni možni hitrosti v vesolju, svetlobni hitrosti (prav tam: 53). Zamislimo si možnost, da se bo človeštvu nekoč uspelo »izogniti se« nekemu uničevalnemu meteorju, podobnemu tistemu, ki naj bi bil odgovoren za izumrtje dinosavrov. Seveda bi tudi to bilo *popolnoma determinirano* dejanje na fizikalnem nivoju, ampak ali bi to res pomenilo, da pojma »izogniti se« niti v tem primeru ne bi smeli uporabiti izven navednic? Ker človeštvo v tistem trenutku ne bo imelo *resnične* izbire? Vsekakor pa bo imelo vsaj več *navidezne* izbire kot dinosavri ali človeštvo samo v preteklosti.

2.3 Razlika med determiniranostjo in neizogibnostjo

Tezo o tem, da determinizem ne implicira neizogibnosti, postavi Dennett v obliki argumenta (2003:56):

- 1) V nekaterih determinističnih svetovih obstajajo izogibalci, ki se izogibajo škodi.
- 2) Torej se v nekaterih determinističnih svetovih izogibalci nekaterim stvarjem izognejo.
- 3) Torej v nekaterih determinističnih svetovih ni vse neizogibno.
- 4) Torej determinizem ne implicira neizogibnosti.

Ta argument nam utegne delovati sumljiv, za kar so po Dennettu najbolj odgovorne skrite predpostavke, ki jih ponavadi ne opazimo. Te predpostavke lahko osvetlimo, če primerjamo stavka (prav tam: 57):

- a) Če je determinizem resničen, potem je vse, kar se zgodi, *neizogiben* izid celotnega seznama vzrokov.
- b) Če je determinizem resničen, potem je vse, kar se zgodi, *determiniran* izid celotnega seznama vzrokov.

Če je neizogibnost samo sinonim za determiniranost, potem prvi stavek ne dodaja nobene vsebine v primerjavi z drugim. Če pa neizogibnosti ne razumemo zgolj kot sinonima za determinizem, potem mora vsebovati nek dodaten pomen. Mogoče se dodatek skriva v neizogibnosti izida. Ampak o neizogibnosti ali izogibnosti nekega izida lahko govorimo samo v primeru akterjev, ki imajo interes izogniti se nečemu. Vendar je sposobnost živih bitij oziroma določenih vrst, da se izognejo nekemu izidu, odvisna od okoliščin, ne od determinizma (prav tam: 57). V nadaljevanju bomo predstavili možne ugovore, za katere Dennett predvideva, da bodo doleteli njegovo teorijo izogibnosti v determinističnem svetu.

Prvi ugovor proti možnosti izogibalcev v determinističnem svetu, ki ga Dennett poskuša zavrnil, je, da izogibalca čaka točno določena usoda, ki mu je bila *od vedno* usojena. Izogibalci tako nimajo »boljše možnosti« za preživetje, možnost nikoli ne »vstopi v igro«: tisti, ki preživijo, preživijo in obratno; vse to je determinirano že od začetka. Na to Dennett odgovarja, da je možnost popolnoma kompatibilna z determinizmom. Med drugim z njo razlagamo evolucijo, ki je tudi kompatibilna z determinizmom. Zakaj vztrajati, da determinirano izogibanje ni resnično izogibanje? Dejstvo, da je celoten *proces* determiniran, nikakor ne zmanjšuje dejstva, da *s časom* ta proces ustvarja vedno več nečesa, kar se vsem zdi kot izogibanje (Dennett, 2003: 58).

Drugi ugovor se glasi, da resnično izogibanje pomeni izogniti se nečemu, kar se *bo zgodilo*. To je po Dennettu odvisno od tega, kaj pomeni »se bo zgodilo.« Nekaj se resnično *ne bo zgodilo* natančno zato, ker se bo sprožil sistem za izognitev. Mi bi pa še dodali, da kritik tega sistema izognitev seveda ne bo upošteval kot lastnosti izogibalca kot akterja, ki ga vsaj v primeru človeka najverjetneje obrav-

nava kot tisto neskončno majhno točko v možganih. Dennett nenehno poudarja, da lahko pozunanjimo vse, če si sebe predstavljamo kot neko matematično točko v možganih.

Tretji ugovor pravi, da izognitev pomeni *spreminjanje izida*. Iz katerega izida v kateri, se sprašuje Dennett. Sama ideja spreminjanja izida je nekonsistentna – razen če to pomeni *predvidevanega* izida, kar se ravno dogaja pri determiniranem izogibanju. Resničnega, dejanskega izida ni možno spreminjati ne v determinističnem ne v indeterminističnem svetu (Dennett, 2003: 59).

Četrti ugovor proti izogibljivosti v determiniranem vesolju trdi, da t.i. izogibalci *neizogibno* imajo samo tiste sposobnosti, ki jih imajo. Dennett odgovarja, da je to ravno definicija neizogibnosti, ki enači neizogibnost in determinizem, tako da ničesar ne dodaja. Če pa hočemo razložiti svet biologije, potrebujemo takšen pojem izogibanja, ki ga lahko svobodno nanašamo na dogodke v zgodovini življenja na Zemlji, ne glede na determiniranost te zgodovine. In ravno v luči zgodovine življenja lahko rečemo, da je determinizem zaveznik in ne sovražnik tistih, ki ne marajo neizogibnosti (Dennett, 2003: 60).

Na tej točki smo seveda še vedno daleč od nečesa, kar bi bilo podobno svobodni volji. Za Dennetta je bilo bistveno dokazati, da determinizem ne implicira neizogibnosti. Kot vidimo, lahko evolucijo razložimo kot popolnoma *determiniran* proces in istočasno ravno kot evolucijo *izogibnosti*. Evolucija je evolucija izogibnosti. Pojem neizogibnosti (ali izogibnosti) ne pripada fizičnemu, osnovnemu nivoju gradnikov, ampak nivoju dizajna,³ na katerem lahko primerjamo neke oblike, vzorce ali stvaritve ravno glede na njihovo zmožnost izogibanja (Dennett, 2003: 62). Prav tako o pojmu avtonomije lahko govorimo samo na nivoju dizajna.

3. Vznik morale

3.1 Začetki življenja

Da bi razumeli pojav človeške svobode, bomo v nadaljevanju z Dennettom sledili Darwinovemu »čudnemu obratu razlaganja«. Na začetku življenja na Zemlji ni bilo prave svobode, razuma, izbire, ampak samo njihove zgodnje oblike: proto-svoboda, proto-razum, proto-izbira (Dennett, 2003: 143).

³ Beseda dizajn (*design*) v besedni zvezi nivo dizajna (*design level*) bi se lahko nanašala na obliko, vzorec, pa tudi na namero in načrt, kot v besedni zvezi *Intelligent design*.

Na začetku življenja je samo ena izbira omogočala preživetje, torej »naredi A ali umri«, danes pa imamo možnosti »naredi A ali B ali C ali, D ali... umri«. Da se je to zgodilo, so morale oblike življenja postati bistveno bolj kompleksne, kot so bile na začetku. Prvi bistven napredek se je zgodil, ko so prve oblike življenja »ugotovile«, da lahko energijo potrebno za življenje, ki so jo do takrat dobile večinoma od sončne svetlobe, dobijo tako, da razgradijo svoje sosede, in dobijo že »sestavljene rezervne dele«. Tako so prvi plenilci in prvi uplenjenci začeli »oboroževalno tekmo«, ki je v približno milijardi let pripeljala do zelo razvitih oblik enih in drugih, do velikega števila načinov preživetja. Tudi večcelični organizmi so se razvili kot posledica te milijardo let trajajoče »državljsanske vojne« (Dennett, 2003: 144).

