

Društvo za analitično filozofijo in filozofijo znanosti

časopis za kritično misel

ANALIZA

01

2016 letnik 20.

AN_αliZA, časopis za kritično misel, številka 1, leto XX., 2016

Izdajatelj in založnik: Društvo za analitično filozofijo in filozofijo znanosti,
Aškerčeva 2, 1000 Ljubljana
<http://www.drustvo-daf.si/analiza/>
e-naslov: analiza@drustvo-daf.si

Naročila sprejemamo na zgornji naslov.

Na leto izidejo trije zvezki Analize (dve enojni in ena dvojna številka).
Cena zvezka v prosti prodaji je 6,00 EUR, za naročnike 4,00 EUR oz. letna
naročnina 12,00 EUR (z vključenim DDV).

Izdajateljski svet:

Bojan Borstner, David Owens, Matjaž Potrč, Marko Uršič, Matthias Varga von
Kibéd, Elizabeth Valentine, Bojan Žalec

Glavna in odgovorna urednica: Maja Malec

Uredniški odbor:

Smiljana Gartner, Božidar Kante, Vojko Strahovnik, Toma Strle, Borut Trpin,
Sebastjan Vörös

Lektor: Borut Cerkovnik

Oblikovanje: K.reg@

Tisk: Demat, d.o.o., Stegne 3, 1000 Ljubljana

Naklada: 200 izvodov

Izid te revije je sofinancirala

JAVNA AGENCIJA ZA KNJIGO REPUBLIKE SLOVENIJE.

Revija Analiza je indeksirana v THE PHILOSOPHER'S INDEX.

ISSN 1408-2969

Kazalo

Epistemologija nestrinjanja

Catherine Z. Elgin Bojim se, da nimate prav	5
---	---

Danilo Šuster Odločnost in spravljivost	11
---	----

Prepričanje in verjetnost

Adam Elga Prepričanja o lastnem položaju in Trnuljičin problem	29
--	----

David Lewis Trnuljičica: Odgovor Elgu	35
---	----

Borut Trpin O nenatančnih stopnjah prepričanj: primer meduzarja, Ellsbergov paradoks in Trnuljičin problem	41
---	----

Sodobno raziskovanje odločanja

Toma Strle Odločanje: od laboratorija do resničnosti vsakdanjega življenja	61
--	----

Možgani in računalniki

Janez Bregant Možgani v primežu računalnikov	87
--	----

Recenzija

Blaž Cotman W. James: Raznolikost religioznega izkustva: Študija človeške narave	117
--	-----

Catherine Z. Elgin

Prevedel Danilo Šuster

Bojim se, da nimate prav*

Nestrinjanje je vsepovsod. Ljudje imamo različna mnenja o vsem, od športa in politike do znanosti in vzgoje otrok. Kadar so izvor nesoglasja očitna nevednost, predsodki ali neumnost enega od razpravljavcev, potem so epistemološko nenevarna. Dejstvo, da se nekdo, ki očitno ne ve, o čem govori, z vami ne strinja, vam ne daje posebnega razloga, da bi ponovno razmislili o svojih stališčih. Toda nekatera nestrinjanja so bolj zaskrbljujoča. Enako inteligentni, izobraženi, razmišljajoči in nepristranski ljudje odprtega duha se pogosto ne strinjajo. Imenujmo takšne strani v sporu intelektualne vrstnike. Ali naj se ob nestrinjanju med intelektualnimi vrstniki zamislimo?

Epistemologi se o tem vprašanju ne strinjajo. Spravljiveži (ang. *conciliatory thinkers*) kot denimo Hilary Kornblith, menijo, da bi se morali. Če Fred prepozna Jurija kot svojega intelektualnega vrstnika, nima nobene osnove, da bi menil, da je njegovo mnenje boljše od Jurijevega (ali pa, da je mnenje Jurija boljše od njegovega). Ko pride do nestrinjanja, menijo spravljiveži, bi se obe strani morali vzdržati mnenja. Zagovorniki odločnosti, kot denimo Thomas Kelly, zagovarjajo politiko »trde roke«. Če naj bi se intelektualni vrstniki, ki se ne strinjajo, vedno vzdržali sodbe, potem grozi skepticizem. Glede na širino področij, kjer se ne strinjamo s svojimi intelektualnimi vrstniki, je zelo malo tega, kar zares vemo. Odločnost je na mestu, ker vsakdo dela napake. Za Freda je odprta možnost, da meni, da ko se ne strinjata, je Jurij v zmoti. V tem primeru ima vso pravico, da zavrne Jurijevo mnenje. Na žalost lahko ima Jurij enako mnenje o Fredu. Odločnost vodi do dogmatizma, saj smo vedno upravičeni, da zavrnemo mnenje intelektualnih vrstnikov, ki se z nami ne strinjajo, če privzamemo, da se motijo. Ne skepticizem ne dogmatizem nista privlačni možnosti. Tretja možnost je, da nestrinjanje med intelektualnimi vrstniki nudi nekaj razlogov, da razmislimo o svojem stališču, vendar ne zahteva, da ga revidiramo ali zavrnemo. V tem primeru bi se obe strani lahko razumno strinjali, da se ne strinjata. Izziv je, kako narediti prostor za to stališče.

* Elgin, C. Z. (2012). »Begging to Differ«. *The Philosophers' Magazine*, 59(4), str. 77-82.

Takšna shematska zastavitev problema lahko zavaja. Različna nestrinjanja vodijo do različnih odgovorov. Če Jurij ve, da je pozorno presodil vsa dokazila, in ima neodvisna dokazila, zaradi katerih lahko sumi, da Fred tega ni storil, potem je lahko zanj razumno, da vztraja v svojem prepričanju, da, denimo, dieta, bogata z ohrovtom, vodi do ledvičnih kamnov, četudi se Fred s tem ne strinja. Ni nujno, da Jurijeva odločnost v tem primeru temelji na splošni domnevi, da so tisti, ki se z njim ne strinjajo, v zmoti. V tem posebnem primeru lahko sklepa, da je Fred prehitro sklepal. Do dogmatizma pride, če to držo posploši.

Kot kaže naslednji primer je včasih na mestu spravljivost:

Matematika: Sedem prijateljev gre na večerjo, pri čemer se strinjajo, da si bodo delili stroške. Skupen račun je 193,91€. Patricija in Miha sta enako dobra v hitrem računanju. Patricija razglasi, da naj bi vsakdo plačal 27,53€. Miha pravi, da mora vsakdo plačati 26,93€.

Zaznava: Patricija in Miha imata enko dober vid. Patricija pravi, da vidi zajca na travniku, Miha pravi, da ga ne vidi.

Izolirano preverljivo dejstvo: Patricija pravi, da se je bitka pri Waterloo-ju zgodila leta 1814, Miha pravi, da je bila bitka leta 1815.

Izolirano nepreverljivo (ali težko preverljivo) dejstvo: Patricija pravi, da je modra ekipa zmagala v dirki z vrečami leta 1993, Miha pravi, da je zmagala bela ekipa. Nobenih zapisov ni več, prav tako sta izgubila vse stike z drugimi udeleženci.

V teh sporih zlahka razsodimo, saj se v osnovi strinjamo o tem, kaj spor razreši. Patricija in Miha se strinjata, da je pravilni odgovor v primeru računa v restavraciji ena sedmina od 193,91€, kolikor to že je, in strinjata se o tem, kako to izračunati. Strinjata se o tem, kako preveriti, ali je določena perspektiva gledanja nezanesljiva. Strinjata se o tem, kako preveriti določeno zgodovinsko dejstvo, in o tem, kako razmišljati, kadar ustrezne informacije ni na voljo. Očitno je, da bi se oba morala vzdržati sodbe, dokler nista izvedla računa, se premaknila na boljše gledišče, preverila spornega dejstva s sklicevanjem na zanesljivi vir. Sodbe bi se morala trajno vzdržati, če takega vira ni.

Včasih pa ni nobene skupne osnove za strinjanje. Kadar se paleozoologi ne strinjajo glede usode stepskega mamuta ali navijači baseballa o moči igralcev notranjega polja, se zelo verjetno ne bodo strinjali niti glede tega, kako razsoditi v teh zadevah. Takšni spori tipično vzniknejo na osnovi sistematično povezanih snopov prepričanj in zavez, ne le o temi sami, ampak tudi o tem, kako o njej sploh razmišljati. Kognitivna znanstvenika Ken in Alica se ne strinjata glede tega, ali flavonoidi krepijo spomin, ker Ken zaupa robustnim izidom dolgoročnih študij, ki hipotezo podpirajo, Alica pa vztraja pri eksperimentih z neodvisno kontrolno skupino, ki pa jih še niso opravili. V sporih med intelektualnimi vrstniki se bodo prepričanja in zaveze verjetno prekrivale. Toda stranki v sporu lahko pripisujeta različno težo dejavnikom, ki jih upoštevata kot relevantne, se razlikujeta v tem,

kateri posebni dejavniki so relevantni, upoštevata različne metode kot relevantne, ali pa se razlikujeta v pragu sprejemljivosti.

Četudi se tekmujoči strani strinjata o tem, kateri dejavniki so relevantni za razrešitev spora, in o splošnem področju, znotraj katerega morata iskati sprejemljivi odgovor, se morda razlikujeta v teži, ki jo pripisujeta različnim dejavnikom. Morda se oba strinjata, da v znanosti iščemo splošnost in natančnost, ne strinjata pa se glede tega, kako ta dva dejavnika med seboj uskladiti. Oba se lahko strinjata, da morajo biti empirični izidi statistično pomembni, vendar se ne strinjata glede meje statistične pomembnosti. Morda se oba strinjata, da mora politična ureditev uravnotežiti svobodo in enakost, vendar se ne strinjata o točki ravnesaja. Naj bodo takšni spori (trenutno) nepresodni. Nobenega konsenza ni o merilu ustrezne razrešitve.

Morda je rešitev v tem, da ustvarimo stične točke. Razrešimo spore, ki so v ozadju, potem pa se premaknemo do točke, ki je predmet razprave. Kako naj to storimo? Ali obstaja objektivno dejstvo o tem, kako dobra morajo biti dokazila, da utemeljijo sklep? Ali obstaja objektivno dejstvo o tem, koliko napačnih pozitivnih ali negativnih izidov naj bi bil raziskovalec še pripravljen sprejeti? Denimo, da je test anemije natančen v 93% primerih. Za dr. Henryja je pozitiven izid dovolj dober razlog za prepričanje, da je njegov pacient anemičen. Ker pa dr. Murphy vztraja pri 95% natančnosti, se vzdrži sodbe v primerih, v katerih je dr. Henry prepričan o izidih. Oba se strinjata o tem, kakšen je izid testa in kako natančen je test. Ne strinjata pa se glede tega, ali je 93% uspešnost testa dovolj dobra, t.j. kakšen je prag sprejemljivosti. Njuno nestrinjanje glede tega, ali dokazila zadoščajo, izvira iz razlik glede sprejemljive stopnje tveganja. Če obstaja določeno dejstvo o tem, kolikšna mera tveganja pri diagnosticiranju anemije je še sprejemljiva, potem je vsaj eden od zdravnikov v zmoti. A v čem bi takšno dejstvo bilo? Ali obstaja jasna meja med tem, ko je kdo že preveč previden ali pa preveč velikodušen v svojem sprejemanju izidov testa? Če je, znotraj določenih meja, vsak iz vrste odgovorov razumen, potem lahko spor o dejstvih temelji na razlikah glede nečesa drugega, kot pa so dejstva.

Denimo, da je Bill prepričan, da je posplošljivost bolj pomembna od natančnosti, Ellen pa je prepričana, da natančnost prevlada nad posplošljivostjo. Njun spor glede učinkov klimatskih sprememb na populacijo metuljev lahko izvira iz spora glede natančnosti in posplošljivosti. Ellen izpelje natančne sklepe, ki so veljavni samo za vrsto severnoameriški monarh, Bill pa izpelje bolj splošne sklepe, ki pa so manj natančni. Ellen zato zavrača nekatere od Billovih sklepov, saj zanje niso dovolj natančni. Ali obstajajo dejstva o tem, kje je v splošnem pravo ravnesaje med natančnostjo in dosegom znanstvenih teorij? Če ne, ali obstaja vsaj dejstvo v tem posebnem primeru? Če da, potem gre morda Bill v svojem entuzijazmu za posplošljivost predaleč ali pa je Ellen v svojem nagnjenju k natančnosti preveč pikolovska. Toda pogosto so stvari, vsaj kolikor lahko rečemo, uravnotežene.

Različna razmerja med natančnostjo in splošnostjo vodijo do teorij, ki so, ko vse upoštevamo, enako zadovoljive.

Morda je rešitev enaka kot pri presodnih sporih: vzdrži se sodbe, dokler niso doseženi bolj zahtevni standardi ali dokler niso oblikovna ustrezna bolj fina merila. V tem primeru je vzdržanje sodbe lahko stalno. Kot smo videli v primeru spora glede dirke z vrečami, je včasih stalno vzdržanje sodbe razumno. Toda takšna strategija nas lahko drago stane. S tem, ko se vzdržijo sodbe, se lahko strani v sporu odrečejo sredstvom, s katerimi bi eventualno lahko razrešile svoj spor.

Biti prepričan v hipotezo ne pomeni le tega, da čutiš, da je resnična. Pomeni tudi to, da si pripravljen uporabiti hipotezo kot premiso v sklepanjih in kot osnovo za delovanje v kognitivno resnih kontekstih. Vzdržati se sodbe pomeni, da nismo pripravljeni uporabiti tega prepričanja v sklepanjih ali delovati v skladu s tem prepričanjem v teh kontekstih. Vsaka sodba, ki se je vzdržimo, nas prikrajša za neko premiso. Če je premisa neresnična ali če je njeno upravičenje bolj majavo, je lahko takšno prikrajšanje tudi prednost, saj prepreči, da bi vključili nepodprte trditve v svoj korpus prepričanj. Toda s tem izgubimo sklepalno moč. S tem, ko se je vzdržala sodbe glede svojega deleža pri računu v restavraciji, je Patricija izgubila informacijo, ki bi jo potrebovala za izračun svojega deleža napitnine. Ker je bil spor presodne narave, je bila ta izguba kratkega veka. Toda včasih ima izguba hujše posledice.

V grobu neandertalcev so našli perforiran košček kosti. Paleontologi razpravljajo o tem, ali gre za primitivno flavto. Če bi jo našli v paleolitskem grobu homo sapiensa, bi jo imeli za košček flavte. V ozadju spora je spor o stopnji nevrološkega razvoja neandertalcev. Nekateri menijo, da neandertalci niso posedovali ročne spretnosti, potrebne za izdelavo in rokovanje s flavto, ter inteligence in domišljije potrebne, da bi jo sploh iznašli. Drugi menijo, da so bili neandertalci bolj nadarjeni. Svoje nasprotnike izzivajo, da pojasnijo, kaj je ta košček kosti, če ni flavta. Malo vemo o neandertalcih. Dokazila so redka in dvoumna. Vsaka od strani v sporu navaja razmisleke, ki so sicer skladni, vendar slabo podprti z nespornimi dokazili, ki so na voljo. Ne gre za površno sklepanje. To je najbolje, kar lahko storimo s sredstvi, ki so na voljo. Če bi bilo tolmačenje koščka kosti edina točka spora, potem bi bilo vzdržanje sodbe smiselno.

Toda paleontologija je prerešetana s kontroverzami o neandertalcih. Njihova stopnja izdelave orodij, zgradba njihove skupnosti, nivo njihovega kognitivnega in umetniškega razvoja – vse to je predmet spora. Če bi se strani v sporu morale omejiti na tisto, kar je za vse strokovnjake čvrsto utemeljeno, potem bi ostalo le nekaj izoliranih dejstev in nobenega načina, da bi jih povezali. Tkanje razumevanja neandertalcev bi bilo prerešetano z luknjami in pretanko, da bi se držalo skupaj. Če pa, po drugi strani, vsaka od strani lahko vsaj provizorično sprejme hipotezo, ki sicer ni gotova, so njuni obeti boljši. Postopoma, s primerjavo in kontrastiranjem ene hipoteze z drugo, lahko spleteta tesno povezan korpus plavzibilnih

trditev, ki vse skupaj potrdijo ali ovržejo hipotezo, da je košček kosti flavta. S tem ko sprejmejo, vsaj začasno in le previdno, hipotezo, ki sicer ni trdno utemeljena, lahko raziskovalci nadaljujejo s svojim preiskovanjem in, z malo sreče, končno razrešijo spor.

Tudi če je ta strategija razumna, bi nekateri lahko trdili, da ji lahko sledimo brez zavezanosti prepričanju. Povsem mogoče je, da obravnavamo neko tezo kot resno delovno hipotezo, ne da bi bili vanjo prepričani. Zato je morda najboljša strategija ta, da smo spravljivi glede prepričanj in nepopustljivi glede delovnih hipotez. Potem lahko obe skupini paleontologov iskreno in prostodušno nadaljujeta s svojimi raziskavami, zbirata dokazila, kolikor jih lahko, jih sistematizirata tako, kot se ji-ma zdi prav, in vsaka zase gradita najmočnejši možni argument za svoje stališče. Le da enostavno ne bi smeli biti *prepričani* v tisto, kar govorita. Če zavzamemo to stališče, potem vloga prepričanja oslabi. Nič več ni povezano s kognitivno odgovornim sklepanjem in delovanjem. Znanstveniki bi bili upravičeni, da uporabijo sporne hipoteze kot podlago za sklepanje in delovanje v svojih raziskavah, da argumentirajo na osnovi teh hipotez in delujejo na njihovi osnovi. Upravičeni bi bili v tem, da jih obravnavajo v svoji znanstveni praksi, kot bi obravnavali hipoteze, v katere so prepričani. Če takšne privzetke (ang. *entitlements*) ohranijo v okoliščinah, v katerih bi epistemologi rekli, da njihovo prepričanje ni upravičeno, bi verjetno rekli: »Toliko slabše za prepričanje.« Takšno vzdržanje sodbe ne bi imelo vpliva na kognitivno odgovorno prakso.

Razlikuje so od vzdržanja sodbe, do katerega pride v presodnih primerih. Tu se strani, ki se vzdržita sodbe, nimata in se ne bi smeli šteti kot upravičeni za to, da uporabita sporne trditve kot osnovo za kognitivno resno sklepanje ali kognitivno odgovorno delovanje. V primeru računa v restavraciji Miha ne misli, da je upravičen dati 26,93\$ (in napitnino) na mizo in potem oditi. Dokler vsega ne preračunata in razrešita spora, ne more reči, koliko dolguje, in ne more plačati tega, kar dolguje.

V nepresodnih primerih se intelektualni vrstniki ne strinjajo, ne le glede dejstev, ampak tudi glede tega, kako naj bi spor razrešili. Četudi se v glavnem strinjajo o tem, katere vrste dejavnikov so relevantne, se mnenja lahko krešejo glede periferije – za eno stran je nek dejavnik pomemben za spor, druga pa misli, da ni. Mor-da se ne strinjata glede tega, kako bi morali pretehtati relevantne faktorje in/ali kakšen je prag sprejemljivosti. Znotraj grobo določenih meja se zdi, da je vrsta stališč enako sprejemljivih. V teh okoliščinah je boljša strategija, kakor da vztrajamo pri tem, da so tisti, ki se z nami ne strinjajo, v zmoti ali pa da se skupaj s kolegi zapišemo večni nevednosti, ta, da se spoštljivo strinjamo v tem, da se ne strinjamo. Strinjanje v tem, da se ne strinjamo, dopušča vsaki strani, da izvaja kognitivno resna sklepanja in se angažira v kognitivno resnem delovanju na osnovi hipotez, ki jim daje prednost. Spoštljivost je v tem, da priznamo, da na osnovi dokazil, ki so na voljo, stališča intelektualnih vrstnikov niso nerazumna. Ne gre za

stališče nespametnega dopuščanja kakršnegakoli mnenja. Na osnovi dostopnih dokazil paleontologi lahko uvidijo zasluge obeh strani v razpravi o tem, ali je košček kosti primitivna flavta. Toda vsi skupaj bodo odločno zavrnili predlog, da gre, denimo, za popolnoma funkcionalen del pozavne na poteg. Biti odgovoren v takšni situaciji pomeni biti usklajen s stanjem razprave na področju raziskav in priznavati, da novi uvidi lahko premaknejo tehtnico v naš prid ali pa proti našemu mnenju.

Osredotočila sem se na znanstvene spore, toda moje ugotovitve lahko razširimo na spore na vseh področjih. Dobro obveščeni športni navijači se lahko ne strinjajo glede tega, ali so Jekleni boljša nogometna ekipa od Patriotov, ker se ne strinjajo o relativnem pomenu dobrih tekaških ali dobrih podajalnih sposobnosti. Oboje je zaželeno, toda malo verjetnosti je, da bi ena sama ekipa bila optimalna v obeh prvinah. Prav verjetno je presoja o tem, katera ekipa je boljša, odvisna od tega, kako se odločimo glede tega vprašanja. Tudi tu se zdi spoštljivo strinjanje o tem, da se ne strinjamo, razumno.

Prej sem omenila, da se epistemologi ne strinjajo o tem, ali naj bomo odločni ali spravljivi. Kaj reči o tem sporu? (Privzemimo, da sta strani v tem sporu intelektualna vrstnika.) Zagovorniki odločnosti menijo, da bi morali biti spravljiveži *odločno* spravljivi, vztrajati bi morali pri vzdržanju sodbe glede tega, katero stališče je pravilno. Toda glede na to, da se odločneži z njimi ne strinjajo, bi se morali spravljiveži *vzdržati* sodbe o tem, ali naj se vzdržimo sodbe o tem sporu. Prepričani so, da pri tem vzdržanju ne morejo vztrajati. Že sam obstoj spravljivežev ustvari naslednjo zadrego: samo tisti, ki tega stališča *ne* zagovarjajo, lahko, glede na merila spravljivežev, trdijo, da lahko sploh kdorkoli to stališče zagovarja. Če je za epistemologe razumno, da se strinjajo o tem, da se ne strinjajo, se lahko izognejo tej dilemi. Tisti, ki dajejo prednost odločnosti, z vso pravico vztrajajo pri zagovoru odločnosti v sporu, za spravljiveže pa je spor razlog za vzdržanje sodbe. Morda bo nekoč dosežena razrešitev spora. Morda pa ne. Ampak zdi se, da je najboljši način za razumevanje narave in epistemološkega pomena spora v tem, da sprejmemo delitev kognitivnega dela, v kateri člani epistemološke skupnosti z največjo resnostjo in zavzetostjo branijo stališča, ki se jim zdijo najbolj plavzibilna, in skušajo ugotoviti, ali so navsezadnje vendarle takšna, da jih je mogoče braniti. Konsenz je morda precenjen.

Danilo Šuster

Filozofska fakulteta Univerze v Mariboru

Odločnost in spravljivost

Nestrinjanje med intelektualnimi vrstniki (ang. *peer disagreement*) je ena od vročih tem v sodobni epistemologiji. Po eni strani gre za nadaljevanje starodavnega problema skepticizma (že Sekst Empirik navaja Agripin modus iz spora kot enega od razlogov za skepticizem), po drugi strani za preseganje klasične individualistične spoznavne teorije v moderni socialni epistemologiji, ki jo zanima tudi vloga drugih v pridobivanju in prenosu spoznanja in njihovem vplivu na naša lastna prepričanja. V članku predstavim osnovna stališča (*spravljivost, odločnost*) in glavne teze v razpravi (*edinstvenost, neodvisnost*). Na koncu obravnavam kot primer spora v filozofiji nestrinjanje glede prepričanja »vem, da je skeptična hipoteza neresnična« in zagovarjam spravljivost.

Ključne besede: spori med intelektualnimi vrstniki, upravičenje, neodvisnost, edinstvenost, skepticizem

1.

Ali je res najboljši način hujšanja ta, da zmanjšaš vnos ogljiko-hidratov in povečaš vnos proteinov? Ali so klimatske spremembe zanemarljive glede na celotno geološko zgodovino Zemlje? Ali naj v času ekonomske krize varčujemo ali spodbujamo potrošnjo? Ali naj država vstopi v NATO? Ali naj ga zapusti? Ali naj poteka meja med državama po sredini zaliva? Ali Bog obstaja? Ali je dopustno ubijanje živali za naše prehranjevanje? Včasih na osnovi enakih dokazil intelektualno enakovredni osebi izpeljeta med seboj nasprotujoča si sklepa. Spori in nestrinjanja so del našega intelektualnega vsakdana, razreševali pa naj bi jih v skladu z normami razumne argumentacije. Ena od njih je gotovo ta, da resno upoštevamo mnenje nasprotne strani. Že Mill je opozoril, da tisti, ki pozna samo *svojo* stran v argumentu, ne pozna dobro niti te. Ampak tudi, ko izključimo nevednost, predsodke ali enostavno neumnost, se včasih zgodi, da se izobraženi, razmišljujoči in nepristranski ljudje odprtega duha o nekem problemu ne strinjajo, čeprav dobro poznajo *obe* strani vprašanja.

Ali sploh obstajajo *razumni* spori med »kolegi«, v katerih sta obe strani enako razumni? In če že, ali naj vztrajamo pri svojem za vsako ceno? Ali ni to *dogmatizem*? Ali naj priznamo moč argumentov druge strani in se ji uklonimo? Ali ni to *konformizem*? Ali naj priznamo, da spora enostavno ni mogoče razrešiti? Bi se za-

to morali odreči velikemu delu svojih prepričanj? Glede na število takšnih sporov se brez globokih spekulacij in bolj ali manj fantastičnih scenarijev (sanje, zli demon, možgani v kadi, Matrica ...), presenetljivo hitro najdemo v *skepticizmu* precej radikalne vrste. Prav res, nestrinjanje med *intelektualnimi vrstniki* (angl. peer disagreements) je morda najboljša ponazoritev, da je filozofski problem *skepticizma* še kako vroč in ne le hladno teoretsko vprašanje o naravi naših spoznavno-teoretskih pojmov.

Največkrat enostavno izberemo eno stran in v skladu s to odločitvijo delujemo, čeprav spor ni intelektualno zadovoljivo razrešen. Ampak želeli bi, da naša odločitev temelji na upravičenem prepričanju. Dejstvo, da se nekdo, ki nam je intelektualno enakovreden, z nami ne strinja, spodkopava razloge, ki jih imamo za svoje prepričanje. Ali pa morda ne? Filozofsko preučevanje sporov med intelektualnimi vrstniki sodi v široko razumljeno spoznavno teorijo, saj nas zanima narava upravičenja, prepričanj in dokazil, pa tudi pravila dialektike (argumentacije) in racionalnosti nasploh. Epistemologija sporov (nestrinjanja) je v trendu modernega preseganja klasične, individualistične spoznavne teorije v *socialni* epistemologiji, ki jo zanima tudi vloga drugih v pridobivanju in prenosu spoznanja in njihovem vplivu na naša lastna prepričanja, sem sodi tudi epistemologija zaupanja (ang. *trust*) in pričevanja (ang. *testimony*). V vseh primerih gre za razmislek o tem, kako na naša lastna prepričanja vpliva poznavanje prepričanj *drugih*.

2.

Intelektualna *vrstnika*, ki razpravljata o nekem problemu, sta enakovredna glede kognitivnih vrlin in veščin ter popolnoma seznanjena z vsemi dokazili in argumenti za eno ali drugo stran. Oba poznata stališče nasprotne strani, vesta, da sta skrbno in dolgo motrila vsa dostopna dokazila, da sta odprtega duha, temeljita, inteligentna... . Bolj natančno, osebi S1 in S2 sta epistemska vrstnika (ang. *epistemic peers*) glede *p*, če in samo če,

- (i) S1 in S2 sta dobro informirana in posedujeta enaka dokazila, argumente ipd. relevantne za *p*;
- (ii) S1 in S2 sta, v splošnem, enako kognitivno sposobna, enako dobra v presoji dokazil glede *p* (enako inteligentna, razmišljujoča, nepristranska ...).

Oznako bi lahko še dopolnjevali: oba zaupata *pričevanju* nasprotne strani. Zaupata temu, da je nasprotnik zares prepričan v tisto, kar trdi, na osnovi tega, kar so zanj zanesljiva dokazila. Prav tako upoštevata pričevanja tistih, ki jih imata za strokovnjake na področju spora (Machuca, 2012: 2). Pa tudi oslabili, saj je v realnih sporih težko izpolniti pogoja (i) in (ii) – očitno gre za idealizacijo, običajno zadošča, da sta osebi S1 in S2 *približno* enakovredni glede vseh dokazil in pose-

dujeta *podobne* kognitivne vrline in veščine. Temeljito sta premislili vse razloge za in proti in se zavedata, da ne ena ne druga ni v epistemski prednosti glede vprašanja, ali je *p* resnična propozicija. Potem pa se izkaže, da je S1 prepričana, da je *p* resnična, S2 pa je prepričana, da je *p* neresnična propozicija. Epistemska vrstnika sta v *sporu* glede *p*. Kakšne so epistemske posledice tega nestrinjanja?

Denimo, da s prijateljem spremljata konjske dirke, eden je prepričan, da je zmagal konj A, drugi pa, da je zmagal konj B (prim. Elga, 2007: 486–487). Vaše dokazilo *prvega* reda za zmago konja A je zaznava, morda elementarna sklepanja (jahač ponosno dviguje pokal ipd.). Vaša dokazila *drugega* reda so dejstva o tem, kako je vrstnik obravnaval dokazila prvega reda. Nasprotno mnenje bi moralo vplivati na trdnost vašega prepričanja. Ali pa morda ne? Ali šteje nestrinjanje kot *novo* dokazilo višjega reda, ki ga moram vključiti med vse druge razloge, ki sem jih upošteval v oblikovanju svojega prepričanja?

V sporih z intelektualnim vrstnikom lahko zavzamemo dve skrajni stališči, lahko smo *odločneži* ali *spravljiveži*. Odločnež (ang. *steadfast*), tudi ne-konformist in zagovornik stališča posebne teže (ang. *extra weight view*), vztraja pri svojem nasproti drugim, ki o predmetu razprave mislijo drugače. Spravljivež (ang. *conciliationist*) ali konformist pa meni, da soočenje z intelektualnim vrstnikom, ki ima nasprotno mnenje, zahteva, da *revidiramo* svoja prepričanja. Zaradi nestrinjanja bi morali biti manj samozavestni glede svojih prepričanj. Koliko manj? Po stališču *enake* teže (ang. *equal weight view*) naj bi veljalo:

Racionalno se zahteva, da v sporu med intelektualnimi vrstniki dodelimo *enako* težo mnenjem vseh strani, kadar ni nobenega razloga, ki bi dajal prednost enemu stališču pred drugimi in bi bil *neodvisen* od predmeta spora. (Machuca, 2012: 3)

O *Neodvisnosti* več kasneje, najprej o tem, kaj sploh pomeni »enaka teža«. Po grobi delitvi so možni samo trije različni odnosi, ki jih lahko imamo glede vprašanja, ali *p*? – po tem, ko smo o tem razmislili, in zadevo, kolikor je v naši moči, preiskali. Prvi je *prepričanje* da (verjetje, ang. *belief*), drugi je *nevérnost* (ang. *disbelief*) in tretji *vzdržanje* sodbe (ang. *suspension*). Prepričanje, da *p*, je zavesten afirmativni kognitivni odnos, v katerem propozicijo *p* sprejemamo kot resnično. Nevéra pomeni, da smo nesposobni, da bi propozicijo *p* sprejeli kot resnično ali da zavračamo sprejetje propozicije kot resnične. *Vzdržanje* sodbe pomeni, da smo v nedoločenem stanju, v katerem propozicije niti ne sprejemamo niti ne zavračamo. Prepričan sem, denimo, da je (bil) Prešeren slovenski pesnik, ne verjamem pa, da obstajajo duhovi. Sodbe o tem pa se ne bi vzdržal (prepričan sem, da jih ni), vzdržim pa se, denimo, sodbe, da bo 26. 4. 2216 v Ljubljani sončen dan (morda bo, morda ne).

Radikalni spravljivež meni, da je vzdržanje sodbe edina racionalna drža, ko izvedemo, da se intelektualni vrstnik z nami ne strinja. To je radikalno tolmačenje, a ni edino. V okviru teorij, ki stopnjo zaupanja v dano prepričanje kvantificirajo, je

mogoče bolj fino določiti popravljene stopnje. Ena stran ima tako visoko stopnjo zaupanja v *p*, druga manjšo, epistemski spor je potem v nestrinjanju glede stopnje zaupanja. Denimo, da sem prepričan z verjetnostjo 0,7 da je zmagal konj A, prijatelj pa je v zmago tega konja prepričan samo z verjetnostjo 0,1. Zaradi nestrinjanja spremeniva svojo stopnjo prepričanja tako, da si razliko »razdeliva«. Popravljen stopnja prepričanja za oba je zdaj 0,4 (jaz za 0,3 zmanjšam stopnjo prepričanja, prijatelj jo za prav toliko zviša). Možna so tudi drugačna pravila zmanjševanja stopnje zaupanja ali sprejemanja danega prepričanja, velja pa, v splošnem:

Racionalno se zahteva, da v sporu med intelektualnimi vrstniki *zmanjšamo* stopnjo svojega zaupanja v sporno prepričanje, kadar ni nobenega razloga, ki bi dajal prednost enemu stališču pred drugimi in bi bil *neodvisen* od predmeta spora.

Elga (2007) navaja zgornji primer konjskih dirk kot motiv za spravljivost. Podoben je Feldmanov (2007: 207–208) primer nestrinjanja glede zaznavnih prepričanj: dve osebi (obe poštene, dobro vidita ipd.) stojita pri oknu in pogledata na dvorišče, pa ena vidi sredi dvorišča dekana, druga pa ne vidi nikogar. Vsaka od njiju ima na začetku upravičeno zaznavno prepričanje. Ko pa ugotovita, da se ne strinjata, bi se morala obe vzdržati sodbe. Obe vesta, da se dogaja nekaj čudnega, a ne vesta, kaj je narobe. Najbolj znan pa je Christensenov (2007: 193) scenarij *Matematika*: pet prijateljev gre na večerjo, pri čemer se strinjajo, da si bodo delili stroške in dali 20 odstotkov napitnine. Patricija in Miha sta enako dobra v hitrem računanju. Patricija razglasi, da naj bi vsakdo plačal 45€. Miha pravi, da mora vsakdo plačati 43€. Patricija in Miha sta epistemska vrstnika, ki sta v sporu (ne strinjata se glede pravega izračuna; vesta, da se ne strinjata, in vesta, da sta *vrstnika*). Ali je lahko Miha še naprej racionalno prepričan, da izračun znaša 43€? Christensen meni, da bi oba morala imeti enako stopnjo zaupanja v *oba* izračuna.

Dekan na dvorišču in *Matematika v restavraciji* sta relativno enostavna scenarija, v katerih je na mestu epistemska *skromnost*. Če osebi, ki sta enako dobri v hitrem računanju, prideta do različnega zneska, bi se morala *vzdržati* sodbe in spor (različen izračun) razrešiti na enega od zanesljivih načinov, ki so v tem primeru seveda na voljo (kalkulator, bližnji pregled dvorišča ipd.). A večina nestrinjanj, omenjenih na začetku, ni tako enostavnih, celo v »empirični« znanosti ne (Elginova, 2012 navaja spor v paleontologiji glede »kulture« neandertalcev: dve skupini raziskovalcev na osnovi enakih dokazil prideta do med seboj nasprotujočih si sklepov).

Pa je nestrinjanje z vrstnikom *vedno* razlog, da prevprašam svoja prepričanja, in »pobijalec«, ki ga moram po mnenju spravljuježa vključiti med dokazila, ki jih upoštevam? Morda lahko odgovorimo pritrdilno, ne da bi zato končali v vsesplošnem vzdržanju sodbe. Epistemologi navsezadnje opisujejo skrajno idealizirane situacije. Ali lahko sploh kadarkoli govorimo o *vseh* dokazilih, ki so enaka za epi-

stemska vrstnika? Ali sploh obstajajo intelektualni vrstniki? Ali ni ena stran vedno v epistemološki prednosti? Toda za epistemološko pomembnost sporov med vrstniki zadošča obstoj dejanskih sporov, v katerih morda ne gre zares za vrstnika, vendar ni jasno, kdo je v boljšem epistemskem položaju (prim. Matheson, 2015).

3.

Ker je dokazilo o nestrinjanju glede p enakovredno drugim dokazilom, ki jih imam za ali proti p , spravljalivost teži k »pomiritvi nasprotij« in vzdržanju sodbe. Glede na število področij, kjer se ne strinjamo s svojimi intelektualnimi vrstniki, ostane zelo malo, kar zares vemo. To je nesprejemljivo, zato *odločnež* meni, da je kljub dokazilom drugega reda (nestrinjanje vrstnika) razumno ohraniti svoje prepričanje. Ali to velja tudi za epistemskega vrstnika, ki zastopa nasprotno stališče? Sva potem oba lahko upravičena, vsak v svojem prepričanju? Ena od predpostavk celotne razprave, ki je za odločneža lahko sporna, je teza o *Edinstvenosti* (Machuca, 2012: 4):

Celota dokazil, ki so na voljo glede propozicije p , upravičuje zgolj *en* epistemski odnos do p (ali zgolj eno stopnjo zaupanja v p) – prepričanje, névernost ali vzdržanje.

Po stališču enake teže se v primeru nestrinjanja nekdo moti (prim. *Matematika*), a ker vrstnika ne moreta ugotoviti, kdo ima prav, se morata oba vzdržati sodbe. Saj ne more biti razumno, da bi na osnovi enakih dokazil intelektualna vrstnika prišla do različnih sklepov!?

Zdi se mi, da podobno kot Wittgenstein dokazuje, da ni privatnega, samo meni dostopnega jezika, teza o edinstvenosti za spravljaliveža predpostavlja, da ni skritih, privatnih dokazil in privatnih izvorov spoznanja, a prav tako ni nekakšne »privatne racionalnosti« (vsak je razumen, na svoj način). Ne morem imeti nekih posebnih, samo meni znanih dokazil, zaradi katerih sem »pametnejši« od nasprotnika (ali pa sva oba pametna, vsak na svoj način), in ne morem imeti posebnega, samo meni lastnega dostopa do dokazil, zaradi katerega sem bolj upravičen v svojem prepričanju od vrstnika (ali pa sva oba upravičena, spet vsak na svoj način). Ne, vse je javno in transparentno, o vsem je mogoče razpravljati in o vsem presojati, toda na koncu je lahko razumno samo *eno* stališče. Če se ne strinjava, potem moram svoje prepričanje revidirati (vzdržati se sodbe), vrstnik, ki se z mano ne strinja, pa prav tako.

Če sprejmemo edinstvenost, potem je eden v zmoti, a kako vem, da nisem jaz? Pomislimo, kaj pa bi pomenilo zanikanje teze o edinstvenosti. V karikirani obliki ga izraža naslednja šala o rabiju:

Benjamin se spre z Baruchom. Rabi meni, da noben od njiju nima prav. Potem mu Benjamin razloži svoje stališče in rabi reče: »Ti imaš prav.« Čez

nekaj časa mu Baruch razloži svoje stališče in rabi mu reče: »Ti imaš prav.« Njegova žena, ki je vse to slišala, se začudi: »Ampak oba vendar ne moreta imeti prav. To je nemogoče.« Potem ji rabi reče: »Tudi ti imaš prav.«

Privzemimo, kar je morda sporno, da se celota dokazil, ki so našemu modremu rabiju na voljo, ne spreminja, kljub temu da jih predočajo različne osebe. V terminologiji epistemologije nestrinjanja rabi manifestira ekstremno *permisivnost*:

V nekaterih primerih je racionalno biti prepričan, da p , vendar je s tem povsem združljivo, da si pri enakih dokazilih popolnoma razumen, če si namesto tega prepričan, da $\neg p$.

Bolj natančno lahko *permisivnost* predstavimo kot zanikanje edinstvenosti (prim. Machuca, 2012: 4):

Celota dokazil, ki so na voljo glede proposicije p , je združljiva z različnimi epistemskimi odnosi do p (ali različnimi stopnjami zaupanja v p).

Edinstvenost je tesno povezana s spravljalivostjo. Če enaka dokazila zares dopuščajo več različnih epistemskih odnosov do danega prepričanja, potem nobena stran nima razloga, da bi svoje prepričanje spremenila – razumno je vztrajati pri svojem, zakaj potem sploh spravljalivost? Ta je smiselna le, če sprejemamo, da je razumno eno samo stališče, nestrinjanje z vrstnikom pa je zdaj dodatno dokazilo, ki govori *proti* našemu izhodiščnemu prepričanju.

V najbolj skrajni obliki *permisivnost* dopušča, da je za poljubno prepričanje in ustrezna dokazila za nekoga racionalno, da sporno prepričanje sprejema glede na ta dokazila, za drugega pa, da ga zavrača. A to *ni* kontekstualizem (oziroma relativizem): vsakdo ima prav (psst! v svojem kontekstu ali relativno glede na »svojo paradigmo«). Po začetni predpostavki gre za skupen kontekst in obe strani v sporu sta *enako* razumni. Nekateri celo menijo, da samo *permisivnost* sploh dopušča obstoj *razumnih* sporov! Ker pa je *permisivnost* zanje epistemološko nevzdržna, *razumnih* sporov med vrstniki ne more biti.

