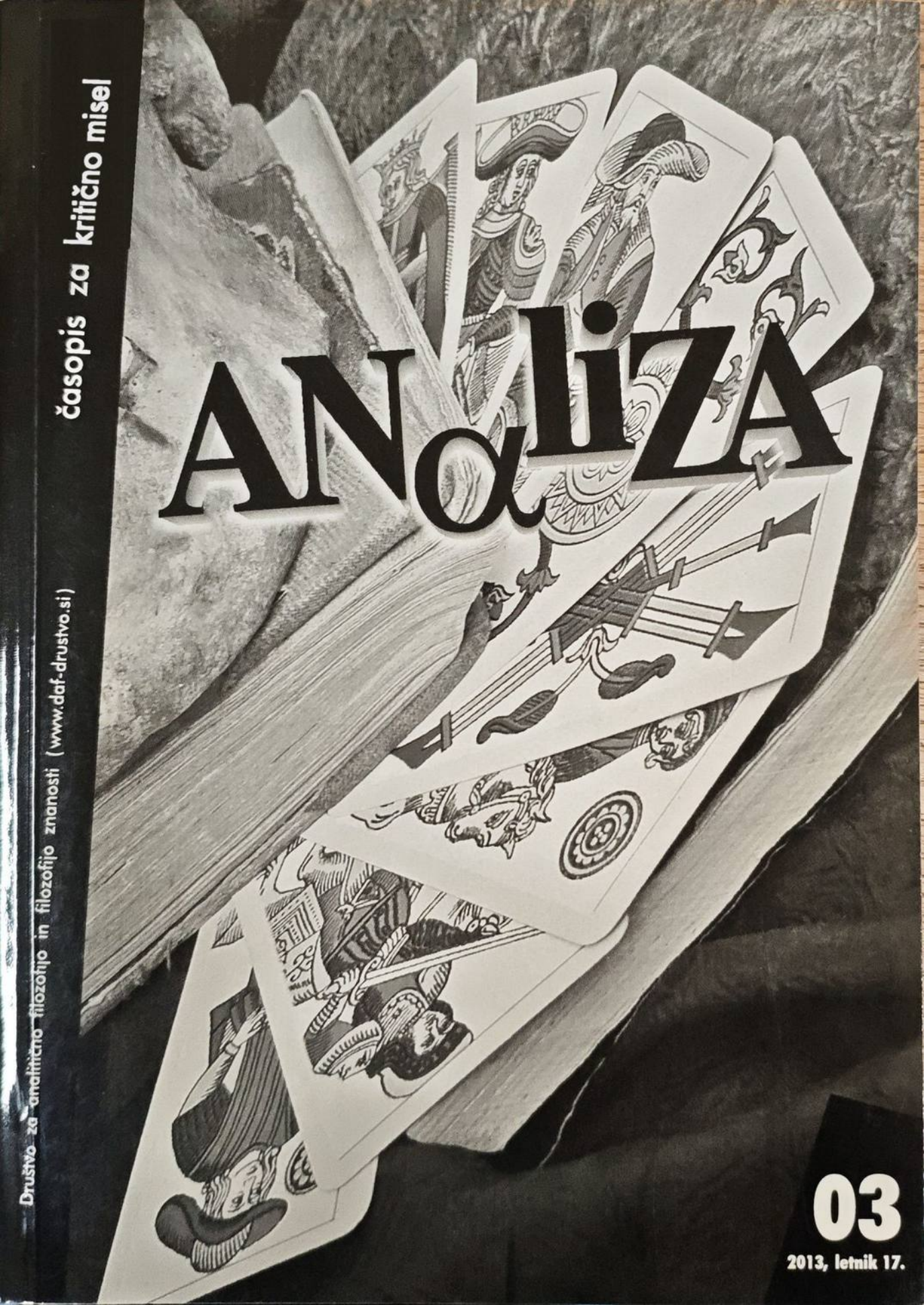
ANaliZA

časopis za kritično misel



Društvo za analitično filozofijo in filozofijo znanosti (www.daf-drustvo.si)

03
2013, letnik 17.

ANαliZA, časopis za *kritično misel*, številka 03, leto XVII., november 2013

Izdajatelj in založnik: Društvo za analitično filozofijo in filozofijo znanosti, Aškerčeva 2, 1000 Ljubljana, www.dafz.si/analiza, e-naslov: analiza@uni-lj.si

Naročila sprejema založba preko spletne strani ali na zgornji naslov. Na leto izidejo štiri številke Analize. Cena številke 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

Glavni & odgovorni urednik: Božidar Kante

Uredniški odbor: Borut Cerkovnik, Friderik Klampfer, Olga Markič, Danilo Šuster, Miklavž Vospernik

Lektor: Borut Cerkovnik

Oblikovanje: K.reG@

Tisk: Grafični studio OK, Maribor, Lavtarjeva ulica 7d, 2000 Maribor, november 2013

Naklada: 200 izvodov

Revija izhaja z denarno pomočjo Javne agencije za knjigo Republike Slovenije.

Revija ANαliZA je indeksirana v The Philosopher's Index.

ISSN 1408-2969

KAZALO

FILOZOFIJA M. DUMMETTA

Andrej Ule
O NEKATERIH DUMMETTOVIH ARGUMENTIH ZA ANTIREALIZEM.....5

Marko Uršič
DUMMETTOVO »DOSEGANJE PRETEKLOSTI«, VPRAŠANJE SUBJEKTA IN MULTIVERZUM.....19

Danilo Šuster
DUMMETT, POGOJNIKI IN FATALIZEM.....31

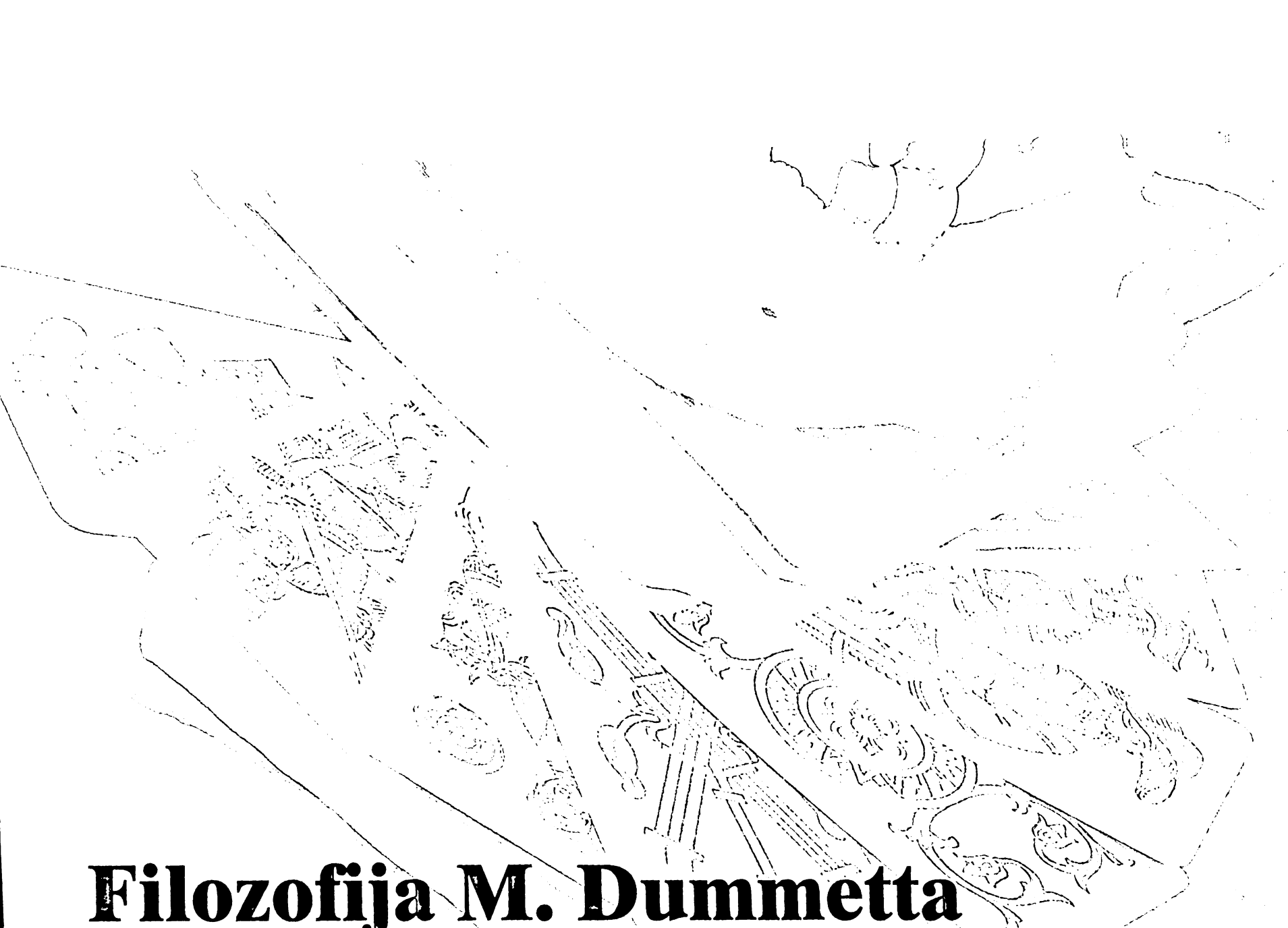
Maja Malec
KRIPKE IN DUMMETT O MOŽNOSTI OBSTOJA SAMOROGOV.....49

FUNKCIONALIZEM IN VPRAŠANJE UMETNE INTELIGENCE

Janez Bregant
FUNKCIONALIZEM, RAČUNALNIKI IN »KITAJSKA SOBA«.....67

IZ ZGODOVINE FILOZOFIJE

Peter Lukan
ARISTOTEL IN POJEM NAKLJUČJA.....99



Filozofija M. Dummetta

Andrej Ule
Filozofska fakulteta Univerze v Ljubljani

O NEKATERIH DUMMETTOVIH ARGUMENTIH ZA ANTIREALIZEM

Analiziram značilne Dummettove argumente za njegovo različico semantičnega antirealizma. Ugotavljam, da so sicer zanimivi, pogosto izzivalni, vendar niso dovolj prepričljivi, kajti problematične stavke lahko po večini preoblikujemo ali tolmačimo tako, da bodisi postanejo povsem deskriptivni ali tako, da se le posredno nanašajo na realnost. Menim, da lahko moder realist najde ustrezne rešitve ali vsaj vizije rešitve zanje.

Ključne besede: antirealizem, semantična nedoločenost, verificljivost, deskriptivnost, kontingenca.

Michael Dummett je verjetno najbolj znamenit zagovornik antirealizma. Dummett je v vrsti svojih spisov in knjig odločno kritiziral realizem, ki ga razume kot tezo, da resničnostni pogoji presegajo možnost verifikacije oz. upravičenja stavkov (Dummett, 1999a, b). Dummet zagovarja svojo različico semantičnega antirealizma, tj. teorije, po kateri nemožnost standardnih semantičnih pogojev za številne, vsaj na pogled deskriptivne stavke, kaže, da se z jezikom praviloma ne nanašamo na realnost, ki biva neodvisno od človeške zavesti in diskurzov, temveč se nanašamo na merila pravilnega ali upravičenega izjavljanja v določenih jezikovnih igrah ali diskurzivnih praksah.

Ali Dummettovi argumenti proti realizmu držijo? Za osnovo si bom vzel njegove argumente za antirealizem iz domnevne odpovedi »realistični«, tj. korespondenčni teoriji resnice v primerih praktične ali načelne nemožnosti verifikacije stavkov. Dummett sklepa, da takšni primeri dokazujejo, da ni mogoče podati resničnostnih pogojev stavkov. V kolikor resničnostni pogoji stavka določajo smisel stavka, potem bi nemožnost določitve resničnostnih pogojev pomenila nedoločenost smisla stavka. Ker pa po Dummettu imamo številne primere stavkov z določnimi smisli brez resničnostnih pogojev, to pomeni, da je določitev smisla stavka z njegovimi resničnostnimi pogoji napačna oz. vsaj preozka. Takšni primeri so npr. stavki o bodočih dogodkih, stavki nekaterih protidejstvenih pogojnikov, stavki kot »Janez je bil pogumen« (če Janez nikoli ni bil v situaciji, ko bi lahko dokazal svoj pogum), stavki o načeloma nejasnih situacijah, nekateri neodločljivi stavki matematike ipd.

To pomeni, da resnice (ali neresnice) stavka ne moremo izenačiti s stanji stvari,

ki se zgodijo, če je stavek resničen (oz. se ne zgodijo, če je neresničen), temveč z nečim drugim, »bolje« dostopnim. Dummett je predlagal verificiranost [verification], kasneje je raje govoril o upravičenosti *[justification]* stavka v okviru ustrezne jezikovne igre. To odmakne resnico stavka in posredno tudi smisel stavka od »sveta«, oz. od celote dejstev, ki obstajajo neodvisno od stavka, govorca in jezika, in jo primakne k jeziku, točneje k merilom »korektne izjavljivosti« izjav v raznih kontekstih. V tem je jedro Dummettovega antirealizma. Menim, da so Dummettovi argumenti, ki jih navaja za svojo različico antirealizma sicer zanimivi, pogosto izzivalni, vendar vsaj zame niso dovolj prepričljivi. Menim, da lahko moder realist najde ustrezne rešitve ali vsaj vizije rešitve zanje, še točneje, Dummettov in podobno Wittgensteinov antirealizem se lahko prikažeta kot dva videza stvari na »realistični« podlagi.