3.2 Teorija iger in evolucija

V nadaljevanju si bomo pogledali, kako poteka evolucijska oboroževalna tekma, kaj sproži in kaj zadrži napad različnih strani v tej tekmi. Napad je seveda evolucijsko veliko bolj razumljiva taktika kot sodelovanje, ki ga je nujno potrebno utemeljiti.⁴ Na tem mestu si bomo pomagali z zelo uporabnim konceptom iz teorije iger. Zapornikova dilema je najbrž najbolj znana igra, ki ima velik razlagalni potencial na področju biologije, psihologije, ekonomije, etike, politologije in ostalih družboslovnih ved. Osnovna ideja je zelo enostavna. Dve osebi sta pripti, ker sta obtoženi nekega prekrška ali zločina. Tožilec ločeno obema obtožencema ponudi isto izbiro. Če noben ne prizna očitane dejanja, vsak dobi kratko kazen. Če en prizna in drugi ne, potem je prvi svoboden in drugi dobi dosmrtno kazen. Če oba priznata, dobita oba srednje dolgi kazni. Vprašanje je torej, kaj se najbolj splača narediti. Če oba ne priznata, bi to seveda bil najboljši scenarij za oba, ker bi v tem primeru dobila najmanjše skupno število let kazni. Če oba priznata, pa je to najslabši izid za oba, ker je skupno število let kazni v tem primeru največje. Seveda bi za vsakega posameznika bil najboljši izid, če prizna, medtem ko drugi ne prizna, in, obratno, najslabši izid, če ne prizna, medtem ko drugi prizna. Mogoče bi bilo najbolje obljubiti si *sodelovanje*, torej da noben ne bo priznal. Kljub obljubi pa bi bila seveda oba soočena z veliko skušnjavo, da priznata in se tako popolnoma izogneta kazni, seveda v primeru, da bo drugi »naivno« držal obljubo. Zlasti v primeru, da se drugi ne bo držal zaobljube, je seveda boljše, da tudi prvi prizna, da sam ne izpade naivnež in si tako »prisluži« dosmrtno kazen. Če oba priznata, dobita torej srednje dolgo kazen. To pa je seveda slabše, kot če bi sodelovala in noben ne bi priznal. Torej bi za oba bilo najboljšo, če bi jima le nekako uspelo preseči svoje sebično razmišljanje in strah pred »izdajo« oziroma sebičnim razmišljanjem drugega igralca.

⁴ To je najpomembnejši razlog (in napaka, kot bomo videli!) za idejo, da bi darvinizem, apliciran na družbo (socialni darvinizem), impliciral skrajne sebično, sovražno, militantno, anti-socialno obnašanje.

Za teorijo je pomembna logična oblika igre, ne njena vsebina. Zato je tako zelo uporabna na različnih znanstvenih področjih, kot smo omenili. Na primer zapornikove dileme naletimo kjerkoli, kjer srečamo takšno strukturo – enostransko odpadništvo/zavrnitev sodelovanja (*defection*) je bolj koristno kot obojestransko sodelovanje, ki pa je bolj koristno kot obojestranska zavrnitev sodelovanja, ki je še vedno bolj koristna od enostranskega sodelovanja. V evolucijski teoriji iger se koristi štejejo v potomcih. Modele iger se prireja na takšen način, da bi prišli do pogojev, v katerih se bolj splača sodelovati. Odpadništvo je izhodiščna zmagovalna teorija, ker se bolj splača od sodelovanja, ne glede na to, kaj drugi igralec naredi. Če torej dva igralca primerjamo po številu potomstva, vidimo, da je evolucijsko veliko bolj utemeljeno odpadništvo. Simulacije igre, kjer je recimo izhodišče takšno, da imamo konstantne odpadnike (*defectors*) in konstantne sodelavce (*cooperators*), pokaže, da odpadniki zelo hitro prevladajo v določeni populaciji, in sicer v času nekaj generacij. (Dennett, 2003: 149) Brez *dodatnih* okoliščin se žal sodelovanje izkaže za zelo slabo strategijo. Najbolj vplivno aplikacijo teorije iger na evolucijsko teorijo je razvil John Maynard Smith s konceptom *evolucijsko stabilne strategije* (ESS). Svet, v katerem vsi zavrnejo sodelovanje, je tako primer ESS, ker alternativne strategije nimajo možnosti, da bi uspešno spremenile trend.⁵ Če bi torej v takšen model postavili nekaj sodelavcev, bi jih takšen svet po zelo hitrem postopku iztrebil. Obstajajo pa posebne okoliščine, ki omogočajo drugačen izid. Ravno takšne okoliščine so bile potrebne za nastanek današnje družbe, kot bomo videli v nadaljevanju.

3.3 Možnosti evolucijske razlage pojava morale

Sledimo torej evoluciji življenja, ki poteka v popolnoma determinističnem okolju. Zgodovina razvoja življenja je zgodovina razvoja vedno bolj sposobnih in vztrajnih izogibalcev, ki jim deterministični fizični zakoni ravno pomagajo pri preživetju. Ko smo na takšen svet naravne selekcije nanesti strukturo zapornikove dileme, smo videli, da simulacija ne podpira sodelovanja kot stabilne strategije, tj. takšne, ki bi se lahko obdržala. V nadaljevanju bomo preverili, če določene specifikacije okoliščin lahko prevesijo tehtnico na stran sodelovanja, če lahko odkrijemo *rojstvo morale iz duha evolucije*.

Kaj bi iz evolucijske perspektive lahko rekli o altruizmu? Altruizem lahko opredelimo kot postavljanje dobrobiti drugih ljudi kot smotra na sebi. (Sober in Wilson, 1998, v Dennett, 2003: 194–195) Tukaj seveda *drugi ljudje* ne pomeni lastni otroci ali družina, za takšen »altruizem« ponuja evolucija zelo prepričljivo razlago.⁶ Če pa je dobrobit drugih *smoter na sebi*, ali potem ne govorimo o pravem al-

⁵ Še en primer ESS je naravna selekcija. Ko se naravna selekcija enkrat utrdi znotraj neke populacije, potem ni možno, da bi jo neka alternativna strategija uspešno premagala.

⁶ David G. Myers tako komentira, da se Carnegiejeva medalja za pogum ne dodeli tistim, ki rešijo sorodnike (Myers, 2008).

truizmu, če se zaradi altruističnega dejanja nekdo počuti dobro sam pri sebi, če mu to dejanje lajša vest? Predpostavimo, da je pojem altruizma vsaj malo bolj konsistenten od tradicionalnega močnega pojma svobodne volje. Ali so altruisti mogoče podobni mulam, ki so možen pojav, čeprav takšen, ki se ni zmožen obnavljati, reproducirati? (Dennett, 2003: 196). Če altruizem primerjamo s sodelovanjem, bi mu, glede na izsledke zgoraj omenjene simulacije, bil usojen propad, če se pojavi v odpadniškem okolju. Ampak tako kot mule, kljub temu da se ne morejo razmnoževati, lahko doživijo razcvet v *določenih* okoliščinah, na primer s pomočjo človeške vrste, morda obstajajo tudi okoliščine, v katerih altruizem ni slepa ulica, strel v nogo, temveč omogoča razvoj posameznika in družbe, v kateri se pojavi.

Najosnovnejši evolucijski model družbe, kot smo videli zgoraj, nam dokazuje, da konstantni sodelavci, recimo jim kar altruisti, neizogibno propadejo. Vendar pa, kolikor več realnih okoliščin dodamo v model, toliko bolj začenjamo prepoznavati obrise nečesa podobnega altruizmu ali vsaj kvazi-altruizmu. Razviti modeli prikazujejo »pogoje, pod katerimi /.../ lahko evolucija določi organizme za sodelovanje ali, bolj natančno, za takšno obnašanje, da so bolj naklonjeni dolgoročni dobrobiti skupnosti kot svoji trenutni dobrobiti« (Dennett, 2003: 197).

Na osnovnem modelu hitro vidimo, kako občutljiv je pojav sodelovanja. Če se znotraj celotne populacije sodelavcev, altruistov, pojavi odpadnik ali zastonjkar, bo imel vsaj enako ali večje število potomcev. Zastonjkar bo namreč deležen istih koristi kot vsi, ne da bi ob tem imel stroške, ki jih imajo sodelavci. Tako bo zelo kmalu znotraj nekoč sodelovalne populacije prevladalo zastonjkarstvo.⁷ Razen če seveda ne intervenira nekaj, kar bi preprečilo razmah zastonjkarjev.