Že Lehrer (1976: 331) je trdil, da so spori med epistemskimi vrstniki racionalno nemogoči. V primerih nestrinjanj med vrstniki gre za nepopolno izmenjavo informacij, individualni dogmatizem ali pa nekomu ni uspelo, da bi dojel matematične implikacije začetnega stanja. Ko bi zares šlo za razumno nestrinjanje, potem obe strani v sporu lahko razumno vztrajata pri svojem stališču. Kako je to mogoče, če pa celota dokazil, ki so na voljo glede proposicije p , po tezi *edinstvenosti* racionalno upravičuje zgolj *en* epistemski odnos do p ? V zares razumnem sporu bi imela lahko oba »prav« – ne v smislu, da sta resnični med seboj nasprotujoči si proposiciji, ampak v smislu, da imata oba dobre razloge (eden se sicer moti, a žal smo lahko razumni in imamo dobre razloge za neresnične proposicije). Za Feldmana (2007: 210) *edinstvenost* dopušča samo dve razrešitvi *navidez* razumnih sporov: (i) stališče trde roke – drugi se enostavno moti in v resnici nima dobrih

razlogov za svoje stališče; (ii) spravljalivost – glede na to, da je možno eno samo stališče, razumen spor ni mogoč, zato ne smeja vztrajati pri dveh med seboj izključujočih si stališčih. Razumno je, da spremenim svoje prepričanje, enako seveda moj vrstnik, zato oba končava v vzdržanju sodbe.

Zavrnitev *edinstvenosti* (ekstremna permissivnost) seveda dopušča razumne spore med vrstniki – glede na dana dokazila sva racionalna lahko oba, jaz, ki zagovarjam p , in ti, ki v p nisi prepričan. Odločnež meni, da je v epistemskem sporu z vrstnikom racionalno vztrajati pri svojem stališču – vrstnika popolnoma ignoriramo ali pa ga sicer upoštevamo, a to ne spremeni pomembno naše stopnje zaupanja v sporno prepričanje. Nekateri odločneži zanikajo edinstvenost (včasih gre za razumne spore, v katerih je racionalno, da vsaka stran vztraja pri svojem), drugi so na strani *trade roke* – samo eden ima prav (torej *edinstvenost*), ampak nasprotnik se pač moti, dejstvo, da se z mano ne strinja, ni dodatno dokazilo, ki bi ga moral upoštevati v svoji presoji. Nekaj takega zagovarja van Inwagen (1996), ki meni, da ima vsak od nas nekakšna privatna dokazila, svoj posebni, drugim nesporočljivi *uvid*, ki ga vrstnik v sporu ni deležen, zato lahko vztrajamo pri svojem.

Zgodbica o rabiju sproži vprašanje notranje konsistence. Logični očitke spravljalivosti je samo-ovrgljivost: spor med vrstniki je prav tako predmet razprave med vrstniki (epistemologiji). Ker imata spravljalivež in odločnež različni mnenji, bi moral spravljalivež revidirati svoje prepričanje in se približati odločnežu. Nenavadno – spravljalivost sama zahteva, da se ji odpovemo?! Elga (2010) zagovarja spravljalivost v primerih spora med vrstniki, *razen* ko gre za vprašanje o tem, kakšno naj bo naše stališče glede epistemskih sporov. Nekateri menijo, da je ta izjema *ad hoc*, toda Elga trdi, mislim, da upravičeno, da mora vsako fundamentalno epistemsko pravilo biti dogmatsko glede svoje *lastne* pravilnosti, sicer postane nekonsistentno. Razprava o tem je logično zanimiva in razvejana (pomislimo samo na razvpiti problem indukcije kot epistemskega pravila).

4.

Še ena od pomembnih tez v razpravi je načelo *Neodvisnosti* (Machuca, 2012: 4):

V razrešitvi nestrinjanja med epistemskimi vrstniki se lahko razpravljavca sklicujeta samo na razloge, ki so neodvisni od njunih prepričanj o sporni zadevi in od sklepanj, na osnovi katerih sta oblikovala svoja prepričanja.

Za spravljaliveža je to eno osrednjih načel – nestrinjanje z vrstnikom zahteva nek *odziv*. Če ohranim svoje izvorno prepričanje, moram imeti dobre razloge za neupoštevanje drugačnega stališča. V odgovor na vrstnikovo nestrinjanje ne morem enostavno reči, *tako* je, kot jaz pravim, in ponoviti svoje izhodiščne razloge, intuicije in sklepanja: P je resnično, vrstnik to zanika, torej se moti, zato bom ohranil svoje prepričanje.

Pogoj *Neodvisnosti* je zapleten, saj kombinira dva tipa dokazil: dokazila prvega reda, ki jih imam za izvorno prepričanje, in dokazila drugega reda, ki zadevajo nestrinjanje in kredibilnost vrstnika. Christensen (2009: 758) predlaga naslednjo formulacijo načela, ki še bolj poudarja, da gre za regulativ, ki predpisuje, katere »poteze« so dopustne kot odziv na epistemsko nestrinjanje:

V presoji epistemskih »poverilnic« (ang. *credentials*) osebe, ki izraža prepričanje, da *P*, se ne smem v določitvi tega, kako (če sploh) spremeniti svoje prepričanje, da *P*, opirati na razloge in sklepanja, ki so vodila do mojega izvirnega prepričanja, da *P*.

Če ohranjam svoje prepričanje, moram pokazati, da je druga stran v nekem spoznavnem primanjkljaju, morda »omamljena,« morda ne pozna vseh dokazil, morda se večina tistih, ki se zares spoznajo, strinja z mano... Bolj splošno: *P* je resnično zaradi razlogov *R* (prvi red), ugotovil pa sem tudi, da *V*, vrstnik se s tem ne strinja (drugi red) in je prepričan, da ne-*P*. *Neodvisnost* zahteva, da se v ohranitvi prepričanja *P* ne opiram na prvotne razloge in sklepanja *R* (prim. Frances, 2014: 162). To pa lahko pomeni, da najdem nek *nov* razlog za *P* ali pa najdem dokazila proti *V* – vrstnik nima zadostnih epistemskih »poverilnic« (slabo vidi, dela napake pri računanju, ne pozna vseh dokazil ipd.). *Neodvisnost* prepoveduje naslednje »utemeljevanje«: Jaz pravim *P* (ker *R*), ti praviš ne-*P*, toda ker je resnično *P* zaradi *R*, ti seveda nimaš prav in jaz svojega prepričanja ne bom spreminjal. *Pravilno* bi bilo, recimo: ugotovila sva, da se ne strinjava, toda (i) dodatna dokazila (sklepanja) podpirajo *P*; ali pa (ii) obstajajo dobri razlogi, ki pojasnijo tvojo pomoto. Zato, vidiš, ne bom spreminjal svojega prepričanja, da *P*. Shema bo bolj razvidna, če navedem nekaj primerov.

Kelly (2013: 40) navaja naslednji pomislek glede načela *Neodvisnosti*:

Veliko dokazil imam za holokavst. Potem pa naletim na osebo, ki holokavst zanika (ker ne pozna zgodovinskih dejstev in dokazil).

Neodvisnost zahteva, da v tem sporu »izločim« izvorna dokazila (pričevanja, arhivi, slikovna gradiva, ...) za svoje prepričanje o holokavstu, a kako lahko potem sploh odgovorim zanikovalcu? Kako mu lahko dokažem, da sem v *boljšem* epistemskem položaju od njega? V tem primeru je odločilen dvom v kredibilnost zanikovalca in njegove epistemske »poverilnice.« Znano mi je, da imam več dokazil od njega (denimo splošno poznavanje zgodovinskih virov) – in ti razlogi za mojo superiornost niso neposredno odvisni od razlogov za prepričanje o holokavstu, torej *neodvisnost* s tem ni ovržena.

Naj navedem še primer, ki sem ga nekoč že uporabil v razpravi o krožnih sklepanjih (Šuster, 2010), saj je *neodvisnost* zares tesno povezana z dialektičnimi in epistemološkimi vprašanji zmote *petitio* (kaj sploh pomeni predpostavljati to, kar še le moraš dokazati). Vzemimo naslednji odlomek iz Platonovega *Timaja*, ki ga si-

cer Copi in Cohen v svojem znanem učbeniku ponujata kot vajo v prepoznavanju zmot (Copi in Cohen, 1991: 110):

Povedati (kaj) o drugih božanstvih in spoznati (njihov) nastanek presega naše moči; treba je zaupati tistim, ki so o tem govorili pred nami; ti so bili – kot so trdili – potomci bogov, in tako so gotovo dobro poznali svoje prednike. Ni pač mogoče, da ne bi zaupali otrokom bogov, četudi govorijo brez verjetnih in nujnih dokazov; slediti moramo postavi in jim verjeti, saj trdijo, da sporočajo družinsko zgodovino. (Timaj 40d-e; Platon, 2006: 1274)

Zaradi nazornosti vzemimo Ahila, ki je po grški mitologiji sin morske nimfe Thetis (boginje vode), in njegov namišljeni dialog z znanim antičnim ateistom, filozofom Ksenofanom:

Ahil: Jaz, Ahil sem sin kralja Peleja in nimfe Tetis. Torej bogovi obstajajo.

Ksenofan: Tole z bogovi je meni malo čudno. Etiopci svoje bogove predstavljajo kot črne s topim nosom, nasproti temu pa Tračani kot modrooke in rdečelase. Če bi krave in konji imeli roke, tedaj bi konji narisali bogove kot konje in krave kot krave. Bogovi so človeške izmišljotine.

Ahil: Saj vendar govorim o dogajanju v moji lastni družini!

Ksenofan: A kako vem, da ti lahko zaupam? Pričevanja niso vedno zanesljiva. Ali ni bolj verjetno, da se motiš *ti*, kot pa to, da bogovi obstajajo?

Ahil: Bom pa ja *poznal* svoje najožje prednike!? Kdo bi lahko vedel bolje od mene?

Gre za klasično nestrinjanje glede vprašanja »bogovi obstajajo«. Ko ateist Ksenofan naleti na teista Ahila, mora presoditi, ali sta epistemska vrstnika – oceniti njegove epistemske »poverilnice,« ali so dovolj dobre, da bi zaradi njih »revidiral« svoj ateizem. Pri tem se (ob predpostavki, da gre za intelektualna vrstnika, ampak dovolimo si to fantazijo) ne sme opirati na izvirne razloge, ki jih ima za svoj ateizem (antropomorfnost bogov). Ksenofan po načelu neodvisnosti pravilno izpodbija kredibilnost pričevanja. Kot je ugotovil Hume (1974: 157), »nobeno pričevanje ni zadostno, da bi ugotovili čudež, razen tistega, katerega neresničnost bi bila bolj čudežna od dejstva, ki ga skušamo ugotoviti.« Ahil bi moral navesti kake dodatne razloge, neodvisne od poznavanja svoje družine (recimo, da z neranljivostjo dokáže svojo božanskost).

Spravljivost zahteva, da se *oba* vzdržita sodb. No, tudi Ksenofan ni bil agnostik, ampak ateist, ki ne bi pristal na spravljivost. Ali bi vseeno zagovarjal *Neodvisnost*? Jasno je, da je *neodvisnost* zid proti dogmatizmu in prazni krožnosti. Jasno pa je tudi, zakaj se mnogi odločneži z njo ne strinjajo: če imam od vsega začetka dobre in zame prepričljive razloge za svoje stališče, zakaj bi moral samo zaradi

nestrinjanja iskati neodvisne razloge za svoje stališče? Odločnost ravno pomeni, da vztrajam pri svojem stališču kljub nestrinjanju.

Tudi odločnež Ahil ne upošteva neodvisnosti – v sporu se sklicuje na sporno prepričanje in prvotne razloge. Vztraja, saj meni, da mu evidenca prve osebe omogoča privilegiran dostop, ki je za Ksenofana v načelu nemogoč. Prav zaradi izvornega upravičenja izpodbija epistemske »poverilnice,« ki jih ima Ksenofan. Ali ni prav perspektiva prve osebe, ki jo poudarja Ahil, odločilna? Moja izkustva, moja dokazila, moja prepričanja ..., imajo vedno večjo težo v oblikovanju stališč, ki jih zagovarjam, kot pa mnenja mojih vrstnikov. Spravljivež res pogosto navaja sklicevanje na razne *zunanje* aparate, denimo moja ura kaže 17.00, tvoja pa 17.10 – če ni razlogov, da je katera pokvarjena, bi morali enakovredno upoštevati obe možnosti (Christensen, 2009: 762). Strinjam pa se s Christensenom, da vedno na nek način razmišljamo iz svoje »perspektive«, pa to še ne pomeni, da lahko kar prezremo možnost, da se morda motimo. S tem enostavno zanikamo celoten problem sporov med vrstniki. Kot že rečeno: racionalnost ni nekaj »privatnega«.

Kaj pa, če imam zares dobre začetne razloge in visoko začetno stopnjo zaupanja v svoje stališče? Spomnimo se še enkrat scenarija *Matematika*: Patricija razglasi, da naj bi vsakdo plačal 45€. Miha pravi, da mora vsakdo plačati 43€. Odločnež Miha meni, da lahko vztraja, ker se je Patricija pač zmotila. A njegov edini razlog, da ji pripiše pomoto, je ta, da je on dobil drugačen rezultat, v katerega pa popolnoma zaupa. Ampak *okrepimo* prvotni razlog in povečajmo zaupanje v pravilni izračun: Miha je račun še enkrat preveril na papirju in s kalkulatorjem ter spet dobil 43€. Patricija je naredila enako in spet dobila 45€. Ali ni v tem primeru racionalno, da Miha vztraja in ne upošteva Patricije, pri tem pa se opira samo na (zdaj sicer okrepljene) razloge, ki jih ima za 43€?

Christensen (2009: 759) še vedno brani *Neodvisnost*, saj meni, da lahko Miha dvomi v Patricijin rezultat zaradi razlogov, ki so *drugačni* od »gotovosti« izračuna, ki je razlog za njegovo prepričanje. Recimo: »Neverjetno je, da bi dva dobra računarja po temeljitem preverjanju prišla do različnih odgovorov, zelo verjetno se dogaja nekaj čudnega (kakšna kognitivna blokada ali pa morda prevara ali šala). Zase vem, da lahko takšne možnosti izločim, torej mora biti nekaj narobe s Patricijo, bolj verjetno je, da je njen odgovor nepravilen.« Razlog za Mihov dvom v Patricijin izračun v tem primeru *ni* njegov prvotni izračun.

Pa lahko vedno najdemo razloge, s katerimi presojujemo stališče vrstnika, ki bodo neodvisni od naših izvornih razlogov v spornem vprašanju? Scenarij *Matematika* nekoliko zavaja – Miha lahko oceni, kako zanesljiva je bila Patricija do sedaj, kako dobra je bila pri matematiki v šoli ipd., in se v presoji njene verodostojnosti ne opira na svoj izračun. V velikih in globokih sporih pa so stališča med seboj povezana in organizirana v velike mreže, nobenih prepričanj ni, ki bi bila *neodvisna* od izvora nestrinjanja, spora ne moremo enostavno »lokalizirati« na eno samo vprašanje. Elga (2007: 492-93) navaja primer Ann in Beth, političnih nasprotnic, ki se

ne strinjata glede tega, ali je splav moralno dopusten. Razpravljali sta tudi o številnih povezanih vprašanjih – ali imajo ljudje dušo, kakšna je vloga pravic v etični teoriji, ali lahko prekinemo zdravljenje bolnikov v dolgotrajni in nepovratni komi itd. Glede velike večine vprašanj se ne strinjata, Ann misli, da se Beth moti, in obratno. Tudi ko Ann izloči problem splava in da v »oklepaj« svoje razloge v *tem* sporu, ostane mreža med seboj povezanih nestrinjanj, zaradi katere Ann dvomi, da bi Beth sploh lahko imela prav glede moralne dopustnosti splava. Po mnenju Elge bo izid ta, da Ann ne bo štela Beth kot svoje epistemske vrstnice (in obratno), zato zaradi njunega nestrinjanja ne bo spreminjala svojih stališč. Spravljalivost tu ni na mestu in problem nestrinjanja med epistemskimi vrstniki kar izgine?!

Vsi poznamo takšne primere velikih nestrinjanj (pomislimo samo na »leve« in »desne«, »liberalce« in »radikalce«, pa »naravoslovce« in »humaniste«, »ekonomiste« in »spiritualiste« ipd.), a nenavadno je, da jih spravljalivež Elga označuje kot »umazane vsakdanje primere« (ang. *messy real-world cases*), ki ne ogrozijo zahteve po spravljalivosti v »čistih« scenarijih (npr. Matematika). Celo več – glavni razlog za odločnost je problem vsesplošnega vzdržanja sodbe, do katerega vodi spravljalivost. Zdaj pa se izkaže, da v velikih sporih nasprotna stran sploh *ni* naša epistemska vrstnica, zato se lahko brez nevarnosti za svoja stališča z njo ne strinjamo. Tisti, ki pa *so* naši vrstniki, imajo podobna prepričanja, zato ni potrebe, da bi jih revidirali in postali agnostiki. Torej spravljalivost v kombinaciji z *Neodvisnostjo* sploh ne vodi do globalnega skepticizma?!

5.

»Umazani vsakdanji primeri« so značilnost velikih (in včasih zelo »nevsakdanjih«) razprav v filozofiji, kjer pa ravno ne bi rekli, da imajo vrstniki *podobna* prepričanja. Včasih res dobiš površni vtis, da v filozofiji v resnici nikdar nismo prepričani, da *p*, ampak samo argumentiramo za *p*, pri čemer se ceni predvsem izvirnost (*moja* teza je, da ...), pri nekaterih samo provokativnost. V tem primeru do epistemološkega problema spora med vrstniki sploh ne bi prišlo. No, bolj kot zanimivost se mi zdi pomembna resnica, pa če je ta še tako trivialna. Filozofi bi naj bili varuhi kritičnega razmišljanja, kar nedvomno vključuje pripravljenost sprejeti *zares* vsako prepričanje, do katerega nas vodi razumna presoja.

A je prav naše področje najbolj prepreženo z velikimi nestrinjanji. Morda je prav to tisto, kar ga definira: dualisti in materialisti, realisti in anti-realisti, kompatibilisti in inkompatibilisti, konsekvencialisti in deontologi, moralisti in avtonomisti, internalisti in eksternalisti, liberalci in anti-liberalci, esteticisti in moralisti, platonisti in fikcionalisti Po letih razmišljanj o nekem filozofskem problemu, poznavanju vseh (proti)argumentov, definicij, razlikovanj, primerov in proti-primerov ..., imamo še vedno različna mnenja o malodane vsakem filozofsko za-

nimivem vprašanju. Večinoma gre za odmaknjene teoretske probleme, a nekateri vendarle imajo pomembne praktične posledice, pomislimo na vprašanja splava, evtanazije, smrtnih kazni, legalizacije drog, odnosa do živali, politične ureditve ...

Christensen (2007) se po mojem mnenju bolje od Elge zaveda razlike med enostavnimi »vsakdanjimi« scenariji (*Matematika*), kjer na koncu obstaja razrešitev, ki jo vsi sprejmejo, in filozofskimi problemi. Zanesljive metode raziskav težijo k temu, da vodijo do soglasja. Dolgotrajni spori v filozofiji (pa tudi religiji, politiki ...) pa po njegovem kažejo, da na teh področjih ne oblikujemo svojih prepričanj na zanesljiv način. Če ste udeleženi v takšni razpravi, potem ne bi smeli imeti trdnih mnenj o spornih temah. Mislim, da ima prav, čeprav to za filozofa ni najbolj privlačno stališče (in v to nisem povsem prepričan!?). Kot zadnji primer bom zato navedel filozofsko razpravo med epistemološkimi vrstniki v znamenitem in starodavnem sporu v teoriji spoznanja.

Problem nestrinjanja med vrstniki je danes ena od vej razprave o skepticizmu, ki se presenetljivo vrača k antičnim virom. Sekst Empirik sistematizira razloge za skepticizem kot moduse *Agripe* in prvi modus je prav naslednja ugotovitev:

Po modusu iz spora ugotovimo, da pride do neodločljivega nesoglasja tako v vsakdanjem življenju kot med filozofi. Zato ne moremo izbrati ali izločiti ničesar in končamo pri vzdržanju sodbe (Sextus Empiricus, 2000: 40–41 (I, 165–169)).

Spomnimo se Descartesa (1988: 51), kako meditira ob kamnu in potem zapiše: »čisto jasno sprevidim, da budnosti nikdar ni mogoče po zanesljivih znamenjih razločevati od spanja, tako da osupnem in mi ta osuplost skorajda potrjuje misel, da spim.« Kartezijski skepticizem temelji na tezi, da za vsako opazovalno prepričanje obstajajo rivalska prepričanja, ki so enakovredno združljiva z dejanskimi čutnimi dokazili za to prepričanje. »Kolikokrat mi nočni počitek pričara vtis običajnih okoliščin: da sem tu, da oblečen v suknjič sedim ob ognju, ko vendar slečen ležim v postelji!« (Descartes, 1988: 50–51). *Skeptična hipoteza* (SH) je alternativna teorija o nastanku naših vtisov in na njih temelječem opazovalnem prepričanju, ki je ne moremo nikdar popolnoma izločiti (sanje, zli demon, možgani v kadi, Matrica ...). Radikalni skeptik iz tega izpelje, da za vsako opazovalno prepričanje obstajajo rivalska prepričanja, ki so enakovredno *podprta* z dejanskimi čutnimi dokazili za to prepričanje, torej nič ne vemo o zunanjem svetu (prim. Devitt, 2002: 28).

Radikalni skeptik o zunanjem svetu je danes teoretska figura, ki jo težko srečamo kot »realno« osebo, s katero bi se zapletli v polemiko. Toda zelo znani epistemologi so v sporu glede tega, ali *vemo*, da je SH neresnična. Dretske in Nozick, oba eksternalista, ki seveda nista skeptika, sta prepričana, da *tega* ne vemo. Moore, Sosa, Devitt, Leplin ..., pa so, nasprotno, prepričani, da to vemo. Oglejmo si spor bolj splošno, kot nestrinjanje med pripadnikoma dveh razredov, SH-vedežem in SH-nevedežem:

SH-vedež:

Možnost radikalne prevare je tako oddaljena in neverjetna, da je moje prepričanje o neresnici skeptične hipoteze popolnoma *varno* – zelo težko bi bilo neresnično glede na to, kakšen svet dejansko je (prim. Sosa, 1999).

Kdo je sploh ta nori znanstvenik, ki me domnevno vara v scenariju možgani-v-kadi? Zakaj nam to dela? Kako to dela? Zakaj se zanima zame? Kako to, da toliko ve o mojem predhodnem duševnem življenju? ... Pojasnjevalne praznine so bistvo neplavzibilnosti. Naznačijo nezadovoljstvo, ki ga občutimo ob goli stipulaciji možnosti, da je SH resnična. Če jih želimo koherentno zapolniti, se moramo vzdržati naravnega skepticizma glede skepticizma (Leplin, 2009: 141).

SH-nevedež:

Nekateri pravijo, da čeprav S ne more ločiti dejanskega sveta od sveta prevare, čeprav bi v svetu popolne prevare še vedno verjel, da ni v svetu prevare, S vseeno ve, da je skeptična hipoteza neresnična. Ampak to je lahko zgolj neko »prazno«¹ védenje, S ve, ne da bi bil sposoben izločiti alternativne scenarije. To so zelo nizka merila! (Heller, 1999: 207).

Mislim, da ne smemo predpostaviti, da je prevara skrajno malo verjetna. Res je skoraj neverjetna glede na to, kar se nam zdi samoumevno, da vemo o sebi in svetu, v katerem živimo. Ampak, glede na hipotezo, o kateri razpravljamo, kaj nam dopušča, da se sklicujemo na to védenje v presoji verjetnosti? ... ali je neverjetnost dovolj, da rečemo, da je hipoteza neresnična? Vem, da je verjetnost moje zmage na loteriji izredno majhna, a to ne pomeni, da vem, da bom izgubil (Dretske, 2005: 44).

SH-vedež:

Zakaj Dretske in Nozick vztrajata pri tem, da moje prepričanje, da je skeptična hipoteza resnična, ni védenje? To prepričanje je (i) resnično; (ii) ni zgolj slučajno resnično; (iii) racionalno (racionalno se sklada s celotnim sistemom prepričanj). Zakaj, natanko, to ni védenje? (Sosa, 1999: 147).

SH-nevedež:

Glede na naš pojem spoznavnega upravičenja je zaznavno prepričanje upravičeno samó, če imam neodvisne razloge za prepričanje, da je neresnična hipoteza, ki je nezdružljiva s tem prepričanjem, vendar bi imel natanko enaka čutna izkustva, ki so vodila do začetnega prepričanja, tudi če bi bila ta hipoteza resnična (prim. Schiffer, 2004: 180).

SH-vedež in SH-nevedež, vsak sicer sestav različnih filozofov, ki pa jih motivirajo podobni razmisleki, sta intelektualna vrstnika, ki se ne strinjata glede prepričanja »Vem, da je SH neresnična.«² Iz razprave je razvidno, da se razlikujeta v pre-

soji posameznih sklopov dokazil, SH-vedež poudarja plavzibilnost, verjetnost, znanstveno kredibilnost SH-nevedež pa meni, da dokazila za zaznavno prepričanje vodijo do védenja le, če *razločijo* med situacijami, v katerih je prepričanje resnično, in situacijami, v katerih ni. Po predpostavki pa dokazila, ki jih imamo na voljo za zaznavna prepričanja o zunanjem svetu, ne izločijo skeptičnih scenarijev. »Oba« filozofa poznata vse razloge, argumente, razlikovanja ... za in proti, vendar se ne strinjata glede sklepa. Morda se vprašamo, kako sta lahko sploh *enakovredna* epistemska vrstnika, ko pa dajeta pri izpeljavi sklepa glede SH *različno* težu razmisleku o nerazločljivosti in presoji o verjetnosti in koherenci? Ali ne bo problem spora enostavno izpuhtel, tako kot meni Elga glede spora med Ann in Beth?

Mislim, da ne bo, saj še vedno ostane vprašanje, kdo ima *prav* v končni presoji! Epistemolog se mora vprašati, kateri sklep *v resnici* podpirajo vsa dokazila. Odločnež trdi, da je popolnoma racionalno, da vsaka stran vztraja pri svojem. Spravljivež meni, da bi morala zaradi nestrinjanja oba postati previdnejša glede svojih stališč. Srce me vleče k odločnosti (seveda vemo!), toda razum mi prišepetava previdnost.

Elginova (2012) na koncu zagovarja nekakšen kompromis, »da smo spravljivi glede prepričanj in nepopustljivi glede delovnih hipotez«. Toda delovanje samo problema spora med vrstniki ne bo razrešilo. Jasno je, da se moramo včasih odločiti in enostavno ravnati tako ali drugače, čeprav je, *epistemsko* gledano, vzdržanje sodbe morda bolj primerno. Feldman (2007: 203) navaja primer razpotja, na katerem se znajdeva s prijateljem na poti na nujen sestanek. Na zemljevidu ga ni, odločiti pa se morava, saj bi bilo najslabše ne storiti ničesar. Jaz izberem levo, prijatelj desno, a enako razumna bi bila drugačna izbira. Epistemsko gledano bi se moral vzdržati sodbe, da je moja izbrana pot pravilna. V nasprotju z Buridanom pa vzdržanje sodbe ne pomeni, da lahko čakava, da se ustrezne okoliščine spremenijo (pojavijo nova dokazila, morda kak domačin, ki ga lahko vprašava), preden se odločiva. Kakorkoli se že to sliši paradoksalno – razumno je ravnati na določen način, čeprav ni razumno biti *prepričan*, da bo izbrani način delovanja uspešen (prim. Feldman, 2007: 203).

Pomislite na svoja razpotja, izbrali ste poti, sprejeli ste svoje odločitve, a koliko je bilo zares *razumnih* v luči dokazil? V nasprotju z ekonomijo, ki se priklanja odločnosti v ravnanju (»najboljša odločitev je hitra odločitev«), filozofija, vsaj tako mislim, narekuje spravljivost v presoji.

Steadfastness and Conciliation

Peer disagreement is one of the hot topics in contemporary epistemology. Pyrrhonians already discussed the so-called mode of disagreement as one of the sources of skepticism. Modern social epistemology tries to overcome the limitations of classical individualistic approaches and introduces the importance of others in framing of our opinions. Basic positions are introduced (*conciliation, steadfastness*) and main theses discussed (*uniqueness, independence*). As a final example of a peer disagreement in philosophy, I present the question whether we can know the skeptical hypothesis to be false. I opt for conciliation.

Keywords: disagreement, peer, justification, independence, uniqueness, skepticism

Literatura

- Christensen, D. (2007). »Epistemology of Disagreement: The Good News«. *Philosophical Review*, 116, str. 187–217.
- Christensen, D. (2009). »Disagreement as Evidence: The Epistemology of Controversy«. *Philosophy Compass*, 4(5), str. 756–767.
- Copi, I. M. in Cohen, C. (1991). *Introduction to Logic*. New York: McMillan.
- Descartes, R. (1988). *Meditacije*. Ljubljana: Slovenska Matica.
- Devitt, M. (2002). »Underdetermination and Realism«. *Philosophical Issues*, 12, str. 26–50.
- Dretske, F. (2005). »The Case against Closure«. V Steup, M. in Sosa, E. (ur.), *Contemporary Debates in Epistemology*, Oxford: Blackwell, str. 13–46.
- Elga, A. (2007). »Reflection and Disagreement«. *Nous*, 41, str. 478–502.
- Elga, A. (2010). »How to Disagree About how to Disagree«. V Feldman, R. in Warfield, T. (ur.), *Disagreement*, New York: Oxford University Press, str. 175–186.
- Elgin, C. Z. (2012). »Begging to differ«. *The Philosophers' Magazine*, 59(4), str. 77–82.
- Feldman, R. (2007). »Reasonable Religious Disagreements«. V Antony, L. (ur.), *Philosophers without Gods: Meditations on Atheism and the Secular Life*, Oxford: Oxford University Press, str. 194–214.
- Frances, B. (2014). *Disagreement*. Cambridge: Polity Press.
- Heller, M. (1999). »Relevant Alternatives and Closure«. *Australasian Journal of Philosophy*, 77, str. 196–208.

- Hume, D. (1974). *Raziskovanje človeškega uma*. Ljubljana: Slovenska Matica.
- Kelly, T. (2013). »Disagreement and the Burdens of Judgment«. V Christensen, D. in Lackey, J. (ur.), *The Epistemology of Disagreement: New Essays*, New York: Oxford University Press, str. 31–53.
- Lehrer, K. (1976). »When Rational Disagreement is Impossible«. *Nous*, 10, str. 327–332.
- Leplin, J. (2009). *A Theory of Epistemic Justification*. Springer.
- Machuca, D. E. (ur.), (2012). *Disagreement and Skepticism*. London: Routledge.
- Matheson, J. (2015). »Disagreement and Epistemic Peers«.
- DOI: 10.1093/oxfordhb/9780199935314.013.13, Oxford Handbooks Online.
- Platon (2006). *Zbrana dela* (prev. Kocijančič, Gorazd). Celje: Mohorjeva družba.
- Schiffer, S. (2004). »Skepticism and the Vagaries of Justified Belief«. *Philosophical Studies*, 119, str. 161–184.
- Sextus Empiricus (2000). *Outlines of Scepticism*. Annas, J. in Barnes, J. (ur.), Cambridge: Cambridge University Press.
- Sosa, E. (1999). »How to Defeat Opposition to Moore«. *Philosophical Perspectives*, 13, str. 141–153.
- Šuster, D. (2010). »O zahtevi po pravem začetku«. *Analiza*, 14(4), str. 5–30.
- van Inwagen, P. (1996). »It Is Wrong, Always, Everywhere, and for Anyone, to Believe Anything, Upon Insufficient Evidence«. V Jordan, J. in Howard-Snyder, D. (ur.), *Faith, Freedom, and Rationality*, Lanham, MD: Rowman and Littlefield, str. 137–154.

Adam Elga

Prevedel Borut Trpin

Prepričanja o lastnem položaju in Trnuljčičin problem*

Poleg negotovosti o tem, kakšen je svet, smo lahko v negotovosti tudi glede lastnega prostorskega ali časovnega položaja v svetu. Rad bi zastavil problem, ki nastane v odnosu med tema dvema vrstama negotovosti, ga razrešil in iz njega izpeljal dva sklepa.

1. Trnuljčičin problem¹

Raziskovalci bodo uspavali Trnuljčico. Njeno spanje bo trajalo dva dneva, vmes pa jo bodo enkrat ali dvakrat za kratek čas zbudili. Ali jo bodo zbudili enkrat ali dvakrat je odvisno od meta kovanca (grb: enkrat, cifra: dvakrat). Po vsakem bujenju jo bodo ponovno uspavali s posebno tableto, ki povzroči, da se svoje budnosti kasneje ne bo spominjala.²

Ko se Trnuljčica prvič zbudi, do katere stopnje je lahko prepričana, da je padel grb?

Prvi odgovor: 1/2, seveda! Na začetku je bila prepričana, da je kovanec pošten in zato je bila njena izhodiščna stopnja prepričanja, da bo padel grb, 1/2. Ko se zbu-

* Elga, A. (2000). »Self-locating belief and the Sleeping Beauty problem«. *Analysis*, 60(2), str. 143–147.

¹ Tako je problem poimenoval Robert Stalnaker (ki se je s takimi primeri srečal v neobjavljenem delu Arnolda Zuboffa). Problem se pojavi kot Primer št. 5 v Piccione, 1997, kjer sta nakazana dva različna odgovora, vendar pa se avtor vzdrži sodbe, kateri odgovor je pravilen (1997: 12–14). Podoben primer z metodo pravične loterije analizira Aumann, 1997. Če njegovo analizo prilagodimo za Trnuljčičin problem, dobimo enak odgovor, kot ga bom zagovarjal v razdelku 2. Za razliko od njegovega argumenta pa moje izvajanje ni vezano na stave.

² Tableta deluje tako, da prepričanja povrne v stanje, v katerem so bila pred začetkom poskusa. Če se zdi obstoj take tablete neverjeten, naj izpostavim, da je mogoče problem zastaviti tudi brez te tablete – relevantno je le, da uspavana oseba verjame, da poskus poteka tako, kot sem ga opisal.

di, ne dobi nobene nove informacije (že ves čas je vedela, da se bo zbudila). Njena stopnja prepričanja, da je padel grb, bi torej morala ostati $1/2$.

Drugi odgovor: $1/3$, seveda! Predstavljajmo si, da se poskus večkrat ponovi. Na dolgi rok se izkaže, da je $1/3$ bujenj t.i. grb-bujenj – bujenj, ko pade grb. Ob vsakem bujenju bi torej morala imeti stopnjo prepričanja $1/3$, da gre za grb-bujenje, in, posledično, stopnjo prepričanja $1/3$, da je padel grb. To velja tudi v trenutnem primeru, ko poskus poteka samo enkrat.

Zagovarjal bom stališče, da je pravilen odgovor $1/3$.

2.

Predpostavimo, da se prvo bujenje zgodi v ponedeljek, drugo (če sploh) pa v torek. Ko se Trnuljčica zbudi, je torej lahko gotova, da je v eni izmed teh treh »situacij«:

G_1 GRB in je ponedeljek.

C_1 CIFRA in je ponedeljek.

C_2 CIFRA in je torek.

Naj izpostavim, da razlika med tem, ali je Trnuljčica v C_1 ali C_2 , ni razlika v tem, kateri možni svet je dejanski, temveč razlika v njenem časovnem položaju znotraj sveta. (Za bolj tehnično obravnavo lahko prevzamemo podobno ogrodje, kot ga predlaga Lewis, 1983. Po tem predlogu osnovne alternative, ki razdelijo stopnje prepričanj, niso možni svetovi, temveč osredinjeni možni svetovi: možni svetovi, ki so vsak zase opremljeni z označenimi posamezniki in časom. V takem ogrodju bi G_1 , C_1 in C_2 predstavljali množice osredinjenih svetov.)

Naj bo P funkcija stopenj prepričanj, ki naj bi jo Trnuljčica imela po prvem bujenju. Po prvem bujenju je gotova o sledečem: v situaciji G_1 je, če in samo če je kovanec padel na grb. Potemtakem je izračun $P(G_1)$ zadosten za razrešitev Trnuljčičinega problema. Najprej bom pokazal, da $P(C_1) = P(C_2)$, in potem, da je $P(C_1) = P(G_1)$.

Če bi (po prvem bujenju) izvedela, da je padla cifra, bi to zadoščalo za to, da spozna, da je bodisi v C_1 ali C_2 . Ker je zanjo situacija, ko je v C_1 , subjektivno povsem enaka kot, ko je v C_2 , in ker so povsem enake propozicije resnične ne glede na to, ali je v C_1 ali C_2 , celo zelo omejen princip indiferentnosti vodi do sklepa, da mora imeti enako stopnjo prepričanja v obe možnosti. Vendar pa bi morala biti njena stopnja prepričanja, da je v C_1 , potem ko izve, da je padla cifra, enaka kot njena pogojna stopnja prepričanja $P(C_1 \mid C_1 \text{ ali } C_2)$ in enako za C_2 . Od tod sledi $P(C_1 \mid C_1 \text{ ali } C_2) = P(C_2 \mid C_1 \text{ ali } C_2)$, in torej $P(C_1) = P(C_2)$.

Raziskovalci mečejo pošten kovanec, da določijo, ali bodo Trnuljčico zbudili enkrat ali dvakrat. To storijo tako, da bodisi

- (1) *najprej* mečejo kovanec in potem Trnuljčico zbudijo bodisi enkrat ali dvakrat glede na izid; ali
- (2) najprej Trnuljčico zbudijo enkrat *in* potem mečejo kovanec, da ugotovijo, ali jo bodo zbudili še enkrat.

Njena stopnja prepričanja (ko se zbudi), da je padel grb, bi morala biti enaka ne glede na to, ali raziskovalci uporabijo metodo (1) ali (2). Brez izgube splošnosti, torej predpostavimo, da uporabijo metodo (2).

Zdaj: če (ko se zbudi) Trnuljčica izve, da je ponedeljek, je to enakovredno spoznanju, da je bodisi v G_I ali C_I . Njena stopnja prepričanja, da je v G_I , predstavlja njeno stopnjo prepričanja, da bo pošten kovanec, ki ga bodo raziskovalci kmalu vrgli, padel na grb. Nerelevantno je, da bodo Trnuljčico zbudili tudi naslednji dan, če in samo če kovanec pade na cifro – v tem primeru bi morala biti njena stopnja prepričanja, da je padel grb $1/2$. Ampak njeno prepričanje, da je padel grb (po tem, ko izve, da je ponedeljek) bi moralo biti enako kot njeno pogojno prepričanje $P(G_I | G_I \text{ ali } C_I)$. Torej $P(G_I | G_I \text{ ali } C_I) = 1/2$, in torej $P(G_I) = P(C_I)$.

Če te rezultate združimo, ugotovimo, da $P(G_I) = P(C_I) = P(C_2)$. Ker se te tri stopnje prepričanj seštejejo v 1, je $P(G_I) = 1/3$

3.

Naj bo G propozicija, da je kovanec padel na grb. Preden Trnuljčico uspavajo, je njena stopnja prepričanja v G $1/2$. Ravno kar pa sem zagovarjal, da bi se njena stopnja prepričanja, ko jo zbudijo v ponedeljek, morala spremeniti v $1/3$. Ta sprememba prepričanja je nenavadna. Do spremembe ne pride, ker bi Trnuljčica prejela nove informacije – že prej je vedela, da se bo zbudila v ponedeljek.³ (Lahko celo predpostavimo, da je Trnuljčica že na začetku poskusa vedela, natančno, kakšna čutna izkustva bo imela, ko se bo zbudila v ponedeljek.) Ta sprememba prepričanj tudi ni posledica neke kognitivne okvare v vmesnem času – spomnimo; tableto za pozabo prejme šele, ko je prvič že povsem zbudjena. Kaj torej upravičuje to spremembo stopnje prepričanja?

Odgovor je, da je Trnuljčica prešla iz situacije, v kateri je bil njen lasten časovni položaj nerelevanten za resnico G , v situacijo, kjer upošteva svoj časovni položaj

³ Ko trdim, da akterka pridobi nove informacije (kot bom trdil tudi v nadaljevanju), trdim, da pridobi podatke, ki izključujejo možne svetove, ki jih prejšnji podatki niso izključevali. Z drugimi besedami: akterka pridobi nove informacije, ko izve resnico propozicije, ki jo lahko izrazimo z večnim stavkom (Quine, 1960: 191) v dovolj bogatem jeziku.

kot relevantnega za resnico G .⁴ Predpostavimo, na primer, da Trnuljčica na začetku poskusa ni prepričana ali je ura 1:01 ali 1:02. V tem trenutku svojega časovnega položaja ne smatra kot relevantnega za resnico G : njena stopnja prepričanja v G ob pogoju, da je ura 1:01, je $1/2$. Prav tako je tudi njena stopnja prepričanja v G ob pogoju, da je ura 1:02, $1/2$.

Nasprotno pa (ob predpostavki, da prepričanja posodablja racionalno), ko Trnuljčico zbudijo v ponedeljek, svoj trenutni časovni položaj smatra kot relevantnega za resnico G : njena stopnja prepričanja v G ob pogoju, da je ponedeljek, je $1/2$, njena stopnja prepričanja v G ob pogoju, da je torek, pa 0. V ponedeljek se njena brezpogojna stopnja prepričanja v G razlikuje od $1/2$, saj predstavlja uteženo povprečje teh dveh pogojnih stopenj prepričanj – se pravi, uteženo povprečje $1/2$ in 0.

Nobeno presenečenje ni, da se to, ali akterka upošteva svoj časovni položaj kot relevantnega za resnico propozicije, skozi čas spreminja. Kar je bolj presenetljivo – in to je prvi sklep –, je, da se tovrstna sprememba lahko zgodi v obdobju, ko akterka niti ne prejme nobene nove informacije niti ne utрпи kognitivne okvare.