Nepopolni stavki in stavki o kontingentnostih

Nekateri od primerov stavkov z nedoločljivimi resničnostnimi pogoji, ki jih navaja Dummett, se dajo hitro »klasično« rešiti, namreč kot nepopolno izražene propozicije. Npr. stavek »Janez je bil pogumen« je preprosto neustrezno oz. nepopolno izražena misel, saj bi morali vsaj grobo navesti kako situacijo, v kateri naj bi bil Janez pogumen. Res, da takšen stavek pogosto uporabljamo brez navedbe kake relevantne situacije, to pa zato, ker nam kontekst govora pomaga določiti situacijo in tako »dopolniti« stavek do popolnega smisla. In tedaj stavek postane bodisi resničen bodisi neresničen. Stavki o prihodnosti, npr. znamenit »aristotelski« primer »Jutri bo v Piranskem zalivu pomorska bitka«, so resnični ali neresnični, vendar za njihovo resničnost ali neresničnost pač danes ne moremo vedeti, to bomo lahko ugotovili šele jutri. Stavek namreč lahko zapišemo v povsem objektivni obliki, brez vsakega indeksikala, npr. »10. novembra 2012 se v Piranskem zalivu zgodi pomorska bitka«. Stavek se pač nanaša na historično situacijo, kakršna bo 10. novembra 2012. To je sicer v prihodnosti glede na datum izjave tega stavka na današnjem simpoziju, vendar je misel, propozicija, ki jo izreka, brezčasna in se brezčasno nanaša na svet, točneje na celoto historičnih situacij, ki zajemajo zaliv, poimenovan »Piranski zaliv«. V ničemer bistvenem se ne razlikuje od stavka »9. novembra 2012 se v Piranskem zalivu zgodi pomorska bitka«, ki je na našo srečo neresničen. Vsak tovrstni stavek je brezčasno resničen ali brezčasno neresničen.

Nesmiselno pa je govoriti o tem, da so tovrstni stavki »že danes« ali bodo »šele jutri« resnični, ker pripis resnice nima nobenega časovnega določila ali modusa. Od tod, da je navedeni stavek resničen ali neresničen, pa nikakor ne izhaja, da je »prihodnost« vnaprej določena, to bi izhajalo kvečjemu od tod, če bi bilo smiselno reči, da je »že danes resnično« ali »že danes neresnično«, ali bo jutri pomorska bitka.

Ta moja ugotovitev ne zavrača poskusov izgradnje večvrednostnih logik za empirične stavke »radikalne kontingence«, tj. takšne, ki jim ljudje realno ali pa celo načelno ne moremo določiti resničnostne vrednosti. Le resničnostno ekvivalenco moramo tedaj nekoliko drugače formulirati, npr. takole:

»Jutri bo v Piranskem zalivu pomorska bitka« je *resničen glede na zgodovinsko situacijo* v Piranskem zalivu, če in samo če glede na zgodovinsko situacijo v Piranskem zalivu jutri v Piranskem zalivu bo pomorska bitka.

Pri tem določimo resničnostni pogoj za relativno neresničnost takole:

»Jutri bo v Piranskem zalivu pomorska bitka« je *napačen glede na zgodovinsko situacijo* v Piranskem zalivu, če in samo če glede na zgodovinsko situacijo v Piranskem zalivu jutri ne bo pomorske bitke.

Seveda je mogoče, da je zgodovinska situacija v Piranskem zalivu takšna, da nikakor ne moremo ugotoviti, ali je stavek »Jutri bo v Piranskem zalivu pomorska bitka« *glede na dano zgodovinsko situacijo* resničen ali napačen, ne pa resničen ali napačen v absolutnem smislu. Tedaj lahko »dodelimo« temu stavku kako tretjo semantično vrednost, npr. »nedoločeno (glede na podano situacijo)« in uvedemo kako trivrednostno logiko z vrednostmi: resnično, napačno in nedoločeno *[glede na podano situacijo)*. Vendar to ne spremeni načelne situacije, namreč da lahko takšne stavke preformuliramo v povsem enolično »deskriptivne« stavke, ki imajo eno od obeh klasičnih resničnostnih vrednosti, le da so nam le te morda nedosegljive.

Stavki o ljudem nespoznatnih okoliščinah

Podobno velja za mnoge druge stavke o ljudem empirično nedosegljivih okoliščinah, npr. »V našem planetnem sistemu je planet, ki ga nikoli ne bomo odkrili«. Stavek namreč jasno govori o možnem empiričnem dejstvu, ki se tiče našega planetnega sistema. Čeprav tega stavka ne moremo verificirati, ga je vsaj v načelu možno falzificirati, če npr. v doglednem času s pomočjo boljših vesoljskih plovil in posnetkov planetnega sistema ugotovimo, da so ti in ti planeti preprosto *vsi* planeti v našem planetnem sistemu. A seveda je povsem možno, da bomo ljudje kmalu propadli ali ne bomo zmožni več dalje raziskovati planetnega sistema, tako da bo preostal kak neznan planet. V tem primeru je stavek resničen. Dejansko sta le dve *objektivni* možnosti: ali bomo v času trajanja naše civilizacije preiskali naš planetni sistem in odkrili vse planete (ni nujno treba, da tudi vemo za to, da smo odkrili vse planete) ali ne bomo odkrili vseh planetov (zato, ker bomo zaradi raznih razlogov opustili ali ne bomo več sposobni za nadaljnje iskanje planetov). Za pojem resnice je pomembno le to, ne pa, ali kako situacijo ljudje lahko prepoznamo ali ne. V kolikor je navedeni stavek resničen, potem je pač primer ljudem nespoznavne resnice. Stavek ima za nas nedosegljive, vendar ne tudi nedoločljivih resničnostnih pogojev.

Tudi v tem primeru bi lahko opredelili neko modifikacijo klasičnega (in tudi tarskijanskega) pojmovanja resnice, če bi vpeljali naslednje semantične ekvivalence:

Stavek »V našem planetnem sistemu je planet, ki ga nikoli ne bomo odkrili« je *resničen glede na spoznavne priložnosti človeštva*, če in samo če je glede na spoznavne priložnosti človeštva tako, da je v našem planetnem sistemu planet, ki ga nikoli ne bomo odkrili.

Stavek »V našem planetnem sistemu je planet, ki ga nikoli ne bomo odkrili« je *napačen glede na spoznavne priložnosti človeštva*, če in samo če glede na spoznavne priložnosti človeštva ni tako, da je v našem planetnem sistemu planet, ki ga nikoli ne bomo odkrili.

Morda bomo v prihodnosti raziskali naš planetni sistem in ugotovili, da so ti in ti vesoljski predmeti vsi planeti našega planetnega sistema. V tem primeru je navedeni stavek napačen glede na spoznavne priložnosti človeštva, morda pa ne bomo nikoli tako dobro raziskali našega planetnega sistema, da bi lahko trdili, da smo prepoznali vse planete. V tem primeru je navedeni stavek pač resničen glede na spoznavne priložnosti človeštva. Toda problem je, da danes v načelu ne moremo reči, kaj od obojega se bo zgodilo, saj ne poznamo niti spoznavnih možnosti in še manj spoznavnih priložnosti človeštva v prihodnosti. V tem primeru navedeni stavek ostaja semantično nedoločen glede na spoznavne priložnosti človeštva. Lahko mu dodelimo kako »tretjo vrednost«, npr. »nedoločeno«. Vendar to ne vpliva na dejstvo, da lahko bolje formuliramo misel, ki jo skuša stavek izreči, tj. da izrečemo »popolno propozicijo«, kot bi morda dejal Frege. Npr. takole:

V našem planetnem sistemu je planet, ki ga glede na spoznavne priložnosti človeštva nikoli ne bomo odkrili.

Ta stavek je preprosto (objektivno) resničen ali (objektivno) neresničen, kajti v našem planetnem sistemu je bodisi vsaj nek planet, ki ga glede na spoznavne priložnosti človeštva v času trajanja človeštva bodisi nikoli ne bomo odkrili bodisi takšnega planeta ni.,

Dummett meni, da smiselni stavek ne more imeti »nedosegljivih« resničnostnih pogojev, kajti razumevanje stavka terja poznavanje njegovih resničnostnih pogojev. Vendar tu Dummett meša nedosegljivost in nedoločljivost resničnostnih pogojev. Prvo pomeni, da ne moremo ugotoviti, ali lahko podani resničnostni pogoji veljajo ali ne, drugo pa, da resničnostnih pogojev stavka ne moremo korektno oblikovati.

Stavki naravnih zakonov

Splošni stavki, npr. stavki znanosti o naravnih zakonih, so domnevno primer stavkov, o katerih ne moremo reči, da so resnični ali neresnični. Vendar, ali imajo ti stavki sploh ambicijo izreči kak opis dejstev? Mislim, da ne, to so tipični psevdodejstveni stavki, tj. stavki, ki dajejo videz opisnih stavkov, dejansko pa navajajo neka pravila gledanja na svet oz. pravila oblikovanja spoznavne mreže dejstev, ne pa kakih posebnih, npr. splošnih dejstev. Zato pri njih ne moremo niti korektno oblikovati resničnostnih pogojev, v *kolikor jih imamo za opise dejstev*.

Stavek zakona »Vsa telesa, ki se v območju težnostnega privlaka gibljejo zgolj pod vplivom težnosti, se gibljejo enakomerno pospešeno« (ali bolj enostavno »Vsa telesa padajo enakomerno pospešeno«) ni stavek o kakem »splošnem dejstvu« in v tem smislu ni deskriptiven stavek in v tem smislu ni empirično resničen ali neresničen. Pač pa izraža *pravilo gledanja* na fizična gibanja, namreč

pravilo, da *vsa prosta gibanja fizičnih teles* pod vplivom težnosti *obravnavamo kot enakomerno pospešena*. *V* kolikor govorimo o resnici tega stavka, gre za ugotovitev, da korektno izraža *pravilo gledanja na fizična gibanja predmetov* v *moderni fiziki* (tj. fiziki, ki traja od splošnega sprejetja Newtonove mehanike dalje).

V tem smislu je stavek le *posredno resničen*, namreč v toliko, kolikor izraža dejstvo, da ga moderna fizikalna skupnost sprejema kot stavek o zakonu, neposredno pa nima (deskriptivne, empirične) resničnostne vrednosti. Lahko bi dejali, da je stavek *znanstveno potrjen*, ne pa *empirično resničen*. Od tega *stavka zakona* pa moramo razlikovati na pogled enak empirični stavek:

Vsa telesa, ki se v območju težnostnega privlaka v *našem kozmosu* gibljejo zgolj pod vplivom težnosti, se gibljejo enakomerno pospešeno.

Ta splošni stavek govori o vseh pojavih določene vrste, ki so se dogajali, se dogajajo ali se bodo zgodili v našem kozmosu. Ta je seveda empirično resničen ali neresničen, saj ima jasno določljive resničnostne pogoje, ki preprosto veljajo ali ne veljajo, pa čeprav so za ljudi nedosegljivi. Vendar ta splošni stavek ni isto kot stavek zakona. Slednji namreč velja povsem splošno in protidejstveno, tj. brez omembe »našega kozmosa« in za vsak možni fizični kozmos.

Seveda morajo dosedanja opazovanja in meritve »potrjevati« zakon vztrajnosti, a to pomeni le naslednje: če se slučajno zgodi, da naletimo na kako gibanje fizičnih teles pod vplivom težnosti, ki je domnevno prosto, pa se ne giblje pospešeno, bomo raje sklepali na to, da je bila domneva »prostega gibanja« napačna, kot pa, da zakon vztrajnosti ne velja. Šele če bi se zgodilo veliko podobnih »odstopajočih« primerov, bi začeli razmišljati o tem, da zakon morda ne velja oz. ne velja povsem, vendar bi tedaj neizogibno iskali nov, boljši zakon, ne bi pa zavrgli starega kot neresnico. To je razlika med stavki o zakonih in empiričnimi hipotezami. Stavke o zakonih je praviloma težko »ovreči«, pravzaprav jih samo spreminjamo, tako da ohranimo njihova »zdrava jedra« in jih vključimo kot posebne primere v nove, boljše zakone ali znanstvene teorije. Znanstvene hipoteze pa lahko zavržemo, če naletimo na protiprimere. Res pa je tudi, da razlika med splošnimi hipotezami in splošnimi stavki o zakonih ni ostra in lahko naletimo na »vmesne« primere, o katerih še ni odločeno, ali jih imamo za dokaj preverjene hipoteze ali za »neovrgljive« zakone. V kolikor govorimo o resnici stavkov o zakonih, je to le okrajšava za ugotovitev, da ti stavki predstavljajo splošno sprejeta znanstvena »pravila gledanja na svet« v ustreznih znanstvenih skupnostih (včasih še širše, v vseh skupnostih, kjer uporabljajo te stavke za razlago ali napoved vedenja določenih objektov). Stavek naravnega zakona, kot je zakon vztrajnosti, torej izraža neko pravilo, še točneje neko »ravnanje« s fizičnimi objekti za vsakogar, ki obvlada moderno fiziko. Neposredno izraža torej neko dispozicijsko znanje, *znanje kako*, ne *znanje da* (Ryle, 1963). Vsaj od Ryla dalje je znano, da stavkov o »znanju, kako« ne moremo brez nadaljnjega prevesti na »znanje, da«, torej na opise empiričnih ali abstraktnih dejstev in v tem smislu *ne izražajo propozicij*.

Če bi splošni stavek o vztrajnosti jemali zgolj kot splošno hipotezo (kot je npr.

domneval Popper), bi v vsakem primeru njegove uporabe ugotavljali, ali (še) velja ali ne. Tu torej prav tako ne moremo govoriti o opisih dejstev in o empirični resnici ali neresnici hipotez. Stavki empiričnih hipotez ne izražajo propozicij, ki bi bile definitivno resnične ali definitivno neresnične. Lahko govorimo le o večji ali manjši empirični potrjenosti hipotez. Tako imamo v primeru stavkov o vztrajnosti tri vrste stavkov:

> Empirični splošni stavek (prej naveden) o vseh pojavih določene vrste v našem kozmosu.
>
> Stavek empirične hipoteze o vedenju pojavov določene vrste v našem kozmosu.
>
> Stavek znanstvenega zakona vztrajnosti.