Odmislimo trenutno neke nadnaravne korektive, kot je na primer božja kazen, in poskusimo poiskati razlago v naravi. Ali lahko evolucija proizvede neko sistematično zaščito pred invazijo odpadnikov? Modeli evolucijske zapornikove dileme *druge generacije* vsebujejo eno »konkretizacijo« več glede na svoje bolj enostavne predhodnike. Dodali so namreč dejavnik »gostote« tako, da se verjetnost stika z nekom, ki je bližje, povečuje z gostoto sveta. Tako se je izkazalo, da *soseščina* igra veliko vlogo v razvoju sodelovanja. Soseščina naredi bolj verjetno dejstvo, da bo nekdo stopil v stik s sebi podobnimi. Sodelovanje tako v povprečju prinaša večje koristi glede na to, da bo obnašanje na drugi strani verjetno recipročno, v primeru recipročnosti pa je seveda sodelovanje boljša strategija. Dodatni kriterij, ki je sledil, je bila *možnost izbire*, tako da so igralci lahko pod določenimi pogoji zavrnilo igro. To je povzročilo oblikovanje rojev podobnih igralcev. Sodelavci so

⁷ Takšen pristop je zlahka mogoče uporabiti pri analizi zgodovinskih okoliščin propada socialističnih sistemov, ki pa očitno lahko ob spremenjenih okoliščinah vzniknejo in se ohranijo tudi »od spodaj navzgor« (*bottom-up*). Takšen primer so med drugim skandinavske države, ki seveda »nominalno« ne štejejo kot socialistične, ampak za njih bi lahko rekli, da je odpadništvo manj pogosto. Naša »evolucijska« teza bi bila, da so bolj ostri pogoji preživetja ustvarili bolj sodelovalno skupnost in posameznike.

rajši poiskali druge sodelavce, tako da so tisti, ki so odklonili sodelovanje, obtičali z drugimi, sebi enakimi (Dennett, 2003: 200). Znotraj roja sodelavcev je torej sodelovanje postalo stabilna strategija. Ampak takšno obnašanje sodelavcev lahko opredelimo za kvazi-altruistično, ker bi od pravega altruističnega sodelavca pričakovali, da bo na svojo lastno škodo vztrajal pri odpadnikih. Seveda pa so tudi znotraj tega modela igralci zelo preprosti, z zelo malo »izbire«. Taktika, ki jo uporabljajo, je zelo podobna taktiki, ki jo najdemo na makromolekularni ravni intragenetskega spopada v prokariontski eri (prav tam). Za simulacijo makromolekule nam torej takšen model zadostuje, za simulacijo človeške družbe pa ga bo seveda potrebno izpopolniti.

3.4 Pomen kazni pri evoluciji sodelovanja

Če naše igralce opremimo s še nekaj sposobnostmi, predvsem možnostjo učenja iz izkušenj, postane model bolj zanimiv. Zastonjkarji so bili zmožni preplaviti neko skupino, ker jih ni nihče zaznal. To lahko popravimo tako, da dodamo možnost, da skupina pridobi informacije o njih in se odzove. Skupina lahko tako uvede določene sankcije proti zastonjkarjem *in* naivnežem, recimo jim resnični altruisti, ki se puščajo izkoriščati zastonjkarjem, saj s tem omogočajo njihov razcvet. Seveda to še ne pomeni utrditve sodelovanja kot ESS, v resnici se oboroževalna tekma s tem šele začne – ali se lahko mutirani zastonjkarji in naivneži dovolj dobro zakrinkajo, da jih nihče ne loči od sodelavcev, in ali lahko skupina razvije nove metode za odkrivanje mutantov. Za evolucijo sodelovanja oziroma družbe je torej bistven pojav kazni, v kakršnikoli obliki že. Rob Boyd in Peter Richardson sta v svojem članku o kulturni evoluciji dokazala, da če so stroški kaznovanja relativno nizki, če lahko tudi tisti, ki sami ne kaznujejo, zanesljivo pričakujejo kaznen, to lahko pripelje do neomejenega konformizma znotraj skupine. Naslov njujnegaj članka je zelo zgovoren: »V dovolj velikih skupinah kaznovanje omogočaja evolucijo sodelovanja (ali česarkoli drugega)« (Dennett, 2003: 200).

Uspelo nam je torej utemeljiti pojav sodelovanja na skrajnje darvinističnih temeljih, na podlagi skrbi za lastno dobrobit in dobrobit svojega potomstva. Sebičnigen se torej lahko izkaže za ne tako skrajnje sebičnega oziroma ne *samo* sebičnega. V besedah Allana Gibbarda: »Oblike sodelovanja, ki so našim prednikom pomagale prenesti njihove gene, da bi ustvarili nas, je vredno hoteti – zaradi boljših razlogov. Darvinistične sile so oblikovale skrbi in čustva, ki jih poznamo, od katerih so nekatera v splošnem moralna« (Gibbard, v Dennett, 2003: 201). Ampak zgoraj opisani svet, kljub nekaterim »moralnim« potezam, deluje zelo hladno in robotsko, kvečjemu spominja na pojave iz živalskega sveta, kjer so nekatere vrste bolj »človeške«, bolj nagnjene k sodelovanju. Za pojav resničnega altruizma bomo našemu modelu seveda morali dodati kakšno značilno človeško potezo. Vsekakor pa smo videli, kako lahko v enostavnem modelu pridemo do družbenega sistema, ki se lahko sam ohranja in ga lahko vidimo kot nujen predpogoj za dolgoročni razcvet resničnega človeškega altruizma.

»Pošast odpadništva visi nad nami vsemi, evolucijski izvirni greh s svojim, večno mamečim, preudarkom: kako je lahko odpadništvo v tem primeru *neracionalno*?« pravi Dennett (Dennett, 2003: 202). Odmislimo primere, ko se ne spleča sodelovati zaradi strahu pred kaznijo⁸ ali izobčitvijo. Na sled svobodni volji in moralni odgovornosti nas pripeljejo ravno težki primeri, ko nam notranji glas govori, da »saj nihče ne bo zvedel, pomisli kaj vse lahko s tem pridobiš« (prav tam: 203). Takšni najtežji primeri skušnjave nam lahko razložijo pojav tradicionalnega koncepta duše kot nesmrtni in duhovne entitete. Poskusili bomo slediti njeni evoluciji iz predhodnih oblik nadzornih sistemov, da bi razumeli, zakaj deluje tako kot deluje.

3.5 Pojav racionalnih akterjev

Ekonomist Robert Frank je postavil tezo, da se evolucijski napredek človeka in družbe zgodi z obratom: »biti dober zato, da bi *dajal videz dobrega*« zamenja »biti dober je bolj pomembno kot dajati videz dobrega«. S tem je človeštvo po Franku doseglo novo raven *evolucije svobode*, tako da so se ljudje naučili soočiti in rešiti *probleme zaveze*, ki »nastanejo, ko oseba iz lastnega interesa sklene obvezujočo zavezo do nekega obnašanja, ki bo v prihodnosti dajalo vtis, da nasprotuje lastnemu interesu« (Frank, 1988, v Dennett, 2003: 204). Vidimo, da ima problem zaveze isto strukturo kot zapornikova dilema – problem je seveda v tem, kako vedeti, kdo bo spoštoval obveznosti, ki izhajajo iz zaveze, in kdo ne. Prvoosebni problem, ki se pojavi z razvojem racionalnih akterjev, ki so zmožni reflektirati razloge za delovanje, pa je dvojne narave: kako naj *sam* spoštujem zavezo in kako naj prepričam *druge*, da jo nameravam spoštovati. Ker so drugi tudi racionalni akterji, ne morem pričakovati, da mi bodo verjeli na besedo. Po Franku pa ravno na tem mestu nastopijo *čustva*, in sicer v dvojni funkciji: preprečujejo nam, da bi bili *preveč* racionalni, in poskrbijo za naš sloves kot *ne preveč* racionalnih. Čustva preprečujejo neželjen izbruh kratkovidne oziroma lokalne racionalnosti, kar je zelo pomembno za razvoj družbeno odgovornega akterja, ki bo zaradi njih težje podlegel skušnjavam, ki bi jih sicer zgolj racionalno lahko upravičil (Frank, 1988, v Dennett, 2003: 204).