Na začetku poskusa je imela stopnjo prepričanja $1/2$ v G . Hkrati pa je bila gotova, da bo ob ponedeljkovem bujenju njena stopnja prepričanja v H padla na $1/3$ – čeprav je bila gotova, da vmes ne bo pridobila nobene nove informacije ali pa utrpela neke kognitivne okvare. Primer Trnuljčice tako predstavlja novo vrsto protiprimera Bas van Fraassenovemu »principu odsevanja« (1984: 244, 1995: 19), celo zelo dodelani varianti, iz katere sledi naslednje:

Vsaka akterka, ki je gotova, da bo imela jutri stopnjo prepričanja x v propozicijo R (in vmes ne bo prejela novih informacij niti ne bo utrpela kognitivne okvare), bi morala že sedaj imeti stopnjo prepričanja x v R .⁵

David Lewis se je nekoč vprašal, »kaj bi se zgodilo s teorijo odločanja, če bi [nandomestili prostor možnih svetov s prostorom osredinjenih možnih svetov]?« in odgovoril: »Ne prav dosti« (Lewis, 1983: 149). Drugi sklep, ki sledi iz Trnuljčičinega problema, pa je, da se nekaj zgodi: poraja se novo vprašanje, kako naj bi racionalni akterji skozi čas posodabljali svoja prepričanja.⁶

⁴ Ko trdim, da akterka upošteva svoj časovni položaj kot relevanten za resnico določene propozicije, trdim, da obstaja tak čas t , da so akterkina prepričanja združljiva s položajem v t , in da se njena stopnja prepričanja v propozicijo ob pogoju, da je v položaju t , razlikuje od njenih brezpogojnih stopenj prepričanj v to propozicijo.

⁵ Za ugotovitev, da je odgovor $1/3$ v nasprotju s principom odsevanja, se zahvaljujem Nedu Hallu.

⁶ Pri nastajanju tega prispevka so mi bili v pomoč Jamie Dreier, Gary Gates, Ned Hall, Vann McGee, Robert Stalnaker, Roger White, Sarah Wright, udeleženci konference na Brown University leta 1999 (tam sem predstavil zgodnejšo verzijo tega prispevka) in anonimni recenzent. Za pomoč se jim zahvaljujem.

Literatura

Aumann, R. J., Hart, S. in Perryi, M. (1997). »The forgetful passenger«. *Games and Economic Behavior*, 20, str. 117–120.

Lewis, D. (1983). »Attitudes *de dicto* and *de se*«. V Lewis, D., *Philosophical Papers, Volume I*, New York: Oxford University Press, str. 133–159.

Piccione, M. in Rubenstein, A. (1997). »On the interpretation of decision problems with imperfect recall«. *Games and Economic Behavior*, 20, str. 3–24.

Quine, W.V. (1960). *Word and Object*. Cambridge, Mass.: The MIT Press.

van Fraassen, B. C. (1984). »Belief and the will«. *Journal of Philosophy*, 81, str. 235–256.

van Fraassen, B. C. (1995). »Belief and the problem of Ulysses and the sirens«. *Philosophical Studies*, 77, str. 7–37.

David Lewis

Prevedel Borut Trpin

Trnuljčica: odgovor Elgu*

1. Problem

Raziskovalci Eksperimentalnega filozofskega laboratorija so se odločili, da izvedejo sledeč poskus. Najprej bodo Trnuljčici povedali vse, kar vam bom povedal v tem odstavku, poskrbeli pa bodo tudi za to, da jim bo povsem verjela. Nato jo bodo v nedeljo uspavali. V ponedeljek jo bodo za kratek čas zbudili. Najprej ji ne bodo povedali, kateri dan je, kasneje pa ji bodo zaupali, da je ponedeljek. Zatem ji bodo izbrisali spomin na budnost. Morda jo bodo za kratek čas zopet zbudili v torek, odvisno od meta poštenega kovanca: če pade grb, jo bodo zbudili samo v ponedeljek, če cifra pa tudi v torek. V sredo se bo poskus zaključil in Trnuljčica se bo lahko zbudila. Ta tri možna kratka bujenja med poskusom bodo nerazločljiva: v ponedeljek bo imela dostop do enakih podatkov ne glede na izid meta kovanca, in če jo zbudijo v torek, bo zaradi ponedeljkovega izbrisa spomina imela dostop do povsem enakih podatkov kot ob ponedeljkovem bujenju. Naučili pa jo bodo, da bo lahko svoje bujenje v sredo, ko se poskus zaključi, razločila od bujenj v ponedeljek in torek, in seveda tudi od vseh prejšnjih in kasnejših bujenj.

Predpostavimo, da je Trnuljčica vzor verjetnostne racionalnosti in ima vedno natančno takšne stopnje prepričanj (subjektivne verjetnosti), kot bi jih morala imeti. Funkcijo njenih stopenj prepričanj bomo preverili trikrat. Naj bo P njena funkcija stopenj prepričanj, takoj ko se zbudi v ponedeljek. Naj bo P_+ njena funkcija stopenj prepričanj, takoj ko izve, da je ponedeljek, in naj bo P_- njena funkcija stopenj prepričanj, tik preden jo uspavajo v nedeljo, ko že ve, kako bo poskus potekal.

* Lewis, D. (2001). »Sleeping Beauty: Reply to Elga«. *Analysis*, 61(3), str. 171–176.

Kakšno stopnjo prepričanja naj bi Trnuljčica, takoj po bujenju v ponedeljek, pripisala hipotezi GRB, da je padel grb? Kakšno stopnjo prepričanja naj pripiše hipotezi CIFRA, da je padla cifra? Adam Elga (2000) trdi, da je $P(\text{GRB}) = 1/3$, $P(\text{CIFRA}) = 2/3$. S tem se ne strinjam in trdim, da je $P(\text{GRB}) = P(\text{CIFRA}) = 1/2$.

Nisem še opredelil, ali se met kovanca zgodi pred ali po ponedeljkovem bujenju. Elgov argument sicer temelji na tem, da kovanec mečejo naknadno; vendar pa meni, in tu se strinjava, da bi moral biti naš odgovor na vprašanje v obeh primerih enak. Moj argument velja za obe možnosti.

2. Skupno izhodišče

Pri najinem nestrinjanju je še posebej zanimivo, da se skoraj v vsem strinjava (vključno z mnogimi stvarmi, s katerimi se marsikdo ne bi). Naj torej najprej orišem najino skupno izhodišče.

Strinjava se, da obstajata dva tipa možnih pripisovanj stopenj prepričanj. Prvi tip se nanaša na to, kateri možni svet je dejanski; drugi tip pa se nanaša na to, kdo in kdaj nekdo je in v kakšni vrsti sveta živi. Tako kot Quine (1969) bomo drugi tip opredelili kot razrede osredinjenih svetov: možni svetovi z označenimi posamezniki-ob-nekem-času znotraj njih. Imenujmo razred osredinjenih svetov osredinjene možnosti. (Prvi tip, neosredinjene možnosti, bi lahko prikazali kot razrede neosredinjenih možnih svetov; vendar to ni potrebno, saj lahko neosredinjene možnosti zaobsežemo pod osredinjene.) Lahko se zgodi, da se dva osredinjena svetova nahajata znotraj istega neosredinjenega možnega sveta: le njuni označeni posamezniki-ob-nekem-času se razlikujejo. V takih primerih ta dva svetova imenujemo *kolocirana*.

Ko se Trnuljčica med poskusom zbudi, so z vsemi njenimi podatki skladne tri osredinjene epistemske možnosti:

G_1 : GRB in je ponedeljek.

C_1 : CIFRA in je ponedeljek.

C_2 : CIFRA in je torek.

Elga zapiše, da »ker je zanjo situacija, ko je v C_1 , subjektivno povsem enaka kot, ko je v C_2 , in ker so povsem enake propozicije resnične ne glede na to, ali je v C_1 ali C_2 , celo zelo omejen princip indiferentnosti vodi do sklepa, da mora imeti enako stopnjo prepričanja v obe možnosti« (Elga, 2000: 144). S »propozicijo« meni neosredinjeno možnost. Razlog, da so enake propozicije resnične ne glede na to, ali je Trnuljčica v C_1 ali C_2 , je v tem, da so osredinjeni svetovi, ki so člani C_1 , kolocirani z ustrežajočimi člani C_2 .

Sprejemam tudi Elgov »zelo omejen princip indiferentnosti«.¹ Strinjava se torej tudi v tem:

$$(1) \quad P(C_1) = P(C_2).$$

Ko Trnuljčica med svojo ponedeljkovo budnostjo izve, da je ponedeljek, bi se morala njena funkcija stopenj prepričanj spremeniti iz P v P_+ . Z Elgom se strinja, da do te spremembe pride ob pogojevanju z njenimi novimi podatki, ki jih lahko izrazimo kot $ne-C_2$ ali kot $(G_1 \vee C_1)$. Torej

$$(2) \quad \begin{aligned} P_+(GRB) &= P(GRB \mid G_1 \vee C_1), \\ P_+(CIFRA) &= P(CIFRA \mid G_1 \vee C_1). \end{aligned}$$

Iz celote njenih podatkov na začetku ponedeljkove budnosti sledi, da je GRB resničen, če je G_1 v tem trenutku resničen; in podobno za CIFRA in $(C_1 \vee C_2)$. Njeno pogojno stopnjo prepričanj lahko enostavno izrazimo kot razmerje brezpogojnih stopenj prepričanj. Strinjava se torej, da

$$(3) \quad \begin{aligned} P(GRB \mid G_1 \vee C_1) &= P(G_1 \mid G_1 \vee C_1) = P(G_1) / [P(G_1) + P(C_1)], \\ P(CIFRA \mid G_1 \vee C_1) &= P([C_1 \vee C_2] \mid G_1 \vee C_1) = P(C_1) / [P(G_1) + P(C_1)], \end{aligned}$$

in dalje, da

$$(4) \quad \begin{aligned} P(GRB) &= P(G_1), \\ P(CIFRA) &= P(C_1 \vee C_2) = P(C_1) + P(C_2). \end{aligned}$$

Strinjava se tudi glede

- (5) $P_+(GRB) = 1/2 = P_+(CIFRA)$ in
- (6) Trnuljčica ne dobi nobenih dodatnih neosredinjenih podatkov, ki so relevantni za GRB proti CIFRA v času od njene funkcije stopenj prepričanj P do funkcije stopenj prepričanj P . Edini podatki, ki jih pridobi, so osredinjeni podatki,² da je trenutno bodisi v ponedeljkovem bujenju ali pa torkovem bujenju, t.j. $(G_1 \vee C_1 \vee C_2)$.

¹ Če ne bi upoštevali kolokacije ustrežajočih članov dveh epistemskih možnosti, bi dobili manj omejen princip indiferentnosti, iz katerega bi sledilo, da $P(G_1) = P(C_1) = P(C_2)$. Na tak način bi po bližnjici prišli do Elgovega sklepa – bistveno preuranjeno in Elga se je tej poti bistro izognil. Sledijo nenavadne posledice: na primer, da za enakovrednost $P(G_1)$ in $P(C_1)$ poštenost kovanca sploh ni relevantna, četudi bi Trnuljčica izvedela, da je kovanec obtežen tako, da je verjetnost padca grba 999 proti 1!

² Naj izpostavim: Elga govori o »novih informacijah«. Vendar pa v njegovi terminologiji »osredinjena informacija« sploh ne služi kot »informacija« (prav tam: 145, op. 3). Kljub temu bi jo vseeno lahko imel za »podatek«.

Tukaj se najino skupno izhodišče zaključí, zato se bom posvetil najinima analizama problema, vključno z najinimi različnimi zaključki glede $P(\text{GRB})$.

3. Elgov argument

Elga upošteva zgolj možnost, ko raziskovalci mečejo kovanec po ponedeljkovem bujenju, in iz kasnejšega sklepa na zgodnejše:

$$(E1) \quad P_+(\text{GRB}) = 1/2 \quad (\text{premisa}).$$

$$(E2) \quad \therefore P(G_1)/[P(G_1) + P(C_1)] = 1/2 \quad (E1, 2, 3).$$

$$(E3) \quad \therefore P(G_1) = P(C_1) \quad (E2).$$

$$(E4) \quad \therefore P(G_1) = P(C_1) = P(C_2) = 1/3 \quad (E3, 1).$$

$$(E5) \quad \therefore P(\text{GRB}) = 1/3, P(\text{CIFRA}) = 2/3 \quad (E4, 4).$$

Quod erat demonstrandum; vendar pa opazimo tudi dve nadaljnji posledici.

$$(E6) \quad \therefore P_+(\text{GRB}) = P(\text{GRB}) + 1/6 \quad (E1, E5).$$

$$(E7) \quad \therefore \text{Spremembe stopnje prepričanja iz } P_- \text{ v } P \text{ niso povzročili novi relevantni neosredinjeni podatki; bodisi je sploh niso povzročili relevantni podatki ali pa so bili osredinjeni podatki } (G_1 \vee C_1 \vee C_2) \text{ pomembni za GRB proti CIFRA} \quad (E5, 5, 6).$$

4. Moj argument

Ob upoštevanju obeh možnosti svoj argument izvajam od zgodnejšega k kasnejšemu.

$$(L1) \quad \text{Zgolj novi relevantni podatki, bodisi osredinjeni ali ne, lahko povzročijo spremembo stopnje prepričanja; podatki } (G_1 \vee C_1 \vee C_2) \text{ pa za GRB proti CIFRA niso relevantni} \quad (\text{premisa}).$$

$$(L2) \quad \therefore P(\text{GRB}) = 1/2 = P(\text{CIFRA}) \quad (L1, 5, 6).$$

Quod erat demonstrandum; opazimo pa pet posledic.

$$(L3) \quad \therefore P(G_1) = 1/2 = P(C_1) + P(C_2) \quad (L2, 4).$$

$$(L4) \quad \therefore P(C_1) = 1/4 = P(C_2) \quad (L3, 1).$$

$$(L5) \quad \therefore P(\text{GRB} \mid G_1 \vee C_1) = 2/3, \\ P(\text{CIFRA} \mid G_1 \vee C_1) = 1/3 \quad (L3, L4, 3).$$

$$(L6) \quad \therefore P_+(\text{GRB}) = 2/3, \\ P_+(\text{CIFRA}) = 1/3 \quad (L5, 2).$$

$$(L7) \quad \therefore P_+(\text{GRB}) = P(\text{GRB}) + 1/6 \quad (L2, L6).$$

5. Oris najinega nestrinjanja

Elga zavrača mojo premiso: njegova (E7) je v protislovju z mojo (L1). Jaz zavračam njegovo premiso: moja (L6) je v protislovju z njegovo (E1). Od tod izvira najino razhajanje. Zavreči morava eno izmed premis. Če obdrživa obe (in najino skupno izhodišče), prideva v protislovje. Kaj še lahko povemo?

Elgov razlog za zavrnitev moje premise je, da gre tja, kamor ga vodi argument. Preseneča ga, da lahko pride do »[spremembe stopnje prepričanja] v obdobju, ko akterka niti ne prejme nobene nove informacije« – t.j. nobenih novih neosredinjenih podatkov – »niti ne utрпи kognitivne okvare« (prav tam: 146). Trdim, da bi bilo prav tako presenetljivo, in prav tako sumljivo, če bi ugotovili, da je osredinjeni podatek (G_1 v C_1 v C_2) relevanten za GRB proti CIFRA; ali če bi odkrili, da je Trnuljčica utrpela kognitivno okvaro. (Izbris spomina je vsekakor kognitivna okvara; vendar pa se to zgodi kasneje, torej je nerelevantno za spremembo med P. in P.) To mu dopustim; bi pa bilo lepše, če bi bila njegova zavrnitev moje premise neodvisno utemeljena.

Trdim, da je moja zavrnitev Elgove premise neodvisno utemeljena. Od kje jo je Elga sploh dobil? V času P_+ , v Elgovem primeru kovanec mečejo za tem, Trnuljčica ve, da bo prišlo do meta poštenega kovanca. Imamo tudi znan princip, da morajo biti stopnje prepričanj o prihodnjih slučajnih dogodkih enake znanim slučajem (glej Mellor, 1971; Lewis, 1980). Prav iz tega principa sva dobila skupno (5), ki ureja tako primere, ko kovanec mečejo pred ponedeljkovim bujenjem in po njem. Po tem principu bi sledilo tudi, da $P_+(GRB) = \text{možnost}^*(GRB) = 1/2$ in da $P_+(CIFRA) = \text{možnost}(CIFRA) = 1/2$, iz česar sledi Elgova premisa (E1).

Naj omenim, da ta princip vključuje predpogoj, ki smo ga izpolnili, ko smo pridobili (5), v Elgovem izvajanju na premiso (E1) pa ni izpolnjen. Zamislimo si, da obstaja prerok, za katerega vemo, da zelo uspešno napoveduje prihodnost, zaradi česar zelo resno obravnavamo hipotezo, da po nekakšni vzvratni vzročni verigi prejema vesti iz prihodnosti. Prerok nam le redko neposredno pove, kaj se bo zgodilo, nam pa pogosto svetuje, kakšno stopnjo prepričanja naj imamo glede nekega izida. Občasno so njegovi nasveti v nasprotju z znanimi možnostmi (ang. *chances*), po katerih bi sicer uravnavali svoje stopnje prepričanj. Kaj naj torej storimo? Če je prerok dovolj pogosto pravilno napovedal prihodnost, potem trdim, da lahko znane možnosti zanemarimo.

Ko torej Trnuljčica med ponedeljkovo budnostjo izve, da je ponedeljek, ali enakovredno ne- C_2 , pridobi podatke – osredinjene podatke – glede prihodnosti: da trenutno ni v njej. To so novi podatki: preden je izvedela, da je ponedeljek, jih ni imela. Naj poudarim, podatkov ne dobi od preroka ali preko vzvratne vzročnosti, hkrati pa tudi ne z usklajevanjem svoje stopnje prepričanja z znanimi možnostmi.

* Op. p.: V izvorniku *chance*, slučaj oz. slučajna možnost.

Vest je relevantna za GRB, ker njeno stopnjo prepričanja dvigne za $1/6$; glej mojo (L7). Elga se strinja; glej njegovo (E6). Predpogoj je torej pomemben in zato ne moremo izpeljati, da $P_+(GRB) = \text{možnost}(GRB)$ in $P_+(CIFRA) = \text{možnost}(CIFRA)$. Priznam, da je to nova in presenetljiva posledica predpogoja, zato se Elgu zahvaljujem, da je problem izpostavil. Kljub tej posledici smatram predpogoj kot prepričljiv, neodvisno od tega, da bi rad šel tja, kamor me vodi moj argument.³

Literatura

Elga, A. (2000). »Self-locating belief and the Sleeping Beauty problem«. *Analysis*, 60(2), str. 143–47.

Lewis, D. (1980). »A subjectivist's guide to objective chance«. V Jeffrey, R. C. (ur.), *Studies in Inductive Logic and Probability*, Vol. 2, Berkeley: University of California Press, str. 263–293. Ponatisnjeno v Lewis, D., *Philosophical Papers*, Vol. 2., Oxford: Oxford University Press.

Mellor, D. H. (1971). *The Matter of Chance*. Cambridge: Cambridge University Press.

Quine, W. V. (1969). »Propositional objects«. V Quine, W. V., *Ontological Relativity & Other Essays*, New York: Columbia University Press.

³ Za pripombe se zahvaljujem Adamu Elgu in uredniku *Analysis*.

Borut Trpin

Filozofska fakulteta Univerze v Ljubljani

0 nenatančnih stopnjah prepričanj: primer meduzarja, Ellsbergov paradoks in Trnuljičin problem

Stopnje prepričanj najpogosteje določimo tako, da jih ponazorimo s stopnjo verjetnosti. »Ostrogledneži« trdijo, da lahko stopnje prepričanj opišemo zgolj z natančnimi verjetnostmi. Na drugi strani pa imamo t.i. »topogledne«, ki trdijo, da so stopnje prepričanj lahko nenatančne in bi jih v primeru nenatančnih (dvoumnih, nejasnih, nasprotujočih si ali pomanjkljivih) podatkov morali ponazoriti z nenatančnimi verjetnostmi. Vendar pa nenatančne stopnje verjetnosti lahko vodijo v neracionalno vedenje, po natančnejšem pregledu pa se izkažejo tudi za pojmovno nesmiselne. Na podlagi instrumentalnega pogleda, ki temelji na koherentnosti, zato branim stališče, da so stopnje verjetnosti lahko opredeljene zgolj z natančnimi verjetnostmi. Ta princip nato uporabim v treh posebnih primerih, za katere se na prvi pogled zdi, da kličejo po nenatančnih stopnjah. To so primer meduzarja, Ellsbergov paradoks in Trnuljičin problem.

Ključne besede: stopnje prepričanj, nenatančna verjetnost, subjektivna verjetnost, negotovost, spoznavna teorija

Uvod

V sodobni spoznavni teoriji najdemo precej zagovornikov stališča, da lahko prepričanja razlikujemo glede na njihovo moč. Ideja je intuitivno sprejemljiva, saj smo glede nekaterih stvari močno prepričani (npr. da bo tudi jutri vžšlo sonce), glede drugih pa zgolj šibko (npr. da bom jutri zadel na loteriji). Na težave naletimo, ko se s prepričanji, ki pridejo v različnih stopnjah, začnemo ukvarjati natančneje. Poraja se več vprašanj - ali stopnje prepričanj obstajajo samo v teoriji ali imajo tudi mentalne korelate? Kako jih analizirati? Kateri principi racionalnosti veljajo za stopnje prepričanj? Kakšne naj bi bile stopnje prepričanj?

En izmed osrednjih pristopov k analizi je preko subjektivne interpretacije verjetnosti, kjer je verjetnost podana kot stopnja prepričanja, t.j. če sem o trditvi pov-

sem prepričan, imam subjektivno verjetnost 1, npr. $P(\text{»Miha je človek ali pa ni človek«}) = 1$, če sem prepričan, da trditev ne drži oz. ekvivalentno, vanjo nikakor nisem prepričan, imam subjektivno verjetnost 0, npr. $P(\text{»Miha je človek in taisti Miha je riba«}) = 0$, za vsa ostala prepričanja pa je stopnja verjetnost nekje vmes med 0 in 1.

Vprašanje, ki ga želim nasloviti v tem prispevku, se naslavlja na koncept nenančnih verjetnosti, ki ima v zadnjih letih vedno več privržencev kot orodje za analizo delnih prepričanj. Trdijo namreč, da so podatki, na podlagi katerih oblikujemo svoja prepričanja, včasih tako neoprijemljivi, da lahko prepričanja, ki nastanejo na podlagi teh podatkov, analiziramo zgolj s pomočjo nenatančne verjetnosti, t.j. verjetnosti, ki ni podana v posamezni vrednosti, temveč v množici vrednosti (običajno gre za intervale). Trdim, da je tak pristop napačen, saj med sabo pomeša dve ravni: raven (delnih) prepričanj in raven (delnih) prepričanj o prepričanjih. Skozi besedilo bom pokazal, da lahko (delna) prepričanja – če kaj takega sploh obstaja – opišemo zgolj z natančnimi posameznimi verjetnostmi. Nenančne stopnje prepričanj, po drugi strani, pa niso stopnje prepričanj kot takih, temveč so natančna delna prepričanja o natančnih delnih prepričanjih prvega reda.

Poglejmo najprej nekoliko natančneje, kaj imamo v mislih, ko govorimo o stopnjah prepričanj. Pogosto se znajdemo v stanju negotovosti: ne vemo, recimo, kako se bo spremenila cena goriva, ali bo jutri deževalo in kdo bo zmagal na volitvah. Zagovorniki teze, da se prepričanja ne razlikujejo zgolj po vsebini, temveč tudi po moči, lahko tovrstna vprašanja razlikujejo na učinkovit način: bistveno močneje sem prepričan, da bo jutri vzšlo sonce, kot da bo jutri deževalo. Kot se je izkazalo, je za opis tovrstnih stopenj prepričanj še posebej dobro orodje verjetnostni račun oz. subjektivna interpretacija verjetnosti.

Kakšna točno je narava stopenj prepričanj, je sicer še vedno odprto vprašanje. Eriksson in Hájek (2007: 212) tako trdita, da so stopnje prepričanj primaren pojem, ki se ga ne da dalje analizirati z drugimi pojmi, in trdita, da »tovrstna teorija zadošča«, kar predstavlja dobro izhodišče za numerično analizo stopenj prepričanj kot stopenj verjetnosti.

Verjetnost običajno opredelimo na intervalu med 0 in 1 oz. kot je bolj domače v vsakdanjem jeziku (ne pa tudi bolj razumljivo – glej Gigerenzer, Hoffrage, 1995) v odstotkih med 0% in 100%. Če na tak način opišemo stopnje prepričanj, pri čemer hkrati sledimo Kolmogorovovim aksiomom verjetnosti, potem – kot dokazujejo zagovorniki bayesijanske spoznavne teorije – so naša delna prepričanja vedno racionalna, če pa ta načela kršimo, nam lahko zviti stavodajalec proda serijo stav, ki se nam v skladu z našimi prepričanji zdijo sprejemljive, a nas privedejo v gotovo izgubo.

Čeprav je v zvezi s tovrstnimi stališči – imenujemo jih probabilizem (Hájek, 2008) – odprtih še kar nekaj vprašanj, še več pa jih s to interpretacijo lahko učinkovito razrešimo, je eno izmed odprtih vprašanj tudi, ali lahko stopnje prepričanj opišemo zgolj z natančnimi ali pa tudi nenatančnimi verjetnostmi. Vpeljava nenatančnih verjetnosti (t.j. družin verjetnostnih funkcij) v analizo prepričanj je tako raziskovalce razdelila v dva tabora. Tiste, ki zagovarjajo tezo, da lahko prepričanja opišemo samo z natančnimi verjetnostmi, imenujmo tabor ostrogledih, njihove nasprotnike pa tabor topogledih.

Glavna motivacija topogledih je, da natančno odmerjenih stopenj prepričanj ne moremo vedno določiti. Če ima akterka na voljo zgolj podatke, ki so bistveno nenatančni (dvoumni, nejasni, nasprotujoči si ali pa pomanjkljivi), potem – tako topogledi – ima lahko glede teh podatkov zgolj nenatančne stopnje prepričanj.

Ideja na prvi pogled deluje prepričljivo, vendar lahko vodi v neracionalno vedenje in ob podrobnejši preučitvi ni smiselna, saj gre za problem drugega in ne prvega reda. Trdim torej, da nenatančni (»topi«) podatki ne zahtevajo nenatančnih verjetnosti, kot to trdijo v taboru topogledih (npr. Levi, 1980; Jeffrey, 1983; Kaplan, 2006; Joyce, 2005).

Kljub zavrnitvi izhodiščne ideje topoglednežev menim, da njihovi izsledki kažejo na pomemben problem, ki se skriva v situacijah z nejasnimi podatki. V teh primerih namreč lahko zavzamemo več različnih stopenj prepričanj, vendar pa ne moremo imeti različnih stopenj prepričanj o isti stvari ob istem času. To ostrogledno stališče bom zagovarjal na treh primerih, za katere se zdi, da kličejo po topogledni obravnavi. To so primer meduzarja, Ellsbergov paradoks in Trnuljičin problem.

Teoretični okvir

Začnimo z nekaj teoretičnimi razmisleki o naravi nenatančnih verjetnosti, nenatančnih stopenj prepričanj in povezanih vprašanj. Nenatančne verjetnosti pomenijo posplošitev natančnih verjetnosti oz. natančneje, gre za množice funkcij verjetnosti. Predstavimo jih lahko kot intervale vrednosti, čeprav to ni edina možnost. Za potrebe tega prispevka nam bo omejitev na intervalne nenatančne verjetnosti zadoščala, saj bo za mojo obrambo natančnih verjetnosti ključna ideja, da ima lahko akterka naenkrat zgolj eno, natančno določeno stopnjo prepričanja, pri čemer bo moj argument veljal tudi za neintervalne množice verjetnosti.

Intervalne nenatančne verjetnosti predstavimo z intervali natančnih verjetnosti, kjer kot meji nastopita t.i. spodnja in zgornja verjetnosti. Natančneje: »Družina verjetnosti P glede X je določena s spodnjo in zgornjo verjetnostjo na množici A . Izrecno:

$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A)$ in $\bar{P}(A) = \sup_{P \in \mathcal{P}} P(A)$.¹ Od tod:
 $\mathcal{P}_{\underline{P}, \bar{P}}(A) = \{P | \forall A \subseteq X \text{ merljiv, } \underline{P}(A) \leq P(A) \leq \bar{P}(A)\}$ «
 (Destercke in Dubois, 2006: 250).

Zamisel, da verjetnosti predstavimo z intervali namesto s posameznimi vrednostmi sicer ni nova (med zgodnejšimi zagovorniki najdemo npr. ekonomista J. Keynesa v dvajsetih letih prejšnjega stoletja; Keynes, 1921), vendar pa se je širše pripoznanje pojavilo šele v zadnjih desetletjih. Razlogov za vpeljavo nenatančnih verjetnosti je več. V mnogih primerih je težko opredeliti natančno verjetnost, zato je – tako topogledneži – bolj primerno, če verjetnost opišemo z intervalom. Zdi se tudi, da nas natančno določena verjetnost v določenih primerih lahko vodi v čudne posledice, npr. če si pomagamo s principom maksimalne entropije (za enega izmed primerov glej van Fraassen, 1989).

V povezavi s stopnjami prepričanj se nenatančne verjetnosti zdijo še posebej primerne zaradi (navidezne) podobnosti s psihološko realnostjo. Pripisovanje natančnih številskih vrednosti stopnjam prepričanj namreč deluje psihološko nepričljivo. Kaj naj bi pomenilo, denimo, da je moja stopnja prepričanja, da bo jutri deževalo, 0,563? Zakaj ravno ta vrednost in ne 0,564? Čeprav so zgodovinski razlogi za vpeljavo nenatančnih verjetnosti ločeni od tovrstnih primerov (prej omenjeni Keynes je nenatančne verjetnosti vpeljal v svoje raziskave logičnih verjetnosti), so nenatančne verjetnosti v zadnjih časih pridobile mnogo zagovornikov ravno v razpravah o (nenatančnih) stopnjah prepričanj oz. v subjektivni teoriji verjetnosti.

Pomembna teoretična razlika med natančnimi in nenatančnimi stopnjami prepričanj se kaže v pogojevanju, t.j. načinu, kako predhodne verjetnosti posodabljammo ob soočenju z novimi podatki, kar je en izmed temeljev bayesianstva (glej Hartmann in Sprenger, 2010).

Pogojevanje v primeru natančnih verjetnosti (vsaj v osnovnem načinu) poteka zelo enostavno: racionalna akterka ima neko predhodno stopnjo prepričanja $P_{\text{predhodnja}}(H)$ in neko pogojno stopnjo prepričanja $P(H|E)$ (stopnja prepričanja v H ob pogoju E). Ko izve E , se njena naknadna stopnja prepričanja spremeni v $P_{\text{naknadna}}(H) = P(H|E)$.

¹ Z oznako *inf* in *sup* označujemo *infimum* in *supremum*, največjo spodnjo oz. najmanjšo zgornjo mejo. *Infimum* je torej največji element delno urejene množice S , ki je manjši od vseh elementov množice S , če tak element obstaja, *supremum* pa najmanjši element delno urejene množice S , ki je večji od vseh elementov množice S , če tak element obstaja.

V primeru posodabljanja z nenatančnimi stopnjami prepričanj je proces v osnovi podoben, vendar računsko zahtevnejši, saj se morajo vse vrednosti intervala (oz. splošneje, množice stopenj prepričanj) posodobiti pogojno glede na nove podatke.

En izmed glavnih razlogov, da se racionalna akterka kljub večji računski zahtevnosti vseeno ravna glede na nenatančne stopnje prepričanj, naj bi bil – tako topogledneži – v večji točnosti (ang. *accuracy*) nenatančnih prepričanj.

Pokažimo s primerom: če je objektivna slučajnost² neke propozicije P natančno 0,341, akterkina stopnja prepričanja v to propozicijo pa brez posebnih razlogov 0,34, akterka krši enega izmed uveljavljenih principov racionalnosti, t.i. glavni princip D. Lewisa, da morajo biti stopnje prepričanj enakovredne objektivni slučajnosti, če nima kakšnih drugih močnejših razlogov. Nenatančna stopnja verjetnosti v intervalu $[3/10, 4/10]$ bi objektivno slučajnost vseeno zajela. To sicer intuitivno deluje neprepričljivo. Kot sta Yaniv in Foster (1997) pokazala v psiholoških eksperimentih, raje sprejememo natančne, a nekoliko zgrešene ocene, kot širše intervale, ki pokrijejo tudi iskano vrednost.

Poleg tega lahko večjo točnost enostavno pokažemo na primeru jutrišnjega dežja, kjer je objektivna slučajnost skoraj gotovo zajeta v intervalu stopenj prepričanj $[1/10, 9/10]$, v natančni vrednosti 0,34 pa le v enem primeru. Kljub temu, da je širši interval v tem smislu točnejši (predpostavimo, da zajema iskano vrednost), je bistveno manj informativen in kot tak bolj ali manj neuporaben.

Topogledneži se tovrstnih ugovorov zavedajo. En izmed standardnih odzivov je, da taka definicija pravih intervalov ni primerna. Meje intervalov nenatančnih verjetnosti niso povsem arbitrarne: spodnja meja mora biti natanko tako nizka, kot to dopuščajo podatki. Enako velja za zgornjo mejo, ki je postavljeno točno tja, kjer podatki dopuščajo zgornjo mejo.

Če je racionalna akterka soočena s povsem dvoumnimi podatki, je tako njena stopnja prepričanja kar cel interval $[0, 1]$. Stopnja prepričanja o jutrišnjem deževju $[1/10, 9/10]$ pa je neupravičena, saj je podatki ne dopuščajo (pod pogojem, da nismo v Angliji). S tovrstno omejitvijo postanejo nenatančne stopnje prepričanj bolj privlačne, vendar pa ostaja odprto vprašanje, kaj se zgodi z mejami, ki vseeno ostanejo ostro zamejene.

Kako natančno zamejiti zgornjo in spodnjo mejo? Zakaj bi neki podatki dopuščali interval $[3/10, 4/10]$, ne pa tudi $[29/100, 41/100]$? V tem se tudi kaže en izmed problemov intervalsko določenih stopenj prepričanj, saj so topogledneži še zmeraj soočeni z ostrimi natančnimi mejami, le da se iz posamezne vrednosti premaknejo na robove intervala (splošneje, v primeru množic pa na binarnost (ne)članstva).

² Pojem objektivne slučajnosti je bil vpeljan kot mera točnosti stopenj prepričanj, t.j. če je stopnja prepričanja točna, potem je enaka objektivni slučajnosti. Za moje izvajanje ni pomembno kaj točno, naj bi objektivne slučajnosti bile oz. ali sploh obstajajo.

Na nek način so torej podvrženi svojim lastnim kritikam ostrogledih. Lahko si zamislimo tudi možen odgovor topogledneža, da so lahko tudi meje intervalov tope (v nekem smislu podobno kot je to predlagal Sturgeon, 2008). Interval stopenj prepričanj bi potemtakem izgledal nekako tako:

$$P(A) = {}_a[{}^b x, y {}^c]{}^d; 0 \leq a \leq b \leq x \leq y \leq c \leq d \leq 1,$$

pri čemer bi bili zgornja in spodnja meja $[a, b]$ in $[c, d]$ opredeljeni kot interval. Vendar pa je na tej točki že jasno, da bi tudi tovrstni obrambi ostrogledi ugovarjali, topogledneži pa bi obrambo ponavljali vedno znova (v višjem in višjem redu). Ostrogledi bi vsakič znova trdili, da sta meji intervalov spodnje in zgornje meje še vedno ostri, topogledneži pa bi vsako ostro mejo v nedogled nadomeščali s topimi intervalsko določenimi posameznimi vrednostmi. Tukaj se tudi že bolj jasno kaže ključna zmeta topoglednežev: to, kar imenujejo nenatančna delna prepričanja, so le delna prepričanja drugega (oz. višjega) reda, v prvem redu pa so vsa delna prepričanja lahko zgolj natančna.

Ena izmed možnosti je tudi, da množice nenatančnih verjetnosti predstavimo z mehкими množicami, kjer ima vsak član določeno stopnjo pripadnosti, ali pa analogno s statistično distribucijo. Zdi se, da se vsaj v teh dveh primerih lahko do neke mere izognemo argumentu ostrogledih, da imajo vse nenatančne verjetnosti vsaj na neki točki še zmeraj natančno zamejene meje. Tudi če dopustimo tovrstno obrambo, bom v svojem argumentu skušal pokazati, da smo še zmeraj soočeni z enakim osnovnim problemom, da je stopnja prepričanja vedno lahko zgolj posamična (in potemtakem natančna), kar pa ne izključuje delnih prepričanj (t.j. prepričanj s stopnjo manjšo od 1) višjega reda.

Za razumevanje, zakaj so lahko stopnje prepričanj zgolj posamične, pogledjmo še, (za)kaj so (predhodne) stopnje prepričanj sploh pomembne. Konek (2013) v razpravi o problemu predhodnih vrednosti (t.j. predhodnih verjetnosti oz. stopenj prepričanj) navede sledeče najvplivnejše poglede:

1. Informacijski pogled: ključno je, da stopnje prepričanj točno odražajo informacijsko vsebino akterkinih podatkov.
2. Subjektivistični pogled: ključno je, da stopnje prepričanj točno odražajo akterkina mnenja o prepričljivosti hipotez (oz. splošneje, propozicij).
3. Praktični pogled: ključno je, da stopnje prepričanj privedejo do smiselnih strategij odločanja v negotovosti.
4. Instrumentalni pogled: ključno je, da nas stopnje prepričanj privedejo do točnih, minimalno odvisnih od sreče, naknadnih stopenj prepričanj, do katerih pridemo preko posodabljanja na podlagi novih podatkov.

Praktični pogled je problematičen, ker najbolj smiselna strategija odločanja »zavisi od epistemske perspektive, s katere jo presoјamo« (Konek, 2013: 9). Večina topogledih sprejema bodisi informacijski ali pa subjektivistični pogled, pri čemer

izpostavimo, da oba temeljita na točnosti, ki je v konfliktu z informativnostjo; manj kot so stopnje prepričanj natančne, manj so informativne, a bolj točne. Instrumentalni pogled (Konekovo stališče) se zdi med navedenimi najbolj obetaven, saj temelji na ideji, da so stopnje prepričanj teoretični konstrukt in da so kot take uporabne predvsem za opredelitev razmerja med predhodnimi in naknadnimi stopnjami prepričanj. Problematično še vedno ostaja dejstvo, da tudi tu stopnje prepričanj temeljijo na (aletični) točnosti, kar je sicer zadosten, ne pa nujen pogoj za racionalnost. Nujen pogoj je, nasprotno, koherentnost.

Težava s točnostjo je, da mnogi raziskovalci, ki so jo postavljali za temelj, implicitno »izpeljejo deduktivno konsistentnost kot normo koherentnosti polnih prepričanj« (Easwaran in Fitelson, 2015). Koherenca, na drugi strani, pa je šibkejši princip kot konsistentnost. Stopnje prepričanj morajo še zmeraj spoštovati zakone verjetnosti, vendar pa popolna deduktivna konsistentnost ni več nujna.

V zvezi z razliko med konsistentostjo in koherentnostjo velja izpostaviti paradoks predgovora. Orišemo ga lahko na sledeč način:

Predstavljal si, da si napisal knjigo. Vsako stran si večkrat preveril, da bi se izognil napakam. Kljub temu se ti zdi vseeno zelo možno, da je vsaj kakšna napaka kljub skrbnemu pregledu ostala. V predgovoru zato zapišeš, da se zavedaš, da so v tekstu ostale določene napake, vendar nenamerno. Tiskarski skratje, kot jim tudi rečemo.

V tem primeru torej lahko govorimo o paradoksu zaradi dveh nekonsistentnih trditev, ki ju tvoji podatki hkrati podpirajo:

1. Tvoji podatki (natančno pregledovanje) kažejo, da v knjigi ni več nobenih napak.
2. Tvoji podatki (izkušnje) kažejo, da smo ljudje zmotljivi, in da je zato v tekstu zagotovo ostala vsaj ena napaka, če ne več.

Do paradoksa pride zaradi zahteve po konsistentnosti, standardnem temelju točnosti. Ena izmed možnih rešitev je, da se vzdržimo obeh trditev, vendar pa Easwaran in Fitelson (2015) prepričljivo pokažeta, da za racionalne akterje to ni prava rešitev. Namesto, da se vzdrži obeh trditev, bi racionalna akterka – tako Easwaran in Fitelson – morala opustiti stremljenje k prepričanju, ki temeljijo na t.i. dominanci točnosti. Povsem racionalno je, da verjamemo tako v (1) kot v (2), čeprav je to netočno (in nekonsistentno). Do tega pride, ker je točnost prestroga omejitev popolnoma racionalnih akterjev; racionalna akterka ne le, da lahko, temveč celo mora poseči po šibkejših principih. Poenostavljeno rečeno (pri tem se odmikamo od Easwaranovega in Fitelsonovega bolj formalnega argumenta, v katerem se ukvarjata s polnimi prepričanji): akterka bi morala opustiti princip dominance točnosti in ga nadomestiti s šibkejšim principom, ki temelji na koherenci stopenj prepričanj (glede na podatke). Tako pridemo do principa koherentnosti (PK):

PK: Množica stopenj prepričanj je koherentna, če nobene druge množice stopenj prepričanj podatki ne podpirajo bolje, ne glede na to, v kateri situaciji smo.