Samo prvi od omenjenih stavkov je empirično resničen ali empirično neresničen, vendar nam ta resnica ali neresnica ni empirično dostopna. Drugi stavek je kvečjemu visoko empirično potrjen, ne pa resničen, tretji pa je le v posrednem smislu resničen, namreč zato, ker korektno izraža način gledanja na fizična gibanja v moderni fiziki, dejansko pa je izraz fizikalnega znanja, kako ravnati s fizikalnimi dogajanji tako, da dobimo znanstveno pravilne fizikalne razlage. To pomeni, da ob njihovi uporabi v znanstvenih razlagah dobimo logično veljavne deduktivne ali visoko verjetne induktivne razlage ali napovedi empirično ugotovljivih dejstev.

Mislim, da moje ugotovitve o nedejstveni naravi določenih splošnih stavkov znanosti nikakor ne potrjujejo antirealizma, prej obratno, potrjujejo realizem znanosti. Znanost nedvomno skuša ugotavljati dejstva, tj. postavljati resnične stavke o vedenju pojavov. Toda moramo razlikovati med vedenjem posamičnih pojavov ali posebnih skupin pojavov, ki jih lahko konkretno določamo, opazujemo, merimo, z njimi eksperimentiramo itn., hipotezami, in med pravili znanstvenega gledanja na svet. Slednja se nanašajo na *pojmovne celote vseh pojavov*, ne na ekstenzionalne celote vseh pojavov določene vrste. Sam pravim, da se tovrstni stavki ne nanašajo na svet (kot celoto tega, kar se zgodi), temveč na stvarnost (kot celoto tega, kar se lahko zgodi) (gl. Ule, 1992). Vendar se na tem mestu ne bi spustil v nadaljnjo raziskavo razmerij med svetom in stvarnostjo in na razpravo o morebitni razširitvi pojma resnice tudi na »ujemanje s stvarnostjo«, ne le na »ujemanje s svetom«.

Protidejstveni stavki

Kako pa je z drugimi protidejstvenimi stavki, ki pri Dummettu in drugih antirealistih tudi nastopajo kot primerki za stavke brez definitnih resničnostnih pogojev, čeprav se zdijo resnični ali neresnični? Vzemimo stavek »Če bi Borut Pahor tega in tega dne ne padel s konja, bi naslednji dan sodeloval na veliki proslavi priključitve Primorske«. Mislim, da bi večina od nas ta stavek razumela kot resničen.

Kako lahko sklepamo, da je ta stavek resničen (ali vsaj zelo verjetno resničen),

čeprav je v antecedensu stavek, ki ni resničen in tudi ne more biti resničen, ker govori o nekem preteklem faktu, ki je lahko le takšen, kot je, namreč, da je Borut Pahor tega in tega dne padel s konja. Morda bi kdo rekel, potem je pač ta pogojni stavek trivialno resničen, saj ima neresničen antecedens. Vendar bi lahko postavili tudi kak primerek protidejstvenega stavka podobne vrste, ki pa ga imamo za neresničnega, npr. »Če Borut Pahor tega in tega dne ne bi padel s konja, potem bi naslednji dan odšel na Triglav«, vsaj če upoštevamo, da je Borut Pahor toliko odgovoren politik, ki »nagovarja« zlasti »levo« polovico politične javnosti, in ne bi šel raje na Triglav kot na omenjeno primorsko proslavo. Ta stavek ima prav tako neresničen antecedens, in če bi to določalo resničnost pogojnih stavkov, bi moral veljati za resničnega.

Kaj tedaj določa resničnost prvega in neresničnost drugega protidejstvenega stavka? Stavka nista deskriptivna oz. sta na videz deskriptivna, kajti nanašata se na določene možnosti, ne dejanskosti. Če se nanašata na možnosti, potem izražata našo večjo ali manjšo pripravljenost sprejeti določeno *možnost kot dejansko*. Nanašata se na situacijo v modusu, *kot da bi bile (oz. ne bi bile)*, ne na situacije, *kot so (oz. niso)*. Če že gre za kaka dejstva, potem gre za dejstva naše pripravljenosti ali nepripravljenosti sprejeti določene možnosti kot realne ali nerealne, tj. kot da so dejanske (ali nedejanske). V našem prvem primeru smo (vsaj večina Slovencev) večinoma močno pripravljeni sprejeti kot realno možno situacijo tisto, kjer Borut Pahor ne pade s konja in naslednji dan odide na proslavo na Primorskem, in smo večinoma močno nepripravljeni sprejeti kot realno možno situacijo tisto, kjer Borut Pahor ne pade s konja in naslednji dan odide na Triglav. Razlogi za takšno gledanje na možnosti so v poznavanju sedanje historične situacije (tj. obdobja predvolilne tekme, kjer je Borut Pahor nastopal kot kandidat za predsednika države, ki nagovarja levi in sredinski del volilnega spektra) in poznavanju značaja Boruta Pahorja (vemo, da ni tako strasten planinec, da bi zanemaril svojo politično priložnost in raje šel na Triglav).

V kolikor tovrstne protidejstvene stavke razumemo kot nepopolne stavke, ki se implicitno izrekajo o (ne)pripravljenosti večinskega dela relevantne govorne skupnosti, da sprejme določene možne situacije, *kot da so dejanske, v* toliko jim lahko določimo neke deskriptivne resničnostne pogoje in v toliko so resnični ali neresnični. Če pa jih razumemo preprosto kot stavke o možnostih kot takšnih, potem jim ne moremo pripisati deskriptivnih resničnostnih pogojev in v tem smislu res nimajo resničnostne vrednosti.

Tak primer bi imeli npr. v stavku »Če Borut Pahor ne bi padel s konja, potem bi naslednji dan prijahal na proslavo na Primorskem na konju«. Nič določnega ne moremo reči o tem stavku, kajti možnost, da Borut Pahor ne pade s konja in naslednji dan prijaha na konju na proslavo se verjetno nekaterim Slovencem zdi verjetna, mnogim pa ne, zato je resničnostna vrednost navedenega stavka nedoločena, je zgolj nekognitivni stavek o določenih možnostih, morda je primeren za kako literarno rabo.

Vtis o resničnosti ali neresničnosti pri mnogih protidejstvenikih izhaja iz

njihovega implicitnega ali eksplicitnega razumevanja kot stavkov o pripravljenosti sprejemanja določenih možnih (a nedejanskih) situacij, kot da so dejanske. Ta ugotovitev prav tako ne nasprotuje realizmu, saj protidejstveniki, v *kolikor imajo resničnostno vrednost*, posredno govorijo o določenih dejstvih (dejstvih govorne skupnosti). V kolikor pa nimajo resničnostne vrednosti, pa se očitno ne nanašajo na svet.

Ponuja se možen ugovor zoper ta sklep, namreč da imamo primere protidejstvenih stavkov, ki so nedvomno objektivni in se ne nanašajo na dejstva kake govorne skupnosti. Vzemimo npr. stavek »Če bi potovali s hitrostjo blizu svetlobne, bi nam čas tekel počasneje kot na Zemlji«. Tudi, če pustimo ob strani relativnost in antropomorfnost govora o »toku časa«, lahko damo temu stavku povsem matematično objektiven pomen, če le določimo hitrost našega gibanja glede na Zemljo. Tu se zdi nesmiselno govoriti o odvisnosti tega stavka od kake govorne skupnosti, celo če je to npr. skupnost sodobnih naravoslovcev, ki poznajo teorijo relativnosti.

Vendar moramo upoštevati, da ta stavek lahko preoblikujemo tako, da ga predstavimo kot logično izpeljan zaključek iz nekaterih zakonov relativnostne teorije, npr. iz zakona (ki mu lahko damo povsem precizno matematično obliko, a to tu opuščam), da je potek časa telesa, ki se giblje na drugo telo, toliko počasnejši glede na tok časa pri drugem telesu, kolikor bližja je hitrost gibanja prvega telesa svetlobni hitrosti in ustreznim robnim pogojev. To ugotovitev lahko preoblikujemo v primer korektne deduktivne znanstvene napovedi o gibanju snovnih teles, ki jo lahko v določeni meri celo empirično preverimo (npr. če merimo razlike med potekanjem ur na satelitih oz. vesoljskih postajah in na Zemlji). V tem smislu je protidejstvenik o poteku časa, če bi se gibali blizu svetlobne hitrosti *posredno resničen*, namreč kot potencialno resničen zaključek preverljive znanstvene napovedi. Vendar tudi tedaj lahko pridemo do protidejstvenikov, ki nimajo resničnostne vrednosti, pač zato, ker nimamo načina, da bi jih preoblikovali v vsaj načelno preverljivo znanstveno napoved. Tak je npr. stavek »Če bi potovali v črno luknjo, bi prešli v drugo vesolje«. Čeprav tako meni vrsta piscev znanstvene fantastike in tudi nekaj kozmoloških teoretikov, vsaj (do nadaljnjega) nimamo možnosti za postavitev znanstveno utemeljene napovedi in še manj, da bi jo empirično preverili. Kot rečeno, protidejstveniki niso argument proti znanstvenemu realizmu, kajti če so preverljivi, predstavljajo *posredne izjave o realnosti*, če pa niso, potem nimajo propozicionalne vsebine.

Stavki o meglenostih

Še en pomemben primer empiričnih stavkov z morda nedoločljivimi resničnostnimi pogoji je, ki jih Dummett pogosto navaja in o njih razpravlja. Mislim na stavke o »meglenih« *[vague]* predmetih, lastnostih in situacijah. Vzemimo kot vzorčni primer stavek kupa [sorites]:

Gruča *n kamnov* je kup.

Preprosto ni nobenega določenega merila, po katerem bi lahko za vsako število *n*

dognali, ali je gruča *n* kamnov kup ali ne. Vemo, da lahko hitro zaidemo v paradoks, če izhajamo npr. iz ugotovitve, da je gruča 10.000 kamnov kup, in privzamemo, da če je neka gruča kup in odvzamemo en kamen, potem je preostala gruča še vedno kup. Zdi se, da to vodi do empiričnega protislovja z »opaženimi« kupi, ker je potem poljubno majhna gruča kamnov kup. Gre pa tudi obratno. Če privzamemo, da kaka gruča kamnov še ni kup in ji dodamo en kamen, potem nova gruča še ni kup. Od tod bi sledilo, da poljubno velika gruča kamnov še ni kup. Tudi to se zdi v nasprotju z opažanji, ker imamo očitne primere majhnih kamnitih gruč, ki so kupi. Prvi sklep od tod je, da navedeni stavki o kupu (Gruča *n kamnov* je (ali ni) kup) niso propozicionalni, tj. ne izražajo polne misli, ki bi bila lahko (definitivno) resnična ali (definitivno) neresnična. Res jim ne moremo dati enoličnih resničnostnih pogojev. Toda od tod ne izhaja, da kupov ni, ali da je govor o njih le jezikovna iluzija itn., kot včasih namiguje Dummett (Dummett, 1975, 1999c). Zadeva je nekoliko podobna kot pri stavkih o zakonih. Dejali smo, da ti stavki neposredno izražajo način gledanja na pojave oz. »ravnanje z njimi« v znanstveni skupnosti, posredno pa dejstvo, da ti stavki predstavljajo znanstveno sprejete protidejstvene povezave med fizikalnimi pojavi. Podobno stavki o meglenih predmetih oz. z meglenimi predikati neposredno izražajo način gledanja na predmete in implicirajo določena ravnanja z njimi v določeni jezikovni skupnosti, npr. gledanje na določeno grmado skal kot kup, a ne zato, ker ima ali nima ustrezno množino kamnov, pač pa iz drugih, običajno bolj pragmatičnih razlogov, npr. zato, ker je ukvarjanje s takšno grmado skal naporno početje. Govor o kupih je npr. povsem razumljiv in enoličen v razmerah (in ustreznih jezikovnih igrah), ko je običajno smotrno govoriti o določenem ravnanju z njimi, npr. o grmadenju ali odnašanju kupov kamenja, o povzpenjanju na kupe kamenja ali sestopanju z njih itn., običajno pa ni razumljiv v razmerah, kjer le preštevamo kamne v gručah kamnov ali primerjamo količine kamenja v raznih gručah. Razlika med kupi in ne-kupi je dejansko neostra, zabrisana, a ker velja tihi dogovor, da predikat »biti kup« uporabljamo daleč od meglenih meja, ne prihaja (točneje rečeno, praviloma ne prihaja) do nobenih nejasnosti.

Govor o kupih, npr. »Premetal sem cel kup kamenja« neposredno govori o tem, da je moje ravnanje s kamenjem terjalo od mene res velik napor, posredno pa o tem, da bi vsak razumen član moje govorne skupnosti priznal, da ukvarjanje z omenjeno gručo kamenja dejansko terja precej napora. Do domnevnega paradoksa pride, če od te »pragmatične« jezikovne igre preidemo k bolj »eksaktni« jezikovni igri, npr. takšni, ki vključuje preštevanje kamnov v gručah ali primerjanje količin kamnov v njih. V tem primeru implicitno predpostavljamo jasne meje med pojmi, konkretno med kupom in ne-kupom. To pomeni, da smo voljni potegniti določene meje, kjer jih prej ni bilo, npr. natančnejše meje med gručami kamenja, ki predstavljajo kupe, in gručami, ki ne predstavljajo kupov. A to pomeni, da izraz (predikat), kot je »biti kup«, uporabljamo izven prvotnega konteksta, nekako v smislu »biti kup, večji ali enak določeni vzorčni gruči kamenja, ki je mejni primer kupa« ali celo »biti kup z mejno vrednostjo 500 kamnov«.