Sloves sodelavca kot »ne preveč pametnega« je pomemben predvsem iz vidika sprejema v skupino sodelavcev, ki je materialno veliko bolj preskrbljena.⁹ Takšna skupina bo hitro postala plen odpadnikov, če ne bo iznašla učinkovitega načina ločevanja med sodelavci in odpadniki. Če to prej na molekularnem nivoju ni bil problem, pa vsekakor postane s pojavom racionalnih akterjev. Verodostojen znak

⁸ Tukaj upoštevamo samo *tuzemsko* kazen. *Onstranska* kazen bi že bila dobra rešitev oziroma stabilna strategija, če bi le vsi verjeli vanjo. Sicer bi pa tej rešitvi lahko očitali isto, kar se očita evolucijskim modelom moralnosti – nemoralen, utilitarističen temelj.

⁹ Spomnimo se razlike med izidi zapornikove dileme v primeru sodelovanja obeh ali v primeru odpadništva obeh.

zanesljivosti nekega akterja ne sme biti takšen, da ga bo »racionalen akter zmožen enostavno ponarediti«¹⁰ (Dennett, 2003: 205). Torej poleg osebnega problema, kako postati zaupanja vreden akter, imamo še problem, kako *dokazati*, da smo vredni zaupanja.

3.6 Samonadzor–vest–altruizem

Po Franku je evolucijska lepota ravno v tem, da čustvom dodeli dvojno nalogo: pomagajo nam namreč pri samonadzoru in hkrati drugim priskrbijo verodostojen znak tega uspeha. Drugi se tako prepričajo, da smo čustvene osebe, ki *strastno* skrbimo za spoštovanje zavez, da smo za ohranitev svoje integritete, poštenosti, pripravljeni plačati iracionalno visoko ceno, seveda z vidika kratkovidnega kritika. Trik za pridobitev ugleda poštenjaka je, da *smo* resnično pošteni (Dennett, 2003: 213).

Zakaj je to najbolj učinkovita rešitev problema, nam bo jasno, ko spoznamo, da je to cena, ki jo plačamo za samonadzor. Nad samonadzorom »nimamo nadzora« – ne poteka selektivno niti natančno. Ne moremo namreč v realnem času nadzirati vsakega ločenega dela svojega odločanja,¹¹ zato se zanašamo na malo bolj grob, nerafiniran pristop, ko se »opremimo z močno emocionalno naravo, ki se razlije čez svoj cilj, in zaradi katere se tresemo od jeze takrat, ko je jeza primerna, ali smo nezmožni brzdati svoje veselje, ko je veselje na mestu...« (Dennett, 2003: 214). Prav *ista* čustva, ki nam pomagajo pri samonadzoru, ko se odpovemo neki svoji kratkoročni koristi zaradi neke svoje dolgoročne koristi, so tista, zaradi katerih se odpovemo svojim kratkoročnim interesom v korist drugih ljudi (prav tam). Takšna nujna altruistična posledica pridobitve samonadzora je v resnici zelo koristna. Brez čustvenih izrazov bi nas obtožili pretvarjanja, pravi Dennett.

O odnosu samonadzora in vesti pravi Frank naslednje: »Lahko bi si predstavljali populacijo, kjer so ljudje z vestjo bolj uspešni kot tisti, ki so brez vesti. Ljudje brez vesti bi sicer varali manj, če bi lahko, ampak enostavno imajo večje težave pri reševanju problema samonadzora. Nasprotno pa so ljudje z vestjo zmožni pridobiti velik ugled in uspešno sodelovati s sebi podobnimi« (Frank, 1988, v Dennett, 2003: 215). Ali lahko tako pridemo do *pristnega altruizma*?

Ljudje s pristnimi moralnimi čustvi so bolj sposobni ravnati v svojem lastnem interesu /.../ Ljudje z visokim ugledom lahko tako rešijo celo nepovnljene zapornikove dileme. Na primer: lahko uspešno medsebojno sodelujejo v poslovnih podvigih, kjer je nemogoče odkriti prevaro. Z drugimi besedami – pristni altruizem lahko vznikne že na podlagi vzpostavitve ugleda zaradi preudarnega ravnanja. (prav tam)

¹⁰ Teza Amotza Zahavija (v Dennett, 2003: 205).

¹¹ Za več o tem glej Dennett (2003: 221) oziroma nevrološke raziskave o zavestnem odločanju.

Frankova teza je torej, da altruizem še zdaleč ni obsojen na propad. Tezo je testiral tudi s poskusom, kjer so študenti igrali večkratno zapornikovo dilemo, v kateri so lahko zaslužili določeno nezanemarljivo vsoto denarja.¹² Študenti so pred začetkom igre dobili možnost, da se na kratko spoznajo. Frank je ugotovil, da so na podlagi tega bili med 60 in 75 odstotkov sposobni napovedati, kdo bo »sodelavec, kdo pa odpadnik« (Frank, 1988, v Dennett, 2003: 216). To po njegovem potrjuje naši intuiciji, da lahko spoznamo neoportunistične ljudi, kar je glavna premisa, na kateri je utemeljen model zaveze:

Iz te premise logično sledi, da se bo neoportunistično obnašanje pojavilo in preživelo celo v neusmiljeno tekmovalnem materialističnem¹³ svetu. Lahko se torej strinjamo, da materialistične sile absolutno upravljajo naše obnašanje, vendar istočasno zavrnemo idejo, da so ljudje vedno in povsod motivirani z lastnim materialnim interesom.« (Frank, 1988, v Dennett, 2003: 216)

V material(istič)nem svetu seveda dominira tisto obnašanje, ki ima za posledico največji materialni uspeh. Ampak »vedno znova ugotavljamo, da najbolj prilagojeno obnašanje ne vznikne iz prizadevanja po materialni koristi. Zaradi pomembnih problemov z zavezo in izvedbo se bo to prizadevanje izkazalo za 'samoporazno'. Zato, da bi nam (materialno) šlo dobro, moramo včasih nehati skrbeti za maksimizacijo materialne koristi.« (Frank, 1988, v Dennett, 2003: 217)

Torej nam Frank ponuja razlago, kako altruizem oziroma moralo spraviti z evolucijo. Altruizem se dolgoročno materialno splača oziroma vsaj ne škodi:

Videti je, da gre altruistom tudi ekonomsko bolje: eksperimentalne študije vztrajno odkrivajo povezavo med altruističnim obnašanjem in socioekonomskim statusom. To seveda ne pomeni, da altruistično obnašanje *povzroča* ekonomski uspeh. Nakazuje pa, da altruistična drža ne more biti preveč resno breme v materialnem smislu. (Frank, 1988, v Dennett, 2003: 217)

¹² Zanimivo dejstvo je, da so bili študenti prvega letnika ekonomije bolj nagnjeni k sodelovanju kot študenti četrtega letnika ekonomije.

¹³ Materialno in materialistično je v tem primeru mišljeno v smislu gmotne koristi in ne filozofskega nauka.