Prav ta napetost med konsistentnostjo (točnostjo) in koherentnostjo (najboljša podpora podatkov) je pomembna za razumevanje našega argumenta v bran natančnim stopnjam prepričanj. PK kaže, da so zahteve po popolni točnosti premočne in da lahko o racionalnosti govorimo tudi v primeru šibkejših principov (kot je PK), ki jim lahko povsem zadostimo z natančnimi prepričanji. Nenatančne stopnje prepričanj so s PK sicer združljive, vendar pa ne dodajo ničesar bistvenega. Še več: popolnoma racionalna akterka se mora nenatančnim stopnjam prepričanj izogibati, saj brez prave motivacije nižajo informativno vrednost natančnih stopenj prepričanj. To nas privede do novega pogleda na naravo stopenj prepričanj, t.i. instrumentalni pogled, ki temelji na koherentnosti (I-K P).

I-K P: Vlogo stopenj prepričanj utemeljimo glede na razmerje med predhodnimi in naknadnimi stopnjami prepričanj glede na posodabljanje. Vloga predhodnih prepričanj je, da nas privedejo do koherentnih naknadnih prepričanj, ki so najbolje podprta s strani podatkov, preko posodabljanja na novih podatkih.

Ta vloga stopenj prepričanj je racionalno izpolnjena zgolj, če so stopnje prepričanj natančne.

Primer meduzarja

V nekaterih primerih se zdi, da so nenatančne stopnje prepričanj edini možen odziv, saj so podatki v teh situacijah še posebej nenavadne narave. Čeprav gre v primeru nenatančnih stopenj prepričanj dejansko za delna prepričanja druge (oz. višje) stopnje, bomo predpostavili topogledno stališče, da gre za prepričanja prvega reda in pokazali, kako nas tudi v teh primerih, torej na njihovem »domačem terenu«, ne privedejo do zaželenih posledic.

En izmed tovrstnih nenavadnih miselnih eksperimentov je primer meduzarja, ki ga je vpeljal Adam Elga:

Na ulici do tebe pristopi neznanec. Iz svoje torbe začne zlagati različne predmete. Prvi trije predmeti, ki jih izvleče, so tuba zobne paste (običajne velikosti), živa meduza in tuba zobne paste (v potovalni velikosti). Do katerih stopnje bi morala biti prepričana, da bo naslednji predmet, ki ga bo postavil iz torbe, še ena tuba zobne paste? (Elga, 2010: 1)

Tipičen odziv je, da v tem primeru ne moremo določiti neke točne stopnje. Situacija je očitno bizarna in zelo nenavadna. Nikakršnih izkušenj nimamo z neznanci, ki se po mestu sprehajajo z meduzo v torbi, torej bi bila kakršnakoli natančna stopnja prepričanja (npr. 0,2, 0,94 ali 0,53) povsem arbitrarna. Edina možna rešitev z natančno stopnjo prepričanja bi lahko bila 0,5, če uporabimo princip maksi-

malne entropije med dvema možnostma: naslednji predmet bo zobna pasta in naslednji predmet ne bo zobna pasta. Vendar je tak odgovor nezadosten – ni razloga, da se odločamo samo med tema dvema možnostma, niti da sta ti dve možnosti enakovredni.

Nesmiselnost takega odločanja se nam pokaže v analogiji z igralno kocko: kakšna je verjetnost, da pri poštenu kocki pade šestica? Če bi ravnali kot v zgornjem predlogu, bi trdili, da je verjetnost $1/2$ – šestica bo padla ali pa ne bo. Očitno je, da dejansko verjetnost padca šestice ni $1/2$, temveč $1/6$. Čeprav se v primeru meduzarja različni izidi ne kažejo tako jasno, ni nobenega razloga, da bi predpostavili samo dva možna enakovredna izida (in tudi če bi to storili, bi lahko zapadli v svojevrstne paradokse; gl. van Fraassen).

Težava torej ni v tem, da ne bi imeli dovolj računskih sposobnosti – popolnoma racionalna akterka ima neomejene računske sposobnosti –, temveč v tem, da zaradi narave podatkov (unikatna in bizarna situacija) ne moremo določiti, po katerem principu naj se ravnamo. Kot izpostavlja Elga (2010: 2), je primer vsekakor umeten, vendar pa si lahko zamislimo tudi bolj realistične scenarije, kjer naletimo na podobne težave. Denimo; kakšna naj bo naša stopnja prepričanja, »da bo cena bakra in raven obresti čez štirideset let višja, ali pa da takrat ne bo nihče več uporabljal ene izmed trenutno popularnih novotarij, ali pa o položaju privatnih lastnikov v družbenem sistemu?« (prav tam).

Zdi se, da tovrstni scenariji, ki so vsekakor pomembni, kličejo po topogledih odgovorih: če so naši podatki bistveno nenatančni, morajo biti nenatančne tudi naše stopnje prepričanj in zdi se, da je to povsem racionalno.

Kot nasprotnik te trditve predlagam, da si pogledamo, zakaj in kje topogledneži skrenejo v napačno smer. Predpostavimo, da je moja stopnja prepričanja, da bo meduzarjev naslednji predmet še ena zobna pasta, opredeljena z intervalom $[3/10, 9/10]$, torej nenatančna. Taka stopnja prepričanja je konsistentna z nenavadnimi podatki, hkrati pa zelo široka. Topogledneži celo trdijo, da je v primerih popolne nejasnosti smiselno stopnje prepričanja razširiti preko celotnega intervala $[0, 1]$. A težava s tovrstnimi nenatančnimi stopnjami prepričanj je, da niso skladne z I-K P principom, kot sem ga navedel zgoraj.

Kako naj tovrstna intervalska predhodna stopnja prepričanja privede v koherentno naknadno stopnjo, ki je najbolje podprta z novimi podatki, v primeru, ko bi stopnjo prepričanja posodobili? Tega preprosto ne zmore. Predpostavimo še, da izvem neke nove podatke; v časopisu npr. preberem, da se po mestu sprehaja gospod, ki v torbi nosi živo meduzo, dve tubi zobne paste in par nogavic. Moja naknadna stopnja prepričanja (še zmeraj intervalska), da bo meduzar izvlekel še eno tubo zobne paste, bi se znižala na $[0, 3/10]$. Ker je meduzar vsekakor nepredvidljiv in ker novice niso povsem zanesljive, torej še vedno ostanem pri intervalu. Vendar ta naknadni interval ni najbolje podprt s podatki (ena izmed zahtev I-K P principa): podatki sami ne zahtevajo široke stopnje prepričanja – omogočajo množico

natančnih verjetnosti, ne pa posamične več-vrednostne stopnje prepričanja. Zdi se celo, da bi bila več-vrednostna stopnja prepričanja (naenkrat) upravičena zgolj v kvantnofizičnih primerih.³

Trdim torej, da mora povsem racionalna spoznavna akterka ob srečanju z meduzarjem dopustiti več možnih stopenj prepričanj (ker so podatki zelo nejasni), vendar pa mora na koncu iz te množice možnosti izbrati posamezno natančno stopnjo verjetnosti, saj podatki ne podpirajo več-vrednostnega odziva. Kaj bi to glede predmeta, ki ga bo meduzar izvlekel, sploh pomenilo? Obstaja sicer nejasnost glede točne posamične natančne stopnje, vendar pa je povsem neracionalno dopustiti množico stopenj v isti stopnji prepričanja, saj tega primer ne zahteva. Tudi če bi to dopustili, I-K P princip v primeru posodabljanja potem odpove. To bo še bolj jasno v našem naslednjem primeru.

Ellsbergov paradoks

Naslednji primer, v katerem se zdi, da si lahko bistveno pomagamo z nenatančnimi stopnjami prepričanj, je sicer nastal preden se je razvila debata med topogledneži in ostrogledimi. Običajno ga imenujemo Ellsbergov paradoks po Danielu Ellsbergu, ki je primer uvedel (Ellsberg, 1961). Paradoks lahko predstavimo na sledeč način:

Pred tabo je večja neprosojna posoda. Znano ti je, da je v posodi 90 žogic treh barv. Ena tretjina (30) žogic je črnih, ostali dve tretjini (60) žogic pa sta bodisi rdeče ali zelene barve. Najprej izbiraš med dvema stavama z nagrado 10 evrov:

1.A) Izvlečena bo črna žogica.

ali

1.B) Izvlečena bo rdeča žogica.

Zatem pa izbiraš še med dvema stavama z enako nagrado:

2.A) Izvlečena bo bodisi črna ali zelena žogica.

ali

2.B) Izvlečena bo bodisi rdeča ali zelena žogica.

O paradoksu govorimo, ker večina sodelujočih izbere odgovora 1.A in 2.B, ki pa sta racionalno gledano nekonsistentna. Če najprej izberemo odgovor 1.A, to pomeni, da smo ocenili, da je v posodi več črnih žogic kot rdečih. To implicira, da je rdečih žogic glede na oceno 30 ali manj, saj je črnih natančno 30 ($\check{C}=30 \geq R$);

³ Kakšna naj bo stopnja prepričanja racionalne akterke, da je Schrödingerjeva mačka živa, preden pogleda v škatlo? V tem primeru se več-vrednostna stopnja prepričanja kaže kot smiselna.

predpostavimo, da je rdečih po naši oceni 29. Ker vemo, da je zelenih ali rdečih žogic 60, to hkrati tudi implicira, da je zelenih žogic vsaj 30. V našem primeru, kjer smo predpostavili oceno, da je rdečih žogic 29, bi to pomenilo, da bi moralo biti po tej oceni zelenih natančno 31. Ko na tej podlagi izbiramo med odgovoroma 2.A in 2.B, vidimo, da možnosti 2.A zadosti vsaj 60 žogic: 30 črnih in vsaj 30 zelenih (v našem primeru skupaj 61), možnosti 2.B pa natančno 60.

Če torej v prvi izbiri izberemo 1.A, bi bilo racionalno, da v drugi izbiri izberemo 2.A. Podobno velja, če najprej izberemo 1.B (in torej ocenimo, da je v posodi več kot 30 rdečih žogic). V tem primeru bi bilo racionalno izbrati tudi 2.B.

Paradoks nastane, ker – kot že omenjeno – večina sodelujočih izbere dva nekonsistentna odgovora 1.A in 2.B. Psihologi in ekonomisti to običajno razložijo z razliko med knightovsko (ne)gotovostjo in (ne)jasnostjo. Možnosti 1.A in 2.B sta sicer negotovi, vendar lahko ocenimo tveganje: 1.A ima natančno 30 od 90 možnosti za uspeh, 2.B pa natanko 60 od 90. Ostali dve možnosti sta nejasni: 1.B ima od 0 do 60 od skupno 90 možnosti, 2.A pa od 30 do 90. Psihološko razlago na stran, trdim, da nam tudi v tem primeru nenatančne stopnje prepričanj ne morejo pomagati, temveč nas kvečjemu vodijo v neracionalno vedenje.

Poglejmo, zakaj nenatančne stopnje prepričanj iz normativnega vidika v primerjavi z natančnimi stopnjami odpovejo. Na prvi pogled se namreč zdi, da Ellsbergov paradoks kar kliče po topi obravnavi. Racionalna stopnja prepričanja glede možnosti 1.A bi tako bila $1/3$, 1.B v intervalu $[0, 2/3]$, 2.A v intervalu $[1/3, 1]$ in 2.B natanko $2/3$.

Čeprav so tovrstne nenatančne stopnje prepričanj (1.B in 2.A) točne, pa nam v konkretnem izbiranju ne podajo racionalnih osnov za odločanje. Kako naj se popolnoma racionalna akterka potem sploh odloči med 1.A in 1.B? Ne more se, kajti na racionalen način možnosti med sabo ne more primerjati, saj je ena možnost natančna, druga pa ne. Prava para izbor (t.j. 1.A in 2.A oz. 1.B in 2.B) bi sicer sledila, če bi lahko opravila prvo izbiro, vendar tega na racionalen način ne more storiti. Če pa bi se naša popolnoma racionalna akterka vzdržala odločitve, bi neodvisno od žreba žogice v vsakem primeru ostala brez denarne nagrade, kar je spet neracionalno. To tudi kaže, kaj se zgodi v primeru kršitev I-K P principa. Če smo v primeru meduzarja govorili o tem, kaj podpirajo podatki, lahko v primeru Ellsbergovega paradoksa govorimo o tem, da nenatančne stopnje prepričanj lahko vodijo do neracionalnega vzdržanja odločitve; posodabljanje iz predhodnih v naknadne stopnje pa se tako sploh ne zgodi, kar je očitna kršitev I-K P.

Če uporabimo natančne stopnje prepričanj, problem izgine. Poglejmo torej eno izmed mnogih možnosti natančnih prepričanj naše racionalne akterke, ki jih podatki podpirajo.

1.A: $1/3$, 1.B: $1/6$; 2.A: $5/6$, 2.B: $2/3$.

Racionalna izbira je v tem primeru jasna: 1.A in 2.A. Da v primeru natančnih stopenj vedno pridemo do koherentnih parov odločitev, lahko tudi formalno dokažemo, vendar to za naše trenutne potrebe niti ni potrebno. Pomembneje je, da izpostavimo, da nenatančne verjetnosti niso povsem nesmiselne, le napačno uporabljene: intervali, ki smo jih navedli, predstavljajo množice možnih natančnih prepričanj. Vendar pa moramo za posodabljanje in primerjanje stopenj prepričanj, ostati pri posameznih stopnjah, ki jih odberemo iz teh množic oz. intervalov. Nenatančnih stopenj prepričanj torej ne bi smeli razumeti kot stopenj prepričanj, temveč kot podatkovno podprte možne natančne stopnje prepričanj, torej nekakšen nabor možnih odzivov.

Trnuljčičin problem

Daniel Jeremy Singer je pred kratkim objavil članek z zgovornim naslovom »Trnuljčica bi morala biti nenatančna« (»Sleeping Beauty Should Be Imprecise«, 2014). Ne sprašuje se, ali bi morala Trnuljčica prestopiti v tabor topogledih, da bi prišla hitreje do prinčevega poljuba, temveč se sprašuje o problemu dveh različnih odzivov na Trnuljčičin problem, kot ju zagovarjata Adam Elga (2000) in David Lewis (2001), ter kopica njunih sledilcev. Oba nasprotujoča si odgovora sta namreč prepričljiva, s čimer Trnuljčičin problem ustvarja vsaj videz paradoksa.⁴

Trnuljčica sodeluje v posebnem spinalnem eksperimentu. Raziskovalci ji razložijo potek eksperimenta: v nedeljo jo bodo spravili v umetno spanje. Med eksperimentom jo bodo enkrat ali dvakrat zbudili, ji postavili vnaprej določeno vprašanje in jo zopet uspavali s posebno tableto, ki povzroči, da se svoje budnosti med eksperimentom kasneje ne bo spomnila. Povedo ji tudi, da se bodo o številu bujenj (eno ali dve) med eksperimentom odločili na podlagi meta poštenega kovanca: če pade grb, jo bodo zbudili enkrat (v ponedeljek); če pade cifra, pa jo bodo zbudili dvakrat (v ponedeljek in torek). V vsakem primeru bo zbujena v sredo, ko se bo eksperiment zaključil in ji dodatnih vprašanj ne bodo več postavljali.

Ko se bo Trnuljčica zbudila, torej ne bo vedela kateri dan je oz. ali so jo pred tem že zbudili. Vprašanje, ki ji ga bodo zastavili pa bo vedno enako: »Do katere stopnje ste prepričani, da je kovanec padel na grb?«

⁴ Ne gre za tipičen paradoks, saj zagovorniki ene izmed običajnih rešitev problema, t.j. bodisi $1/2$ oz. $1/3$, nimajo hkrati konfliktnih odzivov na problem. Videz paradoksa pa se ustvari, ko problem pogledamo »od daleč«, saj sta obe rešitvi intuitivno zelo prepričljivi, čeprav neskladni.

Ponujata se dva različna odgovora:

- 1) $1/2$ – ta odgovor zagovarja David Lewis (2001). Poenostavljeno – Trnuljčica ve, da je kovanec pošten, zato je edina možnost zanjo, da je v padec grba prepričana s stopnjo $1/2$.
- 2) $1/3$ – ta odgovor zagovarja Adam Elga. Predstavlja si serijo ponovljenih metov kovancev. Glede na met kovanca se pokažejo tri možnosti:
 - a) Pade grb – Trnuljčica se zbudi v ponedeljek.
 - b) Pade cifra – zbudi se v ponedeljek.
 - c) Pade cifra – zbudi se v torek.

Vse tri možnosti so enakovredne, ker so odvisne od meta poštenega kovanca. Trnuljčičina stopnja prepričanja, da je ob metu poštenega kovanca padel grb, bi torej morala biti $1/3$, da je padla cifra pa $2/3$.

Čeprav je bilo po odmevnem članku Elga in Lewisovem odzivu predlaganih kar nekaj možnih rešitev in utemeljitev, nikomur ne uspe združiti obeh različnih intuicij, ki se skrivata v ozadju problema. Singer (2014) pa trdi, da lahko oba odgovora spravimo na skupno izhodišče, če se poslužimo nenatančnih stopenj prepričanj, oz. če bi taka prepričanja pripisali Trnuljčici. Njegovo podrobno preučevanje ga tako privede do trditve, da bi Trnuljčičine stopnje prepričanj morali ponazoriti na množici verjetnostnih funkcij, ki za $x \in [0, 1/2]$ pripišejo sledeče stopnje prepričanj:

	ponedeljek	torek
grb	x	0
cifra	x	1-2x

Tabela 1: Singerjeva analiza Trnuljčičinega problema

Singerjev nenatančen pristop tako pokrije oba tradicionalna odziva ($1/2$ in $1/3$), hkrati pa upošteva tudi možnost posodabljanja prepričanj. Vseeno odpove glede na I-K P, saj v podatkih ni ničesar takega, kar bi od Trnuljčice zahtevalo večvrednostno (naenkrat) stopnjo prepričanja v primerjavi z natančno stopnjo. Ali to pomeni, da so njegova odkritja nepomembna? Nikakor.

Pri Ellsbergovem paradoksu smo nabor možnosti (sicer narobe razumljen kot nenatančne stopnje verjetnosti) enostavno določili, v primeru Trnuljčičinega problema pa je ta nabor bistveno težje določiti. Singerjeva analiza je torej uspešna pri določitvi nabora možnih natančnih stopenj prepričanj za Trnuljčico, ki njena pre-

pričanja glede meta kovanca ohranijo koherentna glede na podatke.⁵ Njegova edina napaka je, da množico možnosti zamenjuje z nenatančnimi stopnjami prepričanj in iz tega sklepa, da bi Trnuljčica morala imeti nenatančno stopnjo prepričanj. Njegov odziv na Elgov problem bi torej bil: ne moremo natančno odgovoriti, kakšno stopnjo prepričanja bi morala Trnuljčica imeti, lahko pa njeno stopnjo prepričanj opredelimo z nekim naborom stopenj.

Pravilen odziv pa bi, nasprotno, bil, da imamo množico možnih natančnih stopenj, ki jih Singerjeva analiza zameji z dodatno omejitvijo glede njihovega razmerja, vendar pa lahko Trnuljčica izbere zgolj eno natančno stopnjo, bodisi $1/2$, $1/3$ ali pa drugo stopnjo iz njegove analize. Singerjevi intervali so torej uporabni, če jih, namesto kot nenatančne stopnje verjetnosti, razumemo kot množico možnih natančnih verjetnosti. To pa tudi ponuja dodaten razlog, zakaj stopnje verjetnosti ne morejo biti nenatančne: kar razumemo kot nenatančne verjetnosti so kvečjemu množice možnosti, med katerimi lahko izberemo prave natančne stopnje. Z drugimi besedami, gre za natančna delna prepričanja drugega reda, t.i. (delna) prepričanja o (delnih) prepričanjih: Trnuljčica ima lahko tako-in-tako natančno stopnjo prepričanja (drugega reda), njeno prepričanje prvega reda pa je lahko zgolj natančno in posamično.

Sklep

Čeprav bi lahko ponudili še več primerov, kako nenatančne stopnje prepričanj vodijo v katastrofalne posledice, ti trije posebni primeri zadoščajo za mojo obrambo natančnih stopenj prepričanj v primerjavi z nenatančnimi. Do zdaj je namreč že jasno, da so to, kar topogledneži imenujejo nenatančne stopnje prepričanj, kvečjemu množice možnih natančnih stopenj, ki se ponujajo v primeru dvoumnih, nasprotujočih si, pomanjkljivih ali pa nejasnih podatkov.

Če bi dopuščali obstoj nenatančnih stopenj prepričanj, bi to pomenilo, da so natančne stopnje zgolj poseben primer – podobno kot se polna prepričanja izkažejo za zgolj posebne primere delnih prepričanj, ki presežejo neko mejo verjetnosti. Vendar pa so, kot sta pokazala Hájek in Eriksson (2007), (natančne) stopnje prepričanj primaren pojem, ki se ga ne da dalje analizirati z drugimi pojmi. To sicer ne dokazuje neobstoja nenatančnih stopenj prepričanj, gre pa z roko v roki z mojo obrambo natančnih stopenj.

⁵ Poraja se tudi vprašanje, ali je kriterij koherentnosti Trnuljčičinih prepričanj zadosten za racionalnost njenega odziva na problem. Medtem ko imata stopnji prepričanja v padec grba $1/3$ in $1/2$ dodatno podporo (prva glede na možne scenarije, druga pa glede na poštenost kovanca), si težko zamislimo, zakaj bi bila Trnuljčica npr. prepričana, da je padel grb s stopnjo $1/8$, v scenariju cifra&ponedeljek $1/8$ in v scenariju cifra&torek pa $3/4$. Ad hoc rešitev bi bila, da bi namesto intervala $x \in [0, 1/2]$ uporabili množico z zgolj dvema vrednostma, t.j. $x \in \{1/3, 1/2\}$. Vprašanje na tem mestu puščamo odprto.

Jasno je tudi, zakaj so nenatančne stopnje prepričanj na prvi pogled tako prepričljive: zdi se, da dobro ponazarjajo naša dejanska delna prepričanja. Človeški akterji smo pogosto soočeni z mehкими podatki, iz česar intuitivno sklepamo, da bi taka (mehka, topa oz. nenatančna) morala biti tudi naša prepričanja. Glavna težava je v tem, da v tem primeru izhajamo iz napačnega razumevanja stopenj prepričanj, ki lahko v istem primeru racionalno zavzamejo različne stopnje, vendar pa ne več stopenj naenkrat, kot to sledi iz topogledih stališč. Nenatančni podatki torej omogočajo več racionalnih izbir, vendar pa je neracionalno (in nemogoče, razen v primeru določenih scenarijev kvantne fizike), da bi se v istem primeru hkrati odločili za več različnih vrednosti naenkrat. Gre torej za mešanje delnih prepričanj kot takih in delnih prepričanj o delnih prepričanjih. Dasiravno ima akterka na voljo več delnih prepričanj in je zaradi narave podatkov v nejasnosti glede pravega oz. vsaj dejanskega prepričanja, kar se odraža v nenatančnih stopnjah prepričanj oz., pravilneje rečeno, natančnih stopnjah prepričanj višjega reda, je njena stopnja prepričanja o neki propoziciji lahko zgolj posamična in natančno zamejena.

To stališče je nekoliko sorodno epistemicistom v razpravi o *nejasnih* predikatih (npr. Williamson, 1996) ⁶, kar niti ne čudi, saj so bile nenatančne stopnje prepričanj v razpravo vpeljane prav v povezavi z *nejasnimi*, dvournimi, nasprotujočimi si ali pomanjkljivimi podatki o propoziciji, o kateri akterke zavzamejo neko stopnjo prepričanja. Epistemicisti trdijo, da obstaja neka natančna meja, s katero je nejasen predikat zamejen, vendar pa nikoli ne moremo vedeti, kakšna je ta meja. Naše stališče glede nenatančnih stopenj prepričanj je podobno do te mere, da v primerih, ki kličejo po nenatančni obravnavi (kot so zgornji trije), obstaja neka natančna stopnja prepričanja, ki jo akterka zavzame, vendar ni gotova, katera točno ta stopnja je, kar se odraža v višjem redu, kjer je akterka le delno prepričana o posamezni natančni stopnji. A naše stališče se od epistemicistovega tudi pomembno razlikuje: podatki v teh primerih dopuščajo množico možnih stopenj prepričanj in ne zgolj ene, ki je za dani primer prava (kot bi moralo veljati v analogiji z epistemicistom). Pravih stopenj prepričanj je v takih primerih torej več, vendar pa se lahko udejanji zgolj ena, ki ji racionalna akterka potem tudi sledi in s tem zavrže vse preostale možnosti iz množice.

⁶ Williamson nenatančnost (ang. *inexactness*) v tem delu omeni tudi v povezavi s (polnimi) prepričanji, vendar pa je njegov poudarek drugačen kot naš, saj se nanaša na nenatančnost znanja o nejasnih predikatih.

Do težav pride torej predvsem zaradi pojmovne zmede, saj se teoretični vlogi stopenj prepričanj ne posvečamo dovolj, interpretacija, ki ji sledi posamezen raziskovalec, pa je pogosto nerazdelana ali naivna. Zavedati se moramo, da so stopnje prepričanj (ang. *credences*) teoretičen pojem in kot take igrajo vlogo v navezavi na naše epistemološke principe, ob analizi pa se izkažejo tudi kot primarni pojmi, ki se jih ne da dalje analizirati. Opozoriti moramo, da kljub temu ni nobenega razloga, da stopenj prepričanj ne bi mogli umestiti v dejanske miselne procese običajnih ljudi (Staffel, 2013), vendar bi moja teoretična izvajanja tudi v tem primeru še vedno veljala, saj tudi v dejanskem mišljenju ne moremo imeti več stopenj prepričanj o isti propoziciji naenkrat.

Obrambo natančnih stopenj prepričanj sem večidel izvajal preko I-K P principa, vendar pa imajo topogledneži pred sabo tudi druge težave, ki jih v tem prispevku nisem uspel nasloviti, saj niso vezane na ta princip. Ena izmed takih težav izhaja že iz same narave nenatančnih verjetnosti, ki predstavljajo izhodišče za nenatančne stopnje prepričanj. Vodiijo namreč do verjetnostnih razširitev (*probabilistic dilation*), kar v poenostavljenem jeziku pomeni, da lahko na podlagi irelevantnih podatkov bolj natančna stopnja prepričanja v neko propozicijo preide v manj natančno oceno (za topogledni protiprimer glej npr. Bradley in Steele, 2014).

Da zaključimo bolj pozitivno: nenatančne verjetnosti kot take moja izvajanja ne pokopljejo. Koncept se namreč s pridom uporablja npr. v statistiki, kvantni fiziki in bioinformatiki. Kljub temu pa trdim, da zanje ni prostora v analizi prepričanj, kjer za nenatančne stopnje prepričanj z zelo redkimi izjemami enostavno ni prostora.

On Imprecise Credences: the Jellyfish Guy Case, the Ellsberg Paradox and the Sleeping Beauty Problem

The most common way of representing the strength of credences is by ascribing probabilities to them. Sharpers claim that credences can only be represented by precise probabilities. Non-sharppers, on the other hand, claim that credences may also be imprecise and should be represented by imprecise probabilities when the evidence is essentially imprecise (ambiguous, vague, conflicting or scarce). However, imprecise credences lead to irrational behaviour and do not make much sense when placed under scrutiny. I provide a coherence-based principle which enables me to defend the position of sharpeners. This principle is then applied to three special cases, which are *prima facie* best explained by imprecise credences, i.e. the jellyfish guy case, the Ellsberg paradox and the Sleeping Beauty problem.

Keywords: credences, imprecise probability, subjective probability, uncertainty, epistemology

Literatura

- Bradley, S. in Steele, K. S. (2014). »Uncertainty, learning, and the 'Problem' of dilation«. *Erkenntnis*, 79(6), str. 1287–1303.
- Destercke, S. in Dubois, D. (2006). »A unified view of some representations of imprecise probabilities«. *Soft Methods for Integrated Uncertainty Modelling*, 37, str. 249–257.
- Easwaran, K. in Fitelson, B. (2015). »Accuracy, Coherence and Evidence«. V Szabó Gendler, T. in Hawthorne, J. (ur.), *Oxford Studies in Epistemology*, vol. V.
- Elga, A. (2000). »Self-locating belief and the Sleeping Beauty problem«. *Analysis*, 60(2), str. 143–147.
- Elga, A. (2010). »Subjective probabilities should be sharp«. *Philosophers' Imprint*, 10(5), str. 1–11.
- Ellsberg, D. (1961). »Risk, Ambiguity, and the Savage Axioms«. *Quarterly Journal of Economic*, 75(4), str. 643–669.
- Eriksson, L. in Hájek, A. (2007). »What are degrees of belief?«. *Studia Logica*, 86(2), str. 183–213.
- Gigerenzer, G. in Hoffrage, U. (1995). »How to improve Bayesian reasoning without instruction: frequency formats«. *Psychological Review*, 102(4), str. 684–704.
- Hájek, A. (2008). »Arguments for- or against-probabilism?«. *British Journal for the Philosophy of Science*, 59(4), str. 793–814.

- Hartmann, S. in Sprenger, J. (2010). »Bayesian Epistemology«. V Bernecker, S. in Pritchard, D. (ur.), *Routledge Companion to Epistemology*, London: Routledge, str. 609–620.
- Jeffrey, R. (1983). »Bayesianism with a human face«. V Earman, J. (ur.), *Testing scientific theories*, Minneapolis: University of Minnesota Press, str. 133–156.
- Joyce, J. M. (2005). »How probabilities reflect evidence«. *Philosophical Perspectives*, 19(1), str. 153–178.
- Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan and Co.
- Konek, J. P. (2013). *New Foundations for Imprecise Bayesianism*. Doktorska disertacija, University of Michigan.
- Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge, Mass.: The MIT Press.
- Lewis, D. (2001). »Sleeping Beauty: Reply to Elga«. *Analysis*, 61(3), str. 171–176.
- Singer, D. J. (2014). »Sleeping Beauty should be imprecise«. *Synthese*, 191(14), str. 3159–3172.
- Staffel, J. (2013). »Can there be reasoning with degrees of belief?« *Synthese*, 190(16), str. 3535–3551.
- Sturgeon, S. (2008). »Reason and the grain of belief.« *Noûs*, 42(1), str. 139–165.
- van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.
- Williamson, T. (1996). *Vagueness*. London and New York: Routledge.
- Yaniv, I. in Foster, D. P. (1997). »Precision and accuracy of judgmental estimation«. *Journal of Behavioral Decision Making*, 10 (1), str. 21–32.

Toma Strle

Pedagoška fakulteta Univerze v Ljubljani

Odločanje: od laboratorija do resničnosti vsakdanjega življenja

V članku¹ se osredotočam na kritiko prevladujočih sodobnih empiričnih pristopov k raziskovanju odločanja. Zagovarjam tezo, da le-ti podajajo pomanjkljiv in v nekaterih vidikih napačen pogled na odločanje zaradi naslednjih razlogov: (a) Odločanje pojmujejo deloma napačno, saj premalo poudarjajo subjektivne vidike odločanja in v osnovi odločanja zares ne koncipirajo kot ciljno usmerjenega delovanja. (b) Osredotočajo se na raziskovanje specifičnih in nepomembnih odločitvenih situacij in problemov, ki odločevalcem ne prinašajo kakšnih pomembnih posledic (če sploh) in tako nimajo prav veliko skupnega z odločitvami vsakdanjega življenja. (c) Odločanje raziskujejo skoraj izključno s tretjeosebne vidika, prvoosebne pa večinoma izpuščajo. To ima za posledico, da subjektivni značaj odločanja, ki je bistven in odločanju intrinzičen, ter to, kako odločitvene situacije razumevajo/interpretirajo odločevalci, v najboljšem primeru razumejo zelo slabo. S preučevanjem specifičnih odločitvenih problemov s specifične tretjeosebne perspektive sodobni pristopi k raziskovanju odločanja skratka podajajo pomanjkljivo in do neke mere izkrivljeno podobo odločanja in tako preučevani pojav potencialno reducirajo na nekaj, kar ni.

Ključne besede: odločanje, odločitvene situacije vsakdanjega življenja, prvoosebna perspektiva, subjektivni vidiki odločanja

1. Uvod

O odločanju govorimo tako v kontekstu živih kot neživih sistemov, o odločanju višje razvitih organizmov, kot smo ljudje ali primati (Tomasello et al., 2005), ali nižje razvitih organizmov, kot so npr. čebele (Seeley in Buhrman, 1999) ali celo *C. elegans* (Bendesky et al., 2011). Govorimo tako o odločanju posameznikov kot skupinskega odločanja (npr. Nutt in Wilson, 2010). Nadalje govorimo o odločanju

¹ Članek temelji na prvem poglavju moje doktorske disertacije *Metakognicija kot povezovalac mišljenja in intuicije v procesu odločanja* (Strle, 2015).

v različnih odločitvenih situacijah in o različnih odločitvenih problemih² ter o različnih tipih odločitev, kot so na primer etične ali moralne, potrošniške, ekspertne, politične, družbene idr. Razlikujejo se tudi načini, kako odločanje poteka: od bolj avtomatičnega, nezavednega³ ali intuitivnega odločanja, do bolj razumskega, preudarnega ali celo odločanja z logično analizo. Odločanje seveda vključuje tudi različne povezane kognitivne⁴, čustvene in volicijske procese, kot tudi metakognitivne procese, če se med odločanjem nanašamo tudi na lastne mentalna stanja. Ne nazadnje se razlikujejo tudi sami odločevalci, tako kot osebe, kot v tem, kako razumejo oz. interpretirajo odločitven problem ali situacijo.

Ne glede na mnogotere domene, v katerih govorimo o odločanju, raznovrstne tipe odločitev in različne načine, na katere poteka, je odločanje v splošnem koncipirano kot proces izbiranja med alternativnimi možnostmi oz. poteki delovanja, ki odločevalcu prinaša ugodne izide oz. izide, ki so skladni z njegovimi cilji (glej npr. Baron, 2007; Hardman, 2009; Hastie, 2001; Symmonds in Dolan, 2012). Čeprav nekateri raziskovalci odločanje razumejo širše, drugi ožje, odločanje v splošnem pojmujejo kot proces izbiranja, ki se začne z alternativnimi možnostmi oz. možnimi poteki delovanja in zaključí z izbiro določene alternative in/ali izvedbo nekega dejanja (realizacija odločitve). Odločanje v tem smislu vključuje predvsem primerjanje alternativ in njihovih lastnosti, predvidevanje in presojanje o dogodkih in stanjih zunanjega sveta, zajema pa tudi presojanje in predvidevanje subjektivnih posledic odločitev. Takšna opredelitev in razumevanje odločanja je bolj ali manj splošno sprejeto na različnih področjih kognitivne znanosti, ki se ukvarjajo z raziskovanjem odločanja: od ekonomije in psihologije, do nevroznanosti in računalništva ter njihovih interdisciplinarnih izpeljank, kot je npr. nevroekonomija.

² V članku uporabljam izraz »odločitven problem« predvsem v kontekstu odločanja v laboratoriju in izraz »odločitvena situacija« v kontekstu odločanja v vsakdanjem življenju. Razlog je v tem, da se mnoge odločitvene situacije s tretjeosebne vidika v laboratoriju zdijo kot nekakšen problem, naloga, ki zahteva rešitev (tj. končno izbiro), s prvoosebne vidika odločevalca pa je to odvisno predvsem od njegovega razumevanja/interpretacije »odločitvene« situacije. Možno je npr., da odločevalec »odločitvene« situacije, ki se s tretjeosebne vidika zdi takšna, da zahteva neko rešitev (t.j. končno izbiro), sploh ne razume kot odločitvene ali kot takšne, ki bi zahtevala končno odločitev (rešitev problema).

³ Ne smemo enačiti avtomatičnih in nezavednih odločitev. Sicer se avtomatične odločitve lahko odvijajo nezavedno, vendar temu ni nujno tako. V primeru odločitev, ki smo jih storili že velikokrat, je proces odločanja lahko avtomatičen, tj. takšen, ki ne zahteva kakšnega zavestnega preudarka, vendar to ne pomeni, da se procesa vsaj deloma ne zavedamo. Obratno drži, da se odločitve, s katerimi smo soočeni prvič in ki zahtevajo zavesten preudarek, vsaj deloma odvijajo nezavedno.

⁴ Čeprav se je pojem kognicije tradicionalno nanašal na t.i. višje spoznavne procese, kot so npr. mišljenje, reševanje problemov, odločanje ipd. (glej npr. Gardner, 1985), pa danes ugotavljamo, da so »višji« spoznavni in čustveni procesi pravzaprav neločljivo povezani in tako težko govorimo o njihovem strogem ločevanju (glej npr. Damasio, 1994; Pessoa, 2008; de Sousa, 1987).

V članku se bom osredotočil na kritiko prevladujočih sodobnih empiričnih pristopov k raziskovanju odločanja, predvsem v kontekstu zgoraj omenjenih disciplinarnih področjih, saj skupaj predstavljajo večji del raziskovalnega področja odločanja⁵. Zagovarjal bom tezo, da sodobni empirični pristopi podajajo pomanjkljiv in v nekaterih vidikih napačen pogled na odločanje – tako v razumevanju pojava kot v raziskovalni praksi.

V začetnem poglavju bom analiziral splošno sprejeto opredelitev odločanja in pokazal na nekatere pomanjkljivosti in omejitve takšnega razumevanja odločanja. Trdil bom, da bi morali odločanje razumeti bolj v smislu ciljno usmerjenega delovanja, ki se odvija že pred obstojem alternativ, na primer kot proces identifikacije in/ali kreiranja za odločevalca relevantnih alternativ. Prav tako bom pod vprašaj postavil predpostavko mnogih pogledov na odločanje, ki eksplicitno ali implicitno trdijo, da je odločanje zaključen proces, ki se zaključi s končno izbiro ali dejanjem. V analizi bom poskusil tudi pokazati, da je koncepcija odločanja kot procesa, v katerem se odločevalci večinoma nanašajo na objektivna stanja zunanjega sveta, ne pa toliko na lastna subjektivna stanja, napačna. Nasprotno bom trdil, da so odločitvene situacije (in odločitve) za odločevalca bistveno in neločljivo vpete v to, kaj zanj pomenijo in subjektivno prinašajo.

V drugem poglavju bom primerjal nekatere laboratorijske odločitvene probleme, ki so v takšni ali drugačni obliki mnogokrat uporabljeni za raziskovanje odločanja, z nekaterimi namišljenimi odločitvenimi situacijami vsakdanjega življenja. Opisal bom nekatere ključne razlike med njimi, ki kažejo, da iz spoznanj o prvih ne moremo sklepati na druge ali celo na odločanje kot celoto. V laboratoriju raziskovalci recimo večinoma preučujejo dobro definirane odločitvene probleme, ki vsebujejo malo negotovosti in če jih, te niso posebej relevantne za odločevalce, odločitve, ki so za odločevalca nepomembne itd. Nasprotno pa lahko odločitvene situacije vsakdanjega življenja vključujejo veliko in drugačne vrste negotovosti, so za odločevalce zelo pomembne in vsebujejo subjektivne posledice, ki za odločevalca tvorijo bistvo odločitvene situacije in odločitve. Vsebujejo subjektivne posledice, brez katerih odločitvena situacija za odločevalca pravzaprav ne bi imela smisla in pomena.

⁵ V članku se bom omejil predvsem na individualne, osebne odločitve ljudi, saj obravnava »socialnih kontekstov« odločanja presega obseg članka. V omejitvi na individualne odločitve pa nočem trditi, da so te iztrgane iz socialnega in kulturnega konteksta. Nasprotno, individualne odločitve so v večji ali manjši meri vezane na socialno okolje, iz katerega izhajamo in v katerem delujemo, ter so s tem neposredno vezane na druge odločevalce. Prav tako se ne bom neposredno ukvarjal z vprašanjem, ali bi morali odločanje razumeti in posledično raziskovati v okviru klasične kognitivne znanosti (Thagard, 2005) ali v okviru utelešene (npr. DeCaro et al., 2009; Oullier in Basso, 2010) in/ali udejanjenje kognicije (ang. *enactive cognition*). Bom pa v članku zagovarjal trditev, da je prvooseben vidik odločanja bistven za razumevanje odločanja, v čemer je moj pogled blizu nekaterim pristopom udejanjenje kognicije (npr. Varela, Thompson in Rosch, 1991).

V zadnjem poglavju članka bom nadaljeval z razpravo o problematiki preučevanja odločanja zgolj s tretjeosebne perspektive. Na primeru dveh kvalitativnih študij (Klein, Calderwood in Clinton-Cirocco, 2010; van Manen, 2014) bom pokazal, da preučevanje odločanja s prvoosebne perspektive podaja zelo drugačno podobo odločanja kot preučevanje enostavnih in nepomembnih odločitev v laboratoriju. Izpuščanje prvoosebne perspektive iz preučevanja odločanja vodi tudi v pereč problem, da raziskovalci pravzaprav ne vedo, če gre v laboratoriju – če si odločitvene probleme zamislimo v vsakdanjem življenju – sploh za odločanje. Tretjeosebne laboratorijske študije nam namreč zelo malo povedo o tem, kako odločevalci razumejo/interpretirajo odločitvene situacije, proces odločanja in same odločitve.