Do zmede oz. vtisa paradoksa pride, če pozabimo na te razlike in obravnavamo stavke v tem spremenjenem kontekstu, kot da so stavki prvotnega konteksta. Tako imamo npr. stavek »Gruča *n kamnov* je kup«, ki se zdi kot še en stavek običajnega pragmatičnega konteksta govora o kupih (in ne-kupih), dejansko pa je to stavek iz modificiranega konteksta, ki dopušča (pre)števanje kamnov v kupu in terja jasne meje za pojem kupa. V tem kontekstu implicitno postavljamo določeno ostro mejo med kupi kamenja in ne-kupi kamenja, torej dejansko govorimo o tem, da je navedena gruča kamenja kup, večji ali enak oni gruči kamenja (ki je mejni primer kupa). Paradokse z meglenimi pojmi lahko rešujemo še na kake druge načine, ne zgolj z vlečenjem ostrih meja, npr. tako, da uvedemo modificirano večvrednostno semantiko, npr. tako kot svetuje t. i. metoda supervaluacije, ki jo je svetoval tudi Dummett (1975), lahko pa spremenimo lastnostne predikate v relacijske predikate, npr. namesto »biti kup« govorimo o »biti kvečjemu toliko kupast kot«. Za slednjo seveda velja, da je za vsak *n* res, če je gruča iz *n* kamnov kup, potem je vsaka gruča z manj kamni kvečjemu toliko kupasta kot gruča iz *n* kamnov, vendar od tod ne moremo sklepati, da je takšna gruča kup (Ule, 2008). Stavki o meglenih predikatih ali meglenih predmetih so torej dozdevno deskriptivni, zato ne čudi, da jim v splošnem ne moremo določiti deskriptivnih resničnostnih pogojev. Niso tudi nikakršen argument za ali proti realizmu.

Nedeskriptivni (vrednotni, normativni) stavki

Še veliko primerov stavkov imamo, ki so vsaj dozdevno deskriptivni, vendar jim ne moremo določiti deskriptivnih resničnostnih pogojev. Ne morem pa obravnavati vseh teh primerov, vendar se mi zdi, da bi vedno lahko našli kako »rešitev«, npr. ugotovitev, da so posredno deskriptivni (podobno kot je to primer z znanstvenimi zakoni, s protidejstveniki in stavki o meglenostih). Toda, kako je z izrecno nedeskriptivnimi stavki, ki pa jih vseeno imamo ali čutimo za resnične ali neresnične, npr. z normativnimi ali vrednotnimi stavki? Ali ne gre tam morda za neko zamenjevanje resnice z neko podobno »vrednostjo«, kar pa vendar omogoča prenos te vrednosti v logičnih sklepih? To je povsem mogoče. Vzemimo primer stavka »Vračajte dobro z dobrim«, ki je primer preprostega normativnega stavka, ki ga imamo (vsaj po večini) za resničnega. Vendar noben nabor ali izbor empiričnih dejstev ne ponuja resničnostnega pogoja za ta stavek. Če bi upoštevali le empirična dejstva o človeškem vedenju, bi morali ta stavek verjetno zavrniti, saj se ljudje izjemno pogosto (morda celo bolj pogosto, kot ne) vedejo v nasprotju s tem stavkom. Pa vendar ga imamo trdovratno za etično resničnega.

Lahko bi formulirali neko analogijo z resničnostno ekvivalenco pri deskriptivnih stavkih:

> »Vračajte dobro z dobrim« je *etično veljaven*, če in samo če *naj dobro vračamo z dobrim*.

Pri tem »naj ...« narekuje *etični kontekst razprave*, tj. implicira neko *metateorijo etične veljavnosti*, ki jo sprejema vsak, kdor razpravlja o navedem stavku ali ga

izreka kot resno etično zahtevo. Govorim o etični metateoriji zato, ker pripisuje etično veljavnost ali neveljavnost vsem stavkom iz ustreznega jezika objekta, tj. iz ustreznega diskurza o etičnih vrednostih. Podobno kot pri klasični tarskijanski semantiki bi tudi tu morda lahko začeli z elementarnimi etičnimi ocenami, za katere bi naša etična metateorija enolično poskrbela, da so bodisi etično veljavne bodisi niso etično veljavne. Potem bi ustrezno določili etično veljavnost kompleksnih in kvantificiranih stavkov. Ne bom izvedel tega sicer zanimivega projekta, le ugotavljam, da se mi zdi povsem smiseln. Če bi projekt uspel, bi imeli primer dodeljevanja neke logične vrednosti nedeskriptivnim stavkom, ki bi bila sorodna resničnostnim vrednostim in bi očitno dopuščala, da se ohranja skozi sklepe, ki posnemajo običajne logične sklepe (ali pa bi morda morali veljavne sklepe malce prirediti, a tako, da bi v grobem še vedno spominjali na običajne logične sklepe). Tako bi npr. lahko utemeljili veljavnost sklepa:

Vračajte dobro z dobrim.

Če naj vračam dobro, potem moram upoštevati potrebe drugih ljudi.

Moram upoštevati potrebe drugih ljudi.

Kakorkoli že, normativni in podobni vrednotenjski stavki morda dopuščajo določene pare semantičnih vrednosti stavkov, ki so bolj ali manj podobne »klasični« resnici oz. neresnici. Za primerjavo s klasično resničnostno semantiko je pomembno predvsem vprašanje, ali se v njih ohranja binarnost teh vrednosti. V tem smislu lahko vtis o resničnosti ali neresničnosti teh stavkov kaže na omenjene simulakre klasičnih resničnostnih vrednosti. Ti stavki torej res nimajo resničnostne vrednosti, imajo pa morda »ustrezne« simulakre resničnostnih vrednosti. V kolikor sprejemamo obstoj neke vrste »etičnega sveta«, ki sestoji iz »najstev«, ne pa »dejstev«, bi lahko govorili o etičnem (splošneje o normativno-vrednotenjskem) realizmu, saj bi etično veljavni stavki izražali ustrezna »najstva«, etično neveljavni stavki pa ne bi izražali tovrstnih najstev. Torej ne bi imeli »običajnega« realizma, pač pa etični realizem. Vendar po mojem mnenju za dobro delovanje prej orisane normativno-vrednotenjske semantike ni nujno potrebno sprejemati »etičnega realizma«, čisto dovolj je, če sprejemamo, da je v določeni skupnosti razumno sprejemati določene etične ocene, tj. jim razumno pripisati vrednost »veljavno« in etično metateorijo opredeliti kot celoto dopustnih elementarnih etičnih ocen in njihovih logičnih posledic.

Nepreverljivi (nedokazljivi) matematični in logični stavki

Preostane nam vsaj še en zelo pomemben primer stavkov brez (klasičnih) resničnostnih pogojev, ki jih tako Dummett kot drugi antirealisti navajajo kot paradne primere antirealizma. Gre za nekatere matematične (pa tudi logične) stavke, ki jih ne moremo matematično verificirati ali ovreči. Vsaj pri nekaterih je morda tako, da tega celo v načelu ne moremo storiti, tako da ostanejo trajno nedokazani. Pogosto se navaja stavke, kot je npr. Goldbachova domneva, da je vsako sodo število, večje od 2, enako vsoti dveh praštevil. Ta izjava do danes ni bila dokazana, čeprav se jo na splošno sprejema kot resnično. Vendar, če v

matematiki velja, da je neka trditev resnična natančno tedaj, ko je dokazana, potem Goldbachova domneva pač ni resnična. To se zdi absurdno. A težavo lahko rešimo, če razlikujemo med »do sedaj nedokazanim« in »(efektivno) nedokazanim« stavkom. Od tod, da do sedaj nismo zmogli dokazati Goldbachove domneve, še ne izhaja, da je (efektivno) nedokazana in torej neresnična, saj bi sicer imeli v zgodovini matematike mnogo primerov izrekov, ki bi jih morali imeti za neresnične zgolj zato, ker jih dolgo časa nismo znali dokazati (npr. Fermatov izrek, Poincarejeva domneva, izrek štirih barv itn.). Od tod, da do danes še ne zmoremo dokazati kake matematične izjave, še ne izhaja, da je efektivno nedokazana, torej neresnična. Goldbachova domneva je torej resnična ali neresnična, le dokaza zanjo ali proti njej nimamo.

Situacija bi bila drugačna, če bi se dalo dokazati, da ni mogoč noben dokaz za ali proti tej domnevi.

Resničnostna ekvivalenca za matematične stavke se torej glasi:

> »p« je (matematično) resničen, če in samo če obstaja *dok p*, kjer *»dok p«* pomeni efektivni matematični dokaz stavka *p*.

Če hočemo ohraniti strogo semantično bivalenco moramo privzeti, da je v matematiki običajno tako, da obstaja bodisi dokaz za *p* ali dokaz za *ne-p*, tako da iz dejstva, da ne obstaja dokaz za *p*, potem obstaja dokaz za *ne-p*.

Vendar poznamo nekatere matematične stavke, ki so načeloma neodločljivi. Takšni stavki, kot je npr. Turingov stavek »Ta in ta algoritem privede v končnem številu korakov do rešitve tega in tega problema«, kažejo, da ni vedno tako, da iz dejstva, da ne obstaja dokaz za ta stavek, lahko sklepamo, da obstaja dokaz, da navedeni stavek ne drži.

Poleg tega Gödlovi, Churchovi in Turingovi izreki o nepopolnosti skupno dokazujejo nepopolnost in neodločljivost matematike. V logiki jim ustrezajo stavki logičnih paradoksov. Povsod gre za semantično načeloma neodločljive stavke.

Kaj tedaj? Mislim, da semantično načeloma neodločljivi stavki niso *kognitivno smiselni*, saj sami po sebi ne nudijo nobenega spoznanja, lahko pa dopuščajo razpravo, kaj bi bilo, če bi bili resnični ali ne bi bili resnični. Omogočajo razne pogojne izjave (ki pa niso protidejstvene!, saj ne vemo, če je njihov antecedens neresničen). To ne ukinja pojma resnice v matematiki in logiki, niti je ne zožuje na efektivno dokazljivost stavkov. Iz dejstva, da obstaja dokaz kakega matematičnega stavka, npr. ne nujno izhaja, da imamo efektivni dokaz zanj (pač pa velja obratno). Morda obstaja dokaz Goldbachove domneve, vendar tega dokaza (še) ne poznamo, tj. domneva za nas ni efektivno dokazljiva. Morda nikoli ne bomo dosegli njenega dokaza, vendar je še vedno lahko resnična.

Mislim, da sem s svojimi razlagami »pokril« večino primerov stavkov, ki naj ne bi imeli enoličnih resničnostnih pogojev, oz. domnevno deskriptivnih stavkov, ki jih ne moremo ne potrditi ne ovreči. Pogosto jih lahko preoblikujemo tako, da jim lahko damo kontekstualno specifične resničnostne vrednosti, ki nastopajo kot

dodatne resničnostne vrednosti (npr. v primerih kontingentih stavkov o prihodnosti), posredne resničnostne pogoje (npr. stavki o zakonih), lahko jim določimo neke analogije s klasičnimi resničnostnimi vrednostmi, kot je npr. znanstvena potrjenost (znanstvenih hipotez), normativna veljavnost ali dokazljivost. Te modifikacije nikakor niso v škodo znanstvenemu realizmu, kajti pogosto lahko tovrstne stavke preoblikujemo v standardne deskriptivne stavke. Vsekakor moramo upoštevati, kdaj in v kakšnem oziru kak stavek govori o realnosti in kje se nanjo nanaša le delno in posredno.

Pri nekaterih matematičnih in logičnih stavkih se lahko zgodi, da kak stavek in njegova negacija nimata definitivne semantične vrednosti, kajti dodelitev te vrednosti je vsaj delno odvisna od določenih mentalnih ali fizičnih opravil, ki jih lahko storimo ljudje. Neizvršitev takšnih opravil pri kakem stavku še ni nujno znak tega, da ima takšno vrednost ustrezni negativni stavek, ali da ima stavek »nasprotno« vrednost. To pomeni, da tedaj na splošno ne dobimo bivalence logičnih vrednosti stavkov in da zanje ne veljajo vsi zakoni klasične logike.

V teh primerih bi lahko govorili o nekakšnem antirealizmu, kajti ni treba, da je v svetu kako dejstvo, ki bi jamčilo za to, da ima stavek določeno analogijo z resnico. V teh primerih lahko pritegnem Dummettu v tem, da hipoteza o realizmu ni splošno veljavna a to še ne implicira, da je splošno neveljavna niti to ne, da je po večini neveljavna. Tako sklepam, da Dummettovi argumenti proti realizmu po mojem mnenju na splošno niso veljavni, tako da njegova teza ostaja sicer razumna filozofska domneva ki izziva zagovornike realizma a jih ne more zares »pretresti«.

ON SOME DUMMETT'S ARGUMENTS FOR ANTIREALISM

I analyse some of Dummett's typical arguments for his version of semantic antirealism. I state that his arguments are otherwise interesting, often provocative, but they are not conclusive enough. We can mostly transform the problematic sentences or interpret them in such a way that they become completely descriptive or that they only indirectly refer to reality. I believe that a wise realist can find for them satisfactory solutions or at least visions for such solutions.

Keywords: antirealism, semantic indeterminacy, verifiability, descriptivity, contingency.

Literatura

Dummett, M. (1975). »Wang's paradox«. *Synthese*, 30, str. 301–324.

Dummet, M. (1999a). »Resnica«. *Analiza*, 3–4, str. 61–81.