3.7 Vznik ključnih značilnosti morale iz evolucijske oboroževalne tekme

Poglejmo torej, kam nas je pripeljal poskus naturalistične razlage morale:

Zapletenost družbenega življenja *homo sapiensa* kot bitja, ki ga bistveno določata jezik in kultura, ustvarja serijo evolucijskih oboroževalnih tekem, iz katere vzniknejo ključne značilnosti človeške morale: interes za iskanje pogojev, v katerih se lahko krepi sodelovanje; občutljivost do kaznovanja in groženj; skrb za ugled; nagnjenja na višji ravni k pozitivni samomanipulaciji, ki je ustvarjena zato, da okrepi samonadzor v pričo skušnjave; in sposobnost ustvariti zavezo, ki jo bodo drugi cenili in upoštevali. Takšne inovacije lahko uspevajo skupaj s posebnimi pogoji, s katerimi so se pojavile, ter lahko tako izpodrinejo kratkovidnost oziroma sebičnost bolj enostavnih organizmov. (Dennett, 2003: 218)

Skupaj z Dennettom in Frankom smo torej skicirali pot od začetkov življenja do pojava oseb kot akterjev, ki jim »svoboda predstavlja največjo moč in hkrati največji problem« (Dennett, 2003: 218). Dennett torej na naturalističnih temeljih ponudi precej prepričljivo razlago nastanka morale. Malo manj pa je prepričljiva njegova razlaga svobode. Tako kot se je njegove knjige »*Pojasnjena zaveza*« prijel naslov »*Odstranjena zaveza*«, so nekateri kritiki tudi knjigo »*Vznik svobode*« preimenovali v »*Vznik morale brez svobode*« (Mozes, 2010). Ne gre samo za negativno nastrojene kritike, ki ga vidijo kot dogmatičnega determinista. Številni avtorjev, ki sicer simpatizirajo z njegovim projektom, so mnenja, da Dennett svobodno voljo utemeljuje predvsem zaradi strahu pred negativnimi družbenimi posledicami, ki bi se po njegovem pojavile, če bi ljudje mislili, da nimajo svobodne volje. Dennett sam seveda ponuja ogromno priložnosti za takšno branje.¹⁴

Tudi Dennettova teza, da si ljudje želimo biti obravnavani kot moralno odgovorni, ker si želimo biti polnopravni člani družbe, kaže na pojmovanje svobodne volje in moralne odgovornosti kot neke družbeno dodeljene »zmožnosti«. ¹⁵ In ravno za takšen »družbeno realen« pojav gre – svobodna volja je po Dennettovih besedah pojav, ki se vzpostavi v družbi, tako kot denar ali glasba. Izven družbe je torej po Dennettu nesmiselno govoriti o svobodni volji in moralni odgovornosti.

V nadaljevanju si bomo pogledali, katere pogoje mora izpolniti posameznik, da bi mu družba lahko upravičeno pripisala svobodno voljo in moralno odgovornost, in

¹⁴ Tako na primer v svojem govoru ob sprejemu nagrade fundacije Erasmus Dennett pravi: »Fundacija Erasmus in jaz deliva mnenje, da bi nazor o svobodni volji kot iluziji lahko imel usodne posledice za družbo v primeru, da se ga argumentirano ne ovrže.«

¹⁵ Zdi se, da je to trditev težko spraviti z zgornjo trditvijo o usodnih družbenih posledicah nazora o iluzornosti svobodne volje. Ljudje, ki si *želijo* biti moralno odgovorni, bi torej, soočeni z iluzornostjo moralne odgovornosti in svobodne volje, v takšni meri spremenili svoje obnašanje, da bi to lahko imelo usodne posledice?

ali jih je sploh zmožen izpolniti. Glavna pogoja za takšno pojmovanje svobodne volje in moralne odgovornosti bosta racionalnost in samonadzor. Kriterijev seveda ne bodo zmožni izpolniti vsi posamezniki, vendar ne po lastni krivdi. Veliko vlogo igra namreč tudi sreča, namreč s kakšnimi predispozicijami se oseba rodi in kakšne vzgoje je deležna. Preiskali bomo torej človekovo glavno *differentio specifico* glede na njegovo biološko opredelitev: mišljenje oziroma razum (*sapientia*¹⁶). Iskali bomo možnost človekove avtonomije, ki jo lahko izpeljemo iz naturalistično razumljenega pojma razuma, iz tega, da človek pridobi oziroma ozavešti razloge, ki so pred njegovim nastankom bili zgolj »prosto lebdeči«, kar človeka naredi za *racionalno* žival.

4. Vznik razuma

4.1 Prosto lebdeči in ujeti razlogi

Do zdaj smo evolucijski proces in tako tudi razloge za sodelovanje ali odpadništvo utemeljevali brez kakršnega koli sklicevanja na razumevanje tega procesa in na razloge s strani akterjev, ki jih ta proces oblikuje. Razlogi lahko ostanejo popolnoma skriti svojim živalskim izvrševalcem in še vedno popolnoma učinkoviti v evolucijskem smislu. Lahko so prosto lebdeči oziroma *free floating rationale*, kot se pogosto izrazi Dennett. Čebeli, ki žrtvuje svoje življenje, ni treba vedeti, da je to naredila zato, da bi rešila preostalo populacijo čebelnjaka. Evolucija oziroma narava funkcionira na *need to know* osnovi (Dennett, 1984: 24). Do pojava *homo sapiens*a bi lahko rekli, da noben posameznik v živalskem svetu ni imel potrebe ali koristi od reflektiranja svojih razlogov ali interesov. Evolucija je nemoteno potekala naprej. Avtomatizirano, »programirano« obnašanje in odzivi na okolje so zelo učinkoviti, hkrati pa ne zahtevajo veliko energije, kot jo na primer zahteva pisanje tega stavka. Primeri, ko se nam kakšno živalsko obnašanje zdi nerazumno ali celo smešno,¹⁷ so popolnoma zasenčeni ob navdušenju nad razumnostjo celote stvarstva te »Absolutne nevednosti« (Dennett, 1984: 11). Vendar pa, kot smo videli zgoraj v primeru izogibalcev, predstavlja možnost za inovacije, ki se dogajajo *horizontalno* in s pomočjo »žerjavov«, za razliko od *vertikalnih*, ki potrebujejo več generacij, velik evolucijski napredek. Videti je, da se je investicija v razvoj energetske zelo zahtevnih možganov evolucijsko izredno izplačala.

Medtem ko smo zgoraj z lahkoto izpeljali skoraj celotno evolucijo brez sklicevanja na kakršnokoli vednost, pa vidimo, da že mladiči vrste *homo sapiens*, seveda ob posredovanju kulture in jezika, kažejo zelo veliko vnemo zlasti pri enem opra-

¹⁶ Čeprav se pojem *sapientia* nahaja v človekovi biološki opredelitvi, ga nikakor ne nameravamo samoumevno sprejeti kot dejstvo, ki bi utemeljilo možnost svobodne volje in moralne odgovornosti.

¹⁷ Ko osa iz rodu *Sphex* pred brlog prinese paraliziranega črčika, ga pusti zunaj in opravi pregled brloga. Znanstveniki so izvedli eksperiment tako, da so med njenim pregledom brloga črčika prestavili za nekaj centimetrov. Ko je osa prišla ven, je črčika zvelkla nazaj pred brlog in znova opravila svoj pregled brloga. Črčika so štiridesetkrat prestavili in »uboga« osa je štiridesetkrat, vedno znova, preverila brlog.

vilu: spraševanju »zakaj«, iskanju razlogov oziroma vzrokov. Otroke »že od majhnega spodbujamo, da se nam pridružijo v iskanju in podajanju razlogov ter utemeljevanju prihodnjega početja z razlogi« (Dennett, 2003: 273). Kultura, skupnost in zlasti jezik so najbolj botrovali evoluciji človekove zmožnosti *prepoznati, reflektirati* razloge svojega delovanja.

Jezik je izjemna evolucijska pridobitev, ki je omogočila izjemen razvoj številnih, če ne prav vseh, človekovih sposobnosti. In od nekdaj je jezik služil za komunikacijo razlogov. Če parafraziramo znani biblijski citat, lahko rečemo, da je na *začetku bil razlog*. Poglejmo si namreč, kaj vse so stari Grki razumeli pod besedo *logos*: beseda, govor; pa tudi razlaga, razlog; in prav tako razum. Latinski *ratio* pa obdrži pomena razum *in* razlog, prav tako iz njega nastali francoski *raison* in iz tega nastali angleški *reason*. Od latinščine dalje se izgubi povezava z jezikom, se pa obdrži povezava razlog – razum. Te povezave v slovenščini nimamo, je pa razum bistveno povezan z iskanjem oziroma podajanjem razlogov, z razlaganjem. Slovar slovenskega knjižnega jezika poda naslednjo razlago besede razum: sposobnost vključevati, sprejemati v zavest in ugotavljati vzročne, logične povezave.

4.2 Od razloga do razuma (»From reason to Reason«)

Če torej narava funkcionira na »*need to know*« osnovi, pa je človekov moto ravno obraten: pridobiti čim več relevantnih informacij glede okolja *in* glede sebe in drugih ljudi. Zlasti zaželeno so seveda informacije, med katerimi je možno vzpostaviti vzročno zvezo, torej poiskati vzroke oziroma razloge (»Zakaj?«). Pojem vzročnosti se nam kaže kot konstitutivna komponenta razuma, že če se ozremo na zgoraj navedeno slovarsko geslo, brez filozofskih teorij in kategorij.