2. Nekatere pomanjkljivosti splošnega razumevanja odločanja: subjektivni vidiki, alternative in realizacija odločitev

Hastie (2001) v začetku pregleda raziskovalnih problemov in problematičnih vidikov raziskovalnega področja presojanja in odločanja podaja splošno opredelitev odločanja, kot ga po njegovem v splošnem razume takratno področje. Pravi, da se raziskovalno področje presojanja in odločanja osredotoča na raziskovanje tega, »kako ljudje (in ostali organizmi) združujejo želje (koristi, osebne vrednote, cilje, namene ...) in prepričanja (pričakovanja, znanja, sredstva ...), da izberejo določen potek delovanja« (Hastie, 2001: 655–656).⁶ Takšna opredelitev raziskovalnega fokusa področja da misliti, da obstoj alternativ pravzaprav ni začetna točka odločanja, ampak so to želje in prepričanja odločevalcev. Takšna opredelitev nas prav tako napeljuje na misel, da proces odločanja ni toliko proces preudarjanja o alternativnih možnostih in ostalih objektivnih stanjih sveta, ampak primarno oz. vsaj bistveno vključuje preudarjanje o lastnih ciljih, osebnih vrednotah, namerah, subjektivnih posledicah, kot so subjektivne »koristi«, potencialnih občutjih ipd. Vendar se izkaže, da Hastie (2001), ki opisuje takrat splošno sprejet pogled na odločanje (ki drži še danes), v resnici odločanja ne opisuje kot takšnega »ciljno usmerjenega« delovanja, katerega bistven del tvorijo subjektivna stanja in dogodki.

Na to kaže Hastiejeva (2001) nadaljnja podrobnejša opredelitev odločitev: »Pojmovna (morda definirajoča) šablona odločitve vključuje tri komponente: (a) poteke delovanja (možne izbire in alternative); (b) prepričanja o objektivnih stanjih, procesih in dogodkih v svetu (vključno z izidi in sredstvi za njihovo doseganje); in (c) želje, vrednote ter koristi, ki opisujejo posledice, povezane z izidi vsake kombinacije dejanje-dogodek. Dobre odločitve so torej tiste, ki učinkovito izbira-

⁶ Sam ne sledim Hastiejevi (2001) razdelitvi na želje in prepričanja, saj se mi razdelitev pripadajočih pojmov v dve kategoriji ne zdi ustrezna.

jo sredstva, ki so na voljo v danih okoliščinah, za doseganje odločevalčevih ciljev« (Hastie, 2001: 656). Hastie (2001) odločitve definira kot kombinacije situacije in vedenja, ki vključujejo tri glavne, zgoraj omenjene komponente: alternativne možnosti oz. poteke delovanja, prepričanja o negotovih dogodkih v svetu, katerih del so tudi izidi, in želje, vrednosti (vrednote⁷) ter koristi, ki opisujejo subjektivne posledice, povezane z izidi. Izide opredeli kot »javno opisljive situacije, ki se dogodijo na koncu vsake poti v odločitvenem drevesu« (Hastie, 2001: 657). Posledice kot »subjektivne evalvacijske reakcije (ki jih merimo na skali dobro-slabo, dobitok-izguba), povezane z vsakim izidom« (ibid.). Negotovost (dogodkov) pa je po njegovem mnenju povezana s »presojanjem odločevalca o tendenci dogoditve vsakega pogojnega dogodka« (ibid.), ki je po navadi opisana kot mera »verjetnosti, prepričanosti ali možnosti« (ibid.). Odločanje se tako po njegovem nanaša na celoten proces izbiranja poteka delovanja, vključujoč presojanje, ki se nanaša na dele procesa odločanja, ki so povezani z »ocenjevanjem, predvidevanjem in sklepanjem o tem, kateri dogodki se bodo zgodili in kakšne bodo evalvacijske reakcije odločevalca na te izide« (ibid.). Glede na začetno opredelitev raziskovalnega fokusa področja, prepričanja sledijo iz presojanja (ocenjevanje, predvidevanje, sklepanje) o negotovih dogodkih v svetu, želje, vrednosti in koristi pa opisujejo subjektivne posledice kot reakcije odločevalcev na izide odločitev.

Če podrobneje pogledamo zgoraj povzeto opredelitev, ki relativno dobro ustreza splošno sprejetemu razumevanju odločanja, ugotovimo, da je odločanje koncipirano kot proces, ki se začne z vnaprej danimi alternativami, o katerih se odločevalci odločajo predvsem na podlagi ocenjevanja, predvidevanja in sklepanja o dogodkih zunanjega sveta (vključujoč objektivne izide). O subjektivnih vidikih odločanja – kot je npr. preudarjanje o subjektivnih posledicah, lastnih ciljih, vrednotah, željah ipd. – ni kaj dosti govora. Pogled, ki po mojem mnenju ne odraža »resničnosti« odločanja v vsakdanjem življenju.

V opredelitvi komponent odločitev Hastie (2001) sicer govori o »ocenjevanju, predvidevanju in sklepanju o tem, kakšne bodo evalvacijske reakcije odločevalca na izide«, se pravi o presojanju o potencialnih prihodnjih subjektivnih posledicah odločitev, vendar je to edino mesto, kjer lahko zaznamo omembo nanašanja na in upoštevanja lastnih subjektivnih stanj pri odločitvah. Pogled na odločanje, ki je po mojem mnenju močno različen od (vsaj nekaterih) odločitev v vsakdanjem življenju, kjer so lastna subjektivna stanja in njihovo poznavanje za odločevalce bistven element odločanja, brez katerega odločitve za odločevalce ne bi imele ne smisla, ne pomena. Hastie (2001) nadalje poudarja predvsem pogled na subjektivne posledice kot evalvacijske reakcije na izide in ne odločevalčevega preudarjanja in občutenja le-teh. Prav tako se eksplicitno omejuje na prepričanja o nego-

⁷ V izvirniku: *values*. Ni jasno, ali ima Hastie (2001) v mislih vrednosti ali vrednote. Najverjetneje vrednosti, saj se veliko nanaša na klasične normativne teorije in iz njih izhajajoče deskriptivne pristope k raziskovanju odločanja, ki o (osebnih) vrednotah odločevalcev ne govorijo prav dosti.

tovih dogodkih v zunanjem svetu in ne govori o negotovih dogodkih notranjega sveta in prepričanjih o njih. Izide pojmuje kot javno opisljiva stanja in s tem implicitno nakazuje, da so izidi predvsem, če ne popolnoma, zunanja in ne notranja, subjektivna stanja, s čimer se je težko strinjati. Lahko si zamislimo situacijo, kjer se odločimo, da na neko situacijo ne bomo reagirali tako čustveno, ali situacijo, kjer se odločimo, da bomo poiskali učinkovito tehniko za umirjanje lastnih misli (glej naslednje poglavje), pri čemer je bistven izid lastno notranje stanje (naše misli so npr. bolj ali manj umirjene) in ne neko zunanje stanje. Izid, ki sicer je javno opisljiv – o njem lahko, lahko pa tudi ne, poročamo –, je bistveno subjektivno in ne objektivno stanje. Nadalje Hastie (2001), čeprav v začetku odločanje opredeli kot združevanje želja in prepričanj, o tem, kako odločevalci razumejo lastne odločitve z vidika lastnih želja, prepričanj in ciljev, ter o tem, kako je preudarjanje in občutenje le-teh vključeno v proces odločanja, praktično ne govori. Ne nazadnje ničesar ne pove o tem, kako odločevalci razumejo odločitveno situacijo v kontekstu lastnih ciljev, želja, prepričanj, subjektivnih stanj, preteklih izkušenj ipd. Podoba odločanja, kot ga po Hastieju (2001) razume takratno raziskovalno področje, je v tem smislu omejena in deloma napačna, saj izpušča oz. premalo poudarja nekatere bistvene vidike odločanja (kot bo natančneje razvidno iz primerov naslednjih poglavij).

Poleg premajhnega upoštevanja subjektivnih vidikov odločanja, za Hastieja (2001) – kot mnoge druge raziskovalce – obstoj alternativnih možnosti predstavlja nujen pogoj za to, da lahko govorimo o odločanju. Vendar menim, da lahko o odločanju v nekaterih primerih govorimo, še preden odločevalec pozna oz. zanj obstajajo relevantne alternative. Odločanje se v tem smislu lahko začne z željo, ciljem ali nujo, odločevalec pa mora zanj relevantne alternative, preden lahko med njimi izbira in realizira svojo odločitev, najprej identificirati ali ustvariti. Sam proces določanja relevantnih alternativ lahko v tem smislu koncipiramo kot proces odločanja med tem, katere alternative so in katere niso relevantne za odločevalca. Sam odločanje vsaj v nekaterih kontekstih razumem bolj kot ciljno usmerjeno delovanje, ki šele na neki točki privede do alternativnih možnosti ali možnih potekov delovanja, med katerimi se odločamo.

Ne nazadnje večina raziskovalcev odločanja (tako tudi Hastie) implicitno ali eksplicitno predpostavlja, da se odločanje zaključi z izbiro določene alternative in/ali izvedbo nekega dejanja (realizacija odločitve). Vendar ni nujno, da se proces odločanja zaključi z odločitvijo in/ali realizacijo odločitve. Lahko si zamislimo situacijo, kjer se venomer znova odločamo, da bomo prenehali kaditi, vendar (a) odločitve nikoli ne sklenemo ali pa (b) se odločimo, da bomo prenehali kaditi, vendar odločitve ne realiziramo (tj. ne prenehamo kaditi). Z vidika zunanjega opazovalca situacije, kjer do realizacije odločitve ne pride, običajno ne bi označili kot odločanja, saj niti ne vemo, da je oseba šla skozi kakršenkoli proces odločanja, če opazujemo le vedenje. Po drugi strani se zdi, da bi sam odločevalec navedena primera označil za odločanje, čeprav do odločitve (a) ali realizacije (b) nikoli ni

prišlo. Sam se z razločki med procesom odločanja, odločitvijo in realizacijo odločitve ne bom ukvarjal, vendar le-ti predstavljajo pomemben vidik procesa odločanja, ki ga večina teorij in empiričnih raziskav odločanja izpušča. Razločki recimo vodijo v zanimiva vprašanja, kot so: Kdaj lahko rečemo, da se je nekdo za nekaj odločil? Kako, če sploh, se razlikuje odločanje (in odločitve), ki vodi v realizacijo, in odločanje (odločitve), ki ne? Mar razločki ne kažejo na to, da opazovanje le z zunanje perspektive opazljivih vidikov odločanja pomanjkljivo zajema to, čemur pravimo odločanje? Nekaterih vidikov vprašanj se dotakne v zadnjem poglavju opisana raziskava van Manena (2014), ki kaže, da lahko o odločanju – vsaj s prvoosebne perspektive odločevalcev – govorimo tudi v primerih kjer se odločevalci ne morejo odločiti in/ali realizirati svojih odločitev.

Če nadalje primerjamo odločitvene probleme, s katerimi raziskovalci soočajo odločevalce v laboratoriju⁸ in nekatere odločitvene situacije vsakdanjega življenja, ugotovimo, da je vsaj v nekaterih kontekstih vrzel med laboratorijem in vsakdanjim življenjem pravzaprav zelo široka. Pomislek, ki ga lepo izrazi Galottijeva:

Iz različnih razlogov se mi ne zdi samoumevno, da bi se vsi rezultati, odkriti v dobro nadzorovanih eksperimentih, pojavili tudi v realnem življenju. Vzemimo na primer dejstvo, da odločitve v realnem življenju nikakor niso tako dovršene – vse »dane informacije«, ki jih uporabimo, so le redko, če sploh, lično natisnjene in predane odločevalcu, kot se to po navadi zgodi v tipičnih raziskavah. Velik delež odločanja v realnem življenju je najverjetneje sestavljen prav iz iskanja in pregledovanja tovrstnih informacij. Drugič, odločitve v realnem življenju so za nas velikokrat zelo pomembne in ko jih sprejemamo, uporabljamo tako srce in intuicijo kot tudi glavo in smo motivirani, da jih napravimo pravilno. Tega pa ne moremo vedno trditi za probleme akademskega tipa, ki so predstavljeni udeležencem raziskav, ki morajo praviloma reševati hipotetične scenarije, nepovezane z njihovimi vsakdanjimi izkušnjami... V laboratorijskih študijah odločanja je po navadi pravilen le en odgovor; v realnem življenju pa mnogo pomembnih odločitev najverjetneje nima popolnega izida, temveč je možnih več različnih, ki se razlikujejo po svoji končni zadovoljivosti. V laboratorijskih študijah lahko obstajajo formule, ki se jih da aplicirati na problem in ki bi odločevalca vodile do izbire »pravnega« odgovora; le redke odločitve v realnem življenju posedujejo tako jasno strukturo (Galotti, 2002: ix–x).

3. Od odločitvenih problemov v laboratoriju do odločitvenih situacij v vsakdanjem življenju

V nadaljevanju bom predstavil nekaj primerov laboratorijskih odločitvenih nalog in nekaj namišljenih odločitvenih situacij v vsakdanjem življenju ter pokazal, da

⁸ Na te se nanaša Hastie (2001), ko opisuje težave raziskovalnega področja presojanja in odločanja.

se slednje od prvih razlikujejo v nekaterih bistvenih vidikih, ki pomembno določajo to, kako se odločamo. S tem želim prikazati, da sodobni empirični pristopi k raziskovanju odločanja izpuščajo mnogotero vidikov odločanja in tako podajajo omejeno, če že ne napačne predstave o preučevanem pojavu.

Sledijo štirje primeri laboratorijskih odločitvenih nalog (primeri 1-3, 5) in primer potrošniške odločitve v naravnem okolju (primer 4). Štiri laboratorijske odločitvene naloge predstavljajo klasične primere odločitvenih nalog, ki so v takšni ali drugačni obliki velikokrat uporabljene za preučevanje odločanja v vedenjski ekonomiji, psihologiji in/ali nevroznanosti. Vsi primeri razen prvega predstavljajo ekonomsko ali potrošniško odločanje, ki je eden od najbolj preučevanih tipov odločanja. Prvi primer bi lahko potencialno šteli za odločanje pod pogojem gotovosti⁹, vendar pri njem težko govorimo o izidih ali posledicah. V tem primeru je pravzaprav vprašljivo, če sploh gre za odločanje. Drugi primer je klasičen primer odločanja pod pogojem tveganja. Tretji primer predstavlja klasičen primer odločanja pod pogojem negotovosti. Četrti primer predstavlja odločanje v naravnem okolju pod pogojem negotovosti, če predpostavimo, da končna da/ne odločitev za nakup vsebuje negotove izide in posledice. Petega primera po mojem mnenju ne moremo postaviti v nobeno od klasičnih vrst odločanja, saj ni jasno, ali ima izbira kakršnekoli izide mimo same izbire. Če predpostavimo, da izbira vsebuje negotovost o subjektivnih posledicah, potem bi ta primer načeloma lahko šteli za odločanje pod pogojem negotovosti.

Zaznavno odločanje: odločanje pod pogojem gotovosti. V fMRju gledamo v ekran, na katerem se nam prikazuje kaotično polje majhnih pik, ki se skupno gibljejo proti levi ali desni. Nekatere pike se gibljejo proti levi, druge proti desni, »skupni seštevek« gibanja pa vedno kaže v eno ali drugo smer tako, da to zelo težko zavestno opazimo (če sploh). Večkrat se moramo odločiti, ali skupno gibanje polja točk kaže v eno ali nasprotno smer, npr. proti levi ali desni (naloge diskriminacije smeri gibanja). Odločitev nakažemo s pritiskom na levi ali desni gumb. V drugi različici gledamo na ekran in se moramo večkrat odločiti, ali je prikazana nejasna slika, hiša ali obraz¹⁰.

⁹ Klasično se odločanje deli na odločanje pod pogojem gotovosti, tveganja in negotovosti. O odločanju pod pogojem gotovosti govorimo takrat, ko ima vsaka alternativa oz. potek delovanja le en možen, gotov izid. O odločanju pod pogojem tveganja govorimo takrat, ko so v odločitveni situaciji alternative in izidi poznani, alternative imajo več možnih izidov, odločevalec pa pozna verjetnosti dogoditve posameznih izidov. V odločanju pod pogojem negotovosti pa informacij o tem, kako verjetno ali možno je, da se bo zgodil nek dogodek, povezan z odločitvijo (npr. izid), nimamo, vendar možne izide poznamo (opredelitive po Bermúdez, 2009). O odločanju pod pogojem negotovosti lahko govorimo tudi takrat, ko možnih izidov ne poznamo.

¹⁰ Take ali podobne vrste odločitvenih nalog se uporabljajo predvsem v kontekstu preučevanja t.i. zaznavnega odločanja (ang. *perceptual decision making*) in raziskovanja nevrološke podlage tega, kako organizmi zbirajo in združujejo senzorne informacije in kako to vpliva na njihovo vedenje oz. odločitve (iz Heekeren, Marrett in Ungerleider, 2008).

Denarne stave: odločanje pod pogojem tveganja. V fMRju gledamo v ekran, na katerem se nam najprej prikaže začetna vsota denarja (»Prejmete 50 evrov«). Raziskovalci nas obvestijo, da celotne vsote ne bomo mogli obdržati. Nato izbiramo med dvema alternativama. Če izberemo alternativo A, z gotovostjo *dobimo* 20 evrov, če izberemo tvegano alternativo B, pa z verjetnostjo 0.4 dobimo vseh 50 evrov, a z verjetnostjo 0.6 izgubimo vseh 50 evrov. V skoraj identični odločitveni situaciji spet začasno dobimo 50 navideznih evrov in moramo izbrati med alternativo A, kjer z gotovostjo *izgubimo* 30 evrov, in tvegano alternativo B, kjer z verjetnostjo 0.4 dobimo vseh 50 evrov, a z verjetnostjo 0.6 izgubimo vseh 50 evrov. Nalogo ponovimo večkrat.¹¹

Denarne stave: odločanje pod pogojem negotovosti. V laboratoriju sedimo pred štirimi kupi kart. Raziskovalci nam dajo hipotetično posojilo 2000\$ v obliki kopij bankovcev in nam rečejo, da naj iz kupov karte dvigujemo tako, da bomo izgubili čim manj in dobili čim več denarja (hkrati nam merijo prevodnost kože). Ne da bi vedeli, dvig nekaterih kart v štirih kupih prinaša dobitke (100\$ v kupu A in B in 50\$ v kupu C in D), dvig drugih izgube (velike iz kupov A in B in nizke iz kupov C in D). Zaporedoma dvignemo sto kart. Če dvigujemo karte predvsem iz kupov A in B, bomo skupno na izgubi, če iz kupov C in D, bomo skupno imeli dobiček (to moramo šele ugotoviti skozi ponavljajoče dvigovanje kart).¹²

Izbiranje med potrošniškimi produkti v naravnem okolju: odločanje pod pogojem negotovosti. Pridemo v veleblagovnico s hrano, ki ponuja še posebej veliko število različnih produktov iste vrste. Ko se približamo preizkuševalni stojnici nas eksperimentator, oblečen kot uslužbenec trgovine, povabi, da pokusimo marmelade znamke *Wilkin and Sons*. V enem pogoju je na stojnici šest različnih marmelad iste znamke, drugič štiriindvajset. Prav tako dobimo kupon s popustom enega dolarja za nakup katerekoli marmelade te znamke (ne glede na to ali je bila marmelada na stojnici ali ne), ki ga lahko unovčimo na blagajni.¹³

Potrošniško izbiranje med avtomobili. Sedimo v laboratoriju in naša naloga je izbrati najboljšega med štirimi avtomobili. Najprej preberemo opise vseh štirih av-

¹¹ V eksperimentu so De Martino et al. (2006) uporabili modificirano odločitveno nalogo Kahnemana in Tverskyja (1979) za raziskovanje nevrološke podlage učinka uokvirjanja, vloge čustvenega sistema pri posredovanju pristranosti in individualnih razlik v dovzetnosti za uokvirjanje.

¹² Odločitveno nalogo Iowa hazardersko igro so uporabili Bechara et al. (1997), ki so na podlagi primerjave odločitev zdravih udeležencev in udeležencev s poškodbo ventromedialnega prefrontalnega režnja pokazali, da so nezavedni čustveni procesi (somatski označevalci) bistveni za ugodne odločitve (tj. odločitve, ki prinašajo dobitke), še preden se zavedamo, izbira katere alternative prinaša dobitke in katera izgube. Eksperiment je bil deležen relativno močne kritike, da so Bechara et al. (1997) uporabili slabo metodo merjenja zavedanja tega, kateri kupi prinašajo dobitke oz. izgube (glej npr. Maia in McClelland, 2004).

¹³ To in podobne odločitvene situacije sta v naravnem okolju raziskovala Iyengar in Lepper (2000; opisan je eksperiment 1), ki sta želela ugotoviti, kakšna je korelacija med številom ponujenih alternativ, pogostostjo nakupa in zadovoljstvom z izbiro.

tomobilov, ki imajo v enem pogoju štiri, v drugem pa dvanajst različnih lastnosti (gre za dve ločeni skupini udeležencev). Potem, ko preberemo opise avtomobilov, raziskovalec eni skupini udeležencev naroči, da pred končno izbiro avtomobila o izbiri razmišljajo štiri minute, v drugem pa štiri minute rešujejo anagrame. Po preteklem času nas raziskovalec prosi, da izberemo najboljši/najljubši avtomobil (raziskovalci niso razlikovali med navodilom izbire najboljšega ali najljubšega avtomobila).¹⁴

Zdaj si oglejmo nekaj hipotetičnih odločitvenih situacij v vsakdanjem življenju, ki se od prej opisanih laboratorijskih v različnih aspektih močno razlikujejo.

Izbira okusa sladoleda. S prijateljico se odpravim na sladoled. Kateri sladoled naj izberem? Naj izberem jagodni, malinov ali čokoladni sladoled? V mislih si predstavljam okuse, lastni odziv in razmišljam, da sem ta teden pojedel že preveč sladkega.

Odločanje za študij. Vpisan sem v četrti letnik gimnazije in okolje name že nekaj časa pritiska, da se odločim za študij. Čeprav se o študiju ne želim odločati, pa vem, da bom neki študij vsekakor vpisal. Po pregledu možnih študijev na Univerzi v Ljubljani, pogovorih s starši in prijatelji ugotovim, da sta zame najprimernejša študija filozofije in psihologije. Kateri študij naj izberem? Po dolgem razmisleku se odločim, da izbiro prepustim naključju in vržem kovanec, saj nikakor ne morem ugotoviti, kateri študij je zame najbolj primeren, kateri najbolj ustreza mojim željam in ciljem itd.

Odločanje za karierno pot. Zaključujem študij ekonomije in razmišljam o karierni poti. Vem, da imam žilico za podjetniške zadeve, a imam s podjetništvom zelo malo izkušenj, obenem pa čutim, da me je strah negotovosti, ki jo prinaša samostojnost podjetnika. Naj grem raje v službo ali vseeno odprem podjetje in preizkusim svoje sposobnosti na tekmovalnem trgu? Če podjetje, kakšno podjetje naj sploh odprem?

Odločanje o tem, kdaj imeti otroka. S partnerico se zadnje čase veliko pogovarjava o otrocih. Ona si otroka želi že kar nekaj časa, meni pa občutek pravi, da še nisem pripravljen na starševstvo. Po drugi strani razumsko vem, da se počasi bliža čas, ko bova dovolj stara, in vem, da si nekoč sam želim otrok. Naj ugodim želji partnerice ali pa odločitev o tem, kdaj imeti otroka, odlagam še nekaj časa?

Odločanje o tem, kako umiriti misli. Že nekaj časa čutim, da velikokrat nisem dovolj osredotočen na to, kar počnem. Opazujem, kako mi včasih misli begajo, kako

¹⁴ Odločitveno nalogo so uporabili Dijksterhuis et al. (2006), ki so želeli pokazati, da v kontekstu kompleksnih izbir nezavedno mišljenje vodi v boljše izbire kot zavestno preudarjanje. Kompleksne odločitve so raziskovalci definirali kot izbiranje med avtomobili z dvanajstimi lastnostmi, preproste kot izbiranje med avtomobili s štirimi lastnostmi. Raziskave je bila deležna številnih kritik (npr. Mamede et al., 2010; Nieuwenstein et al., 2015; Strle, 2013), vendar je močno vplivala na nadaljnje predstave o »moči« nezavednega odločanja.

preskakujejo z ene vsebine na drugo. Včasih se mi zazdi, da tok misli in povezani občutki niso več podobni valu oceana, ampak razburkanem morju. Kaj naj storim, da bodo moje misli spet osredotočene? Naj spremenim zunanje okoliščine ali naj poskusim spremeniti kaj v lastni duševnosti? Kje naj se sploh lotim problema?

Sledi primerjava opisanih laboratorijskih odločitvenih nalog in hipotetičnih odločitvenih situacij vsakdanjega življenja z naslednjih vidikov: obstoja alternativnih možnosti in poznavanja izidov; stopnje in vrste negotovosti; pomembnosti odločitev za odločevalce; trajanja odločanja; pogostosti odločitev; odločevalčevega razumevanja/interpretacije odločitvene situacije; in na koncu podrobnejša primerjava različnih vidikov izbiranja med denarnimi stavami (De Martino et al., 2006) in odločanja o karierni poti.

Alternativne možnosti in izidi

V podanih primerih laboratorijskih odločitev so alternative dane vnaprej, v mnogih odločitvah vsakdanjega življenja pa odločevalcem niso poznane, odločevalci ne ve, katere so zanj sploh relevantne, in če jih približno pozna, niso dobro definirane. Tako je npr. v primerih odločanja za študij, karierno pot in odločanje o tem, kako umiriti misli. Tudi izidi niso vnaprej opredeljeni, lepo natisnjeni na belem papirju ali prikazani na ekranu, kar močno poveča negotovost odločevalcev. Čeprav recimo v eksperimentu Bechare et al. (1997) ne vemo, iz katerega kupa naj potegnemo karto, da zaslužimo denar, in ne vemo, koliko denarja bomo zaslužili, pa vemo, da obstajata le dva možna negotova izida: da izgubimo ali zaslužimo namišljen denar. V vsakdanjem življenju izbira določene alternative ali poteka delovanja prinaša mnoge izide, ki jih niti ne moremo dobro predvideti (to velja za vse odločitvene situacije, razen morda za izbiranje med okusi sladoledov).

Negotovosti

Negotovosti, če jih opisani laboratorijski odločitveni problemi sploh vsebujejo, so relativno majhne in so predvsem negotovosti o zunanjem svetu. Odločitvena naloga Bechare et al. (1997), ki naj bi po njihovem simulirala odločanje realnega življenja v načinu vključevanja negotovosti, nagrad in kazni, recimo vključuje negotovosti, ki pa za odločevalca nimajo posledic oziroma so te zanj nepomembne. Kot rečeno v prejšnjem poglavju, prav tako vključuje le negotovosti o dveh možnih izidih zunanjega sveta. Lahko bi si sicer zamislili, da izbira najboljšega avtomobila v raziskavi Dijksterhuisa et al. (2006) prinaša veliko zadovoljstvo, vendar menim, da v tako nezavezujočih situacijah subjektivne posledice za odločevalce večinoma niso relevantne. Odločitve v vsakdanjem življenju po drugi strani prinašajo več in drugačne vrste negotovosti. V nekaterih odločitvah vsakdanjega življenja negotovost ni vezana le na nepoznavanje časovno in prostorsko bližnjih izidov in njihovih verjetnosti, ampak tudi na nepoznavanje najrazličnejših bolj dolgoročnih dogodkov, vezanih tako na dogodke v zunanjem svetu, kot na stanja lastne duševnosti. V zgoraj omenjenem primeru odločanja o študiju ni

jasno, ali bomo čez deset let bolj zadovoljni z izbiro, ker na primer ne vemo, kaj za naše življenje (karierno pot, zadovoljstvo, ugled?) v resnici pomeni, če se odločimo za študij psihologije ali za študij filozofije. Prav tako ne vemo, kakšne izide in subjektivne posledice prinaša odločitev, da otrok še ne bomo imeli, kot tudi odločitev, da jih bomo imeli v kratkem. To so le nekateri primeri negotovosti, s katerimi se soočamo v vsakdanjem življenju, ki pa v laboratoriju, zaradi narave odločitvenih problemov, niti niso prisotne.

Pomembnost odločitev

V nasprotju z opisanimi laboratorijskimi odločitvenimi problemi so lahko nekatere odločitve v vsakdanjem življenju veliko bolj pomembne, morda eksistencialno pomembne, saj v nasprotju s prvimi prinašajo izide in subjektivne posledice, ki so lahko potencialno nevarni, (moralno) zavezujoči (glej tudi tretje poglavje članka za empirično podporo) ali takšni, ki spreminjajo velik del našega življenja. To lahko drži za odločanje za študij, karierno pot, otroka ali odločanje o tem, kako osredotočiti lastne misli. Po drugi strani so odločitve v vsakdanjem življenju lahko tudi relativno nepomembne, kot je morda izbira sladoleda. Vendar je pomembnost odločitev bistveno odvisna od samega odločevalca, tj. od tega, kako razume/interpretira odločitveno situacijo, od njegovih želja, ciljev in osebnih vrednot. Izbira imena otroka je na primer lahko za nekoga izjemno težka, dolgotrajna in nadvse pomembna, za drugega morda sploh ne predstavlja odločitvenega problema, saj je bilo osebi vedno jasno, kako bo otroku ime.

Trajanje odločanja

Nadalje, opisano laboratorijsko in potrošniško odločanje večinoma traja zelo kratek čas. Lahko si sicer predstavljamo, da odločanje med marmeladami v četrtem predstavljenem primeru poteka dlje, vendar najverjetneje ne tako dolgo kot npr. razmislek o tem, ali odpreti podjetje ali poiskati varnejšo službo, spremeniti ali nadaljevati študij ipd. Nekatere odločitvene situacije – tako v laboratoriju kot v vsakdanjem življenju – od nas zahtevajo odločanje pod časovnim pritiskom, kot je odločanje med denarnimi stavami (De Martino et al., 2006), druge dopuščajo, da se odločamo dlje (npr. odločitev za študij, otroka). Če se o določeni stvari odločamo dlje, odločitvenega procesa ne moremo prikladno zajeti v neko »točko«, kot to lahko storimo z večino odločitev v laboratoriju. Dalj trajajoče odločanje lahko recimo vključuje posvetovanje z drugimi osebami, preudarjanje o lastnih željah in ciljih (posebej, če odločitev interpretiramo kot pomembno), resnejši razmislek o tem, kakšne subjektivne posledice bo imela odločitev za nas in druge osebe ter razmislek o tem, kako smo sprejemali podobne pretekle odločitve. Vsaj nekatere odločitve v vsakdanjem življenju v nasprotju z opisanimi laboratorijskimi odločitvami skratka puščajo oz. odpirajo več prostora za zavestni razmislek o svetu in lastni duševnosti.

Pogostost odločitev

Odločitve se razlikujejo tudi po tem, kako pogosto jih sprejemamo. Odločitve, ki smo jih opravili že velikokrat, kot je izbira sladoleda, po navadi storimo bolj rutinsko in o njih ne preudarjamo prav veliko, če sploh. O odločitvah, ki so nam neznane in s katerimi smo soočeni prvič ali pa jih opravljamo zelo poredko (kot so npr. odločitve za študij, karierno pot ali otroka), po navadi več preudarjamo, tako z vidika potencialnih izidov kot z vidika subjektivnih posledic. A čeprav smo odločitev storili že velikokrat, to nujno ne pomeni, da o njej ne preudarjamo. V kontekstu odločanja o marmeladah si lahko zamislimo potrošnika, ki močno premleva o tem, katero marmelado naj kupi, čeprav je marmelado kupil že nešteto krat. Ali potrošnika, ki dlje časa premleva o tem, ali naj unovči pridobljeni kupon za nakup marmelade ali naj raje malo privarčuje. Lahko pa si zamislimo drugačnega potrošnika, ki namerno ne nakupuje v tako velikih trgovinah, kot je ta, kjer so raziskovalci napravili svojo študijo, in se nasploh ne ustavlja ob promocijskih stojnicah. Številnih možnosti, ki jih ponujajo trgovine, sploh ne vidi kot alternativ: kot po navadi gre v manjšo, lokalno trgovino in s police vzame marmelado, ki jo je navajen (odločati se mora šele takrat, ko nehajo prodajati marmelado, ki jo vedno kupuje). V takih primerih pravzaprav ne gre za odločanje, ampak le za sledenje navadi. V četrtem opisanem primeru potrošniške odločitve v naravnem okolju raziskovalci take vrste potrošnikov recimo sploh niso zajeli.

Perspektiva odločevalca

Nadalje (kar po mojem mnenju veja za vsakršno odločitev) je bistveno tudi to, kako odločevalec razume/interpretira odločitveno situacijo. Za nekoga je izbira sladoleda lahko zelo lahka, rutinska, ker vedno izbere čokoladni ali jagodni sladoled (in tako mnogokrat sploh ni odločitev), za drugega morda zelo težavna, saj se nikakor ne more zediniti, kateri okus si tisti trenutek želi. Težkost odločitve lahko za nekoga tiči že v tem, da se mora v tistem trenutku sploh odločati. Po drugi strani si lahko predstavljamo odločevalca, ki se »slasti« ob izbiranju med mnogoterimi možnostmi okusov sladoleda in mu proces odločanja prinaša ugodje, ne pa anksioznosti, ki jo morda nekdo drug povezuje z odločanjem. Za koga je lahko odločitev o izbiri sladoleda zelo pomembna in morda prav zaradi tega težka. Take vrste odločitve po navadi sicer ne bi označili kot pomembne, vendar je primer dober ravno zato, ker dopušča, da ga razumemo v nasprotju z našo (avtorjevo) intuicijo. Če se npr. odločimo, da bomo shujšali, in zato sledimo strogi dieti, vendar si – ker imamo sladoled zelo radi – enkrat na mesec vseeno privoščimo sladoled, je za nas odločitev za izbran okus lahko zelo pomembna. Morda pomembnejša in dolgotrajnejša od odločitve o tem, ali imeti otroka ali ne.

Izbiranje med denarnimi stavami in odločanje o karierni poti: podrobnejša primerjava

Če podrobneje primerjamo odločanje o denarnih stavah (De Martino et al., 2006) in odločanje o karierni poti, hitro ugotovimo, da sta odločitvi zelo različni. V prvem primeru ima odločevalec na voljo le dve alternativni, o katerih pa ima načeloma vse informacije in pozna vse možne izide in verjetnosti, ki bodo sledile izbiri določene alternative. Odločitven problem tako ne vključuje nobenih negotovosti o dogodkih zunanjega sveta. Odločitveni problem tudi ne vsebuje relevantnih subjektivnih posledic, o katerih bi moral odločevalec preudarjati. Odločevalec mora, če želi izbirati ugodno, sicer stremeti k izbiri, ki mu prinaša največjo denarno korist, če predpostavimo, da si to tudi želi, vendar mu ne bi nič koristilo, če bi začel preudarjati o lastnih željah, ciljih ali osebnih vrednotah. Odločevalčevo poznavanje lastnih preferenc, potencialnih čustvenih stanj, občutkov, želja, prepričanj ali ciljev za takšno odločitev skratka ni relevantno. In čeprav bi imel odločevalec na voljo neomejeno časa, da bi preračunal, katera alternativa je resnično najboljša, si ga zaradi pomanjkanja motivacije najverjetneje niti ne bi vzel. Težko si predstavljam, da bi neki posebej prekanjen udeleženec sredi eksperimenta dvignil roko in zahteval najnovejšo literaturo o odločanju, da bi lahko prišel do optimalnega odgovora na problem, ki so mu ga zastavili eksperimentatorji. Po drugi strani je odločanje o karierni poti bistveno drugačno od odločanja med denarnimi stavami. Prvič, odločevalec vsaj na začetku alternativ (kakšno podjetje naj odpre) ne pozna dobro in jih mora šele poiskati oz. ustvariti ter ugotoviti, katera so zanj sploh relevantne. Prav tako odločitev vsebuje mnogo negotovosti o dogodkih zunanjega sveta. Odločevalec ne more vedeti, ali bo, če se odloči proti odprtju lastnega podjetja, želeno službo sploh dobil. Ne more vedeti, kako hitro bo napredoval in zaslužil toliko denarja, kot si ga želi, in ali ne bi bilo s tega vidika bolje, da ustanovi lastno podjetje ipd. Prav tako ne ve, ali je v obzir vzel vse možne relevantne izide in če je pravilno ocenil »verjetnosti« dogoditve negotovih dogodkov. Negotovost odločevalca izvira tudi iz tega, da morebiti slabo pozna subjektivne posledice odločitve in ostale negotove dogodke v lastni duševnosti. Le-ti predstavljajo bistven element odločitve, saj brez vsaj delnega poznavanja subjektivnih posledic, lastnih želja, ciljev in prepričanj, odločevalec sploh ne bi mogel ugotoviti, kaj je zanj primerna izbira. Se bom v izbrani službi, če jo dobim, zares počutil tako dobro, kot sem si predstavljal? Bom odločitev, da ustanovim lastno podjetje, morda obžaloval? Kako se bom počutil v službi, če si predstavljam, da tam delam že deset let? Sem dovolj pameten in iznajdljiv, da uspem na tekmovalnem trgu? Je odločitev za odprtje podjetja v skladu z mojimi cilji (kot je na primer več denarja)? Katera je tista prava izbira? V primeru denarne stave je veliko lažje določiti, kaj predstavlja dobro izbiro: najboljša je najverjetneje tista, ki pač prinaša največ denarja. V primeru odločanja o karierni poti – če odločevalec sploh lahko ugotovi, katera je boljša – je to veliko težje. Je boljša odločitev tista, ki vsebuje najboljše razmerje med vloženim trudom in zasluženim denarjem? Morda potencialno zadovoljstvo z določeno karierno potjo? Kaj to sploh

pomeni za odločevalca? Čim manj stresa? Pravo razmerje stresa in zadovoljstva? Je to stabilnost zaposlitve? Morebiti sreča? Umirjenost? Če naj si odločevalec pride na jasno o tem, katera možnost je boljša in katera slabša, mora vsaj nekaj razmisleka nameniti temu, kakšne sploh so njegove želje, sposobnosti in/ali cilji¹⁵.

S podano primerjavo nekaterih odločitev v vsakdanjem življenju in odločitev v laboratoriju sem želel prikazati, da to, kakšne odločitve preučujemo, na katere vidike se osredotočamo in iz katere perspektive jih preučujemo, bistveno določa, kaj o odločanju izvemo. Če se npr. vprašamo, kakšne vrste procesi »vodijo« naše odločitve – so to bolj intuitivni, večinoma nezavedni, bolj preudarni, zavestni ali metakognitivni procesi? –, odgovor ni odvisen le od tega, kako pač deluje naš kognitivni aparat, ampak od same odločitvene situacije in samega odločevalca, ki odločitve sprejema: kako odločitveno situacijo interpretira, jo interpretira kot zase pomembno in je zato motiviran, da o njej bolj zavestno preudarja ter se sprašuje o tem, kaj si res želi ipd. To pomeni, da bi morali za razumevanje odločanja kot celote raziskovalno energijo usmeriti tudi v preučevanje prvoosebnega vidika odločanja v vsakdanjem življenju. In čeprav bi se večina raziskovalcev najverjetneje strinjala s tem, da so odločitvene situacije lahko med seboj zelo različne in da jih različni odločevalci razumevajo na različne načine, pa pozornosti odločitvam v vsakdanjem življenju večinoma ne posvečajo; posebej ne s prvoosebne perspektive.

Seveda je ena izmed možnih raziskovalnih strategij, da se omejimo na specifične odločitve in dobro kontrolirane eksperimentalne pogoje v upanju, da bomo lahko kasneje, ko pridobimo dovolj podatkov, izsledke prenesli na drugačne vrste odločitev v vsakdanjem življenju. Ta strategija je najverjetneje vsaj v specifičnih primerih uspešna, vendar večina laboratorijskih odločitvenih problemov po mojem mnenju nima veliko skupnega z bolj osebnimi in pomembnimi odločitvami vsakdanjega življenja. Posploševanje iz dvomljivih študij potrošniških odločitev na veliko kompleksnejše in drugačne vrste odločitve, kot so na primer družbene ali politične (kar npr. počnejo Dijksterhuis et al., 2006), je potencialno etično sporno. Podobno je vprašljivo, če na podlagi opisanih izsledkov in podobnih študij laboratorijskih odločitev kreiramo družben prostor odločanja vsakdanjih odločevalcev, pa četudi z namenom pomoči pri odločanju¹⁶. Podoba odločanja, ki je močno omejena in še močnejše izkrivljena, če v zakup vzamemo dejstvo, da odločanje, ki je bistveno prvooseben, subjektiven pojav, sodobna empirična znanost o odločanju preučuje skoraj izključno s tretjeosebne perspektive, a se v svojih interpretacijah teh podatkov ne omejuje na tretjeosebne, objektivne vidike odločanja. Osredo-

¹⁵ Takšne vrste odločitev lepo kažejo na pomembno vlogo metakognicije v odločanju.

¹⁶ Glej npr. program odločitvenih spodbud Thalerja in Sunsteina (2008) ali izvršni »ukaz« predsednika Združenih držav Amerike za uporabo izsledkov vedenjskih ved v kreiranju državnih politik in programov (The White House, 2015).

točnost na raziskovanje specifičnih odločitev s specifične tretjeosebne perspektive, ki mu sledi večji del raziskovalnega področja odločanja, tako po mojem mnenju vodi v izkrivljeno, močno pomanjkljivo podobo odločanja, na kar lepo kažeta tudi v naslednjem poglavju predstavljeni kvalitativni študiji.