Dummett, M. (1999b). »Metafizične razprave o realizmu«. *Analiza* 3–4, str. 106–122.

Dummett, M. (1999c). »Wittgensteinova filozofija matematike«. *Analiza* 3–4, str. 82–98.

Ryle, G. (1963). *The Concept of Mind*. London: Peregrine Books.

Ule, A. (1992). *Sodobne teorije znanosti*. Ljubljana: ZPS.

Ule, A. (2008). »The *Sorites* paradox and the change of language games«. V Ule, A., *Circles of Analysis: Essays on Logic, Mind and Knowledge*, Berlin: LIT Verlag, str. 60–79.

Marko Uršič
Filozofska fakulteta Univerze v Ljubljani

DUMMETTOVO »DOSEGANJE PRETEKLOSTI«, VPRAŠANJE SUBJEKTA IN MULTIVERZUM

Izhodišče razprave je vprašanje, kaj pomeni razlikovanje med *spreminjanjem* preteklosti in *vplivanjem* na preteklost, pri čemer je Dummettovo »doseganje preteklosti« [*bringing about the past*] varianta slednjega. Vsakršno dejansko doseganje preteklosti pa postane problematično ne le zaradi možne nekonsistentnosti dogodkov (»dedov paradoks« ipd.), ampak tudi tedaj, če pri »časovnih potovanjih« upoštevamo *zavest*, še posebej spomin tistega *subjekta*, ki naj bi potoval nazaj v preteklost. Navezujoč se na znani Gödlov matematično-fizikalni model »sklenjene časovnostne krivulje« [*closed time-like curve*], ki jo omogoča Einsteinova splošna teorija relativnosti, avtor tega članka dokazuje, da je mogoče ohraniti konsistentnost tega in podobnih modelov »časoplova« samo z uvedbo multiverzuma, posledično pa tudi »dvojnikov« [*counterparts*] tistega subjekta, ki naj bi potoval v preteklost. S tem se odpre tudi epistemološko vprašanje spomina in osebne identitete: ali lahko takšen »razdeljeni subjekt« prehaja iz enega sveta v drugega, iz ene veje multiverzuma v drugo, ter pri tem vendarle ohranja *sebe*?

Ključne besede: Michael Dummett, vnazajšnje povzročanje, Kurt Gödel, David Lewis, multiverzum, subjekt, čas.

Uvod

> Če govorimo o »vnazajšnjem povzročanju« [*backward causation*], bi pravzaprav morali najprej opredeliti samo povzročanje *[causation]* oziroma definirati pojem vzročnosti [*causality*]. Kot vemo, pa to ni ravno lahka naloga, saj imamo že pri Davidu Humu dve bistveno različni opredelitvi vzročnega odnosa. Zadeve se še

bolj zapletejo, če poskušamo uskladiti naše intuitivno razumevanje vzročnosti s fizikalnim, saj lahko vzročni odnos fizikalno interpretiramo na več različnih načinov (npr. kot prenos energije ali kot ohranitev fizikalnih količin ali kot interakcijo sil ipd.), še posebej pa zato, ker so temeljni fizikalni zakoni invariantni glede na smer časa, medtem ko povzročanje, vsaj intuitivno razumljeno, ni invariantno (simetrično) glede na smer časa. Toda namen mojega današnjega prispevka ni analiza in definicija same vzročnosti, ampak le krajša razprava o konceptu domnevnega »vnazajšnjega povzročanja«, ki ga Dummett previdno imenuje *bringing about the past*, kar ne prevajam več s sintagmo »povzročanje preteklosti«, tako kot sem ta pojem prevedel v svoji zgodnji knjigi *Matrice Lagosa* (Uršič, 1987: 133 isl.), ampak »doseganje preteklosti« (še bolj dobesedno bi bilo »nazaj/ali naokoli/prinašanje preteklosti«, tako kot pomeni *to bring about* v navtiki »obrniti ladjo«), Dummett formalno opredeli »doseganje preteklosti« kot vzročni ali vsaj »kvazivzročni« odnos, pri katerem učinek nastopi časovno *pred* svojim vzrokom. Prvo razpravo na to temo je Dummett objavil že leta 1954 pod naslovom *Ali lahko učinek predhodi svojemu vzroku?* in je bila pozneje ponatisnjena v znanem zborniku njegovih razprav *Resnica in druge enigme* (Dummett, 1978: 319–32).

Spreminjanje preteklosti ali »zgolj« vplivanje na preteklost?

Najprej je treba poudariti – čeprav je morda to samoumevno – da pojma »vnazajšnje povzročanje« in »časovno potovanje v preteklost« ne pomenita istega. Modeli časovnega potovanja v preteklost so praviloma povezani s »sklenjeno časovnostno krivuljo« [*closed timelike curve*], ki pa ni nujna v domnevnem vnazajšnjem povzročanju oziroma v Dummettovem »doseganju preteklosti«; po drugi strani pa pri Časovnih potovanjih, vsaj v referenčnem okviru samega »časoplova« *[time machine]* oziroma v *lastnem* času časoplovca, ni nujno vnazajšnje povzročanje (običajno ga v teh modelih sploh ni). Skratka, Dummettovo doseganje preteklosti ne implicira časovnega potovanja v preteklost, niti obratno, časovno potovanje v preteklost ne implicira doseganja preteklosti v pomenu vnazajšnjega povzročanja. Dummett pravzaprav ne govori kaj dosti o možnosti časovnega potovanja, razen izjemoma, na primer ko v članku »Vzročne zanke« (»Causal Loops«) piše o časoplovu iz Wellsovega znanstvenofantastičnega romana (gl. Dummett, 1986: 135 isl.). Jasno pa je, da sta zamisli vnazajšnjega povzročanja (oz. doseganja preteklosti) in Časovnega potovanja vsebinsko povezani, pri čemer so po mojem mnenju najbolj zanimivi tisti primeri oziroma modeli, ki zahtevajo razmislek o obojem, predvsem o vprašanju *kompatibilnosti* časovnega potovanja z doseganjem (ali povzročanjem) preteklosti.

V diskusijah o konsistentnosti modelov časovnih potovanj v preteklost se pogosto poudarja domnevna razlika med *spreminjanjem* nečesa v preteklosti in *vplivanjem* na nekaj, kar se je zgodilo/kar se dogaja/v preteklosti *[influencing on what happens in the past)*. Naj navedem citata dveh znanih avtorjev, Davida

Lewisa (1976) in Paula Horwicha (1987), ki ugotavljata to razliko:

> »Ali lahko časoplovec spreminja preteklost? Zdi se, da ne: dogodki nekega preteklega trenutka se ne morejo spremeniti nič bolj, kot bi se lahko [spremenila] števila. Vendar pa se zdi, da bi [časoplovec] lahko, tako kot kdorkoli drug, storil nekaj, kar bi spremenilo preteklost, če bi to storil.« (Lewis, 1993:141)

> »... bistveno je prepoznati razliko med spreminjanjem preteklosti in vplivanjem na preteklost. Časovno potovanje bi dovolilo človeku, da vpliva na preteklost, ne pa, da jo spreminja. Spreminjanje preteklosti je res logično nemogoče. Vendar pa ni nobenih takšnih protislovij v ideji vplivanja na preteklost.« (Horwich, 1987:116)

Tudi Jan Faye se v članku »Vnazajšnje povzročanje« (»Backward Causation«) pridružuje temu razlikovanju, ko pravi: »Koherentno pojmovanje vnazajšnjega povzročanja zahteva zgolj to, da ima prihodnost lahko vpliv na to, kar se dogaja v preteklosti« (Faye, 2010). Kajti zamisel o vnazajšnjem povzročanju kot spreminjanju nečesa v preteklosti je problematična že zaradi tistega znanega ugovora, ki se ga je prijelo ime »argument preklica« [*bilking argument*]; prvi ga je formuliral Max Black (1956), Anthony Flew pa ga je uporabil v polemiki proti Dummettu: če neko spremembo v preteklosti (A) povzroči časovno poznejši vzrok (B), potem v času med A in B lahko nastopi še neki drugi dejavnik C, ki spremembo A *prekliče*, torej B ne more povzročiti A.

Očitno nekonsistenten primer takšnega preklica je »paradoks deda«, znan predvsem iz članka Davida Lewisa »Paradoksi časovnega potovanja« (1976):

> »Dogodki leta 1921 [ko naj bi Tim ubil svojega deda] so dvojno locirani v Timovem (razširjenem) osebnem času, kakor nadvoz *[trestle]* na železniški progi [tj. most/viadukt na progi v obliki zanke, npr. osmice], toda ‘originalno’ leto 1921 in ‘novo’ leto 1921 sta eno in isto. Če Tim ni ubil deda v ‘originalnem’ letu 1921, potem [drži], da če ubije deda v ‘novem’ letu 1921, mora ubiti in obenem ne ubiti deda v letu 1921 [kar je protislovje] – v enem in edinem letu 1921, ki je oboje, ‘novo’ in ‘originalno’ leto 1921. Logično nemogoče je, da bi Tim spremenil preteklost, tako da bi ubil deda v letu 1921. Torej Tim ne more ubiti deda« (Lewis, 1993: 142).

Malce drugače, bolj preprosto rečeno: Tim v *enem (edinem) in istem letu* 1921 ni ubil (ni mogel ubiti) svojega deda, ker Tima takrat niti še ni bilo na svetu; *če* pa bi takrat že bil na svetu, bi na dogodke v tistem letu lahko samo vplival, ne bi pa jih mogel spreminjati »za nazaj«. – Ampak v katerem času je potemtakem mišljena tista »dvojna lociranost« dogodkov? V katerem Timovem »razširjenem« osebnem času je sploh smiselno razlikovati (četudi zgolj nominalno) med ‘originalnim’ in ‘novim’ letom 1921? Premislimo še enkrat: Lewis pokuša rešiti »paradoks deda« (in podobne nekonsistentnosti pri domnevnem »doseganju

preteklosti«) s tem, da se odpove *spreminjanju* preteklosti: paradoksa ni več, če ne vztrajaš pri spreminjanju preteklosti, ampak se zadovoljiš z možnostjo vplivanja na preteklost. Toda – Analiza 3 2013 21 kako naj bi bilo mogoče *vplivati* na preteklost in je obenem ne spreminjati? Lewis špekulira, da se časoplovec lahko vrne samo v tisto leto, v katerem je *že* bil, namreč takrat, leta 1921, *prvič in edinič*, takrat, ko je vplival na dogodke v tem letu, ni pa s svojim vplivom na takratno dogajanje spreminjal preteklosti, saj je bila tista »preteklost« zanj pravzaprav sedanjost. Toda takšna rešitev je zgolj *tavtološka*, vsebinsko prazna, saj »po ovinku« ne pove nič drugega kot to, da preteklosti ne moreš ne spreminjati, ne vplivati nanjo, pravzaprav niti teoretsko (hipotetično) ne moreš potovati v preteklost, če želiš ohraniti svojo »osebno identiteto«.

Da bi bolje razumeli, v čem je tavtološkost Lewisove rešitve »paradoksa deda«, malce spremenimo časovne koordinate te zgodbice, deda zamenjamo z očetom ter obenem ažurirajmo letnice: najprej recimo, da bo časoplov izumljen v naslednjih letih, najkasneje do leta 2021, in potem poglejmo dve varianti. Varianta (1): Tom, ki je bil rojen, denimo, leta 1951, se pri svojih sedemdesetih letih, leta 2021 vkrca v časoplov in odpotuje nazaj v leto 1944 ter v drugi svetovni vojni ubije svojega očeta – to, skupaj z Lewisovo predpostavko, da obstaja *eno in isto* leto 1944, očitno ni mogoče, kajti, če naj v tem enem in istem letu živi tudi *en in isti* Tom (kakopak drugače, če naj bo leto res eno in *isto*?), ta Tom *ni* živel leta 1944, saj takrat še ni bil rojen: *ergo*, ni mogel ubiti svojega očeta. Tavtologija je v tem, da nekdo, ki ga sploh ni, ne more storiti ničesar, kaj šele ubiti očeta. Torej tudi paradoksa ni, ostaja pa seveda teoretsko vprašanje možnosti samega Tomovega potovanja v leto 1944: *kako* bi lahko en in isti Tom potoval v čas pred svojim rojstvom, a če bi vendarle nekako prišel tja, *kaj* bi zadržalo njegovo roko, da ne bi mogel ubiti očeta itn. Tudi varianta (2) je tavtološka, le nekoliko drugače: Denimo spet, da je bil Tom rojen leta 1951 in da se bo leta 2021 vkrcal v pravkar izumljeni časoplov, ki pa ga ne bi ponesel v čas pred njegovim rojstvom, ampak, recimo, v leto 1969, v vietnamsko vojno, kjer bi kot osemnajstletni mladenič (če naj bo Tom v letih 2021 in 1969 *en in isti)* ubil nekega veterana, za katerega ne bi vedel, da je njegov oče (podobno kot Ojdip). Sledeč Lewisu, bi bilo to dejanje časovnega popotnika, ki pride v leto 1969 iz leta 2021, logično mogoče »zgolj« kot *vplivanje* na preteklost, torej le tedaj, če bi bil Tom že ‘originalnega’ leta 1969 *dejansko* tam, v vietnamski vojni, kamor naj bi se »vrnil« s časoplovom iz leta 2021, se pravi v ‘novo’ leto 1969, pri čemer pa naj bi po Lewisu ‘originalno’ in ‘novo’ leto 1969 bili *eno in isto leto*. Če v tem edinem letu 1969 mladenič Tom *ni* ubil svojega očeta, potem ga ne more (zaradi nekonsistentnosti) ubiti niti kot »povratnik«, kot časovni popotnik, ki pride v leto 1969 iz leta 2021- *razen* če gre za *dve* Tomovi »repliki« [*counterparts*] in posledično, *nota bene*, tudi za dve *različni* leti 1969 (vsaj glede Toma), za ‘originalno’ in ‘novo’ leto 1969. Druga verzija tavtologije je torej v tem, da če nečesa *nisi* storil, potem tega ne moreš storiti pozneje, »za nazaj«, iz prihodnosti (in pri tem ohraniti svojo identiteto).