Človeku je zlasti pomembna refleksija njegovih in tujih razlogov za neko delovanje. Zato je tudi bil skozi zgodovino svojega razvoja, kot smo zgoraj omenili, nagnjen pripisovati razloge in interese tudi naravnim silam in neživim stvarim.¹⁸ Tudi ko se je tega sčasoma odvadil, pa je bil še vedno prepričan, da pozna razloge svojega delovanja. Zato je bil najbrž najtežji od treh udarcev¹⁹ za njegov ego prav Freudov: večina razlogov, interesov, motivov, miselnih procesov, spominov ni dostopna introspekciji. Tisto, kar je dostopno, predstavlja samo vrh ledene gore.²⁰ Ne potrebujemo torej nobenih zlobnih nevroznanstvenikov iz filozofskih miselnih eksperimentov na temo svobodne volje in moralne odgovornosti. Tukaj imamo precej benevolentnega nevroznanstvenika, ki nam pravi, da ne poznamo večine vzrokov za svoja dejanja.²¹ Ampak to je samo pol problema, s katerim nas sooči

¹⁸ Zanimiva je komična »filozofska smer« rezistencializem, katere glavni slogan je »Stvari so proti nam«.

¹⁹ Prvi je Kopernikov udarec, da Zemlja ni središče vesolja, drugi pa Darwinov, da imajo človek, opica in vse ostalo življenje na Zemlji skupnega prednika, da človek ni narejen po podobi boga.

²⁰ John Bargh trdi, da se je Freud zmotil za vsaj decimalno, gre bolj za snežinko na vrhu ledene gore.

²¹ Še dva »benevolentna« nevroznanstvenika povezana s problemom svobodne in zavestne volje sta Benjamin Libet in Daniel Wegner.

Freud. Če je naše nezavedno dostopno psihoanalitikom, bi lahko torej z njihovo pomočjo dobili bolj pravilno in celovito sliko svojih lastnih vzgibov. Za možnost racionalnega delovanja in načrtovanja zahtevamo čimbolj natančno poznavanje vseh svojih razlogov, motivov in interesov, tako da lahko razumsko izbiramo med bolj in manj upravičenimi.

Druga polovica problema je, da je bil Freud prepričan, da tudi če smo seznanjeni z nezavednimi nagoni, ki vodijo naše ravnanje, nimamo možnosti bistveno spremeniti vzorcev svojega obnašanja. Vzorci, ki jim potem sledimo celo življenje, so po Freudu namreč bili fiksirani v zgodnjem otroštvu. Človek torej nima pomembne možnosti spremeniti navade, ki jih oceni za negativne, tudi če se jih zaveda. Navedimo primer, ki naj bi dokazoval, da so ozaveščeni razlogi kljub Freudovega mnenju lahko »praktični«. Ženska, ki se zaveda, da jo privlačijo nasilneži, se lahko zavestno začne izogibati krajem, kjer je večja verjetnost, da spozna nasilne moške. Freud bi najbrž odgovoril, da je to možno, ampak da bo najverjetneje kljub svojim frankfurtovskim željam drugega reda sledila svoji želji prvega reda (Frankfurt, 2007). Verjetnost, da spremenimo vzorce svojega obnašanja zaradi novih informacij, je po našem vsekakor večja, kot je bila pred pridobitvijo teh informacij. Vprašanje pa je seveda, ali je takšna verjetnost bistveno večja, zlasti v primeru najbolj trdovratnih vzorcev.

4.3 Svobodna volja in racionalnost

Ob zgoraj obravnavanih težavah se nam možnost navezati svobodno voljo na princip *popolne racionalnosti* zdi zelo problematična. In vendar je ideja podreditve volje nareku razuma zelo pogost pojav v literaturi o svobodni volji. Kant je mogoče res utemeljil svojo celotno moralno filozofijo na tej ideji, ampak nikakor ni bil osamljen niti ne prvi, ki je svobodo vezal na razum, med drugimi so podobno trdili Spinoza, Descartes in celo Locke. Povezavo med racionalnostjo in svobodno voljo pa na podoben način ugotavlja tudi John Searle. Po njegovem teorije racionalnosti skoraj neizogibno predpostavljajo neko stopnjo svobodne volje. Kakšen smisel bi namreč imela zmožnost racionalno narediti najboljši načrt za delovanje, če ne bi bili zmožni spremeniti svojega obnašanja tako, da sledimo temu najboljšemu načrtu. Brez tega bi, po njegovem, zavestni um (*conscious mind*) bil nemočen, depresiven opazovalec lastnih iracionalnih dejanj, medtem ko bi ugotavljal, da bi neko drugo dejanje bilo boljše (Searle, v Baumeister, 2008: 71). Vendar, kot smo videli pri Freudu, tudi takšna pesimistična razlaga ni izključena. Verjetno se dogaja, da se »opazujemo« pri ravnanju, za katerega smo ocenili, da ni najbolj racionalno, smotno. Ampak bi se tudi v tem primeru strinjali s Searlom, da bi *možnost* racionalnega ravnanja implicirala možnost svobode. Racionalno dejanje bi bilo bolj svobodno od iracionalnega, če naj svoboda karkoli pomeni, pravi Locke:

Če je resnična svoboda v osvoboditvi od vodstva razuma in v želji po omejitvi razmišljanja in presojanja, ki nam preprečujeta, da bi izbrali ali naredili slabše dejanje, potem so norci in neumneži edini svobodni ljudje. Menim, da se nihče ne bi odločil biti nor zavoljo takšne svobode, razen tistega, ki je že nor. (Locke, 1690, v Dennett, 1984: 51)

Po Freudu vsi kažemo nekatere psihopatološke znake v vsakdanjem življenju – vsi smo nekoliko nori, če se izrazimo bolj neposredno. Ampak ali ne nastopa ravno to kot omejitev naše svobode? Svobodo ravno vidimo v možnosti, tudi če popolnoma nerealni, ubežati ravnanju, za katerega ocenjujemo, da je manj racionalno in manj pod našim zavestnim nadzorom. Ali mogoče tudi ne?

Ideja svobode kot popolne racionalnosti je izražena v filozofski literaturi tako pogosto, pravi Dennett, da mora biti na njej nekaj resničnega. Ampak po drugi strani je izražena tako vehementno, da se v njej mora kot protiutež skrivati neka napaka, nekaj zaskrbljujočega in nasprotujočega naši intuiciji. Intuicija ali nekaj drugega očitno ustvarjata takšen odpor do te ideje, da se tem avtorjem zdi tako pomembno razgllašati svoj vpogled²² (Dennett, 1984). Vzrok je najbrž v strahu pred avtomatizmom, izgubo spontanosti, individualnosti, ustvarjalnosti. Ampak ta strah predpostavlja nekatere zelo neverjetne lastnosti (človeške) racionalnosti. Ena takšna lastnost je, da bi v določeni situaciji bilo najbolj racionalno izbrati točno eno in isto dejanje, ne glede na ponovitve te situacije. Vprašanje je torej parafrazirano vprašanje svobodne volje: če popolna racionalnost obstaja, *ali bi lahko ravnal drugače?* Ali ne bi bilo najbolj racionalno *vsakič* ravnati enako, če predpostavimo identične okoliščine oziroma nekaj takega kot identično preteklost v primeru svobodne volje? Temu lahko ugovarjamo, da bi bilo nemogoče predvideti vse posledice vsakega od možnih dejanj, tudi če bi lahko poznali »vse relevantne podatke« (*all! things considered*). Odločitev se na koncu vedno zgodi tako, da premišljevanje oziroma racionaliziranje na neki točki prekinemo in se odločimo, ne pa da ga pripeljemo do absolutnega konca, ki bi zahteval neskončno časa. V takšni bitki s časom, in ob nepopolnih podatkih, bi se najbrž tudi najbolj popoln razum ob ponovljeni izbiri različno odločal. Takšna »spontanost« ima veliko prednosti. Največja je v tem, da na ta način lahko razvijamo vedno nove in bolj učinkovite načine odzivanja. Ljudje smo sicer popperjanska bitja, zmožna virtualno testirati različne scenarije in se tako izogniti posledicam slabih strategij, ki bi v velikih primerih bile skrajne pogubne. Ampak nekaterih »simulacij« enostavno ne moremo »zagnati« in izpeljati samo v možganih, oziroma vsaj celote njihovih posledic ne moremo predvideti samo z možgansko simulacijo.²³ Različnost odzivov na enake

²² Je v temelju te intuicije ravno strah, da nismo kompetentni racionalni akterji? Najbrž ne, spomnimo se Descartesove šaljive opazke, kako je očitno razum najbolj pravično razdeljena dobrina. Intuicija se vsekakor ne napaja iz Freuda.