4. Prvoosebne perspektive odločanja

Implicitna predpostavka mnogih raziskav, ki odločanje preučujejo s tretjeosebne perspektive, je, da sta perspektivi odločevalca in raziskovalca o tem, ali neko dejanje je ali ni odločanje, ter o tem, kako odločanje poteka, med seboj relativno dobro usklajeni. Vendar nekatere kvalitativne študije (Klein et al., 2010; Kordeš, 2009; van Manen, 2014) kažejo, da odločanje v okolju vsakdanjega življenja, če nanj pogledamo s perspektive odločevalcev, v nekaterih primerih poteka zelo drugače, kot se zdi, da poteka v večini laboratorijskih tretjeosebnih raziskav. Študija Kleina et al. (2010) kaže, da odločitve izkušenih odločevalcev na nobeni točki ne vključujejo alternativnih možnosti in odločanje tako nikakor ne poteka kot primerjanje alternativ in preudarjanje o njih. V tej študiji je vprašljivo, ali sploh gre za odločanje. Raziskava van Manena (2014) pa kaže na različne drugačne načine, kako odločevalci doživljajo in sprejemajo pomembne etične odločitve. Izsledki, do katerih bi težko prišli na podlagi tretjeosebnega raziskovanja nekompleksnih odločitvenih problemov v laboratoriju, ki za odločevalca nimajo nikakršnih resnejših posledic.

Klein et al. (2010) so v raziskavi odločitev izkušenih gasilcev v naravnem okolju pod časovnim pritiskom pokazali, da se ti večinoma ne odločajo s primerjanjem in ocenjevanjem alternativnih možnosti oz. možnih potekov delovanja. V več kot 80% t.i. točk odločitve (glej spodaj za razlago) niso poročali o odločanju kot preudarjanju med alternativami, saj z njihovega vidika alternative večinoma sploh niso obstajale. Raziskovalci pravijo: »Na ta način so na točki odločitve enostavno implementirali svoje delovanje, ne da bi sploh razmišljali o kakršnihkoli drugih možnostih. To smo v intervjujih zelo skrbno preučili, PGEji [poveljniki gasilskih enot] pa so bili nadvse jasni, da se v tistih trenutkih niso zavedali drugih možnosti. Zato tudi niso čutili, da so se odločali« (Klein et al., 2010: 194; oglate oklepaje dodal avtor). Zanje je obstajala le ena »prava« alternativa oz. možen potek delovanja, ki pa pravzaprav sploh ni bila alternativa. Namesto izbiranja med alternativami so gasilci situacijo prepoznali kot prototipski primer, s katerim so bili soočeni že velikokrat, čemur je sledilo ustrezno primerno delovanje: »Ključ, da lahko kaj takega sploh deluje, je, da morajo v večini primerov takoj identificirati učinkovito možnost, izkušeni PGEji pa učinkovito možnost identificirajo tako, da najdejo ujemanje trenutne situacije s prototipom in jo tako prepoznajo kot tipično in dovzetno za tipične postopke« (Klein et al., 2010: 201). Klein (1999) je na podla-

gi študij ekspertnega odločanja predlagal model odločanja kot prepoznave,¹⁷ ki je očitno drugačno od klasičnega pojmovanja odločanja kot izbiranja in preudarjanja med alternativami.

Zanimivo je tudi, kako so raziskovalci sploh identificirali t. i. točke »odločitve«: »Da smo zagotovili, da je res šlo za točko odločitve, smo prosili udeležence, naj identificirajo potencialne možnosti. Udeleženci jih po navadi niso znali najti. V teh primerih smo jih povprašali o možnostih, na katere bi morda nespametno pomislil pripravnik. Včasih smo morali možnosti predlagati sami. Če smo našli vsaj dve možnosti, smo to imeli za točko odločitve« (Klein et al., 2010: 194). Raziskovalci po eni strani tako trdijo, da gre za odločanje, po drugi strani pa poročila gasilcev jasno kažejo na to, da se sami – vsaj v smislu zavestnega izbiranja med možnimi alternativami – niso odločali. Vendar vprašanje, ali v takih primerih lahko govorimo o odločanju ali ne, ni le vprašanje definicije odločanja, ampak kaže na osnovni »razkol« med perspektivo raziskovalca, ki povečini poskuša zavzemati tretjeosebno, objektivno perspektivo, in perspektivo odločevalca, ki je primarno prvoosebna, izkustvena. Vprašanje, ali gre pri odločitvah gasilcev za odločanje ali ne, je težavno tudi zaradi tega, ker če navkljub prvoosebnim poročilom gasilcev, njihovo delovanje opredelimo kot odločanje (kot to naredijo avtorji študije), pristanemo na vprašanju, katero delovanje potemtakem ne predstavlja odločanja (hoja?; proces umivanja zob?; nakupovanje produktov, ki jih vedno kupujemo? ...). Raziskava lepo kaže vsaj na to, da preučevanje tretjeosebne perspektive odločanja nikakor ni zadostno. Če bi »podobno« raziskavo izvedli v laboratoriju, kjer bi gasilce soočili s podobnimi kriznimi situacijami gašenja požarov, jim pri vsaki dali na izbiro nekaj alternativnih potekov delovanja in jim naročili, naj izberejo najboljšega, bi najverjetneje mislili, da so se gasilci med temi alternativami pač odločali, morda o njih preudarjali, ipd. Zaključek, ki bi bil zelo drugačen od tega, kar kaže raziskava Kleina et al. (2010).

Kvalitativna raziskava van Manena (2014) nadalje lepo prikazuje različne načine, kako odločevalci, ki niso eksperti, doživljajo subjektivno zelo pomembne odločitvene situacije v vsakdanjem življenju. Van Manen (2014) je preučeval izkustvo¹⁸ etičnega (ne)odločanja staršev nedonošenčkov ali novorojenčkov, ki so ob rojstvu potrebovali intenzivno nego v bolnišnici. Diagnoze otrok, ko so jih sprejeli v bolnišnico, so bile ekstremna nedonošenost, kompleksne prirojene napake, nevrološke poškodbe ipd. Starši, ki so sodelovali v raziskavi (štirinajst staršev), so se morali odločati o prekinitvi terapij življenjske podpore (npr. mehanske ventilacije,

¹⁷ Ang. recognition-primed decision model.

¹⁸ Podobno kot pri nekaterih ostalih praksah empirične fenomenologije (glej npr. Petitmengin, 2006) je bistvo van Manenove (2014) fenomenološke metode v tem, da raziskovalec ne sprašuje po mnenjih, stališčih in prepričanjih »soraziskovalcev«, ampak poskuša pridobiti podatke o dejanskem »živetem izkustvu« (metoda se, kot sodobne empirična fenomenološke prakse, opira na filozofsko »metodo« fenomenološke redukcije, ki jo je razvil Husserl).

kardiovaskularne podpore, umetnega hranjenja) ali sprejemati odločitve o spremembi smeri zdravljenja (npr. operaciji srca, presaditvi organa ali traheotomiji). Intervjuju s starši so potekali v času, ko so bili otroci na oddelku za intenzivno nego.

Van Manen je v svoji raziskavi prišel do petih glavnih »tem« izkustva zgoraj opisanih odločitev: »(a) odločitev, ki nikoli ni bila izbira; (b) odločitev kot iskanje poti stran; (c) odločitev kot razmišljanje in občutenje samega sebe skozi posledice; (d) odločitev kot neodločenost; in (e) odločitev kot nekaj, v kar pademo« (van Manen, 2014: 282).¹⁹

Odločitev, ki nikoli ni bila izbira. Samovi starši so se morali odločiti, ali naj prekinijo življenjsko podporo svojemu novorojenčku. Sami so poročali o tem, da so se počutili prisiljene v odločitev o prekinitvi podpore, vendar tega, da prispevajo k smrti novorojenčka in mu ne omogočijo možnosti preživetja, nikakor niso videli kot možnosti, čeprav so npr. vedeli, da ima novorojenček velike možnosti, da bo prizadet, če preživi. Situacije, v kateri so zdravniki od njih zahtevali »odločitev«, niso izkušali ali videli kot takšne, v kateri bi se bilo sploh mogoče odločati. Situacija skratka zanje ni bila odločitvena situacija.

Odločitev kot iskanje poti stran. Odločanje kot iskanje poti stran van Manen (2014) opisuje kot iskanje dokazov (informacij), ki bi odpravile negotovost odločitve. van Manen navaja pričevanje staršev, ki sta se morala na hitro odločiti med tem, ali naj otroka pustita na umetni podpori dihanja brez zdravil, kar bi vodilo v počasno smrt, ali z zdravili, kar bi potencialno vodilo v možganske poškodbe. Prekinitev terapije zanju ni bila možnost. Starša se nista spraševala le o tem, kakšne posledice ima lahko zdravlilo za možgane in dihanje, ampak sta se v svoji odločitvi spraševala o tem, kakšno bi bilo otrokovo življenje, kaj bi to pomenilo zanju ipd. Na koncu sta se odločila, da pred izbiro naročita pregled, ki bi pokazal, ali ima otrok možgansko krvavitev: če je nima, bosta izbrala drugo možnost (podpora dihanja z zdravili), če jo ima, bosta izbrala možnost brez zdravil. Nemogočo odločitev sta tako preložila na izid ultrazvoka glave. Pravita: »Čutila sva, da bi bila to pot stran od tega, da se morava odločiti« (van Manen, 2014: 283).

Odločitev kot razmišljanje in občutenje samega sebe skozi posledice. Tretji primer lepo prikazuje, kako starša, ki sta se morala relativno hitro odločiti, ali naj prekineta življenjsko podporo ali ne, negotovosti nista odpravljala z »racionalnim« premislekom o tem, kaj je najbolje za njunega otroka. Namesto tega sta si podrobno predstavljala in poskušala občutiti, kaj bi zanju in otroka pomenilo, če bi, sicer prizadet, preživel. Starša recimo poročata: »Mislim, da nisva preudarjala

¹⁹ V izvirmiku: "(a) a decision that was never a choice; (b) a decision as looking for a way out; (c) a decision as thinking and feeling oneself through the consequences; (d) a decision as indecision; and (e) a decision as something one falls into" (van Manen, 2014: 282).

in primerjala zdravstvenih alternativ, vprašala sva se le, ali to zmoreva?» (van Manen, 2014: 283).

Odločitev kot neodločenost. Neodločenost po van Manenu v kontekstu situacij, ki jih je raziskoval, izvira iz tega, da starši izbire, v katere so prisiljeni, občutijo kot nemogočo odgovornost. Oče Isaiaha, ki bi bil, če bi preživel, močno prizadet, je razmišljal o tem, kaj bi to pomenilo za njune druge otroke, kaj bi morali za to žrtvovati ipd. Vseeno pa je vedel, da je Isaiah le njun otrok. Odločitev se je tako prevesila v neodločenost, o kateri van Manen (2014) pravi, da je za očeta morda edina odgovorna pot delovanja.

Odločitev kot nekaj, v kar pademo. Mama, o kateri poroča van Manen (2014), je živel v stalnem stanju odločanja. Čeprav je odločitev že sklenila, odločitve ni izkušala kot končane, zaključene: »V resnici sem v katerikoli dejavnosti – ko jem, pijem, vozim, kar koli – še vedno v tem trenutku odločitve. Čeprav je izčrpavajoče, svoje odločitve enostavno nisem pripravljena sprejeti. ... Je tako, kot da je odločitev prostor, v katerem se ne morem ustaliti, a iz katerega prav tako ne morem oditi« (van Manen, 2014: 285). Po daljšem obdobju »ujetosti v odločanje« je mama poročala, da je kar naenkrat začutila »ustalitev odločitve«: začutila je, da mora hčeri dati možnost preživetja, in namesto paliativne nege, počakati na presaditev srca, ne glede na možne usodne posledice.

Opisane odločitve staršev so močno povezane z občutki, odgovornostjo in najglobljimi osebnimi vrednotami; so odločitve, kjer izidov in subjektivnih posledic ni mogoče dobro predvideti; odločitve, za katere starši pravzaprav ne vedo dobro, kaj prinašajo; »odločitve«, ki jih nekateri starši nikakor ne morejo videti ali občutiti kot odločitev, jih ne želijo ali ne morejo sprejeti; odločitve, o katerih starši ne morejo prenehati razmišljati, čeprav so »odločitev« že sprejeli, razrešitev odločitve pa izkusijo kot realizacijo nekakšne nujnosti tega, kar je prav. Opisane situacije in načini odločanja so radikalno drugačni od hipotetičnih potrošniških in drugih nepomembnih ter enostavnih laboratorijskih odločitev, opisanih v prejšnjem poglavju. Seveda se močno razlikujejo tudi od mnogih nepomembnih odločitev v vsakdanjem življenju, a podani primeri lepo prikazujejo, kako o takšnih odločitvah ne bi mogli prav dosti izvedeti, če bi se omejili na tretjeosebno preučevanje enostavnih in nepomembnih odločitev v laboratoriju. Tulving (1994) lepo ubesedi pomanjkljivost takšnih tretjeosebnih pristopov, ki niso daleč od tega, kar so počeli behavioristi: »Mentalni procesi, s katerimi so novo nastali kognitivni znanstveniki začeli polniti 'črno škatlo', so bili bolj abstrakcije opazovalcev kot pa zavestno izkustvo posameznikov. To preučevanje duševnosti s tretjeosebnega vidika se v tem smislu ni močno razlikovalo od osnovnih pogledov behavioristov« (Tulving, 1994: viii). Razumevanje prvoosebne perspektive odločevalca v različnih odločitvenih situacijah vsakdanjega življenja je tako bistveno za razumevanje odločanja kot celote – na to lepo kažeta obe predstavljeni raziskavi.

5. Zaključek

V članku sem želel pokazati, da prevladujoči sodobni pristopi k raziskovanju odločanja podajajo pomanjkljiv in v nekaterih vidikih napačen pogled na odločanje – tako v razumevanju pojava kot v raziskovalni praksi. Na podlagi primerjave nekaterih laboratorijskih odločitvenih problemov in namišljenih odločitvenih situacij v vsakdanjem življenju ter kvalitativnih raziskav odločanja v naravnem okolju sem pokazal, da je raziskovanje odločanja s specifične tretjeosebne perspektive, v specifičnih in nepomembnih odločitvenih situacijah, nezadostno in potencialno vodi v izkrivljeno podobo odločanja kot nekakšnega »preračunavanja« večinoma objektivnih stanj sveta, kjer občutja, želje, cilji in osebne vrednote odločevalcev niso prav relevantne.

Nasprotno, če se želimo izogniti prehitrim in varljivim posplošitvam iz laboratorija na vsakdanje življenje in odločanje razumeti bolj celostno, moramo raziskovalno energijo usmeriti v preučevanje različnih vrst odločanja v vsakdanjem življenju in v znanost o odločanju pripustiti prvoosebno raziskovanje subjektivnih perspektiv odločanja, brez katerih odločitve za odločevalce pravzaprav nimajo smisla. Soočiti se moramo s tem, da je v empiričnem raziskovanju odločanja treba stopiti ven iz laboratorija in ugotoviti, kako se ljudje v svojem naravnem okolju odločajo v različnih situacijah vsakdanjega življenja: od tistih, ki se pojavljajo »vsak dan«, pa vse do enkratnih, morda nadvse težkih, pomembnih in moralno zavezujočih. Nazadnje se moramo soočiti s tem, da čeprav tretjeosebna objektivna znanost o duševnosti prinaša mnoge zanimive in uporabne izsledke, enostavno ni zadostna za pojasnitev tega, kako in zakaj se odločamo, kot se odločamo, in delujemo, kot delujemo.

Decision-Making: From Lab to Real-Life

In the paper, I focus on criticising mainstream contemporary approaches to decision-making research. I argue that such approaches are severely limited and in certain aspects provide a false account of decision-making since: (a) their conception of decision-making is partly wrong since they do not emphasise enough the subjective aspects of decision-making, and do not truly conceive decision-making as a goal-oriented activity; (b) they focus on researching specific, inconsequential and unimportant decision situations and problems that do not have much in common with decisions of everyday life; and (c) they research decision-making almost entirely from the third-person perspective, leaving out the first-person perspective. Consequently, the subjective character, being essential, even intrinsic to decision-making, and how decision situations are understood/interpreted by decision-makers, are poorly understood at best. Contemporary approaches to decision-making research thus lead to a severely limited and to some degree perverted image of decision-making, potentially reducing the researched phenomenon to something that it is in fact not.

Keywords: decision-making, decision situations of everyday life, subjective aspects of decision-making, first-person perspective

Literatura

Baron, J. (2007). *Thinking and deciding* (4. izd.). New York: Cambridge University Press.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). »Deciding advantageously before knowing the advantageous strategy«. *Science*, 275(5304), str. 1293–1295.

Bendesky, A., Tsunozaki, M., Rockman, M. V., Kruglyak, L., & Bargmann, C. I. (2011). »Catecholamine receptor polymorphisms affect decision-making in *C. elegans*«. *Nature*, 472(7343), str. 313–318.

Bermúdez, L. (2009). *Decision Theory and Rationality*. Oxford: Oxford University Press.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Sons.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). »Frames, biases, and rational decision-making in the human brain«. *Science*, 313(5787), str. 684–687.

De Sousa, R. (1987). *The Rationality of Emotion*. Cambridge: MIT Press.

- DeCaro, D. A., Bar-Eli, M., Conlin, J. A., Diederich, A., Johnson, J. G., & Plessner, H. (2009). »How do motoric realities shape, and become shaped by, the way people evaluate and select potential courses of action? Toward a unitary framework of embodied decision making«. V Raab, M., Johnson, J., & Heekeren, H. (ur.), *Mind and Motion: The Bidirectional Link Between Thought and Action*, New York, Oxford, Amsterdam: Elsevier, str. 189–203.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). »On Making the Right Choice: the Deliberation-Without-Attention Effect«. *Science*, 311(5763), str. 1005–1007.
- Galotti, K. M. (2002). *Making Decisions that Matter: How People Face Important Life Choices*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books, Inc., Publishers.
- Hardman, D. (2009). *Judgment and Decision making: Psychological Perspectives*. BPS Blackwell Publishing Ltd.
- Hastie, R. (2001). »Problems for Judgment and Decision Making«. *Annual Review of Psychology*, 52(1), str. 653–683.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). »The neural systems that mediate human perceptual decision making«. *Nature Reviews Neuroscience*, 9(6), str. 467–479.
- Iyengar, S. S., & Lepper, M. R. (2000). »When choice is demotivating: can one desire too much of a good thing?«. *Journal of Personality and Social Psychology*, 79(6), str. 995–1006.
- Kahneman, D., & Tversky, A. (1979). »Prospect Theory: An Analysis of Decision under Risk«. *Econometrica*, 47(2), str. 263–291.
- Klein, G. (1999). *Sources of Power: How People Make Decisions* (2. izd.). Cambridge, Mass.: The MIT Press.
- Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). »Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript«. *Journal of Cognitive Engineering and Decision Making*, 4(3), str. 186–209.
- Kordeš, U. (2009). »The Phenomenology of Decision Making«. *Interdisciplinary Description of Complex Systems*, 7(2), str. 65–77.
- Maia, T. V., & McClelland, J. L. (2004). »A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task«. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), str. 16075–16080.

- Mamede, S., Schmidt, H. G., Rikers, R. M. J. P., Custers, E. J. F. M., Splinter, T. A. W., & van Saase, J. L. C. M. (2010). »Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert«. *Psychological Research*, 74(6), str. 586–592.
- van Manen, M. A. (2014). »On Ethical (In)Decisions Experienced by Parents of Infants in Neonatal Intensive Care«. *Qualitative Health Research*, 24(2), str. 279–287.
- Nieuwenstein, M. R., Wierenga, T., Morey, R. D., Wicherts, J. M., Blom, T. N., Wagenmakers, E.-J., & van Rijn, H. (2015). »On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage«. *Judgment and Decision Making*, 10(1), str. 1–17.
- Nutt, P. C., & Wilson, D. C. (ur.) (2010). *Handbook of Decision Making*. John Wiley & Sons, Ltd.
- Oullier, O., & Basso, F. (2010). »Embodied economics: how bodily information shapes the social coordination dynamics of decision-making«. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1538), str. 291–301.
- Pessoa, L. (2008). »On the relationship between emotion and cognition«. *Nature Reviews Neuroscience*, 9(2), str. 148–158.
- Petitmengin, C. (2006). »Describing One's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness«. *Phenomenology and the Cognitive Sciences*, 5(3-4), str. 229–269.
- Seeley, T. D., & Buhrman, S. C. (1999). »Group decision making in swarms of honey bees«. *Behavioral Ecology and Sociobiology*, 45(1), str. 19–31.
- Strle, T. (2015). *Metakognicija kot povezovalac mišljenja in intuicije v procesu odločanja*. Doktorska disertacija, Univerza v Ljubljani, Filozofska fakulteta.
- Strle, T. (2013). »Uganka izkustva: vloga prvo- in drugoosebnih metod v kognitivni znanosti«. *Analiza*, 17(1/2), str. 41–60.
- Symmonds, M., & Dolan, R. J. (2012). »The Neurobiology of Preferences«. V Dolan, R., & Sharot, T. (ur.), *Neuroscience of preference and choice: Cognitive and neural mechanisms*, London: Elsevier (Academic Press), str. 3–31.
- Thagard, P. (2005). *Mind: Introduction to Cognitive Science* (2. izd.). Cambridge, Mass.: The MIT Press.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven & London: Yale University Press.

The White House. Office of the Press Secretary. (2015, 15. september).
»Executive Order -- Using Behavioral Science Insights to Better Serve the
American People«. [5. maj 2016] <https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american>.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005).
»Understanding and sharing intentions: the origins of cultural cognition«. *The Behavioral and Brain Sciences*, 28(5), 675–691.

Tulving, E. (1994). »Foreword«. V Metcalfe, J., & Shimamura, A. P. (ur.),
Metacognition: Knowing about Knowing, Cambridge, Mass.: The MIT Press, str.
vii–x.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, Mass.: The MIT Press.

Janez Bregant

Filozofska fakulteta Univerze v Mariboru

Možgani v primežu računalnikov

Članek s pomočjo enostavne dileme zagovarja prepričanje, da smo ljudje oziroma naši možgani v nekem določenem smislu računalnik. Copeland (1993) trdi, da empirične razlike med računalnikom in možgani ne uspejo zavrniti klasičnega pristopa k enačenju možganov in računalnika, t.i. močne simbolne hipoteze, ki zagovarja njuno identifikacijo na osnovi simbolnega predstavljanja znanja o svetu, saj te dokazujejo zgolj to, da možgani niso tradicionalni, von Neumannovski računalnik. Tudi konekcionistični pristop k pojasnjevanju naše kognicije se tukaj ne odreže nič bolje, saj kljub temu, da stavi na nesimbolno predstavljanje informacij o svetu, ki naj bi bilo računalnikom tuje, njihovega enačenja z možgani ne uspe preprečiti. Še vedno smo lahko računalniki, a neke druge vrste, ki ni zgolj načeloma možna, kar dokazuje najbolj razvit hibridni model naše spoznavne ACT-R, ki združuje oba pristopa. In celo če zavoljo argumenta sprejmemo prepričanje, da računalnik kot manipulator s simboli ni združljiv z nevronskimi mrežami, nas to še vedno ne obvaruje pred tem, da ne bi bili računalniki v nekem splošnem smislu.

Ključne besede: močna simbolna hipoteza, konekcionistične mreže, ACT-R, dinamični sistemi, jezik misli

0 Uvod

V 18. stol. so mnogi ljudje menili, da nismo nič drugega kot utelešeni duhovi, tj. duhovi, ki jim telo rabi zgolj kot sredstvo za pridobivanje inteligence, znanja in izkušenj, naši najbližji nečloveški sorodniki pa naj bi bili angeli.

V 19. stol. se je z Darwinom, Freudom in Marxom to spremenilo do te mere, da nismo točno vedeli, kaj smo. Z Darwinovo izdajo prelomne študije *O nastanku vrst* leta 1859, se je zdelo, da smo najbolj razvite živali, ki so do te stopnje prišle tekom evolucije, botroval pa ji je naravni izbor. Da je naš najbližji sorodnik človeku podobna opica, je bila za tisti čas seveda bogokletna trditev, o posledicah katere ni treba izgublјati besed, segale pa so vse od cerkvenega izobčenja do prikazovanja Darwina kot opice. Pri Freudu se zdi, da nas določata spolni gon – libido in težnja k smrti oziroma samouničenju. Vzroki zanju naj bi se skrivali v naši podzavesti, potlačene misli in občutki pa privrejo na dan, ko smo podvrženi hipnozi. Z Marxovim raziskovanjem ekonomije in njenim povezovanjem s filozofijo pa se je zdelo, da smo subjekti in objekti hkrati, vendar šele potem, ko presežemo

njuno odtujitev. »Dejstvo, da je človek *telesno*, dejansko, čutno, objektivno bitje z naravnimi zmožnostmi, pomeni, da ima za svojo naravo *dejanske, čutne predmete* kot predmete svojega življenjskega izražanja, ali da lahko svoje življenje izrazi zgolj v dejanskih čutnih predmetih« (Marx, 1885/1997: 325).

V 20. stol. pa je bilo že veliko ljudi prepričanih, da smo računalniki, naši najbližji sorodniki pa naj bi bili proizvodi HP-ja, Appli ali Toshibe. Nekateri so se za to možnost glasno izrekli, drugi je niso želeli obešati na veliki zvon, še več pa se jih je s to idejo v obliki fraz, kot je npr. »Vsi smo sprogramirani s strani družbe«, spogledovalo.

Namen tega članka je pokazati, da smo ljudje, čeprav obstajajo empirični in teoretični razlogi za zavrnitev ideje, da so naši možgani računalnik v klasičnem von Neumannovem smislu, kljub vsemu računalnik. Osnovna ideja, ki se za tem skriva, je, da smo v vsakem primeru fizični stroji, vprašanje je zgolj kakšne vrste (Copeland, 1993: 180). Mogoče je to presenetljivo, ampak o tem, da smo ljudje stroji, je pisal že La Mettrie v svojem delu *L'homme machine (Strojni človek)* leta 1748: »Človek je stroj, zgrajen na takšen način, da si je to nemogoče predstavljati in ga zato tudi nemogoče definirati«¹ (La Mettrie: 1748/1996: 5). Zgodovinsko gledano je bilo vedno tako, da so možgane zaradi njihove kompleksnosti enačili s trenutno tehnološko najrazvitejšo napravo, v enem trenutku celo s telegrafom in telefonom. Vendar dejstvo, da znanost o možganih rada sledi »modnim zapovedim« in jih danes skuša razumeti in predstaviti kot računalnike, še ne spodkoplje trditve, da so možgani računalnik, pa čeprav se je v zgodovini analogija med tehnološkimi izumi in delovanjem možganom vedno izkazala za zgrešeno.

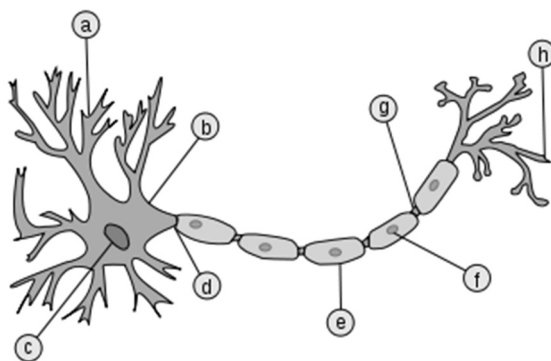
V prvem delu je predstavljen klasični pristop k enačenju možganov in računalnika, t.i. močna simbolna hipoteza, ki zagovarja njuno identifikacijo na osnovi simbolnega predstavljanja znanja o svetu, v drugem pa empirične razlike med računalnikom in možgani, ki naj bi pokazale, da pri tem ni uspešna. V tretjem delu je predstavljen konekcioniistični pristop k pojasnjevanju naše kognicije, ki temelji na nesimbolnem predstavljanju informacij o svetu in tako preprečuje enačenje možganov in računalnika, v petem pa alternativna modela naše kognicije, hibridni, ki združuje najboljše iz simbolnega in konekcioniističnega sveta, ter dinamični, ki poudarja vpetost človeka v okolje in spremembe skozi čas. Peti del predstavlja sklepno dilemo, ki nas zaradi svoje enostavne zgradbe pouči, da smo v vsakem primeru računalniki, res pa je, da zelo splošne sorte.

¹ Pri tem ne pozabi omeniti, da je o tem pisal že Descartes v *Razpravi o metodi* leta 1637: »Nihče ne zanika, da je ta slavn filozof naredil mnogo napak, je pa razumel živalsko naravo in bil prvi, ki je dovršeno pokazal, da so živali stroji« (La Mettrie, 1748/1996: 35).

1 Simbolna hipoteza

Zgradba nevrona

Možgani so sestavljeni iz dveh različnih skupin celic, pri čemer so za razumevanje njihovega delovanja pomembnejši in zanimivejši *nevroni*, katerih funkcija je prenos informacij preko električnih in kemičnih impulzov (manj pomembne in manj zanimive, vsaj kar se tiče možganskega delovanja, so *nevroglijske* celice, katerih funkcija je zgolj oskrba nevronov s hranili in njihova izolacija). Spodnja slika prikazuje (unipolarni) nevron:



Slika 1: Zgradba unipolarnega nevrona

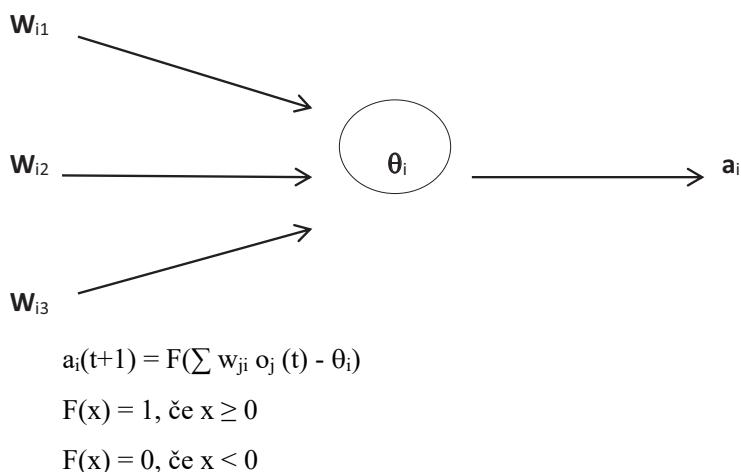
Nevron je sestavljen iz dendritov (a), telesa ali *soma* (b), jedra (c), nevrita ali *aksona* (d), mielinske ovojnice (e), Shwannove celice (f), Ranvierovega zažetka (g) in živčnih končičev (h).

Nevroni so med seboj povezani tako, da se dendriti enega držijo živčnih končičev drugega, zaradi česar tvorijo nepredstavljivo kompleksno celoto oziroma mrežo, ki obsega nekje tisoč dendritov na nevron in 10^{12} nevronov na možgane.

Že La Mettrie je predpostavljal naslednje: »Posebna lastnost našega stroja je, da vsako njegovo vlakno, tudi najmanjše, oscilira. To naravno nihanje je kot ura, ki včasih zamre in se mora potem obnoviti. Če postane šibko, se mora okrepiti, če pa premočno, oslabiliti.« (La Mettrie, 1748/1996: 31) Kljub temu so šele leta 1843 odkrili, da je delovanje živčnega sistema v osnovi sprejemanje in prevajanje električnih impulzov. V somi nevrona se sprosti električna energija, ki potuje po nevritu do sosednjih z njim povezanih nevronov. Do sprostitve električne energije pride v telesu nevrona (celici) takrat, ko je vzdraženih dovolj njegovih dendritov oziroma ko je dosežen ali presežen prag. Takrat »zagori« cel nevron in preko svojih živčnih končičev samodejno vzdraži vse dendrite sosednjih nevronov, s katerimi je povezan. Delovanje nevronov je v nekem smislu krožno: nevron pošilja signale svojim sosedom, da bi ti pošiljali signale svojim sosedom, da bi ti pošiljali signale svojim sosedom itd. Tako nam delovanje nevronov, ker je neinformativ-

no, saj sploh ne vemo, kaj je namen njihovega vzajemnega proženja, ne pove še ničesar o delovanju možganov. No, vsaj tako je bilo v letih med 1840 in 1940, kljub temu da so bile vmes do podrobnosti raziskane kemične značilnosti nevronov in njihov električni potencial.

Potem pa sta leta 1943 McCulloch (nevrofiziolog) in Pitts (matematik) objavila prelomni članek z naslovom »A Logical Calculus of the Ideas Immanent in Nervous Activity«, ki je korenito razburkal in spremenil področje raziskovanja možganov, saj je pokazal, »[k]ako lahko operacije nevronov in njihove povezave z ostalimi nevroni modeliramo s pojmi logike« (Markič, 2010: 30). Dokazovala sta, da je edino, kar je pri nevronih pomembno to, da so sproženi (»prižgani«) ali nesproženi (»neprižgani«), pri čemer lahko prvo označimo z npr. 1 ali »da«, drugo pa z npr. 0 ali »ne«, to, ali so sproženi ali nesproženi, pa je odvisno od tega, ali je dosežen ali presežen njihov aktivacijski prag. Slika njunega idealiziranega nevrone kot osnove za modeliranje operacij pravih nevronov je naslednja:



Slika 2: Delovanje idealiziranega nevrone

Gre za binarno enoto, ki izračuna vsoto vhodnih podatkov iz drugih binarnih enot – $F(\sum w_{ji} o_j(t) - \theta_i)$ – in kot izhodni podatek odda a_i , tj., 1 ali 0, odvisno od tega, ali vsota presega vrednost θ_i , ki predstavlja aktivacijski prag. Utež w_{ji} je moč sinapse, ki povezuje nevron i z nevronom j , lahko pa so vzbujevalne ali zaviralne ali pa so enake nič, kar pomeni, da med nevronoma ni stika. (Markič, 2010: 30)

Tako ni nevron nič drugega kot naprava, ki lahko fizično realizira eno izmed obeh stanj. Vse ostalo, kot je npr. spreminjanje kemične zgradbe nevrone ali intenzivnost električne eksplozije v somi, naj ne bi bilo pomembno. Z drugimi besedami, skupine nevronov delujejo kot neke vrste preprosti manipulatorji s simboli, med seboj pa so povezani tako, da ko se sproži npr. prva dva, se sproži tudi tretji, oziroma ko se pri prvih dveh pojavi znak, da je vključen (1), se pojavi znak, da je

vklučen (1) tudi pri tretjem. »Zakon 'vse ali nič', kateremu je podvržena živčna aktivnost, je zadosten za to, da je lahko aktivnost kateregakoli nevrona predstavljena kot propozicija« (McCulloch, Pitts, 1943/1990: 23–24). To pomeni, da si lahko nevrone zamislimo kot propozicije, ki so resnične (1) ali neresnične (0), obtežene povezave med njimi pa kot logične veznike in tako v pojmihi logike izračunamo vse, kar lahko jezikovno opišemo.

Zanimivo je, da tudi računalnik deluje po istem principu preklapljanja med 1 in 0 kot nevron (deli, ki mu to omogočajo, se imenujejo »logična vrata«). Recimo, če v računalnik preko tipkovnice vnesem »mačka« (vhod ali vnos), računalnik preklaplja med 0 in 1, odvisno od kode, ki jo vsaka črka ima, dokler ne pride do binarnega zapisa, npr. 1000011 1000001 1000111 1001111 1000001, ki pomeni izpis »mačka« na mojem ekranu (izhod ali iznos). Tako tudi računalnik ni nič drugega kot naprava, ki lahko fizično realizira eno od obeh stanj. Ker je tako tudi sam podobno kot nevron manipulator s simboli, so ljudje, ko so se prvi računalniki pojavili, te povezali z modelom nevrona, ki sta ga razvila McCulloch in Pitts, ter prišli do sklepa, da so možgani (in s tem tudi ljudje) računalniki. To, kako smo pri enačenju možganov z napravami napredovali, pa lepo ponazarja naslednji odlomek iz sredine 18. stoletja: »Ne motim se; človeško telo je ura /.../ ogromna in inteligentno zgrajena /.../, njen izdelovalec pa je njen nov prebavni sok.« (La Mettrie, 1748/1996: 34, 31)

Močna in šibka simbolna hipoteza

V splošnem simbolna hipoteza pravi, da je mišljenje računanje, ki vključuje manipulacijo s simboli, ta pa je lahko realizirana v sistemih, ki jo izvajajo s preklapljanjem med 0 in 1 kot elementoma binarnega sistema. Če mislimo, da je ta ideja nova, se motimo, »Mišljenje ni nič drugega kot računanje« pravi že Hobbes leta 1651 v *Leviathanu* (Hobbes, 1651/2006: 32). Njegove besede, kot sem dejal že drugje,² vsebujejo dve domnevi: (i) mišljenje je mentalni proces, ki obsega manipulacijo s simboli, ki se dogaja v možganih, zaradi česar so primerki misli v bistvu primerki možganov, ali, kot jih imenuje Hobbes, »fantazme«. »Ko človek razmišlja, ne dela ničesar drugega kot, da si zamišlja končno vsoto, ki jo dobi s seštevanjem delov, ali ostanek, ki ga dobi z odštevanjem ene vsote od druge. /.../ Te operacije se ne dogajajo samo pri številih, ampak pri vseh načinih stvari, ki jih lahko združujemo ali odvezujemo« (Hobbes, 1651/2006: 32). In (ii) mišljenje je najbolj čisto, ko zvesto sledi metodičnim pravilom, zaradi česar lahko rečemo, da je mišljenje mehanski proces, je nekaj takega kot upravljanje mentalnega abaka: fantazme premetavamo sem in tja skladno s pravili razuma, podobno kot na pravem abaku skladno z računskimi pravili sem in tja premikamo kroglice (Haugeland, 1986: 23).

² Glej Bregant (2010).

Vendar danes znotraj simbolne hipoteze ločimo dve smeri, prva, imenujmo jo *šibka simbolna hipoteza*, je predpostavka, da *je načeloma možno zgraditi univerzalni simbolni sistem (računalnik), ki misli*. Recept, kateremu moramo slediti, vsebuje pet točk: (i) uporabo primerne kode za predstavitev stvarnih predmetov; (ii) kreacijo spomina, v katerem je shranjeno znanje o svetu; (iii) izdelavo ustreznih vhodnih naprav, ki pretvorijo dražljaje iz okolja v simbole; (iv) razvoj naprav, ki omogočajo uporabo osnovnih operacij na proizvedenih in shranjenih simbolnih strukturah, in (v) proizvodnjo izhoda v obliki simbolne predstave ustreznega (tudi govornega) obnašanja (Copeland, 1993: 80).

Druga, imenujemo jo, *močna simbolna hipoteza*, pa je predpostavka, da *so zgolj univerzalni simbolni sistemi (računalniki) zmožni misliti*. Krila ji daje ideja o *večvrstni realizaciji simbolov*, ki vsebuje domnevo, da so lahko simboli oziroma njihovi tipi realizirani na različne načine: z barvo, ogljem, svinčnikom, elektromagnetnimi stikali (releji), elektronkami, silikonskimi polprevodniki (tranzistorji) itd. Že Turing je menil, da večvrstna realizacija simbolov dokazuje, da računalnik v svojem bistvu ni električna naprava:³ »Pogosto se misli, da je pomembno, da so sodobni digitalni računalniki električni in da je tudi živčni sistem električen. /.../ ker so vsi digitalni računalniki v nekem smislu enakovredni, vidimo, da uporaba elektrike z vidika teorije ni pomembna« (Turing, 1950: 439).

Od razumevanja računalnikov na ta način pa ni več daleč do njihove izenačitve z možgani. Skladno z močno simbolno hipotezo, namreč noben umetni sistem, ki ni računalnik, ne more misliti. Še več, vsako naravno bitje, kateremu pripisujemo sposobnost mišljenja, se po natančni preiskavi izkaže za računalnik. Res je, da so pri računalnikih in naših možganih manipulacije s simboli verjetno realizirane z bistveno drugačnimi fizičnimi operacijami, vendar to, kot smo videli, pri tem, kaj smo, nima nobene vloge. Tako ravno močna simbolna hipoteza pooseblja idejo, ki se je pojavila, ko so nekateri avtorji povezali McCullohovo in Pittsovo model nevronskega računalnika, da smo ljudje (oziroma naši možgani) računalniki. Vnet zagovornik močne simbolne hipoteze je Pylyshyn: »Semantična vsebina [prepričanj in namer] je kodirana s strani možganskih lastnosti na isti splošni način, na katerega so semantične vsebine računalniških predstav kodirane s strani fizično uprimerjanih simbolnih struktur« (Pylyshyn, 1984: 258). Meni, da so naše mentalne operacije v bistvu manipulacije s simbolnimi reprezentacijami, ki so zgolj podobne običajnim stavkom kateregakoli govorjenega jezika: »/.../ [simbolne reprezentacije] so simbolni izrazi v notranjem, fizično uprimerjanem simbolnem sistemu, ki se včasih imenuje 'mentalese' ali 'jezik misli'«⁴ (Pylyshyn, 1984: 194). Poudariti velja, da kljub enačenju možganov in računalnika, med njimi skorajda ne

³ Analitični stroj Charlesa Babbaga, razvit med leti 1823 in 1833, ki je bil prvi programiran računalnik, je bil denimo »narejen« (v resnici ni nikoli zaživel, ker bi bila za tisti čas njegova izdelava tehnološka prezapletena) izključno iz mehanskih delov: koles in zobnikov.

⁴ Za jezik misli glej tudi Fodor (1975), Markič (2010), Bregant (2013).

moremo pričakovati drugih podobnosti razen manipulacije s simboli, pa še ta lahko poteka na povsem različne načine, in jih močna simbolna hipoteza tudi ne predpostavlja, kot bomo videli v nadaljevanju.

Kakorkoli, z vidika identifikacije možganov z računalnikom imamo na voljo tri možnosti: (i) da jo vzamemo dobesedno, in sicer v smislu, da smo zgolj manipulatorji s simboli (kot v primeru močne simbolne hipoteze), (ii) da jo vzamemo kot koristno metaforo ali provokacijo in (iii) da jo vzamemo kot resnično v smislu, da sicer smo (neke vrste) računalniki, nismo pa manipulatorji s simboli.