Seveda pa ne smemo spregledati, da Lewis pravi, da so dogodki »dvojno locirani« v Timovem »razširjenem« *osebnem* času. Če za zdaj pustimo ob strani časoplovčev *lastni* čas v fizikalnem pomenu (tj. v pomenu teorije relativnosti), o katerem bomo govorili v nadaljevanju, ter razumemo njegov »osebni čas« kot njegov *psihološki* čas, tedaj lahko Lewisu, navezujoč se na našo varianto zgodbice (2), zastavimo neko dokaj preprosto vprašanje: kako to, da se časoplovec v letu 2021 (in že prej, preden odpluje v preteklost) ne *spominja* tistega leta 1969, v katerem naj bi »že bil« in v katerega naj bi se iz leta 2021 s časoplovom vrnil, namreč v *eno in isto* leto 1969? Kako to, da se »povratnik« ne spominja vsaj tega, da je *že* bil v vietnamski vojni, tudi če bi tam, recimo, utrpel delno amnezijo zaradi psihične travme očetomora? Splošnejše rečeno, če bi se lahko vrnili v pretekle čase in bi nanje lahko *vplivali* (čeprav jih ne bi mogli spreminjati), potem bi se vendarle morali vsaj deloma spominjati svojih 'originalnih' doživetij iz preteklosti, morali bi jih prepoznati v spominu, tedaj ko bi jih doživljali na 'novo', se pravi, le *navidezno* na novo? Ali pa bi naš »povratek« v preteklost bil omejen samo na že doživeto *osebno* preteklost? Na potovanje zgolj v osebnem spominu? Ampak za takšno potovanje v preteklost ne potrebujemo kakega novega časoplova, za to nam zadostujejo, na primer, albumi starih fotografij. Naj povzamem: če lahko (kot menijo Lewis, Horwich, Faye idr.) na preteklost zgolj vplivamo, ne moremo pa je spreminjati – ali gre pri tako razumljenem vplivanju na preteklost, ki je podobno vplivanju na prihodnost (saj tudi prihodnosti ne moremo v emfatičnem pomenu spreminjati, seveda pa lahko nanjo vplivamo), sploh še za *potovanje* v preteklost? Lewis nam odgovarja, da je tisto leto 1921, kamor potuje (se vrača?) njegov Tim, za Tima *osebno* v prihodnosti, zato se ga pač ne more spominjati ... ampak ta odgovor nikakor ni zadovoljiv, saj nemudoma sledi novo vprašanje: če je časoplovec *zares* odpotoval v preteklost, kateri *dejavnik* pa je izbrisal njegove prejšnje spominske sledi na to *eno in isto* preteklost, ki naj bi bila »zanj osebno v prihodnosti« – in zakaj? Spominske sledi naj bi bile, vsaj tako običajno menimo, nekaj objektivnega, »materialnega«, nekaj, kar zaradi nekega vzroka nastane in kar je, če je, zaradi nekega drugega vzroka izbrisano. Toda *kaj* je tisto, kar izbriše Timove spominske sledi iz njegove sedanjosti, tedaj ko se s časoplovom napoti v preteklost, in *kaj* zbriše spominske sledi njegovega vplivanja na preteklost, ko živi v sedanjosti? Zdi se mi, da pri iskanju odgovora na to vprašanje tavamo v temi.

Tudi Dummett se v dilemi med spreminjanjem preteklosti in vplivanjem na preteklost (ali delovanjem *iz* prihodnosti *v* preteklosti?), podobno kot Lewis, Horwich, Faye idr., bolj nagiba k drugi možnosti, *vplivanju* na preteklost, čeprav eksplicitno ne uporablja izraza *influencing on the past*, ampak le *bringing about the past*, »doseganje preteklosti« (namreč iz sedanjosti ali prihodnosti). Njegovo stališče do tega vprašanja je morda najbolj razvidno takrat, ko govori o »inteligentnih drevesih«, ki zgolj opazujejo svet, kakor je (bil) v preteklosti; ta pasus najdemo v članku »Doseganje preteklosti« (»Bringing About the Past«, 1964) in tudi pozneje v »Vzročnih zankah« (»Causal Loops«, 1986). V prvem

članku pravi:

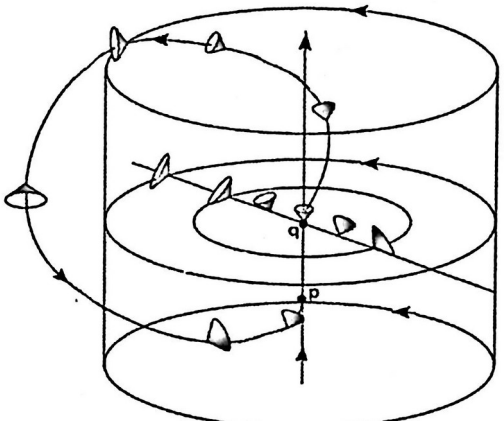
> »Če si zamislimo, da smo kakor inteligentna drevesa, ki opazujejo takšen svet [tj. svet, v katerem je časovni red dogodkov obrnjen] in ki medsebojno komunicirajo, vendar ne morejo posegati v tok dogodkov – je jasno, da bi imeli velike težave pri odkrivanju ustreznih vzročnih razlag za dogodke v okviru procesov, ki so bili *privedli k* njim *[which had led up to them]*. Če pa si [ob tem] zamislimo, da smo [tudi sami] agenti in da so vzročni zakoni, ki vodijo dogodke, takšni, da lahko v njih posegamo, se zdi, da predstava o vnazajšnji vzročnosti [*backward causality*] poraja nesmisle« (Dummett, 1978: 334–35).

Toda problem ne nastane samo tedaj, če bi bil časoplovec v takšnem časovno »narobe svetu« *agent* dejavnosti, ampak tudi (oziroma že) tedaj, če bi bil v njem zgolj *opazovalec*, »inteligentno drevo«, ki bi zgolj z opazovanjem »dosegalo preteklost« (v Dummettovem pomenu *to bring about the past)*, kajti spet se lahko vprašamo: kaj bi se po vrnitvi iz »časovnostne zanke« zgodilo s časoplovčevim *spominom* nanjo, s spominskimi sledmi, brez katerih ne bi bil niti »inteligentno drevo«? Zakaj in kako bi bile te spominske sledi izbrisane ali bi bile dejavne šele v »osebni« prihodnosti, kot špekulira Lewis? Tudi pri interpretaciji Dummetta se moramo znova vprašati, ali je sploh smiselno razlikovati med spreminjanjem preteklosti in vplivanjem na preteklost, tudi če slednje razumemo zgolj kot *vračanje* v preteklost v vlogi opazovalca, »inteligentnega drevesa«. Ja kam pa bi v tem primeru izginile časoplovčeve spominske sledi iz prihodnosti, recimo iz leta 2021, namreč tedaj, ko bi obiskal (se vrnil) v leto 1969 in se tam vêdel le kot inteligentno drevo? Kako naj opazuje (čeprav »zgolj« opazuje) leto 1969 brez spominskih sledi, ki jih prinese iz prihodnosti? Kako naj bi bil v letu 1969 *prvič in edinič*, kot špekulira Lewis, tj. *popolnoma*, brez preostanka, tudi brez »prejšnjih« osebnih in zgodovinskih spominov na to leto, če pa prihaja iz leta 2021? To je, vsaj zame, nerazumljivo. In bojim se, da je »povzročanje preteklosti« v *tem* pomenu res eno izmed tistih »filozofskih odkritij«, ki so kakor *Luftgebäude*, »gradovi v oblakih«, pred katerimi nas je posvaril Wittgenstein v *Filozofskih raziskavah* (§ 118) in kot ugotavlja tudi Hanoch Ben-Yami v članku »Nemožnost vnazajšnjega povzročanja« (*The Impossibility of Backwards Causation*, gl. Ben-Yami, 2007).

Sklenjene časovnostne krivulje, razvejanje subjekta in multiverzum

Toda recimo, da je potovanje v preteklost načeloma kljub vsemu možno, sklicujoč se na modele »sklenjenih časovnostnih krivulj« [*closed timelike curves*] v fiziki. Najbolj znan je še vedno prvi takšen model, ki ga je matematično opisal Kurt Gödel v članku »Pripomba k odnosu med relativnostno teorijo in idealistično filozofijo«, ki ga je podaril Einsteinu ob njegovem 70. rojstnem dnevu (1949). Gödel je odkril takšno rešitev Einsteinovih enačb splošne teorije relativnosti, ki omogoča »sklenjeno časovnostno krivuljo«, ali preprosteje rečeno,

»časovno zanko« [*time loop*]. Einstein baje ni bil ravno navdušen nad tem darilom, saj je, bolj pompozno rečeno, Gödel odkril fizikalni model časoplova, potovanja v preteklost. A čeprav je ta model ne samo konceptualno vprašljiv zaradi možne kavzalne nekonsistentnosti, ampak je tudi precej nerealističen (predvsem zaradi vrtenja univerzuma na način togega telesa), je že zaradi fizikalne *možnosti* potovanja v preteklost vreden naše pozornosti, tudi še dandanes. Poglejmo si torej Gödlov model v preprosti grafični obliki:





Levo: skica Gödlovega »časoplova« (1949). Valj se hitro vrti v nasprotni smeri urinega kazalca, pri čemer se svetlobni stožci v tem vrtečem se prostoru-času *nagibajo* (to nagibanje je ponazorjeno tudi na prečni premici) in s tem omogočijo, da se časovnostna svetovnica (*time-like worldline*) časoplova (ponazorjena s krivuljo), *vrne* v časovno točko-trenutek *p*, ki je v času matične ladje, katere ravna svetovnica je ponazorjena z navpično premico (ker se ladja »giblje« le v času, ne v prostoru), *pred* točko-trenutkom *q*. Povratek časoplova na ladjo (*q*) je njenem času torej *pred* njegovim odhodom (*p*), čeprav sam časoplov potuje v svojem *lastnem* času po vsej svoji svetovnici iz preteklosti v prihodnost. – *Desno*: Kurt Gödel in Albert Einstein v Princetonu leta 1950 (foto: Imagno, Wienbibliothek im Rathaus, na spletu).

V tem modelu je v referenčnem okviru »časoplova« oz. v *lastnem* času »časoplovca« pomembno in upoštevano (samo) dvoje: lokalno ohranjanje smeri vzročnosti od preteklosti k prihodnosti in omejitev hitrosti časoplova, ki ne preseže (niti doseže) svetlobne hitrosti *c*. O zavesti in spominih časoplovca tu ni govora (morda se o tem v matematični fiziki niti ne bi spodobilo govoriti). Toda pravi filozofski problemi se začnejo tedaj, če v Gödlov model vključimo časoplovčevo zavest in spomine, njegovo »osebno identiteto«, tj., če upoštevamo *subjekt* časovnega potovanja.[1] Označimo ta subjekt, ki potuje po »časovni zanki«,

[1] Gödel se je zavedal tega problema: »Zdi se, da to stanje stvari implicira neki nesmisel, saj človeku omogoča, da potuje nazaj v bližnjo preteklost tistih krajev, kjer je on sam živel. Tam bi našel osebo, ki bi bila on sam v nekem svojem zgodnejšem življenjskem obdobju. In zdaj bi tej osebi lahko storil nekaj, za kar s svojim spominom ve, da se ni zgodilo njemu« (cit. po Horwich, 1987:117). Pa še to: pozneje so nastali še drugi fizikalni modeli časoplovov, zlasti na osnovi kvantne mehanike, prehoda skozi »črvine« (*wormholes*) ipd. (za podrobnejše »tehnične« opise teh predlogov gl. Nahin, 1998; za filozofsko analizo pa, med drugim, gl.

s črko *s* in si zamislimo njegovo *vrnitev* na prostorsko mirujočo matično ladjo (tj. gibajočo se v njenem lastnem času le po časovni dimenziji), in sicer vrnitev v trenutku *p*, ki je na navpični časovni osi *pred* trenutkom njegovega odhoda z matične ladje *q*. Časoplovec *s* je v trenutku *p* že/še tam, na matični ladji, saj je z nje odpotoval/odpotuje/bo odpotoval šele v poznejšem trenutku *q*! Torej, recimo, med približevanjem matični ladji ali ko spet stopi vanjo, časoplovec *s* tam uzre samega sebe (morda malce mlajšega, vendar se v njem/sebi prepozna, saj se *spominja* svoje podobe v zrcalu).

Da bi bila ta zgodba konsistentna, namreč glede osebne identitete časoplovca, ga moramo »razcepiti« v dve repliki, dvojnika (*counterparts* v Lewisovem pomenu), označimo ju z indeksoma s_1 in s_2, saj bi sicer *isti* časoplovec *s* hkrati *ne imel spominov* na prepotovano časovno zanko *in jih imel*, kar je očitno protislovje. Tako pa lahko rečemo, da je potovanje po časovni zanki za še-ne-časoplovca s_1 šele v *prihodnosti* (morda v nameri, pričakovanju), za že-časoplovca s_2 pa je že v *preteklosti* (v spominu), in s to podvojitvijo subjekta se izognemo protislovju.