²³ Podobno kot računalniške simulacije imamo namreč »težavo«, ker je vsakič potrebno okoliščine idealizirati, jih poenostaviti.

situacije je racionalna tudi zaradi večje nepredvidljivosti, kar je bistvenega pomena, glede na evolucijsko oboroževalno tekmo (Dennett, 1984: 71). Racionalnost torej nikakor ne nasprotuje spontanosti, individualnosti in ustvarjalnosti.

Torej skrb, da bi kot popolni racionalni akterji izgubili pristne človeške značilnosti in zdrsili na nižjo razvojno ali, človeku prav tako pomembno, moralno stopnjo,²⁴ ni utemeljena. Vseeno pa ima filozofska tradicija od Sokrata dalje težave, kako uskladiti človeka in razum. Tisti Searlov racionalni opazovalec lastnega iracionalnega početja ima po našem dve možni »racionalni« razlagi te diskrepance: *šibkost volje* ali *samoprevara*. Šibkost volje je torej prepoznanje diskrepance, »nemočno« opazovanje tega, kako naše želje prvega reda premagujejo želje drugega reda. Samoprevara pa je najbrž še večji udarec predpostavki racionalnega akterja. Beseda racionalizacija je že dolgo nazaj postala sopomenka upravičevanju tudi kot opravičevanju svojega početja drugim in samemu sebi. Mogoče pa samoprevara ne bi smeli tako kritično obravnavati. Možno bi bilo namreč tudi zagovarjati njene pozitivne evolucijske in psihološke vplive. Tudi če samo za trenutni namen predpostavimo, da je na primer občutek svobodne volje samoprevara, lahko trdimo, da gre za pozitivno samoprevaro, za katero smo že omenili, da ima veliko pozitivnih učinkov na človekovo življenje.

4.4 Neskončna lahkost avtomatiziranosti

Po zgoraj omenjenih iracionalnih »pošastih iz Ida«, kot Kihlstrom imenuje Freudove nezavedne motive naših dejanj, si zdaj pogledajmo novo verzijo *adaptivnega nezavednega* (*adaptive unconscious*) kognitivne psihologije (Kihlstrom, 2008: 165). John Bargh v svojem zagovoru *avtomatiziranosti* človekovega delovanja trdi, da se je Freud utegnil zmotiti za celo decimalko, da je zavestno delovanje, bolj kot vrh ledene gore, zgolj snežinka na vrhu le-te (Kihlstrom, 2008: 162) ali, kot sam nekoliko v šali pravi, avtomatično delovanje predstavlja 99,44% vsega človekovega delovanja (Bargh, 1997, v Kihlstrom, 2008: 162). Avtomatiziranost ne implicira nujno iracionalnosti delovanja, razen če seveda za racionalnost ne zahtevamo ozaveščenost, temveč delovanje na *avtopilotu*, brez vplivov zavestnega preudarjanja. Po Barghu namreč nismo ne tako zavestni ne tako svobodni, kakor smo si do zdaj mislili (Bargh, 1997, v Kihlstrom, 2008: 162), zato se moramo sprijazniti z *neskončno lahkostjo avtomatiziranosti*. Adaptivno nezavedno namreč po Barghu odlično opravlja različna pomembna dela, kot so zaznavanje okolja, opozarjanje pred nevarnostjo, zastavljanje ciljev in sprožanje dejanj na sofisticiran in učinkovit način, in predstavlja nujen in obsežen del visoko učinkovitega psihičnega aparata (Bargh, 1997, v Kihlstrom, 2008: 162). Bargh trdi, da je težko ubežati napovedi, da bo z napredkom znanosti o psiholoških pojavih vedno manj prostora za svobodno voljo in celo zavestno odločanje samo (Bargh, 1997, v Kihlstrom, 2008: 1).

²⁴ Glej Frankova izvajanja o racionalnosti popolne sodelovalne (moralne) države v zgornjem razdelku.

Ob takšnih raziskovalnih ugotovitvah, kot so zlasti Barghove in Wegnerjeve, in podobnih nadaljnjih napovedih, se zdi problem svobodne volje bistveno manj pomemben kot problem zavestnega odločanja. Če je namreč sama *zavestna volja* iluzija, kot se glasi naslov Wegnerjevega temeljnega dela (*The Illusion of Conscious Will*, 2002), ali ni s tem odpravljena tudi sama svobodna volja, ali pa bi lahko govorili o nezavedni svobodni volji? V primeru, da se Barghova teza o 99,44% avtomatiziranosti vsega človekovega delovanja in Wegnerjeva teza o iluzornosti zavestne volje izkažeta za resnični, s tem nikakor ne bo ogrožena oziroma odpravljena samo možnost svobodne volje. Obe tezi se bistveno dotikata tudi vprašanja mentalne vzročnosti. Zlasti Wegnerjeva teza o iluzornosti zavestne volje, ki implicira epifenomenalizem psihičnih stanj, ogroža možnost vzročnosti zavestne volje. Isti determinizem, ki ogroža možnost svobodne volje, predstavlja tudi problem za mentalno vzročnost zaradi predpostavke sklenjene vzročne verige vseh fizičnih pojavov. Vendar sta determinizem in vzročna sklenjenost fizičnega sveta skoraj neizogibni znanstveni predpostavki. Zdi se, da lahko tudi psihologija pridobi resen znanstveni status s pomočjo avtomatiziranosti in nezavednosti delovanja in mišljenja ter epifenomenalnosti psihičnih pojavov²⁵ (Kihlstrom, 2008: 172).

Poleg takšnih precej črnogledih, bolj teoretičnih vidikov so mogoče praktični vidiki tega, kar bi lahko imenovali kompatibilistična svoboda, ravno tako precej ogroženi s številnimi dognanji socialne psihologije, ki jih po Kihlstromu lahko razumemo kot trditev »Ljudje so neumni« (Kihlstrom, 2008). Če odmislimo tezo, po kateri bi nas popolnoma racionalno delovanje naredilo nesvobodne in nas ravno iracionalnost tako naredi bolj svobodne, predstavljajo ugotovitve o bistveno »zmanjšani« racionalnosti zelo resno grožnjo predpostavljeni svobodi racionalnega akterja. Poskusi kažejo, da se ljudje v vsakdanjem življenju močno zanašamo na predsodke, hevrstike in ostale podobne procese, ki pripeljejo do velikega števila napak v sodbah.²⁶ Nekatere primere bi sicer lahko razložili na podlagi omejene racionalnosti (*bounded rationality*), ki se zaradi kompleksnosti problema, časovne omejitve in omejenih informacij zateče k uporabi določene hevrstike, ven-

²⁵ Kihlstrom ugotavlja, da socialna psihologija kaže bistvene podobnosti behaviorizmu. Po Barghu je socialni psihologiji uspelo ravno tam, kjer behaviorizmu zaradi zanikanja nujnih znotrajosebnih psiholoških razlagalnih mehanizmov ni uspelo.