2 Možgani in računalnik: empirične razlike

Vprašanje je, zakaj že nekateri pionirji umetne inteligence niso verjeli tega, da je človek računalnik, npr. von Neumann (1958), odgovor na to pa je sedaj verjetno še bolj na dlani kot v 50. in 60. letih 20. stoletja. Danes vemo o možganih več kot kdajkoli prej, ta nevrofiziološki napredek pa postavlja identifikacijo človeka z računalnikom, vsaj v smislu močne simbolne hipoteze, po kateri smo oboji zgolj manipulatorji s simboli, pod velik vprašaj. Empirične razlike med možgani in računalniki lahko strnemo v sedem točk:⁵

(i) možgani so v primerjavi z računalnikom katastrofalno počasni: čas, ki poteče od vzdraženja dendritov do sprostitve elektrike v njegovih živčnih končičih (prehod iz stanja 0 v stanje 1), je tisočinka sekunde, kar je milijonkrat dlje od časa, ki ga porabi računalnik, da v logičnih vratih preklopi med 0 in 1;

(ii) nevroni so preko dendritov in živčnih končičev povezani z ogromnim številom drugih nevronov: tudi do 100.000 jih vpliva na to ali »gorijo« ali ne »gorijo«, kar močno prekaša povezave med deli računalnika, ki so odgovorni za preklapljanje med 0 in 1, preštejemo jih lahko na prste ene roke;

(iii) možgani so v nekem smislu odporni na poškodbe: njihova zmogljivost upada postopoma, tj. glede na količino možganskega tkiva, ki ga izgubimo; tudi huda poškodba, npr., ne pomeni nujno tega, da možgani ne zmorejo več opravljati svojih bistvenih funkcij, medtem ko računalnik povsem odpove že ob najmanjši fizični poškodbi elektronskih delov procesorja;

(iv) v možganih obstaja več različnih nevronov: zdi se, da je njihova oblika, npr. piramidni, zvezdnati, vretenasti itd., odvisna od njihove funkcije, pa tudi določene oblike se vedno pojavljajo na istih mestih, tako da tega ne moremo označiti kot nekaj nepomembnega, medtem ko je računalniški procesor sestavljen iz omejenega števila istih elementov;

⁵ Glej npr. Churchland (1986), Copeland (1993).

(v) možgane obravnavamo kot paralelne (vzporedne) procesorje, tj. takšne, ki hkrati izvajajo različne operacije, računalnike pa kot sekvenčne (zaporedne) procesorje, tj. takšne, ki izvajajo operacije eno za drugo, kar pomeni, da možgani zaradi velikega števila med seboj povezanih nevronov v tisočinki sekunde izvedejo toliko različnih operacij, kot bi jih računalnik kljub svoji hitrosti v nekaj letih ali stoletjih;

(vi) človeški spomin deluje po načelu dostopa preko vsebine, tj. glede na vsebino/opis vprašanja se bomo spomnili odgovora nanj, računalniški spomin pa po načelu dostopa preko naslova (pri tem je razmerje med številko in vsebino podobno razmerju med hišno številko in njeno notranjo opremo): glede na številko naslova v spominu, se bo računalnik »spomnil«⁶ odgovora na vprašanje;

(vii) možgani ne shranjujejo vsakega podatka na svojem mestu, ampak je to porazdeljeno med določeno število nevronov (domneva se, da pri shranjevanju posameznega podatka sodeluje od 700.000 do 7 milijonov nevronov), zaradi česar npr. v poskusih podgane, ki jim, potem ko so se naučile pot skozi labirint, odvzamemo del možganov (do 10% in kjerkoli), pot do cilja kljub vsemu najdejo,⁶ medtem ko se v računalniku vsak drobec informacij nahaja na točno določenem in njemu lastnem mestu.

Na kaj lahko iz navedenih razlik sklepamo? Churchland meni, da iz tega preprosto sledi, da ljudje nismo računalnik: »/.../ te razlike povzročijo, da metafora, ki identificira možgane z računalnikom, izpuhti«⁶ (Churchland, 1986: 459). Copeland (1993) se z njo ne strinja in trdi, da je največ, kar lahko iz tega izpeljemo, da možgani niso to, kar se razume pod pojmom von Neumannovski računalnik. Seznam lastnosti, ki jih je von Neumann predpisal za to, da bi naprava štela kot računalnik, se namreč povsem sklada z opisanimi razlikami med njim in možgani. Posplošeno so zahtevane lastnosti štiri: (i) naprava je zaporedni procesor: ukaze izpolnjuje enega za drugim, (ii) vsak niz simbolov v njej je shranjen na svojem posebnem mestu v spominu, (iii) dostop do shranjenih podatkov poteka preko njihovih oštevilčenih naslovov in (iv) nadzor se izvaja preko zgolj ene, t.i. centralno procesne enote (CPE) (Copeland, 1993: 193). Res je, da je njegov seznam še danes osnova vsakega računalnika, vendar gre kljub temu zgolj za en način, kako lahko izdelamo univerzalni simbolni sistem, iz česar pa še ne sledi, da bi takšni morali biti možgani, pa čeprav bi se izkazalo, da so resnično zgolj manipulatorji s simboli. In močna simbolna hipoteza načina zgradbe univerzalnega simbolnega sistema, ki je edini sposoben misliti, niti ne določa: zaporednost in vzporednost, način shranjevanja in dostopa do podatkov, izvedba kontrole nad izvajanjem operacij itd. na to, ali je naprava univerzalni manipulator s simboli, nimajo nobenega vpliva in tako močne simbolne hipoteze niti ne zavrnejo niti ne potrdijo.

⁶ Glej Lashley (1960).

Vprašanje, na katerega moramo odgovoriti, da bi prišli do tega, ali je močna simbolna hipoteza resnična ali ne, je, ali je »mentalese«, ki naredi možgane za manipulator s simboli, resnično edini način, s katerim lahko razložimo našo kognicijo. Tipično se ji pripisujejo štiri lastnosti, *produktivnost*, *sistematičnost*, *kompozicionalnost* in *sklepalna koherentnost*. Produktivnost misli pomeni, da lahko iz končnega števila propozicij tvorimo neomejeno število misli, npr. vedno bi lahko končnemu naboru vseh misli, ki jih proizvedemo, dodali še eno in tako dalje v neskončnost (če nas ne bi prej prehitela smrt). Sistematičnost misli pomeni, da če razumemo stavek, ki ima določeno obliko (je zgrajen iz danih elementov), potem razumemo tudi stavek, ki ima podobno obliko (je zgrajen iz istih, a drugače povezanih elementov), npr., če razumemo stavek »Peter ljubi Mojco«, razumemo tudi stavek »Mojca ljubi Petra«. Kompozicionalnost misli pomeni, da sintaksa, tj. elementi stavkov in slovnična pravila, določa pomen, ker so, npr., besede »Peter«, »ljubiti« in »Mojca« neodvisne od konteksta, lahko iz tega, kako so povezane, brez da bi morali vedeti karkoli drugega, takoj izpeljemo pomen danega stavka, v katerem nastopajo. Sklepalna koherentnost ali sistematičnost sklepanja pa pomeni, da če lahko izvajamo sklepanje, ki je določene oblike, potem lahko izvajamo tudi sklepanje, ki je podobne oblike, npr. iz »Peter ljubi Mojco, stanuje v Mariboru, dela na Filozofski fakulteti in rad smuča«, lahko izpeljemo »Peter ljubi Mojco«, to pa lahko izpeljemo npr. tudi iz »Peter ljubi Mojco in stanuje v Mariboru«⁷ (Fodor, Phylyshyn, 1988: 33–50).

Fodor in Pylyshyn (1988) trdita, da je pojasnitev opisanih štirih lastnosti kognicije, ki jo lahko imenujemo *razlagalna naloga* (Markič, 2010), možna zgolj v primeru, ko možgane razumemo kot manipulator s simboli, ki temelji na fizičnem simbolnem sistemu, imenovanem jezik misli, zaradi česar je močna simbolna hipoteza resnična. Takšen *simbolni model* naše duševnosti, ki ga Beer opredeli kot »/.../program, katerega vhod je simbolni opis problema, ki ga mora rešiti, kar stori tako, da preko svojega (simbolno predstavljenega) znanja z vhodnim simbolnim opisom tako dolgo manipulira (pri čemer se drži izključno sintaktičnih pravil), dokler ne pride do izhodne rešitve« (Beer, 2010: 96), in ki ga imenujemo tudi *kognitivizem*, je, kot sta se slikovito izrazila, »edina igra v mestu«. Zanimivo je, da je resničnost njune izjave postala vprašljiva tako rekoč takoj zatem, ko sta jo izrekla, saj so se že v sredini 80. let 20. stoletja pojavili prvi *konekcionistični modeli kognicije*, ki so svoj navdih prav tako našli v že omenjenem prelomnem članku McCullocha in Pittsa. Tako se naenkrat pojavi nov igralec, ki je prav tako sposoben opraviti razlagalno nalogo in to brez sklicevanja na jezik misli kot simbolni sistem, kar postavlja resničnost močne simbolne hipoteze pod velik vprašaj in s tem spet oživlja idejo, da možgani kljub vsemu niso računalniki oziroma vsaj ne

⁷ Za povzetek njune razlage štirih lastnosti kognicije in možnosti njihove realizacije v računalnikih glej Markič, 2010: 59–65.

takšni, kot se je mislilo, tj. manipulatorji s simboli. Če se lahko dostojno zoperstavi simbolnemu modelu kognicije, pa bomo videli v nadaljevanju.

3 Konekcionistični model

Analogija med računalnikom in možgani temelji na kognitivistični predpostavki, da poteka pojasnjevanje obnašanja stroja in človeka na treh istih ravneh, ki jih je prvi razdelal Marr (1982), danes pa so znane pod imeni (i) semantična (vsebinska), (ii) sintaktična (algoritemska) in (iii) strukturna (fizična). Prva se ukvarja s tem, kaj je cilj računanja z mentalnimi predstavami, druga s tem, kako to računanje poteka, tj. z odkrivanjem algoritmov, s katerimi se iz vhodnih podatkov dobijo izhodni, tretja pa s tem, kako to računanje fizično realizirati. Videli smo, da močna simbolna hipoteza poudarja prvi dve ravni, tj. semantično in sintaktično, zanemarja pa tretjo, tj. strukturno raven, s čimer se spogleduje s teorijo duha, znano kot *funkcionalizem*.⁸ Ta poudarja, da sta za pojasnitev našega obnašanja, t.i. psihološko razlago, pomembni zgolj prvi dve ravni, ne pa tudi tretja, saj razume mentalno kot »[m]nožico duševnih reprezentacij, nad katerimi potekajo sintaktično določene transformacije« (Markič, 2010: 39). Ni čudno torej, da sta močna simbolna hipoteza in funkcionalizem zaveznika, kar se ne nazadnje kaže tudi v tem, da so zagovorniki simbolnega modela kognicije ravno filozofi, ki prisegajo na formulacijo duha v pojmihi njegovih tipičnih vzrokov in učinkov, tj. njegove funkcije.

Posledica prepričanja, da takšno zanemarjanje tretje, tj. strukturne, ravni pri pojasnjevanju mentalnih pojavov ni na mestu, pa je bil vzpon novega modela naše kognicije, ki na razlagalno hipotezo ne odgovarja v pojmihi manipulacije s simboli, pa tudi načina, kako je obdelava podatkov izvedena, povsem ne zanemarja. Njegov vzpon se je začel v sredini 80. let 20. stoletja, ko so simbolni modeli padli v krizo, saj je postajalo vse bolj očitno, da lahko z njimi dobro ponazorimo zgolj višje kognitivne procese,⁹ npr. izpeljevanje, kjer iz pridobljenega znanja sklepamo na dogodke v svetu, nižjih kognitivnih procesov,¹⁰ kot je npr. zaznavanje, kjer čutni podatki zgolj opisujejo dogodke v svetu, pa ne. Nov spoznavni model, ki naj bi to odpravil, sicer ni povsem nov, saj ima svoje korenine v modeliranju nevronskih mrež že omenjenih McCullocha in Pittsa, se je pa v času razcveta simbolnega modela nanj gledalo z omalovaževanjem in se ga skorajda povsem ignoriralo.¹¹ Zgleduje se po tem, kako so zgrajeni in kako delujejo naši možgani, ter sku-

⁸ Zanj glej npr. Block (1980/2007), Kim (1996), Bregant (2013).

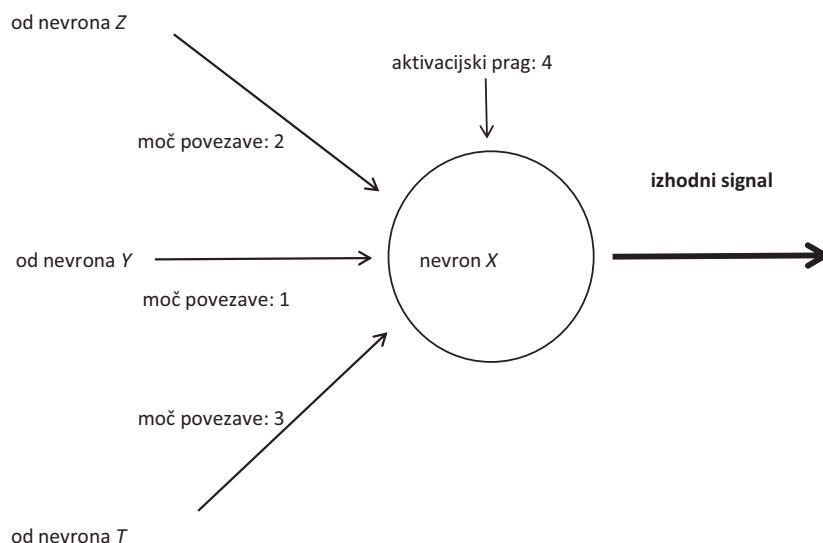
⁹ Ang. *top-down*.

¹⁰ Ang. *bottom-up*.

¹¹ Rosenblatt (1962) se takšnemu pristopu ni nikoli odrekel, zaradi česar je bil v zlati dobi simbolnih modelov (60. in 70. leta 20. stoletja) tarča pogostih kritik.

ša medsebojno povezanost njihovih osnovnih enot, nevronov, ki tvorijo prepleteno mrežo povezav, prenesti na bolj abstraktno raven. Z idealiziranim opisom nevronov v smislu njihovega prepletanja in delovanja nam ponudi novo orodje za modeliranje mentalnih procesov oziroma odgovarjanje na razlagalno nalogo, ki pa računalnika ne dojema več kot manipulatorja s simboli, ampak kot virtualno mrežo medsebojno povezanih enot. Takšen model naše kognicije, ki za vzor sicer jemlje zgradbo naših možganov, zanemarija pa podrobnosti delovanja nevronov in živčnega sistema, je dobil ime *konekcionistični model* (včasih se uporablja izraz nevronska mreža, vendar lahko ta pomeni tudi model na ravni dejanske, možganske zgradbe in ne abstraktne). Ker zagovarja nesimbolno procesiranje informacij, tako spodkopava idejo, da smo računalniki, vsaj v primeru, če slednje razumemo v klasičnem smislu, tj. kot napravo za obdelovanje predstav, prikazanih s simboli.

Konekcionistični model kognicije jemlje za vzor, kot smo omenili, delovanje dejanskega nevrona, zanj pa so z vidika delovanja bistvene naslednje štiri stvari: (a) da je povezan z drugimi nevroni, (b) da so te povezave različno obtežene (močne), (c) da obstaja prag, ki določa, ali je aktiviran ali ne in (d) da je tako »izključen« (0) ali »vključen« (1). Skratka, biološki nevroni tvorijo prepleteno mrežo povezav, ki so različno močne, slednje je deloma odvisno od premera vlaken, ki so povezana, deloma pa od kemične zgradbe stika (sinaps). Ko vsota vhodnih signalov določenega nevrona doseže ali preseže njegov aktivacijski prag, nevron »zagori« (se sproži) ter posreduje izhodni signal naprej svojemu najbližjemu sosedu. Oglejmo si to na naslednji poenostavljeni sliki dejanskega nevrona (Copeland, 1993: 209):



Slika 3: Delovanje dejanskega nevrona

Nevron X posreduje elektrokemične signale sosednjemu nevronu, tj. je »vključen« oziroma v stanju 1, ko vsota njegovih vhodnih signalov doseže ali preseže njegov aktivacijski prag, ki je 4, kar se lahko zgodi, npr., ko sta sprožena nevrna Y in T (skupna moč vhodnega signala je potem točno 4) ali ko so sproženi vsi trije nevroni, Y , T in Z (vsota moči vhodnega signala je takrat 6). Vse opisane nevrnske povezave, ki prispevajo k vhodnemu signalu nevrna X , so bile v tem primeru vzbujevalne ali *ekscitirajoče*, tj. nevron X so spodbujale k proženju. Lahko pa so vhodne nevrnske povezave tudi takšne, da sosednji nevron odvrčajo od proženja, zaradi česar jih imenujemo zaviralne ali *inhibirajoče*, npr. če so sproženi vsi trije nevroni, Y , T in Z , vendar povezava med nevronoma X in Z ni vzbujevalna, ampak zaviralna (ne več +2, pač pa -2), nevron X ne bo sprožen, saj skupna moč vhodnega signala ne bo več 6, temveč 2, s čimer ne bo dosežen ali presežen njegov aktivacijski prag.

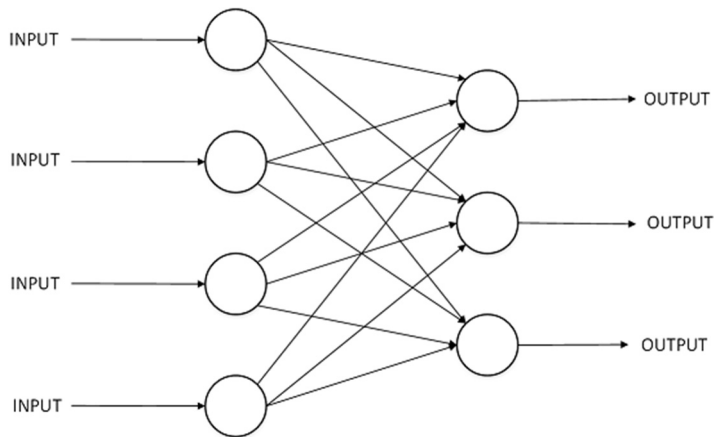
Zgradba konekcionističnih mrež

Podobno je zgrajen tudi konekcionistični model, pri čemer so njegovi osnovni gradniki, t.i. idealizirani nevroni, preproste, neinteligentne enote, ki so lahko vklopljene ali izklopljene. Takšni umetni nevroni so med seboj povezani, vsak izmed njih pa ima določeno aktivacijsko vrednost, ki jo preko vezi posreduje drugim enotam in s tem pripomore k povečanju (ekscitiranje) ali zmanjšanju (inhibiranje) njihove aktivacijske vrednosti, kar vpliva na to, ali se sprožijo ali ne.¹² Konekcionistični modeli so lahko zgrajeni na različne načine, ki se med seboj, kot pravi Markič (2010), po kateri jih povzemam, razlikujejo glede na (i) število plasti mrež, (ii) smerjo povezave med njenimi enotami, (iii) vrsto aktivacijske in izhodne funkcije in (iv) izbrano učno pravilo.

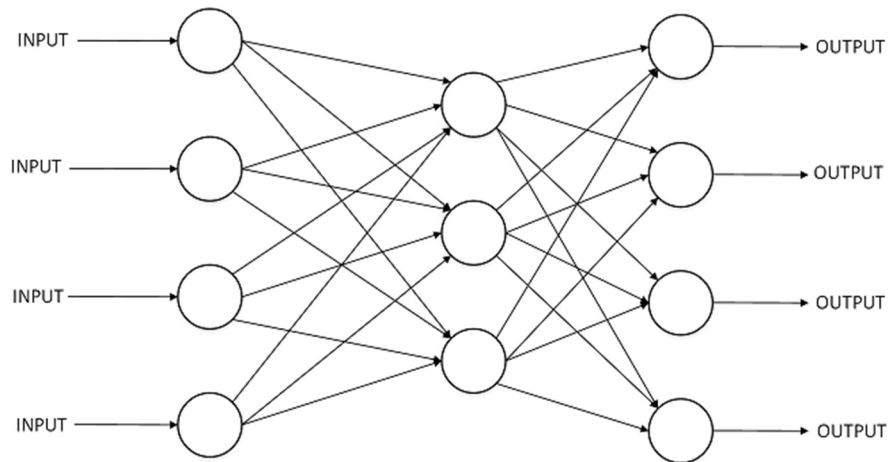
Z vidika števila plasti mrež ločimo *enoplastne*, kjer je vsaka enota hkrati vhodna in izhodna, *dvoplastne*,¹³ kjer prvi nivo enot predstavlja vhodne podatke, drugi pa izhodne, in *večplastne*, kjer je med vhomom in izhodom še več »skritih« ravni.

¹² Glej sliko 2.

¹³ Npr. Rosenblatt (1962) in njegov »Perceptron«; danes je to kar drugo ime za nevrnsko mrežo.

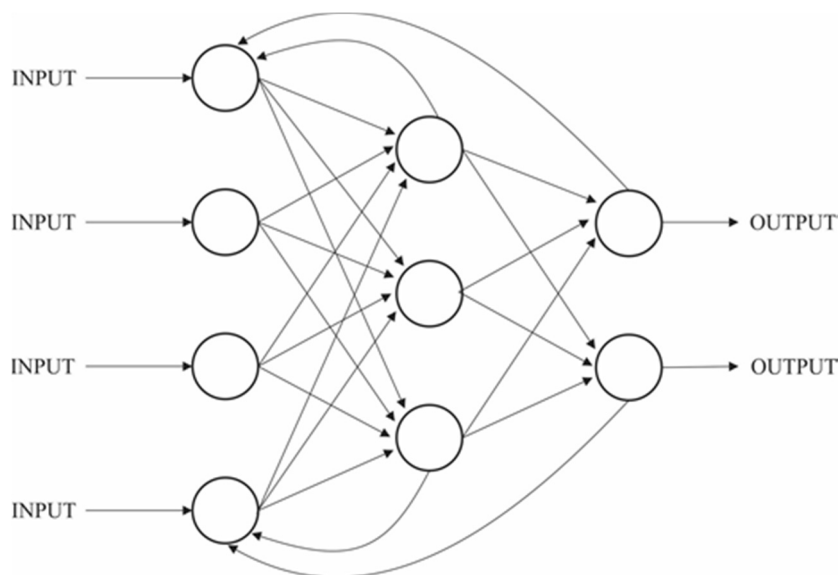


Slika 4: Dvoplastna mreža



Slika 5: Troplastna mreža

Z vidika smeri povezave so lahko slednje *enosmerne*, kjer so enote nižje ravni povezane samo z enotami višje ravni (ang. *feedforward*) in informacije potujejo zgolj v eni smeri (sliki 4 in 5), tj. naprej (od vhoda do izhoda), in *povratne*, kjer so tudi enote višje ravni povezane z enotami nižje ravni (ang. *recurrent*), kar omogoča pretok podatkov v obeh smereh, tj. naprej in nazaj.



Slika 6: Povratna mreža

Aktivacijska funkcija, ki določa način preslikave vhodnih podatkov v aktivacijsko stanje, je za a_i npr. lahko $\sum w_{ji} * a_j + \theta_i$ (w_{ji} je moč povezave med i in j , a_j stanje j enote, θ_i pa prag i enote), lahko pa so to tudi druge funkcije, npr. kvadratna $\sum w_{ji} * a_i^2 + \theta_i$ ali sferična $\sum (a_i - w_{ji})^2 - \theta_i$. Izhodna funkcija, ki določa način preslikave aktivacije enote v njen izhod, pa je lahko deterministična, v tem primeru vhod enolično določa izhod, ali stohastična, kjer vhod zgolj z večjo ali manjšo verjetnostjo določa izhod.

Učno pravilo, ki ga izberemo, da bi se sistem prilagajal okolju, pa je lahko *nenadzorovano* ali *nadzorovano*. V prvem primeru mreža sama postopoma okrepi vezi med dvema enotama, da bi v bodočnosti dosegala ustrezen izhod. Tukaj sta najbolj znani pravili *Hebbovo*, kjer se okrepijo povezave med istočasno aktivnimi enotami, ostale se oslabijo, in *tekmovalno*, kjer enote med seboj najprej tekmujejo, katera bo prva postala aktivna, tista, ki ji to uspe, pa potem okrepi svojo zvezo s tistimi, ki so ji do zmage pomagale, z ostalimi pa jo oslabi. V drugem primeru pa mrežo učimo tako, da izhodne podatke primerjamo s pravilnimi odgovori, zato ga imenujemo tudi učenje z učiteljem. Tukaj so najbolj znana pravila *delta*, kjer učenje poteka s pomočjo učnih primerov, tj. z znanimi pari vhodnih in izhodnih podatkov, pri čemer se pri danem vhodu najprej izračuna izhod, potem pa še razlika med dobljenim in želenim izhodom, skladno z njo pa se potem ustrezno zmanjšajo ali povečajo moči povezav med enotami.¹⁴

¹⁴ Za podrobnosti glej npr. Bechtel in Abrahamsen (2002), Peruš (2001), Markič (2010), Guid in Strnad (2007).

Pomembna značilnost konekcioniistične mreže je ravno učenje, ki ji omogoča samostojno posploševanje iz vzorcev (brez pomoči programerja), kar sta pokazala že Rumelhart in McClelland (1987), ko sta jo naučila tvoriti preteklik angleških pravilnih in nepravilnih glagolov. Učenje je potekalo preko mehanskih prilagajanj moči povezav med enotami in je med drugim vključevalo, npr., dodajanje končnice »ed« vhodnim podatkom pri formulaciji izhodnih odgovorov v primeru pravilnih glagolov, konekcioniistična mreža pa je potem sama tvorila preteklik tudi za tiste glagole, ki jih v procesu učenja ni spoznala. Začela sta s tipičnimi glagoli, ki se jih naučimo najprej, npr. priti, imeti, dati, in jih kot nedoločnike predstavila na vhodni ravni, potem pa sta primerjala dobljene in želene izhode ter glede na razliko med njimi spreminjala moči povezav med enotami, kot to določajo pravila delta¹⁵ (Copeland, 1993: 215).

Prednosti konekcioniističnih sistemov

Privlačnost konekcioniistične obdelave podatkov tiči v njeni podobnosti človeškemu procesiranju informacij, konekcioniistične mreže pa imajo po Copelandu (1993) šest lastnosti, ki jih najdemo tudi pri človeku. Ni presenečenje, da so na tem seznamu ravno tiste, v katerih se možgani razlikujejo od klasičnih računalnikov, saj konekcionizem, kot smo omenili, govori prej v prid trditvi, da jim ljudje nismo podobni, kot pa, da smo jim. Prva je *biološka podobnost*, ki pa se odvija zgolj na abstraktni ravni, saj konekcioniistične mreže npr. zanemarijajo razlike, ki obstajajo med dejanskimi nevrottransmiterji, ne izkazujejo kompleksnosti vzorcev bioloških nevronske povezav ali ne simulirajo zapletenega procesa dejanskega vzbujanja, vendar »[t]e razlike niso problematične, v bistvu so celo zaželene, če se osredotočamo na konekcioniistične modele kot na kognitivne in ne biološke modele« (Bechtel, Abrahamsen, 2002: 57). Druga je *prepričanje, da poteka učenje v možganih na podoben način kot v mrežah*, tj. s spreminjanjem uteži povezav med nevroni preko izkušenj, kar možganom in mrežam omogoča učenje ter prilagajanje novim in odprtim vsakdanjim situacijam. Pomembno je ločiti med aktivacijsko vrednostjo enote, ki predstavlja njeno trenutno spremembo stanja glede na vhodne podatke, in utežmi povezav med enotami, ki so posledica ponavljajočih se sprememb ter konekcioniistični mreži omogočajo obdelavo vhodnih podatkov tako, kot se je naučila, tj. da znanje, ki je shranjeno v povezavah, uporabi pri odgovoru na novo vprašanje (Markič, 2010: 55). Tretja je *vzporedno procesiranje*, saj so v mrežah izhodni podatki posledica motnje vhodne plasti, ki povzroči z vidika aktivacije enot masovno vzbujanje in zaviranje, v katerem sodelujejo vse hkrati, kar konekcioniističnim mrežam zaradi velikega števila med seboj povezanih enot omogoča, da v istem času izvedejo (približno) isto število operacij kot možgani. Četrta je *razpršeno shranjevanje podatkov*, kar pomeni, da znanje v mreži ni shranjeno na enem mestu (čeprav obstajajo tudi take), kot npr. v računalniku, am-

¹⁵ Za praktične primere glej npr. Mcleod, Plunkett, Rolls (1998), Peruš (2001), Guid in Strnad (2007).

pak je porazdeljeno po povezavah tako, da vsaka izmed njih hrani različne drobce iste informacije. V omenjenem primeru učenja tvorbe preteklika, znanje, ki je konekcionistični mreži omogočalo pravilno tvorbo preteklika danega nepravilnega glagola, ni bilo shranjeno v eni povezavi, ampak je bilo razpršeno po celem vzorcu povezav, ki so bile aktivne v času učenja in so se tako okrepile (Copeland, 1993: 222). Peta je *prepričanje, da so mreže v nekem smislu odporne na okvare*: če pride do prekinitve nekaterih povezav med enotami, ali celo do uničenja enote, konekcionistična mreža še vedno deluje, vendar se poveča možnost izhodnih napak. Z drugimi besedami, njena učinkovitost opravljanja kognitivnih funkcij pada s številom uničenih enot (dokler uspešno nadomešča manjkajoče podatke s približki, še nekako gre), podobno kot sposobnost izvajanja kognitivnih operacij možganov pada s številom njihovih uničenih nevronov. »V sistemih, ki uporabljajo porazdeljene reprezentacije, lahko izločimo precejšnje število enot, pa bo obnašanje takšnih sistemov še vedno zgolj malenkostno izkrivljeno« (Bechtel, Abrahamsen, 2002: 62). Šesta pa je *vsebinsko naslavljanje spomina*, ki mreži, ko se ne spomni takoj zahtevanega podatka, omogoča dostop do njega preko shranjenih namigov ali ključev, kot to počnemo ljudje, npr. če vemo, da smo knjigo posodili študentu, ne spomnimo pa se kateremu, si pri njegovi identifikaciji pomagamo s tem, da se skušamo spomniti, ali je bil npr. visok, svetlolas, debel itd. »Mreža je lahko narejena tako, da deluje po načelu vsebinskega naslavljanja spomina v smislu, da če je vhodni podatek zgolj *del* zapomnjenega vzorca, mreža kljub temu kot izhodni podatek proizvede popoln zapomnjen vzorec« (Copeland, 1993: 222).

Med dejanskimi nevroni in konekcionističnimi mrežami sicer ni popolne podobnosti in ta z vidika kognicije, kot pravita Bechtel in Abrahamsen (2002), tudi ni potrebna, zaradi česar nekatere razlike med možgani in konekcionističnimi sistemi niso pomembne, npr. različne vrste oziroma oblike bioloških nevronov, ki glede na to opravljajo določene in različne funkcije, česar konekcionistični modeli ne upoštevajo, prispevek nevrotansmitorjev, ki s kemičnim prispevkom premestijo špranjo (sinapso) na stiku med dvema povezavama, o čemer v konekcionističnih mrežah ni govora, in urejena ter specificirana geometrija možganov, ki jo sestavljajo prefinjeno urejeni popolni vzorci nevronske povezave, ki temeljijo na naloženih plasteh vodoravno položenih živčnih vlaken in tako tvorijo različne geometrijske oblike, kar v geometrijsko nedoločenih konekcionističnih sistemih zaman iščemo. Vendar pa vse ni tako rožnato. Strinjam se namreč s Copelandom (1993), da se konekcionistični modeli in dejanski možgani kljub vsemu razlikujejo v eni točki, ki je z vidika nesimbolnega pojasnjevanja naše kognicije s pojmi proženja in neproženja formalnih enot pomembna, in to je proces učenja. Mreža se uči izključno iz izkušenj oziroma števila ponovitev iste naloge, zanjo ne obstaja noben drug način, da bi se dokopala do znanja, kar pa v primeru človeka ne drži. Nekdo z dobrim spominom lahko nekaj zna že po prvem branju in mu v procesu učenja ni treba izgubljati časa in energije z neskončnimi ponovitvami, da bi si pridobil zadostno mero izkušenj. Poleg tega se mreže, ki uporabljajo nadzorovano učenje,

nečesa naučijo izključno s pomočjo učitelja, kar pa za naše možgane ne velja; ti so v tem pogledu veliko bolj samozadostni. Ko se učimo obvladati neko večščino ali spretnost, naši možgani nimajo predhodnega znanja o tem, kaj je pravi izhod, in ne računajo na zunanjega vsevednega znanstvenika, ki bo s specifikacijo poskrbel za proženje ustreznih nevronov, da bi ga proizvedli.

Searle (1990) pa z miselnim eksperimentom, imenovanim »kitajska telovadnica«, izpostavlja nekaj drugega, zaradi česar nobena konekcionistična mreža naj ne bi bila sposobna kognicije. To je njena nezmožnost razumevanja pomenov, s katerimi operira, »kitajska soba«, ¹⁶ zdaj preurejena v telovadnico, pa naj bi to nedvoumno pokazala. Zamislite si telovadnico namesto sobe, v kateri imamo veliko zgolj angleško govorečih ljudi, ki izvajajo iste operacije kot enote in povezave v konekcionistični mreži. Način, kako ljudje v ekipi simulirajo enote v mreži, ponudi Copeland (1993): posredovanje zelenih žetonov sosedom predstavlja vzbujevalno vez, rdečih zaviralno, število posredovanih žetonov enega sosedu drugemu pa predstavlja moč te povezave; trener oskrbi sodelujoče z navodili, v katerih so določene količine žetonov, ki jih je treba predati, in igralci, ki jih morajo dobiti, da bi izvedli neko akcijo (aktivacijski prag), kar predstavlja specifikacijo povezav; da pa bi ekipa dosegla zelene rezultate, igralci v času treninga spreminjajo oziroma prilagajajo navodila skladno s trenerjevimi zahtevami, kar predstavlja način učenja mreže. Searle trdi, da je rezultat na koncu enak kot v sobi, kjer s kitajskimi simboli s pomočjo knjig s pravili manipulira zgolj en človek. V ekipi (mreži) nihče ne govori kitajsko, čeprav se zunanjim opazovalcem tako zdi, in ne obstaja noben način, da bi se celota (mreža) naučila pomena kitajskih besed (Searle, 1990: 22).

Kakorkoli, tudi konekcionistični model – ki ga Beer definira kot »/.../ večplastno mrežo preprostih, nevronom podobnih elementov, ki so naučene številčno vhodno predstavo pretvoriti v številčno izhodno« (Beer, 2000: 96) – naj bi bil sposoben rešiti razlagalno nalogo, tj. odgovoriti na omenjene štiri lastnosti naše kognicije: produktivnost, sistematičnost, kompozicionalnost in sklepalno koherentnost, kot poudarjajo njegovi zagovorniki (npr. Pollack, 1990, Chalmers, 1990, Markič, 2010), kar pa pomeni, da simbolni modeli niso več »edina igra v mestu«. ¹⁷ Ni več potrebno, da obdelava predstav naše spoznave (podatkov) poteka izključno preko programske vodenih operacij s simboli, ki temelji na fizičnem simbolnem sistemu, imenovanem »mentalese«, ampak lahko procesiranje informacij naše kognicije poteka tudi brez tega, tj. kot ekscitiranje in inhibiranje medsebojno povezanih umetnih nevronov. Sicer je res, da močna simbolna hipoteza pušča vprašanje na-

¹⁶ Za miselni eksperiment, imenovan »kitajska soba«, glej Searle (1980/1990), Preston (2002), Bregant (2013).

¹⁷ Za ugovore temu, da lahko konekcionistični modeli kognicije resnično pojasnijo vse štiri lastnosti, ki jim jih pripisujemo, glej npr. Fodor, Pylyshyn (1988), Ramsey, Stich, Garon (1991), Bechtel, Abrahamsen (2002).

čina manipulacije s simboli povsem odprto, pravi zgolj, da so samo univerzalni simbolni sistemi (računalniki) zmožni misliti, a pojava konekcionističnih modelov kljub temu ne preživi. Z njimi se namreč pojavi alternativna razlaga tega, kako možgani funkcionirajo, ta pa ne vključuje nobenih simbolov in programov, ki vodijo in nadzirajo operacije z njimi, ampak zgolj kompleksne mreže med seboj povezanih enot, ki ena drugo vklaplajo in izklaplajo.

Tako imamo naenkrat napravo, tj. možgane, ki so sposobni misliti, niso pa računalniki v klasičnem smislu, kar postavlja trditev zagovornikov simbolnih modelov, da je pojasnitev opisanih štirih lastnosti kognicije možna zgolj v primeru, ko možgane razumemo kot računalnik, na laž, s čimer se spet kot feniks iz pepela dvigne prepričanje, da smo drugačni od njih.

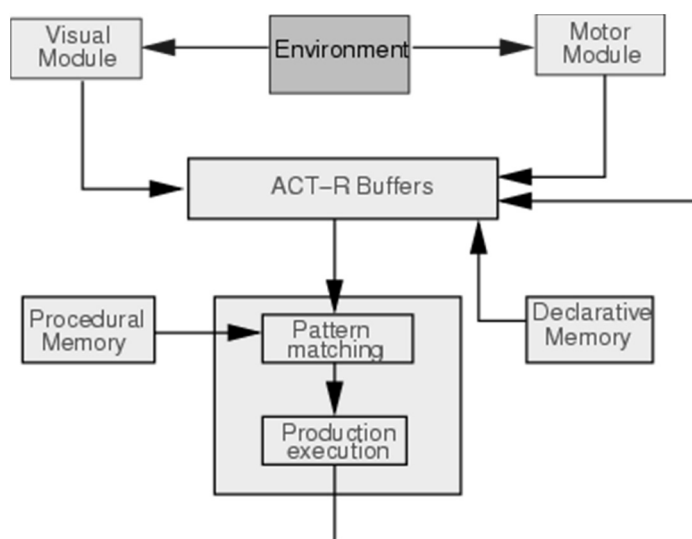
4 Drugi modeli kognicije

Hibridni model

Če na kratko povzamemo prednosti in slabosti simbolnih modelov in konekcionističnih mrež, vidimo, da je prednost simbolnih sistemov njihova uporabnost pri modeliranju višjih kognitivnih operacij, npr. izpeljevanja, njihova slabost pa posnemanje nižjih spoznavnih funkcij, npr. zaznavanja. Dodaten problem predstavlja njihova togost, na kar je opozoril že H. Dreyfus (1979), ki je posledica množice točno določenih navodil za manipuliranje s simboli, ki jim morajo slediti, ko modelirajo našo kognicijo, kar jim preprečuje posnemanje našega vsakodnevnega razmišljanja v nejasnih in odprtih situacijah. Na to, da je s simbolnimi modeli umetna inteligenca dosegla kritično točko in da pri izdelavi inteligentnega stroja ne napreduje več, je v obliki naslednje metafore opozoril S. Dreyfus: »/.../[napredek v tem], da nam bo uspelo narediti inteligentne računalnike, je podoben napredku moža, ki je za to, da bi bil bližje luni, splezal na drevo« (Dreyfus, Dreyfus, 1986: 10). Nasprotno je prednost konekcionističnih sistemov njihova primernost za modeliranje nižjih kognitivnih funkcij, npr. prepoznavanje vzorcev, slabost pa posnemanje nižjih spoznavnih operacij, npr. igranje šaha. Njihova dodatna težava je velika poenostavitev v primerjavi z biološkimi nevronske mrežami, kar postavlja pod vprašaj njihovo primernost za pojasnjevanje kognicije, ki črpa svojo verodostojnost iz dejanske zgradbe in delovanja možganov. Zato ne preseneča, da so se pojavile težnje po izgradnji kognitivnega modela, ki bi združeval prednosti simbolnega in konekcionističnega pristopa ter tako izničil njune posamezne slabosti.

Najbolj znan poskus povezovanja simbolnega modeliranja, ki v splošnem stavi na to, da ni obdelava podatkov nič drugega kot manipulacija z nizi simbolov in njihova transformacija, s konekcionističnim, ki v splošnem razume obdelavo podatkov kot vklapljanje in izklapljanje medsebojno prepletenih formalnih enot (umet-

nih nevronov), je t.i. ACT-R (ang. *Adaptive Control of Thought – Rational*), katerega avtor je Anderson (2007). Takšne modele imenujemo *hibridni*, zanje pa je značilno, da vsebujejo tako simbolno (zaporedno), kot tudi podsimbolno (vzporedno) obdelavo informacij, in ACT-R ni s tega vidika nobena izjema. Sestavljen je iz treh osnovnih elementov: (i) modulov, (ii) medpomnilnikov (ang. *buffers*) in (iii) usklajevalca vzorcev (ang. *pattern matching*). Spodnja slika ponazarja njegovo osnovno zgradbo:¹⁸



Slika 7: Osnovni deli ACT-R

Obstajata dve vrsti modulov: (a) *zaznavno-motorni*, ki je odgovoren za interakcijo z resničnim okoljem oziroma z njegovo simulacijo – tukaj sta najbolj razvita vizualni in manualni, obstajata pa še, npr., govorni in slušni –, in (b) *spominski*, ki je sestavljen iz *deklarativnega spomina* (ali znanja, da) – ta hrani podatke tipa »Ljubljana je glavno mesto Slovenije«, »Človek je sesalec« ali »5+5=10« –, in *proceduralnega spomina* (znanje, kako), ki proizvaja t.i. produkcijska pravila, ki niso nič drugega kot znanje o tem, kako delamo stvari, npr. znanje o tem, kako se vozi, kako se kuha, kako se seštevata itd. Medpomnilniki se uporabljajo za dostop do modulov (razen do proceduralnega spomina) in predstavljajo vmesnik med njimi in sistemom. Njihova vsebina v danem trenutku predstavlja trenutno stanje sistema. Usklajevalec vzorcev išče produkcijsko pravilo, ki se ujema s trenutnim stanjem oziroma zahtevami medpomnilnikov, vsebuje pa navodilo, kako naj se sistem v tem trenutku obnaša oziroma kaj naj naredi. Pomembno je, da je lahko v

¹⁸ Vir: <http://act-r.psy.cmu.edu/about/>.

vsakem koraku, ker deluje ta modul kot zaporedni procesor, aktivno le eno produkcijsko pravilo, ki mu sistem sledi, kar povzroči prilagoditev medpomnilnikov in spremembo stanja sistema. Vprašanje je, kako se ACT-R odloči, katero produkcijsko pravilo bo za izvedbo določenega koraka izbral oziroma uporabil. Pri tem se usklajevalec vzorcev ravna po načelu najvišje koristi za celoten sistem v danem trenutku, ki izhaja iz *teorije racionalne izbire*,¹⁹ imenu ACT-R pa prispeva črko »R«, ki predstavlja »racionalnost«. Skratka, usklajevalec vzorcev izvaja neke vrste analizo stroškov in koristi, t.i. *podsimbolno računanje*, ki vključuje uporabo ustreznih enačb (simbolov) in na osnovi tega brez kakršnegakoli stika z osrednjo nadzorno enoto (ki je sploh ni) sam od sebe izbira najbolj učinkovita produkcijska pravila.