Iteracija zanke pa bi zahtevala nadaljnje cepitve subjekta v mnoštvo subjektov-dvojnikov: s_1, s_2, s_3 … s_n.[2] Toda, če razmišljamo dalje, kako bi bila subjekta-dvojnika s_1 in s_2 na krovu *iste* matične ladje (označimo jo s črko *w*), saj bi potemtakem ta ladja v časovnem intervalu med trenutkoma *p* in *q* tako rekoč »prevažala protislovje«? Da bi ohranili to zgodbo konsistentno, moramo tudi matično ladjo podvojiti na ladji-dvojnici w_1 in w_2, pri čemer prva prevaža še-ne-časoplovca s_1, druga pa že-časoplovca s_2. Iteracija sklenjenih časovnih zank časoplovcev pa bi zahtevala nadaljnja podvajanja v mnoštvo matičnih ladij: w_1, w_2, w_3 … w_n. »Matična ladja« pa je lahko tudi Zemlja ali kar ves Univerzum, in tako za ohranitev konsistentnosti časovnih potovanj v preteklost oz. sklenjenih časovnih zank, po katerih potujejo časoplovci množice:

$$S = \{s_1, s_2, s_3 \ldots s_n\},$$

potrebujemo *multiverzum*

$$W = \{w_1, w_2, w_3 \ldots w_n\}.$$

Drugače rečeno, časoplovec se ne vrača na svojo prvotno matično ladjo, v svoj prvotni svet, ampak se njegov dvojnik »vrača« na neko drugo matično ladjo, v neki drugi svet, v neki drugi univerzum multiverzuma.[3]

Lockwood, 2005: 124–177), vendar za naš kontekst analize »doseganja preteklosti« zadostuje Gödlov model, saj se v poznejših variantah pojavljajo enaki ali vsaj podobni epistemološki problemi.

[2] Ob tem se lahko spomnimo tudi na »sobo preobrazb«, v katero pride na koncu svoje vesoljske poti David Bowman, glavni junak slavnega filma *Odiseja 2001* (gl. tudi Uršič, 2010: 158–174).

[3] Takšna je navsezadnje tudi Lewisova rešitev »paradoksa deda«, kajti njegova hipoteza o »enem in edinem letu 1921« ne more zadovoljivo odgovoriti na protidejstveno vprašanje, kaj pa bi se zgodilo, če bi Tim vendarle (lahko) ubil svojega deda. Zato se Lewis na koncu citiranega članka spraševanje in tudi odgovarja: »Kakšna bi bila razlika, če bi Tim potoval v razvejanem času *[branching time]?* [...] To ni zgodba, v kateri Tim spreminja preteklost. Leto 1921 in poznejša leta vsebujejo dogodke obeh vej, ki nekako koeksistirajo brez interakcije. Pri tem ostaja resnično v vseh osebnih časih Timovega življenja, celo po uboju, da ded živi v eni veji in umre v drugi« (Lewis, 1993: 146). Ta razlaga je seveda v skladu z Lewisovim modalnim realizmom (možnih) svetov.

Multiverzum sicer lahko imenujemo tudi »mnoštvo zgodovin«, tako kakor Dummett, ko v eseju »Resničnost preteklosti« (»The Reality of the Past«, 1969) pravi, da za antirealista –

> »... to pomeni, da ne obstaja samo *ena* pretekla zgodovina sveta: vsaka možna zgodovina, ki je kompatibilna s tem, kar je sedaj *[what now is the case]*, stoji na istem (*stands on an equal footing*, je enakovredna). [...] Toda v vsaki takšni možni zgodovini sveta bo vsaka posamezna trditev [*statement*] o preteklosti bodisi resnična ali neresnična« (Dummett, 1978: 367).

– kar bi v našem kontekstu pomenilo, da se časoplovec vselej »vrača« v neko *drugo* »zgodovino« glede na zgodovine prejšnjih matičnih ladij, v katerih so pluli njegovi dvojniki, zlasti pa v drugo zgodovino glede na tisto *prvotno*, v kateri bi plul, če ne bi odšel na časovno potovanje. Vendar ob tem dvomim, da s preimenovanjem mnoštva univerzumov, tj. multiverzuma v mnoštvo »zgodovin« povemo kaj bistveno novega. Kakorkoli že poimenujemo to mnoštvo, se moramo sami odločiti, ali gre zgolj za različne možne zgodovinske *opise* enega *dejanskega* sveta/univerzuma, ali pa govorimo o *realno* obstoječem multiverzumu. Paradoksom se gotovo lažje izognemo z *realnim* multiverzumom. To so seveda dobro znane teme iz modalne semantike, v našem kontekstu pa so povezane tudi z Dummettovo glavno dilemo med realizmom in antirealizmom; najbrž so razprave o multiverzumu celo odvisne od rešitve te dileme.

Zanimivo je, da se antirealist Dummett v svojem poznem delu »Resnica in preteklost« (»Truth and the Past«, 2004) začne nagibati nazaj k realizmu, kar v našem kontekstu pomeni (hipo)tezo, da sta »preteklost in prihodnost obe resnični« (Dummett, 2004: 86 isl.).[4] S tem se približa »brezčasnim teorijam časa« [*tenseless theories of time*], »realnemu času« iz razprav Hugha Mellorja ali Nathana Oaklanderja (gl. zbornik Oaklander & Smith, 1994), posredno pa seveda tudi relativističnemu brezčasnemu konceptu prostora-časa, »časovni pokrajini« [*timescape*] ipd. Dummett v tej knjigi govori o »mentalni konstrukciji [prostorsko-časovne] mreže« *(ibid*., 48 idr.), ki sega iz sedanjosti nazaj v preteklost, govori o »senčni eksistenci« preteklih dogodkov *(ibid*., 52) ter po svoji dobri stari navadi o vlogi jezika pri vsem tem, obenem pa vztraja pri (sicer revidiranem) antirealizmu in »justifikacionizmu« v teoriji pomena, tudi glede stavkov o preteklih dogodkih, čeprav »je bila teorija pomena modificirana tako, da se je približala realizmu« *(ibid*., 70). Pozni Dummett nas torej vrača k starim dilemam glede (anti)realizma, skoraj bi lahko rekli, da nas vrača k izhodišču, na tisto točko, kjer je pred nekaj desetletji začel premišljevati o »resnici in drugih enigmah«.

Toda vrnimo se h glavnemu poudarku naše razprave, k vprašanju konsistentnosti

[4] »V realnosti ne obstaja razlika med preteklostjo in prihodnostjo; v realnosti ni nobenega *zdaja*. Resnična pa je razlika med eno časovno smerjo in drugo; smer prej-k-pozneje je smer vzročnega učinkovanja« (Dummett, 2004: 86).

»doseganja preteklosti« z uvedbo multiverzuma: tudi če s tem formalno razrešimo paradokse in ohranimo konsistentnost »zgodb« o časoplovcih Timu ali Tomu idr., pri tem ostaja nerešen problem *razcepa* subjekta (zavesti, spomina).[5] Kajti ne izognemo se vprašanju, ali pri teh »časovnih potovanjih« v multiverzumu vendarle lahko govorimo o *prehodu* istega Subjekta iz enega sveta/univerzuma v drugega (se pravi, o ohranjanju kontinuitete Zavesti, Spomina na neki »višji« ravni)? Drugače rečeno: *Kdo* je pravzaprav tisti, ki bi zmogel potovati v preteklost in se tega zavedati, spominjati? *Kdo* bi lahko prehajal skozi »vzporedne« (tj. ločene) svetove/univerzume v Multiverzumu? Zdi se, da ni mogoče določiti »osebne identitete« takšnega Subjekta, da ga ni mogoče »identificirati« (razen če ne sežemo v teologijo). O teh vprašanjih obširneje razpravljam v (Uršič, 2012).

Pa še pomisel za konec: uvedba multiverzuma v dummettovski razpravi o »doseganju preteklosti« je v nekem smislu analogna, čeprav *inverzna*, uvedbi multiverzuma pri konceptualnem reševanju »problema meritve« v kvantni fiziki (Hugh Everett, David Deutsch idr.): *kako* se zgodi to, da se množtvo superpozicij realizira v eni sami dejanskosti – in obratno, pri časoplovnem multiverzumu: *kako* se ena sama zgodovina sveta razcepi v množtvo zgodovin? Predvsem pa – za koga, za kateri *Subjekt* se ta razvejitev zgodi, *če* se zgodi?

DUMMETT'S "BRINGING ABOUT THE PAST", THE QUESTION OF THE SUBJECT, AND MULTIVERSE

The starting point of this paper is the question what does it mean to distinguish between changing of the past and influencing on the past, whereas Dummett's "bringing about the past" is considered as a variant of the latter. However, every actual bringing about the past turns out to be problematic, not only because of possible inconsistency of events ("Grandfather's paradox" etc.), but also if we take into account the consciousness of the "time traveler", especially the memory of the subject who allegedly goes into the past. On the basis of the well-known mathematical model of a "closed time-like curve", proposed by Kurt Gödel in 1949 as a possible solution of Einstein's equations of General Relativity Theory, the

[5] Tudi Lewis v zadnjem odstavku »Paradoksov časovnih potovanj« ugotavlja: »Tim potuje ne le v času, temveč tudi iz ene veje v drugo« (Lewis, 1993: 146) – vendar to samo omeni, ne problematizira pa samega vprašanja *subjekta*.

author of this paper argues that the only way to save the epistemological consistency of this and similar models of "time machines" is by introducing a multiverse, and consequently also to consider the "counterparts" of the subject who allegedly travels to the past. It means that questions concerning memory and personal identity are raised as well, for example: is it possible that such a "divided" subject passes from one universe to another, from one branch of the multiverse to another, and nevertheless remains the same?

Keywords: Michael Dummett, backward causation, Kurt Gödel, David Lewis, multiverse, subject, time.

Literatura

Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.

Dummett, M. (1986). »Causal Loops«. V Flood, R. & Lockwood, M. (ur.), *The Nature of Time*, Oxford: Basil Blackwell, str. 135–169.

Dummett, M. (2004). *Truth and the Past*. New York: Columbia University Press.

Ben-Yami, H. (2007). »The Impossibility of Backwards Causation«. *Philosophical Quaterly*, 57, str. 439–455.

Faye, J. (2010). »Backward Causation«. V *Stanford Encyclopedia of Philosophy*, na spletu: http://plato.stanford.edu/entries/causation-backwards/.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge, Mass.: The MIT Press.

Lewis, D. (1993). »The Paradoxes of Time Travel«. V Le Poidevin, R. & MacBeath, M. (ur.), *The Philosophy of Time*, Oxford: Oxford University Press.

Lockwood, M. (2005). *The Labirinth of Time*. Oxford: Oxford University Press.

Nahin, P. (1998). *Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction*. New York: AIP Press, Springer.

Oaklander, N. & Smith, Q. (1994). *The New Theory of Time*. New Haven, London: Yale University Press.

Uršič, M. (1987). »Povzročanje preteklosti. Problem retroaktivne vzročnosti pri Michaelu Dummettu«. V Uršič, M., *Matrice Lagosa*, Ljubljana: Državna založba Slovenije, str. 133–142.

Uršič, M. (2010). *Daljna bližina neba. Človek in kozmos (Štirje časi – Jesen)*. Ljubljana: Cankarjeva založba.

Uršič, M. (2012). »Mind and Cosmology, Searching for a ‘Third Way’«. V Uršič, M. & Markič, O. & Ule, A., *Mind in Nature, from Science to Philosophy*, New York: Nova Science Publishers, str. 1–78.

Danilo Šuster
Filozofska fakulteta Univerze v Mariboru

DUMMETT, POGOJNIKI IN FATALIZEM*

V člankih »Bringing about the past« (1964) in »Causal Loops« (1986) Dummett zagovarja možnost retroaktivne vzročnosti. Argument proti takšni možnosti je zanj sicer prepričljiv, ampak zmoten. Starodavni »leni argument« proti možnosti vplivanja na prihodnost ima podobno obliko in je prav tako zmoten, vendar zaradi primerjave z argumentom proti vnazajšnji vzročnosti vreden resne obravnave. Fatalist začne z »E nastopi ali ne. Če E ne nastopi, potem ne nastopi, četudi storiš A. Če E nastopi, potem nastopi, četudi opustiš A«. In sklene, da je vsak ukrep A, s katerim bi vplivali na E, neučinkovit ali nepotreben. Dummett (1964) sprva trdi, da smisel pogojnika, v katerem lahko sklepamo iz »E nastopi, torej četudi ne storiš A, še vedno E«, ne dopušča, da izpeljemo nepotrebnost A. Ukrep A bi bil nepotreben, samo če *bi* E nastopil tudi brez A, pogojnika v veznem naklonu pa fatalist nima. Dummett (1986) s to diagnozo ni več zadovoljen, ker temelji na pogojnikih v veznem naklonu, ki se mu zdijo nejasni in slabo razumljivi. Zato predlaga »čistejšo« rešitev, v kateri dokazuje, da se fatalist moti, ko trdi, da ni racionalno ukrepati preventivno. Toda Dummett s takšno rešitvijo samo na zapleten način ponavlja, da je fatalistov sklep nerazumen (kar so vedeli že v antiki, saj so argument razumeli kot *reductio* stoiškega nauka o usodi) ali pa po ovinku spet pripelje problem pogojnikov, tokrat »odprtih« pogojnikov v povednem naklonu. Bolj prepričljiva se mi zdi rešitev, po kateri ob predpostavki, da nastopi želeni učinek E, zares sledi nepotrebnost ukrepa A, zmoten pa je končni sklep konstruktivne dileme na absolutno nesmiselnost (nepotrebnost ali neučinkovitost) vseh ukrepov. Takšno rešitev z uvedbo razlikovanja med razumnim in veljavnim sklepanjem uvaja Stalnaker in z njo pojasni tudi uspešnost Dummettove prvotne rešitve.