²⁶ Logične napake žal nikakor niso nujno moralno nevtralne, ampak so lahko zelo povezane s predsodki. V znanem poskusu Kahnemana in Tverskyja je 85% udeležencev zagrešilo »napako konjunkcije«. Udeleženci so dobili naslednje besedilo: »Linda je stara 31 let, je samska, odprta in zelo bistra. Diplomirala je iz filozofije. Kot študentko jo je zelo skrbela diskriminacija in družbena pravičnost. Prav tako je sodelovala v protijedrskih demonstracijah. Potem so morali odgovoriti, katera od spodnjih trditev je bolj verjetna:
Linda je blagajnik.
Linda je blagajnik in je aktivna v feminističnem gibanju.
V nekaterih izvedbah poskusa se je celo 85% udeležencev odločilo za drugo možnost. Stephen J. Gould je bil očaran ravno nad tem primerom. Zanimivo mu je bilo, kot pravi, ravno to, da kljub temu, da ve, da je konjunkcija dveh trditev manj verjetna kot vsaka posamična, nek človeček v njegovi glavi še vedno skače gor in dol in se dere nanj: »Saj ne more biti samo blagajnik – preberi si opis!«

dar obstajajo tudi dokazi, da se to dogaja zaradi avtomatiziranosti in nezavednosti teh procesov. »Ne gre samo za to, da ne razmišljamo dovolj intenzivno, ampak tudi za dejstvo, da nismo dovolj pozorni na dogajanje okrog nas in na naše lastno početje« (Kihlstrom, 2008: 173). Ob takšnih ugotovitvah in podobnih napovedih behaviorizma s kognitivnim obrazom, kot Kihlstrom imenuje socialno psihologijo, se nam iracionalnost zdi veliko večja ovira naši svobodi kot determinizem fizičnega nivoja.

5. Zaključek

V pričujočem članku smo na naturalističnih temeljih poskusili razviti kompatibilistično različico svobodne volje in moralne odgovornosti. V prvem delu smo raziskali evolucijo izogibalcev znotraj determinističnega sveta in dokazovali, da je smiselno govoriti o *determiniranem izogibanju*, da (ne)izogibnost ni lastnost osnovnega, fizičnega nivoja, ampak pojem znotraj višjega nivoja, s pomočjo katerega primerjamo zmožnosti preživetja različnih bioloških vrst. Takšna rešitev seveda ne bo prepričala tistih, ki determiniranost in neizogibnost obravnavajo kot sinonima in potemtakem tudi kot pojava z istega nivoja. Vendar v tem primeru ne smejo zahtevati, da neizogibnost pomeni karkoli več kot determiniranost.

V drugem delu smo s pomočjo aplikacije zapornikove dileme na evolucijo utemeljili možnost za naturalistično razlago morale, kar je prav tako dolgo časa veljalo za težko izvedljivo. Moralno obnašanje se je izkazalo za evolucijsko stabilno strategijo, ki omogoča večji in hitrejši napredek določene družbe. Za kompatibilistično razpravo o svobodni volji in moralni odgovornosti je to pomembno, ker dokazuje, da je možno utemeljiti moralo tudi brez močnega pojma svobodne volje in moralne odgovornosti.

Zadnji del članka je posvečen razumu, ki je najverjetneje najpomembnejši kompatibilistični vidik svobodne volje. Na razum je bistveno navezana *svoboda reflektivnega samonadzora*, kot Robert Kane (2005) imenuje zmožnost razumeti in reflektivno *ovrednotiti* razloge in motive, po katerih želimo ravnati ali naj bi ravnali, in zmožnost *nadzorovanja* našega vedenja v skladu s tako reflektiranimi razlogi. Kot smo pokazali, ima človek probleme pri obeh zahtevah te svobode.

Prvi problem se nanaša na zmožnost samoovrednotenja, zmožnost spoznanja razlogov in motivov, po katerih ravnamo ali naj bi ravnali. Prvi pomemben udarec predpostavki racionalnega akterja je zadal Freud s teorijo nezavednega. Socialna in kognitivna psihologija ter nevroznanost pa so, po nekaterih precej prepričljivih

interpretacijah obsežnega raziskovalnega materiala, pokazale na veliko stopnjo avtomatiziranosti in celo iracionalnosti delovanja psihičnega aparata.²⁷

V prvem delu obravnave smo videli, da se je zmožnost samoovrednotenja izkazala za problematično, kar lahko samo po sebi predstavlja problem za možnost reflektiranega samonadzora. Vendar bi tudi ob predpostavki popolne zmožnosti samoovrednotenja srečali probleme pri samem samonadzoru, kot je recimo primer odvisnika, ki ni zmožen delovati v skladu s sicer pravilnim samoovrednotenjem. V vsakdanjem življenju pa poznamo takšne primere pod imenom šibka volja, pa tudi primere, kjer zaradi neke nezavedne želje prilagodimo naš vrednostni sistem oziroma kognitivno disonanco blažimo s pomočjo samoprevare.

Vsekakor sta ta dva vidika zelo pomembna, in to ne samo za kompatibilistično svobodno voljo. Ne glede na obstoj neke višje, metafizične vrste svobodne volje, nam je namreč pomembno, da so naša spoznanja samih sebe in sveta pravilna in da je naše ravnanje skladno s temi spoznanji.

Naturalistic Aspects of Free Will and Moral Responsibility

Free will and moral responsibility can be assessed on two levels. On the ultimate level of evaluation they are difficult to reconcile with contemporary science and might even be incoherent. On the compatibilist level they can be defined as evolved capacities of humans as biological beings. The determinism of the basic level does not contradict such concepts of free will and moral responsibility. It even enhances the evolution of avoiders on the design level. Moral behavior can be established as the evolutionary stable strategy with the application of the prisoner's dilemma to the evolutionary theory, even without the strong metaphysical conceptions of free will and ultimate moral responsibility which compatibilism cannot provide. Compatibilist freedom requires the ability to understand and reflectively evaluate the reasons one wants to act upon, and to control one's behavior in accordance with such reflectively considered reasons. Contemporary psychological and neuroscientific research indicates that rational agents face problems with both capacities. It also shows that humans are highly susceptible to cognitive dissonance, self delusion and the like.

Keywords: free will, moral responsibility, compatibilism, evolution, rational agent

²⁷ Filozofija in naravoslovne znanosti že dolgo obravnavajo introspekcijo kot zelo sumljiv vir informacij o psihičnih procesih. Kljub umiku behavioristične psihologije pa je videti, da zgoraj navedene znanosti poskušajo popolnoma zaobiti zavest kot predpostavko, ki je ne potrebujejo, če parafraziramo slavni Laplakov komentar. Zavest tako ne samo da ni gospodar v hiši, pač pa je – če do konca potegnemo sklep iz visoke stopnje avtomatiziranosti in iracionalnosti – človek celo avtomat, pa še to bolj neumen.

Literatura

- Baumeister, R. F. (2008). »Free Will, Consciousness, and Cultural Animals«. V Baer, J. (ur.), *Are We Free? Psychology and Free Will*, Oxford: Oxford University Press, str. 65–85.
- Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Mass.: MIT Press.
- Dennett, D. C. (2003). *Freedom Evolves*. New York: Viking/Penguin.
- Frankfurt, H. G. (2007). »Svoboda volje in pojem osebe«. V Šuster, D. (ur.), *O svobodni volji: Od Leibniza do Libeta*, Maribor: Založba Aristej, str. 117–131.
- Kane, R. (2005). *A Contemporary Introduction to Free Will*. Oxford: Oxford University Press.
- Kihlstrom, J. F. (2008). »The Automaticity Juggernaut – or, Are We Automats After All«. V Baer, J. (ur.), *Are We Free? Psychology and Free Will*, Oxford: Oxford University Press, str. 155–180.
- Locke, J. (1690/1959). *An Essay Concerning Human Understanding* (A.C. Fraser Edition). New York: Dover.
- Mozes, E. (2010). »The Dogmatic Determinism of Daniel Dennett«. Dostopno na: <http://atlassociety.org/objectivism/atlas-university/deeper-dive-blog/3712-the-dogmatic-determinism-of-daniel-dennett>.
- Myers, D. G. (2008). »Determined and Free.« V Baer, J. (ur.), *Are We Free? Psychology and Free Will*, Oxford: Oxford University Press, str. 32–43.
- Smilansky, S. (2000). *Free Will and Illusion*. Oxford: Clarendon Press.
- Wegner, D. M. (2002). *The Illusion of Conscious Will*. Cambridge, Mass. in London: MIT Press.