Vidimo, da vsebuje ACT-R različne module, ki izvajajo različne kognitivne naloge, tip procesiranja v njih pa je odvisen predvsem od tega, kakšne vrste so. Glede na to izvaja *simbolno* in *podsimbolno*²⁰ obdelavo podatkov, pri čemer ga prvo dela za simbolni, drugo pa za konekcionistični sistem. Na simbolni značaj ACT-R tako kaže način zgradbe proceduralnega ali implicitnega spomina, ki se nanaša na obvladovanje pravil in postopkov, kjer je znanje predstavljeno kot procesna pravila v obliki parov *če-potem* (ang. *if-then*), ter način delovanja usklajevalca vzorcev, ki z obdelavo informacij korak za korakom funkcionira kot zaporedni procesor. Na njegov konekcionistični značaj pa kaže način zgradbe deklarativnega ali eksplicitnega spomina, ki vsebuje podatke o dejstvih in dogodkih, kjer je znanje predstavljeno v obliki semantičnih mrež, način delovanja njegovega priklica, tj. ali bo neko dejstvo priklicano in kako hitro, kar je odvisno od njegove relevantnosti v danem kontekstu in njegove uporabnosti v preteklosti, ter način izbire najbolj učinkovitega produkcijskega pravila, ki se odvija na osnovi izračuna stroškov in koristi: produkcijska pravila so res predstavljena simbolno in potek njihovega delovanja je vnaprej določen, ni pa vnaprej določeno, katera se bodo izvajala in v katerem primeru (večino časa namreč obstaja množica pogosto med seboj podobnih produkcijskih pravil, ki bi lahko bila aktivna, če bi bila izbrana).

Hibridni model kognicije je tako poskus združiti najboljše obeh svetov, simbolnega in konekcionističnega, v želji dokopati se do bolj celovitega razumevanja našega duha, ki nosilce informacij ali *gradnike*, obravnava kot simbolne predstave, katerih delovanje je določeno s procesnimi pravili tipa *če-potem* in ki so skrite znotraj deklarativnega spomina. Gradniki morajo biti za to, da bi bili učinkoviti, iz njega priklicani in nameščeni v medpomnilnik, kar pa urejajo od simbolne zgradbe gradnikov povsem neodvisne podsimbolne (porazdeljene) enačbe, ki do-

¹⁹ *Teorija racionalne izbire* ali *racionalna teorija ukrepanja* je stališče, ki kot kriterij pri izbiri poti za doseg nekega cilja poudarja in upošteva njeno učinkovitost oziroma koristnost, ki se odraža na delovanju celotnega sistema.

²⁰ Ali *porazdeljeno*.

ločajo aktivacijski nivo vsakega gradnika kot funkcijo njegove uporabnosti v preteklosti in koristnosti za trenutno situacijo.

Dinamični model

Dinamični pristop k modeliranju duha ali dinamična teorija, ki jo Horgan in Tien-son opredeljujeta kot »[s]kupek matematičnih pojmov, tehnik in rezultatov /.../« (Horgan, Tienson, 1992: 27), se pomembno razlikuje od simbolnega in konekcionističnega pogleda. Naše kognicije ne skuša pojasniti s pojmom mehanizma, ampak se pri odgovarjanju na vprašanja, ki se zastavljajo v zvezi z njo, po pomoč spet zateče k t.i. deduktivno-nomološkemu modelu razlage. Tukaj je eksplanandum izpeljan iz eksplanansov, ki vsebujejo vsaj en naravni zakon, zaradi česar sledi nujno, hkrati pa je njegova razlaga tudi znanstvena,²¹ z drugimi besedami, s tem modelom nek pojav razložimo tako, da pokažemo, kako ga je mogoče izpeljati iz začetnih pogojev in naravnih zakonov.

Dinamični pogled ni nekaj resnično novega, saj je z nami že kar lep čas – in sicer od začetkov kibernetike, ki segajo v sredino 20. stoletja – kot alternativa računalniško-algoritemskemu (simbolnemu) in informacijsko-procesualnemu (konekcionističnemu) pristopu, na svoji priljubljenosti pa hitro pridobiva predvsem v zadnjem času (van Gelder, 1998, Spivey, 2007). Dinamični sistemi razumejo psihične procese kot kompleksne in nelinearne procese, ki vzniknejo iz preprostejših stanj ter so sposobni samoorganizacije. Razvijajo se skozi čas in predstavljajo zgolj opis ene izmed možnih alternativ, ki se, če se pojavi v stvarnosti, uresniči oziroma udejanji. Običajen razlog, zakaj naj bi bila dinamična teorija najboljši način za pojasnitev naše kognicije, se skriva v tem, da je zelo uspešna pri razlagi drugih naravnih pojavov, in ker so tudi naši možgani in naša mentalnost del narave, se zdi razumno pričakovati enak uspeh tudi pri pojasnitvi psihičnih procesov, mogoče celo večjega, kot smo ga dosegli s simbolnimi in konekcionističnimi modeli.

Bermudez (2010) opredeljuje dinamične sisteme na splošno kot katerekoli sisteme, ki se razvijajo po določenih zakonih skozi čas, ali, kot pravi Beer (2000), gre za matematične objekte, ki nedvoumno opisujejo, kako se stanje sistema spreminja s časom. Formalno jih lahko zapišemo kot trojko, ki jo sestavljajo urejen časovni niz T , prostor stanja S in razvojni operator ϕ_t , ki spremeni začetno stanje sistema, X_0 , po preteku določenega časa v neko drugo stanje sistema, npr. X_t (Beer, 2000: 92). Horgan in Tienson pravita, da je njihova tipična značilnost to, da vsebujejo specifikacijo dejanskih in hipotetičnih stanj, v katerih se lahko skozi čas znajdejo, tako opisana možna stanja sistema pa oblikujejo njegov abstraktni prostor stanj, v katerem slednje predstavljajo točke. Dinamični model je zato zbirka

²¹ Iz eksplanansa kot začetnih pogojev $C_1, C_2 \dots C_n$ in univerzalnih zakonov $L_1, L_2 \dots L_n$ sledi eksplanandum, npr. pojav P , kot znanstveno razložena deduktivna posledica.

smeri gibanja v času, ki jih ustvarijo različne točke prostora in ki bi jim sistem lahko sledil, zaradi česar ga lahko v nekem smislu identificiramo z nizom trajektorij, ki izhajajo iz stanj prostora, podobno kot lahko simbolni model enačimo z nizom njegovih teoremov (Horgan, Tienson, 1992: 30).

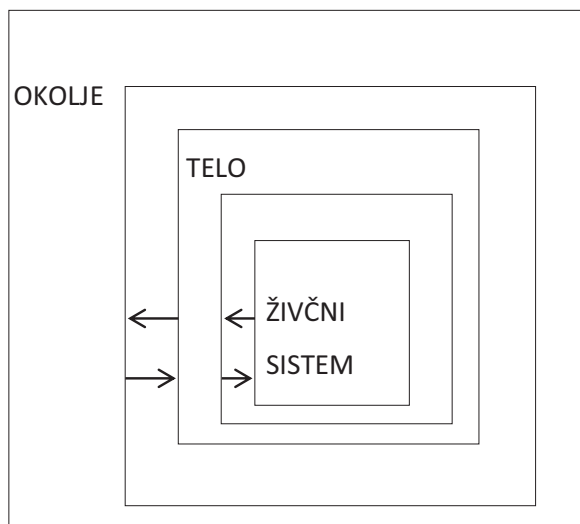
Beer (2000) navaja kot primer dinamičnega sistema *jezik*, za katerega se sicer misli, da je paradigmatični primer simbolnega sistema in tako neprimeren za konekcionistično in kakršnokoli drugo nesimbolno modeliranje, kar pa po njegovem ne drži. Pri tem se sklicuje na Elmana (1990), ki naj bi s posplošenim delta pravilom (ang. *backpropagation*) povratno naprej usmerjeno večslojno mrežo naučil predvidevati naslednjo besedo v stavku, pri čemer je uporabljal nabor 10000 kratkih stavkov, oblikovanih iz 29 samostalnikov in 21 glagolov. Vprašanje je, zakaj njegov poskus pooseblja dinamični in ne konekcionistični model. Beerov odgovor je, da zato, ker sta razumevanje in proizvodnja jezika v svojem bistvu časovna dogodka, kjer predhodne besede močno vplivajo na tolmačenje in izbiro tistih, ki jim sledijo, ta pomembnost časa pa je v Elmanovem modelu zaobjeta v kontekstualnih enotah, ki hranijo stanje skritih enot v prejšnjem časovnem koraku in jim ga v obliki signalov posredujejo v naslednjem ter tako vplivajo na oblikovanje izhodnih odgovorov sistema. »Takšne mreže so neavtonomni časovno-diskretni dinamični sistemi« (Beer, 2000: 91).

Kljub temu, da lahko npr. dinamične sisteme opišemo v simbolnem jeziku, da so, kot smo videli v opisanem Elmanovem primeru, nekateri konekcionistični modeli dinamični in da lahko nekatere dinamične sisteme predstavimo kot konekcionistične mreže, razlike med njimi obstajajo. Beer (2000) trdi, da se nanašajo na teoretični besednjak in slog pojasnitve, ki sta v njih uporabljena pri reševanju problemov, pred katere nas postavlja kognicija. V simbolnem modelu se pojasnitev osredotoča na vsebino uporabljenih reprezentacij in naravo ter učinkovitost uporabljenih algoritmov, v konekcionističnem na arhitekturo mreže, učni algoritem in razvijanje vmesnih porazdeljenih predstav, v dinamičnem pa na zgradbo prostora možnih poti/stanj in notranje in zunanje sile, ki z vzajemnim delovanjem oblikujejo tisto, ki se udejanji skozi čas.

Van Gelder (1998) je bil že pred več kot desetletjem prepričan, da bo kognitivna znanost prejkoslej ugotovila, da so kognitivni akterji dinamični sistemi, ki jih lahko izrazimo kot niz diferencialnih in/ali diferenčnih enačb,²² ki opisujejo, kako se njihova stanja spreminjajo skozi čas. Do sedaj se to sicer še ni zgodilo, saj se zdi, da sta glavna tekmovalca pri pojasnjevanju naše kognicije še vedno simbolni model in konekcionistična mreža, morda še Andersonov ACT-R kot kombinacija (hibrid) obeh, dinamični sistemi, čeprav so združljivi z nečim, kar imenujemo *ute-*

²² Diferencialna enačba je enačba, ki omogoča spremljanje zveznih sprememb sistema, npr. pospeševanje padajočega predmeta, diferenčna enačba pa enačba, ki omogoča spremljanje diskretnih sprememb sistema, npr. spreminjanje biološke populacije v času enega leta.

lešena kognicija (Varela et al., 1991), s čimer mislimo na to, da je kognicija odvisna od izkustev, ki jih priskrbi naše, v okolje umeščeno, telo,²³ kar izgleda razumno in privlačno, pa še vedno stopicajo v ozadju. Spodnja slika prikazuje dinamični pogled na utelešeno kognicijo:²⁴



Slika 8: Živčni sistem, telo in okolje so pojmovani kot dinamični sistemi, ki so v stalni interakciji.

Fussela (2013) meni, da dokler ne bo prišlo do spremembe temeljne paradigme, od tega, da smo digitalni računalniki, ki operirajo kot zaporedni procesorji simbolov, v to, da smo večdimenzionalni računalniki, ki se spreminjajo skozi čas, sestavljeni iz v naravo umeščenega in od nje neločljivega telesa, kognitivna znanost ne bo pripravljena na napredek.

5 Sklep: kaj smo?

Vrnimo se po tem izletu na področje alternativnih kognitivnih modelov k našemu začetnemu vprašanju, tj. ali resnično nismo nič drugega kot računalnik, pri odgovoru na katerega igrata glavno vlogo simbolni in konekcionistični model, ter si oglejmo, do kod nas je naša razprava pripeljala. Videli smo, da močna simbolna hipoteza možgane razume kot manipulatorje s simboli, zaradi česar naj ljudje ne bi bili nič drugega kot računalnik v klasičnem smislu, tj. naprava za obdelovanje predstav, prikazanih s simboli; konekcionistična hipoteza pa kot večplastne pre-

²³ Kognicija je tako posledica interakcije med telesom in okoljem.

²⁴ Slika je iz Beer, 2000: 97.

pletene mreže, zaradi česar naj ljudje ne bi bili računalnik, vsaj ne v tradicionalnem smislu. Vse lepo in prav, problem se pojavi, ko ugotovimo, da lahko računalnik simulira ali reproducira tudi obnašanje oziroma delovanje umetne nevronske mreže, saj se naenkrat pojavi sum, da sta oba, baje tako različna, pristopa k pojasnjevanju naše kognicije, simbolni in konekcionistični, v resnici združljiva. Ali to drži?

Zagovorniki močne simbolne hipoteze odgovarjajo, da vzporedno procesiranje informacij z možnostjo vsebinskega naslavljanja spomina, kjer je medij izmenjave podatkov vzbujanje in zaviranje aktivnosti umetnih nevronov, ne pomeni zavrnitve simbolno-programskega pristopa k umetni inteligenci, saj bi bilo načeloma možno narediti (neklasičen) računalnik, nek hibrid, ki bi bil vzporedni procesor z informacijami, predstavljenimi s simboli, in bi deloval po načelu manipuliranja z njimi. »Vzporedni stroji, ki jih imamo v mislih, bi podatke še vedno hranili v simbolnem zapisu, njihove operacije pa bi še vedno temeljile na strukturno občutljivih manipulacijah z gradniki tega zapisa« (Copeland, 1993: 243). (Nekaj takega danes vsaj deloma pooseblja predstavljeni hibridni model naše kognicije ACT-R s svojo simbolno zgradbo proceduralnega spomina in zaporednim delovanjem usklajevalca vzorcev na eni strani, ter konekcionistično zgradbo deklarativnega spomina in vzporednim načinom izbire najbolj učinkovitega produkcijskega pravila na drugi.) Oglaševalci konekcionistične hipoteze se s tem ne strinjajo in trdijo, da takšen program, ki bi lahko do popolnosti simuliral obnašanje možganov, človeku, ki bi mu ga npr. vsadili v glavo, sam po sebi še ne bi omogočal mišljenja, saj bi slednjega zgolj simuliral, simulacija mišljenja pa še ni dejansko mišljenje, tako kot simulacija mleka še ni pravo mleko. Z drugimi besedami, simboli programa se ne nanašajo na mišljenje, ampak zgolj na enote v mreži, njihov aktivacijski prag in prepletenost, zato pove program zgolj nekaj o mrežah, nič pa o mišljenju. Sutherland (1986) to podkrepi s tem, da četudi bi bilo s programom možno simulirati mišljenje, takšen program ne bi bil misleč nič bolj, kot bi bil vetoven program, ki bi simuliral orkan.

Predpostavimo zaradi argumenta, da simbolni modeli in konekcionistične mreže resnično niso združljivi in da nekaj takega, kot je vzporedni procesor z informacijami, predstavljenimi s simboli, ki bi deloval po načelu manipuliranja z njimi, npr. hibridni sistem, ne premosti ključnih razlik med njima, ki bi obema pristopoma dovoljeval identifikacijo možganov z računalnikom. Kljub temu bi naši možgani in s tem tudi mi sami še vedno lahko bili računalnik, ne sicer v von Neumannovem smislu, ki ga opredeljuje kot zaporednega procesorja informacij, zapisanih kot niz simbolov, pa tudi ne v hibridnem smislu, ki ga definira kot vzporednega obdelovalca podatkov, predstavljenih v simbolnem jeziku. Računalnik lahko namreč razumemo tudi širše, in sicer kot *napravo, ki obdeluje kakršnokoli vrsto predstav (ne zgolj simbolne) na kakršenkoli način (ne zgolj zaporedno)* (Copeland, 1993: 205). Če posvojimo to definicijo, se tudi v primeru, ko mislimo, da kognitivizem in konekcionizem nikakor nista združljiva in da je pravi pristop k

pojasnjevanju naše kognicije drugi, ne pa tudi prvi, ne moremo izogniti temu, da naši možgani, s tem pa tudi mi sami, ne bi bili v nekem tretjem, tj. splošnem, smislu računalnik. »Posledica tega, kar sem do sedaj povedal, je, da četudi bi možgani bili sestavljeni iz vzporedno procesirajočih mrež [v katerih znanje ni predstavljeno posredno, tj. s simboli, ampak neposredno], bi še vedno lahko bili računalnik« (Copeland: 1993: 246). Brezizhodnost situacije, v kateri se znajdejo zgovorniki tega, da možgani s tem pa tudi ljudje nismo računalniki, ponazarja naslednja enostavna dilema:

1. Ali je resnična močna simbolna hipoteza ali konekcionizem.
 2. Če je resnična močna simbolna hipoteza, potem smo računalniki (v von Neumannovem smislu).
 3. Če je resničen konekcionizem, potem smo računalniki (v splošnem smislu).
- ∴ 4. *Ljudje smo računalniki.*

Sklepno misel prepustimo La Mettriju, ki sicer v mislih ni imel računalnikov, je pa že v 18. stoletju materialistično zapisal, da nismo nič drugega kot stroji: »Zaključimo, da je človek v grobem stroj in da je celotno veselje zgrajeno zgolj iz ene substance /.../. Ali bi bila potem organizacija takšnega stroja dovolj, da bi lahko pojasnili vse? Na to bom še enkrat odgovoril z da.« (La Mettrie, 1748/1996: 39, 26).

Brains in the hands of computers

The article defends with the help of the simple dilemma the belief that people, i.e. our brains, are in some particular sense a computer. Copeland (1993) claims that empirical differences between a computer and a brain do not reject the classic approach to the identification of a brain with a computer, i.e. the strong symbol hypothesis grounded in symbolic representation of our knowledge of the world. He argues that these differences merely show that our brains are not the traditional von Neumann computer. The connectionist approach to the explanation of our cognition fails in this as well. Despite applying the non-symbolic way of representing information about the world, which is supposed to be alien to computers, it does not succeed in preventing their identification with brains. We may still be computers, but of some other sort, a sort which is not possible only in principle as it is demonstrated by the most developed hybrid model of our cognition ACT-R, which combines both approaches. And even if we, for the sake of argument, accept the belief that a computer as a symbol manipulator is not compatible with neural networks, we are still not safe to claim that we are not computers in some general way.

Keywords: strong symbol hypothesis, connectionist networks, ACT-R, dynamical systems, language of thought

Literatura:

- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe*. Oxford: Oxford University Press.
- Bechtel, W., Abrahamsen, A. (2002). *Connectionism and the Mind*. Oxford: Blackwell.
- Beer, R. D. (2000). »Dynamical approaches to cognitive science«. *Trends in Cognitive Science*, 3, str. 91–99.
- Bermúdez, J. J. (2010). *Cognitive Science*. Cambridge: Cambridge University Press.
- Block, N. (1980/2007). »Uvod: Kaj je funkcionalizem?«. V Markič, O in Bregant, J. (ur.), *Narava mentalnih pojavov*, Maribor: Aristej, str. 147–166.
- Bregant, J. (2010). »Ali lahko stroj misli?«. *Analiza*, 4, str. 55–72.
- Bregant, J. (2013). »Funkcionalizem, računalniki in 'kitajska soba'«. *Analiza*, 3, str. 67–95.
- Chalmers, D. J. (1990). »Syntactic Transformations on Distributed Representations«. *Connection Science*, 2, str. 53–62.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge: The MIT Press.
- Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.
- Darwin, C. (1859/2009). *O nastanku vrst*. Ljubljana: ZRC SAZU.
- Descartes, R. (1637/2007). *Razprava o metodi*. Ljubljana: ZRC SAZU.
- Dreyfus, H. L. (1979). *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper and Row.
- Dreyfus, H. L., Dreyfus, S. E. (1986). *Mind Over Machine*. New York: Macmillan.
- Elman, J.L. (1990). »Finding Structure in Time«. *Cognitive Science*, 14, str. 179–211.
- Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
- Fodor, J. A., Phylyshyn, Z. W. (1988). »Connectionism and Cognitive Architecture: A Critical Analysis«. *Cognition*, 28, str. 3–71.

- Fusella, P. V. (2013). »Dynamic Systems Theory in Cognitive Science: Major Elements, Applications, and Debates Surrounding a Revolutionary Meta-Theory«. *Dynamical Psychology*, <http://dynapsyc.org/2013/Fusella.pdf>.
- Guid, N., Strnad, D. (2007). *Umetna inteligenca*. Maribor: FERi.
- Haugeland, J. (1986). *Artificial Intelligence: The Very Idea*. Cambridge: The MIT Press.
- Hobbes, T. (1651/2006). *Leviathan (Revised Student Edition)*. Cambridge: Cambridge University Press.
- Horgan, T., Tienson, J. (1992). »Cognitive Systems as Dynamical Systems«. *Topoi*, 1, str. 27–43.
- Kim, J. (1996). *Philosophy of Mind*. Boulder: Westview Press.
- La Mettrie, Julien Offray de (1748/1996). *Machine Man and Other Writings*. Cambridge: Cambridge University Press.
- Lashley, K. S. (1950/1960). »In Search of the Engram«. V Beach, F. A., Hebb, D. O., Morgan, C. T., Nissen, H. W. (ur.), *The Neuropsychology of Lashley*, New York: McGraw Hill, str. 478–505.
- Markič, O. (2010). *Kognitivna znanost*. Maribor: Aristej.
- Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.
- Marx, K. (1885/1997). »Capital II: The Process of Circulation of Capital«. V *Collected Works of Karl Marx and Frederick Engels*, New York: International Publishers.
- McCulloch, W., Pitts, W. (1943/1990). »A Logical Calculus of the Ideas Immanent in Nervous Activity«. V Boden, M. A. (ur.), *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press, str. 22–39.
- McLeod, P., Plunkett, K., Rolls, E. (1998). *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford: Oxford University Press.
- Peruš, M. (2001). *Biomreže, mišljenje in zavest*. Maribor: Satjam.
- Pollack, J. (1990). »Recursive Distributed Representations«. *Artificial Intelligence*, 46, 77–105.
- Preston, J. (2002). »Introduction«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 1–50.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Towards a Foundation of Cognitive Science*. Cambridge: The MIT Press.

- Ramsey, W., Stich, S. P., Garon, J. (1991). »Connectionism, Eliminativism, and the Future of Folk Psychology«. V Ramsey, W., Stich, S. P., Rumelhart, D. (ur.), *Philosophy and Connectionist Theory*, London: Erlbaum, str. 199–228.
- Rosenblatt, F. (1962). *Principles of Neurodynamics*. Washington: Spartan.
- Rumelhart, D. E., McClelland, J. L. (1987). »Learning the Past Tenses of English Verbs: Implicit Rules or Parallel Distributed Processing«. V MacWhinney, B. (ur.), *Mechanisms of Language Acquisition*, Mahwah: Erlbaum, str. 194–248.
- Searle, J. (1980/1990). »Duhovi, možgani in programi«. V Hofstadter, D. R. in Dennet, D. C. (ur.), *Oko duha*, Ljubljana: Mladinska knjiga, str. 361–379.
- Searle, J. (1990). »Is the Brain's Mind a Computer Program?«. *Scientific American*, 262, 20–25.
- Spivey, M. J. (2007). *The Continuity of Mind*. Oxford: Oxford University Press.
- Sutherland, N. S. (1986). »Removing the Mystery from Intelligence«. *The Times Higher Education Supplement*, Feb. 28, str. 6.
- Turing, A. (1950). »Computing Machinery and Intelligence«. *Mind*, 59, str. 434–460.
- van Gelder, T. (1998). »The Dynamical Hypothesis in Cognitive Science«. *Behavioral and Brain Sciences*, 21, str. 615–628.
- Varela, F., Thompson, E., Rosch, E. (1991). *The Embodied Mind*. Cambridge: The MIT Press.
- von Neumann, J. (1958). *The Computer and the Brain*. New Haven: Yale University Press.

Blaž Cotman

William James: Raznolikost religioznega izkustva: Študija človeške narave

Založba Krtina, Ljubljana 2015, str. 464; prev. Sebastjan Vörös

»Sto ljudi, sto čudi« ali Jamesov mnogoplastni izkustveni univerzum

Raznolikost religioznega izkustva (*The Varieties of Religious Experience*, 1902),¹ klasika eminentnega ameriškega psihologa in filozofa Williama Jamesa (1842–1910), ki je nedavno naposled izšla v – zgledno dognanem – slovenskem prevodu Sebastjana Vörösa, ostaja verjetno še dandanašnji, torej več kot stoletje po njenem izidu, najsubtilnejša (psihološka) tematizacija človekovih religioznih občutij, nagnjenj in vzgibov. Gre za skrajno dragoceno, eruditivno in večplastno razpiranje vpogleda v pomemben drobec – mnogovrstne – človeške narave,² ki ga nemalo sodobnih protireligioznih gibanj, vehementno prepričanih, da je religija zgolj in samo »atavistični zdrs v načine mišljenja, ki jih je človeštvo v svojih bolj razsvetljenih fazah že zdavnaj preraslo« (James, 2015: 422)³, spregleduje ali vsaj znatno podceni. James je, nasprotno, izpostavil nespregljiv pomen, ki ga ima religija za »človekovo delovanje in trdoživost« –, celo v tolikšni meri, da jo velja prištevati med »človekove najpomembnejše biološke funkcije« (str. 436).

Jamesovo temeljno delo o religiji, nekaj poglobitvenih poudarkov katerega bom v splošnih obrisih začrtal v nadaljevanju, je sestavljeno iz 20 predavanj, ki jih je avtor med letoma 1901 in 1902 izvedel v okviru slovitih Giffordovih predavanj o

¹ Poslej *Raznolikost*.

² Podnaslov knjige se glasi *Študija človeške narave*.

³ Odslej, če ni posebej označeno, se vse navedbe nanašajo na recenzirano knjigo.

naravni religiji na Univerzi v Edinburgu na Škotskem; v knjižni različici jih uokvirjata kratek predgovor in razjasnjevalni pripis.

Ker meni, da religiji »najbrž ne bomo našli enega samega bistva«, saj gre za »kollektivni naziv« (str. 29), se James v svojih predavanjih osredini na enega izmed njenih – za zasledovano problematiko najrelevantnejših – pomenov, ki »se navezuje na občutke, dejanja in izkustva posameznikov v osami, ki verjamejo, da so v odnosu z nečim, kar sami pojmujejo kot božanstvo« (str. 33).⁴ Najgloblji izvor religioznosti prepozna v občutenju, raznovrstni filozofski, teološki in drugi »zunanji« vidiki religije pa zanj predstavljajo samo nekaj naknadno pridane, ne- kakšne »sekundarne produkte«, ki se najverjetneje sploh ne bi pojavili, če pred tem ne bi bilo religioznih čustev in potemtakem »živega odnosa z nevidnim božanstvom« (str. 386). Dvomi o tem, »da bi brezstrastno intelektualno motrenje univerzuma, izvotljeno duševne potrnosti in hrepenenja po odrešitvi na eni ter mističnega čustva na drugi strani, moglo privedi do nastanka filozofij religije, kot jih poznamo dandanes« (str. 373). Še več, racionalistični intelektualizem, ki se je, kakor opaza, »razpasel v religiji« (str. 375), bi morali po njegovem prepričanju brzovito zavreči, saj slejkoprej uteleša neupravičen »prezir do občutenja« (str. 375) – del našega duševnega življenja, ki ga je ta filozofska drža sposobna zapopasti, je namreč »razmeroma neznaten« (str. 68).

Ni torej pretirano presenetljivo, da je v *Raznolikosti* velikanska večina pozornosti namenjena osebni religiji kot tisti najosnovnejši in najprvobitnejši. Jamesov cilj je predvsem izbrskati in proučiti »goli minimum, ki tvori jedro vseh religij« (str. 433). Ta naj bi bil v religioznem vedênju in občutkih, ki so »skorajda vedno enaki«, čeravno je polje religije »idejno zelo raznoliko« (str. 434). »Če želimo dojeti bistvo religije, bomo trajnejše prvine našli med občutki in vedênjem,« pravi James (str. 434), svoj ostroviden pogled pa posledično najraje upira v »izvorna« religiozna izkustva, na katera naletimo samo »pri posameznikih, za katere religija ni dolgočasna navada, ampak akutna vročica« (str. 12).

Za Jamesovo obdelovanje religioznega doživljanja je precej ključno razlikovanje med dvema vrstama sodb, ki ga inavgurira že v svojem prvem predavanju. Eksistencialne (opisne, razlagalne) sodbe v njem skrbno razmejuje od duhovnih (vrednostnih) sodb ter hkrati opozarja, da prve niso izpeljive iz drugih in obratno. Eksistencialna dejstva »sama po sebi nikoli ne zadoščajo za določitev vrednosti« (str. 12), zaradi česar tudi njegovo lastno motrenje religioznih oblik življenja z biološke in psihološke perspektive v ničemer ne implicira njihovega razvrednotenja.

⁴ Božanstvo pri tem razume kot »izključno tisto prvobitno resničnost, na katero se posameznik čuti primoranega odzvati na svečan in resnoben način« (str. 39).

Na podlagi zarisane distinkcije kritično pretrse »medicinski materializem«, po katerem z razkrinkanjem materialnega in/ali patološkega izvora nekega pojava sočasno spodkopljemo tudi njegov duhovni pomen. Problem takšnega – v srži eliminativističnega – sklepanja tiči v tem, da je »povsem nelogično in samovoljno« (str. 19). James upravičeno izpostavlja, da se v »naravoslovnih in tehniških znanostih [...] nikoli ne primeri, da bi kdo skušal neko mnenje ovreči s sklicevanjem na avtorjevo nevrotično konstitucijo« – »[v]eljavnost vsakokratnega mnenja se na teh področjih dosledno preverja z logiko in eksperimentom« (str. 22). In nobenega razloga ni, da bi bilo z religioznimi prepričanji kakorkoli drugače. To pa pomeni, da bi se veljalo tudi v tem primeru posluževati zdravorazumskih in empiričnih kriterijev, kot so »neposredna jasnost, filozofska razumnost in нравna koristnost« (str. 22). Če se hočemo prebiti do sodbe o pomenu in vrednosti tega ali onega religioznega fenomena, se nikakor ne smemo »zadovoljiti s površnim 'zdravniškorečjem'« (str. 357), temveč moramo nadrobno preiskati njegove celokupne praktične sadove.⁵ Resničnost religije po Jamesovem – paradigmatično pragmatičnem – naziranju »najbolje izpričujejo argumenti, ki temeljijo na njeni koristnosti: na koristih, ki jih daje verujočemu posamezniku, in koristih, ki jih tak posameznik daje svetu« (str. 395).

Pripomniti velja, da je James prepričan, da nas »obširno poznavanje posamičnosti dela veliko modrejše kakor posedovanje abstraktnih formul« (str. 7). Zavaljo tega bralca zasiplje z nepreštevimi konkretnimi primeri, ki mu omogočajo ponazarjanje bolj abstraktnih konstruktov in poant s konkretnejšimi (izkustvenimi) realnostmi, nemalokrat v vsej njihovi raznoterosti in čutni otipljivosti. V knjigi tako kar mrgoli impresivnih pričevanj in poročil, še posebej v tistih njenih razdelkih, ki predočujejo »zdravodušni« in »bolnodušni« značjski tip, spreobrnitev, svetniškost in mistična stanja zavesti. Ko James očrtuje zdravodušje, s katerim označuje nagnjenje, ki »vidi v vsem dobro« (str. 80) ter »žalosti, bolečini in smrti odreka vsakršno pozitivno in dejavno vlogo« (str. 143), se denimo med drugim navezuje na Buckov izčrpen in prelep opis neusahljivega religioznega optimizma pri Waltu Whitmanu kot prvovrstnem primeru »nezmožnosti začutiti zlo« (str. 77). Še bolj pretresljiva in pomenljiva je Jamesova razgrnitev bolnodušja, ki zlo dojema kakor »vseprežemajočo dejanskost sveta, v katerem živimo« (str. 145). Tu razločno nakazuje nepogrešljivost določenih vrst religije za določene značjske tipe in se pri takne tega, kar poimenuje »srčika religioznega problema«, ki jo kar se le da lapidarno oriše z vzklikom: »Pomagajte! Pomagajte!« (str. 142). Pesimistične »bolne duše«, kot sta bila npr. John Bunyan in Lev Nikolajevič Tolstoj, lahko »dosežejo srečo samó, če so rojene dvakrat« (str. 145). Bolnodušni značaj po Jamesovi pro-

⁵ Jeremy Carrette (2002) je v tem kontekstu opozoril, da je *Raznolikost* osrednji tekst za (boljše) razumevanje sodobne psihologije religije, saj osvetljuje njene slepe pege, ki ostajajo – med drugim zaradi trdovratne težnje k pozabljanju vsega preteklega, kar bi utegnilo razkriti negotovost in krhkost disciplinarne vednosti, ter fetišiziranja vsega novega, kar utegne tako ali drugače utrditi hegemonijo (te) znanosti – bolj ali manj spregledane.

dorni koncepciji izrašča iz izrazito nepopolnega, neubranega in heterogenega jaza, ki ga zaznamujejo silovite notranje vihre, religija pa predstavlja »eno od številnih poti, ki vodijo k poenotenju« (str. 153). Zanj je značilno, da trpko bedo, ki je predhodila procesu spreobrnitve, »nemalokrat na preprost, dokončen in uspešen način preobrazi v najglobljo in najtrdovratnejšo srečo« (str. 153), spreobrnjenega posameznika pa navda z dotlej povsem nesluteno ravno »duhovne vitalnosti« (str. 207) in ga porodi v »novo kraljestvo biti« (str. 246). To, da so religije v tem oziru tako nenadkriljivo učinkovite, za Jamesa ni nobeno naključje, saj te zadevajo »posameznikovo skrb za lastno usodo« (str. 422) in so »že od nekdaj v službi trenutno prevladujočih potreb« (str. 287).

Ena izmed nedvomno najbolj vznemirljivih tematik, obravnavanih v *Raznolikosti*, je vloga podzavestnega jaza v religiji. James poudarja, da se pri spreobrnitvi »naš zavestni jaz prepusti na milost in nemilost silam, ki nas presegajo po popolnosti in pripomorejo k naši odrešitvi« (str. 184), in pripominja, da je obstoj teh sil nedvomen tako za religijo kot tudi za psihologijo.⁶ Razlika med njunima pristopoma je le v tem, da jih slednja »imenuje za podzavestne in njihove učinke označi kot posledice 'zorenja' ali 'možganske dejavnosti'« ter s tem »namiguje, da so te sile pravzaprav del naše osebnosti« (str. 185). A to samo po sebi »ne izključuje ideje o nadnaravnem posredovanju« (str. 210), poleg tega pa je popolnoma irelevantno za njihov globlji duhovni pomen in vrednost, ki se pogosto izkazuje v notranjem spokoju, ki ga je deležen »odrešeni« posameznik, in v njegovem korenito predru-gačenem, ljubečem odnosu do sobitij in dejanskosti.

Kot je namigoval že moj dosednji očrt, razum in znanost po Jamesu še zdaleč ne izčrpata razplastene, pluralistične stvarnosti, v kateri prebivamo. Izkušnja »jasno kažejo, da je univerzum veliko bolj mnogostranski, kot je pripravljena priznati katerakoli sekta, tudi znanstvena« (str. 109), religija pa lahko, podobno kot znanost, ponudi »pravi ključ za odklepanje svetovne zakladnice« (str. 109). Avtor v zaključnem poglavju samega sebe celo izrecno označi za supernaturalista in dokaj nedvoumno izrazi svoje (nad)prepričanje, da »preko stika z idealnim v naš svet vstopajo nove sile, ki imajo realne učinke« (str. 448), in da religija pravzaprav »ne meče zgolj nove luči na že znana dejstva«, temveč »tudi sama postulira nova dejstva« (str. 445).

⁶ Najpomembnejše dognanje psihologije, ki »meče povsem novo luč na številne z religioznim življenjem povezane pojave« (str. 202), razpozna v njenem odkritju »zavesti, ki se nahaja zunaj duševnega polja« (str. 202) in se torej izmika našemu zavedanju.

Nabor dilem, problemov in vprašanj, s katerimi se – in nas – sooča James, je, skratka, obilen in nesporno fascinanten, v tej recenziji sem ga lahko samo neznan-sko površ(i)n(sk)o nakazal. Toda vprašanje je, v kolikšni meri mu uspe te dileme, probleme in vprašanja tudi ukrotiti. Menim, da mu to večkrat spodleti. K temu odločilno prispevajo pretirana nejasnost, nezadostna teoretska razdelanost in nekon-sistentnost prenekaterih njegovih izvajanj. Na tem mestu naj omenim le dve, re-cimo temu, spotakljivosti: omejenost Jamesove individualistične, za širši, družbe-no-zgodovinski kontekst zvečine nezainteresirane zastavitve problema in trenje med docela pragmatičnim pristopom začetnih poglavij, kjer se James večkrat eks-plicitno zavzema za vrednotenje religioznega življenja »zgolj na podlagi njegovih rezultatov« (str. 25), ter precej bolj ambivalentno, ontološko in metafizično obte-ženo držo, ki jo privzema v nekaterih drugih, zlasti poznejših, sklepnih pasusih (gl. Rorty, 2004; Gale, 1999). Bolj ali manj posledično tudi epistemski status, ki ga pripisuje mističnim stanjem, v katerih koreninijo religiozna izkustva, ni vedno (dovolj) jasen (gl. Kitcher, 2004) – po eni strani so ta po Jamesu sicer zavezujoča *zgolj in le* za tiste, ki so jih doživeli, po drugi pa naj bi scela spodrezala »zavezu-jočnost nemistične oziroma racionalistične zavesti, ki temelji izključno na razumu in čutilih« ter razprla »možnost, da obstajajo tudi druge ravni resnice« (str. 365). Če ne drugega, se zdi, da bi takšne in podobne izpeljave zahtevale nekolikšne modifikacije in predvsem natančnejšo, filozofsko bolj prefinjeno obrazložitev, kot so jo deležne v knjigi.

Raznolikost je navkljub naznačenim pomanjkljivostim, ki se v najslabšem primeru zazankajo v aporije, nadvse imenitno in navdihujoče delo. Je vsebinsko-tematsko bogato, intelektualno ingeniozno, jezikovno izpiljeno in pripovedno živahno. Na ta način kar najodličneje izpričuje mišljenjsko pretanjenost vélikega psihologa in filozofa, znanstvenika religij in metafizika. Neusmiljeno preizprašuje privajene podobe sveta, nas beza iz dogmatičnega dremeža, kakršnemu se vsaj občasno predaja sleherni med nami, ter vztrajno uči nečesa, česar v splošnem primanjkuje celo v filozofskih logih – kritične odprtosti duha (v neznano, nedoumljivo, pre-sežno ...). Skrajni čas je torej, da se Jamesa (s tem pa kajpak tudi *Raznolikost*) »spet (ali pa sploh) začne brati« (Šimenc, 2002: 224). In, ne nazadnje, pogloblje-no premlevati.

Literatura

- Carrette, J. (2002). »The Return to James: Psychology, Religion and the Amnesia of Neuroscience«. V James, W., *The Varieties of Religious Experience: A Study in Human Nature*, (Centenary Edition), London: Routledge, str. xxxix–lxiii.
- Gale, R. M. (1999). *The Divided Self of William James*. Cambridge: Cambridge University Press.
- James, W. (2002). *Pragmatizem in Volja do verovanja*. Ljubljana: Krtina.
- James, W. (2015). *Raznolikost religioznega izkustva: Študija človeške narave*. Ljubljana: Krtina.
- Kitcher, P. (2004). »A Pragmatist's Progress: The Varieties of James's Strategies for Defending Religion«. V Proudfoot, W. (ur.), *William James and a Science of Religions: Reexperiencing The Varieties of Religious Experience*, New York: Columbia University Press, str. 98–138.
- Rorty, R. (2004). »Some Inconsistencies in James's *Varieties*«. V Proudfoot, W. (ur.), *William James and a Science of Religions: Reexperiencing The Varieties of Religious Experience*, New York: Columbia University Press, str. 86–97.
- Šimenc, M. (2002). »William James, pragmatizem in resnica«. V James, W., *Pragmatizem in Volja do verovanja*, Ljubljana: Krtina, str. 209–224.