* Razširjeni prispevek s simpozija o delu Michaela Dummetta, Društvo za analitično filozofijo in filozofijo znanosti, Ljubljana, 9. 11. 2012.

Ključne besede: Dummett, fatalizem, leni argument, pogojniki, Stalnaker

1.

Po znamenitem Dummettovem programu, ki je bil nekaj let eno od glavnih osišč dogajanja v analitični filozofiji, metafizična vprašanja v principu razrešimo v teoriji pomena. Realizem glede preteklosti je, denimo, odvisen od tega, ali smo pripravljeni propozicijam o preteklosti pripisati realistične pogoje resnice – vsaka propozicija je (določeno) resnična ali neresnična, neodvisno od tega, ali in kako lahko njene pogoje resnice (pre)poznamo. Dummett sam je kolebal glede odgovora na to vprašanje, v svojih zadnjih delih je zagovarjal tako realizem (Dummett, 2004) kot antirealizem (Dummett, 2006). Razlog za njegovo nihanje so morda tudi religiozna prepričanja, ki jih antirealist včasih težko zagovarja. Ena od metafizičnih tem, o katerih je razmišljal, je pojmovna *možnost*, da učinek v vzročni zvezi nastopi časovno *pred* vzrokom. Zagotovo njegovo zanimanje za to temo izvira iz zanimanja za učinkovitost retrospektivne molitve, meni eden od piscev nekrologov.[1] Denimo, da po radiu izvem, da se je pred dvema urama na Atlantiku potopila ladja, na kateri je bil moj sin. Poročajo o nekaj preživelih, jaz pa molim, da je med njimi tudi moj sin. Ali ni ta molitev nesmiselna, se sprašuje Dummett (1964: 341), saj moja namera, glede na to, da je njena izpolnitev v preteklosti, metafizično ne more biti izpolnjena?

Nesmiselnost vplivanja »za nazaj« bi lahko podkrepili z naslednjim argumentom: »Tvoj sin je utonil ali pa ni. Če je utonil, potem tvoja molitev ne more biti uslišana. Če pa ni utonil, potem je molitev odveč. V vsakem primeru pa je molitev brez smisla, saj ničesar več ne spremeni, pa naj je utonil ali ne« (Dummett, 1964: 344–345). Dummett zagovarja *smiselnost* učinkovanja molitve, ki je »uperjena« v času nazaj, zato mora ta »retro« argument zavrniti. No, argument o nemožnosti vplivanja na preteklost ima popolnoma *enako* zgradbo kot argument nemožnosti vplivanja na prihodnost, v antiki znan kot »leni« argument. Dummett (1964: 345) ga navaja v obliki konstruktivne dileme prebivalca Londona, ki v času nemškega bombardiranja med drugo svetovno vojno razmišlja o tem, ali bi stekel v zaklonišče ob zvoku siren, ki oznanjajo letalski napad: »Bomba me bo zadela ali pa me ne bo zadela. Če me bo bomba zadela, potem nima smisla, da bežim. Če me pa bomba ne bo zadela, potem ni nobene potrebe, da tečem v zaklonišče. Previdnosti ukrepi zato nimajo smisla«. Že v antiki se je zdel sklep tega *tipa* argumenta nesmiseln, ker pa naj bi sledil iz stoiške doktrine o usodi, so argument uporabljali kot reductio proti fatalizmu (prim. Bobzien, 1998: 180). V zgodovinskih premenah se je prvotna dialektika izgubila, zato je leni argument danes razumljen kot argument *za* fatalizem. Tako ga razume tudi Dummett (ali pa, denimo, Berčič, 2002). Za Dummetta je fatalizem zgrešen, možnost »vnazajšnjega« vplivanja molitve pač ne, zato skuša pokazati, da na isti način, na katerega zavrnemo leni argument, zavrnemo tudi

[1] Moore, A. W. »Sir Michael Dummett obituary.« *The Guardian*, Wednesday 28 December 2011. Prim. http://www.guardian.co.uk/world/2011/dec/28/sir-michael-dummett, pridobljeno 3. 5. 2013.

argument zagovornika fiksne preteklosti. Tako kot lahko, »lenemu« fatalistu navkljub, vplivamo na prihodnost, lahko vplivamo tudi na preteklost.

Dummett označi fatalizem kot stališče, po katerem je intrinzično absurdno storiti nekaj z namenom, da bi se kasneje pripetilo nekaj drugega. Vsako tako delovanje je nujno brez vsakega smisla.[2] Po tradicionalni filozofski oznaki je fatalizem stališče, da je vse, kar se dejansko pripeti, neogibno in mi na to nimamo *vpliva*. Ne glede na to, kar storimo, bo rezultat isti. Klasični primer je kralj Ojdip, ki izpolni svojo usodo prav z dejanji, s katerimi se ji skuša *izogniti*. Fatalist te vrste (»trdi« fatalist) se mi zdi kot kak razvpiti duhovitež na filozofski zabavi – nihče ga ne jemlje preveč resno, toda vsi ga vabijo, ker želijo z njim razpravljati in ga zavrniti. Obstajajo pa tudi drugačne, »mehkejše« oznake, tako naj bi fatalist trdil »le« to, da nihče nima sposobnosti, da bi ravnal drugače kot dejansko ravna (prim. Bernstein, 2002). Nezmožnost drugačnega ravnanja naj bi veljala tudi v deterministočnem univerzumu in zdaj je težje razločiti posebno grožnjo, ki jo svobodi postavlja fatalizem od grožnje determinizma. Kot pravi Taylor (2007: 16): »Kdor pritrjuje trditvi o univerzalni vzročnosti in teoriji vzročne določenosti vsega človeškega obnašanja, je neke vrste fatalist ali pa bi to vsaj moral biti, če želi biti konsistenten«.

Antični viri (Origen, Ciceron, prim. Bobzien, 1998: 182) leni argument navajajo kot argument *proti determinizmu – fatalizmu* stoiške šole. Nastopa pa v nekoliko drugačni formulaciji:

(1) Če ti je usojeno, da boš okreval po bolezni, potem boš okreval, če se posvetuješ z zdravnikom ali pa se ne posvetuješ.

(2) Toda prav tako: če je usojeno, da ne boš okreval po bolezni, potem ne boš okreval, če se posvetuješ z zdravnikom ali pa se ne posvetuješ.

(3) Ampak usojeno je, da boš okreval po bolezni, ali pa je usojeno, da ne boš okreval.

(4) Zato je posvet z zdravnikom jalov [angl. *futile*].

Ker je (4), nesmiselnost vsakega načrtovanja in presoje med alternativami, nesprejemljiva, naj bi leni argument predstavljal resen napad na stoiški determinizem, meni Brennan (2005: 270). Stoike naj bi prepričal v *neresnico* tretje premise in s tem v zavrnitev univerzalne doktrine o usodi. No, razlika med tezo o usodi in tezo determinizma pri stoikih ni povsem jasna, tako naj bi bili po Brennanu stoiki prepričani, da je za vsak dogodek, ki se pripeti v vesolju, usojeno, da se pripeti, *in* determinirano, da nastopi (Brennan, 2005: 235), Bobzien (1998) pa uporablja za stoike kar oznako *deterministi usode* [angl. *fate-determinsts*].

Razlikovanja (trdi/mehki fatalizem, determinizem) so pomembna, saj raziskave

[2] »I here characterize fatalism as the view that there is an intrinsic absurdity in doing something in order that something else should subsequently happen; that any such action – that is, any action done with a further purpose – is necessarily pointless« (Dummett, 1964:345).

(prim. Baumeister et al., 2011) sodobnih eksperimentalnih filozofov kažejo, da veliko ljudi »z ulice« zares enači fatalizem (kot življenjsko držo) in determinizem (za vsak dogodek obstajajo predhodni pogoji tako, da je njihov nastop zadostni pogoj za nastop tega dogodka). Dummettova oznaka je blizu tradicionalnemu pojmovanju fatalizma kot drže vdanosti v usodo, ker naj bi bili dogodki, ki šele prihajajo, neodvisni od naših namer in odločitev. Že antični viri so argument s takšnim sklepom označili kot sofizem. To naj bi pokazal Hrizip v svoji obrambi stoiškega nauka o usodi. Leni argument trdi, da se usojeni dogodki pripetijo ne glede na karkoli. No, bi rekel Hrizip, morda je usojeno, da imaš vnuka, ampak to se *ne* bo pripetilo ne glede na *karkoli*. Vnuka imaš lahko samo, če imaš sam potomca, ki ima potomca. In morda je usojeno, da boš ozdravel, ampak to se bo pripetilo samo, če se prej posvetuješ z zdravnikom. Četudi ti je ozdravitev usojena, je bolje zate, če si prizadevaš, da do nje pride in greš k zdravniku (kljub temu, da je tudi to prizadevanje »usojeno«).

Morda obstajajo formulacije lenega argumenta, ki zahtevajo bolj poglobljeno obravnavo od Hrizipove. Presoja o nekem dejanju je mogoča samo, če to dejanje ni že vnaprej fiksirano, meni Brennan (2005: 284) in če ti je že usojena ozdravitev, ti je *usojena* z vsemi svojimi predpogoji. Kako lahko tedaj razmišljaš o tem, ali bi se posvetoval z zdravnikom, če pa to, po tezi o univerzalni usodi, nikdar ni »odprta« možnost? Že Spinoza je ugotavljal, da ne moremo takrat, ko o nečem še presojamo (se odločamo med X ali ne-X), obenem napovedati, kaj se bo zgodilo (recimo X). Kaj to pomeni za determinizem, je posebno vprašanje, ki presega okvir te razprave.

Tudi Dummett včasih označi leni argument kot očiten sofizem,[3] a večinoma ga obravnava kot resen argument *za* trdi fatalizem. V to ga sili simetrija med lenim argumentom in retro argumentom – zavrnitev »tehtnega« argumenta proti retroaktivni vzročnosti pač ne more biti trivialna. Če je domnevni argument za trdi fatalizem (»kar *bo* pa *bo*, mi nič ne moremo glede tega«) očitno zmoten, zakaj se nam zdi taisti argument v preteklem času (»kar je *bilo je bilo*, mi nič ne moremo glede tega«) tako sprejemljiv?

2.

V tem članku se ne bom ukvarjal z metafizično vznemirljivo temo retroaktivne vzročnosti, čeprav ima Dummettov zagovor te možnosti nekaj sodobnih simpatizerjev. ki se sklicujejo na fenomene kvantne mehanike (Bellov teorem, prim. Price, 1996: 173). Strinjam se s presojo antičnih avtorjev, da je sklep lenega argumenta *očitno* nesprejemljiv. Ampak, kje je napaka? Je ta zares povezana z naukom o usodi? *Usoda* stoikov ni tako daleč od sodobne teze o *determinizmu*, ampak zdi se mi, da možnosti kompatibilizma – teze o združljivosti svobode in determinizma – ne moremo kar takoj izključiti. Neresnica sklepa zato ne more biti odvisna od (ne)resnice teze o determinizmu.

[3] »... the fatalist argument is a blatant piece of sophistry ...« (Dummett, 1993: 353–54).

Že v antiki nastopajo formulacije lenega argumenta brez predpone »usojeno je, da«:

P1 Če A, potem A, ne glede na to, ali ti storiš X.

P2 Če ne-A, potem ne-A, ne glede na to, ali ti storiš X.

P3 Toda, A ali pa ne-A.

S Torej je (glede na A), jalovo storiti X.

'A' je propozicija o prihodnosti. Bobzienova (1998: 190) domneva, da bi v tej obliki izvor lenega argumenta lahko iskali pri megariških logikih, Filonu ali Diodoru Kronosu. Za Filona pogojnik »če *p*, potem *q*« pomeni zgolj to, da se tak sestavljeni stavek ne začne z resnico in konča z neresnico. Po tem pojmovanju je logično resnično: »Če *p*, potem, če *q*, tedaj p,« iz česar takoj sledi resnica premis (Pl) in (P2): »Če *p*, potem, če X ali pa ne-X, *p*.« S tem smo že pri moderni formulaciji argumenta in njegovi povezavi z razumevanjem pogojnikov.

Bombardiranje nam danes (na srečo) ni več blizu, znane pa so nam drugačne katastrofe, zato si oglejmo naslednjo različico, ki je, tako kot antična, povezana z vzročno vlogo zdravstvenih ukrepov pri preprečitvi bolezni. V obvezno zdravstveno zavarovanje so danes vključena določena obvezna cepljenja otrok, tudi proti davici. Znameniti *Zakon za uravnoteženje javnih financ* (ZUJF, maj 2012) je zapovedoval varčevanja na vseh koncih in krajih, zato si zlahka zamislimo, da bi dobili zdravniki navodilo, naj pripravijo ljudi do tega, da bi za določene prej brezplačne zdravstvene posege zdaj posebej plačevali. Denimo, da pripelje oče v času veljavnosti ZUJF osemletno deklico na zadnje cepljenje proti davici. Zdravnik pa mu reče:

> Deklica bo davico dobila ali pa ne. Če je bo dobila, jo bo dobila, četudi jo cepite. Če je ne bo dobila, je ne bo dobila, četudi cepljenje opustite. Če davico dobi, potem je cepljenje, če ga izvedemo, brezplodno. Če pa davice ne dobi, potem je cepljenje, če ga izvedemo, odveč.

V tej obliki argument navaja Dummett.[4] Še krajše ga povzamemo v obliki konstruktivne dileme: Deklica bo davico dobila ali pa ne. Če je bo dobila, potem je cepljenje neučinkovito. Če je ne bo dobila, potem je cepljenje nepotrebno. Torej je cepljenje neučinkovito ali pa nepotrebno. (Če pa res vztrajate, lahko posebej plačate in bomo ta poseg izvedli).

Očetova *impulzivna* reakcija je predvidljiva: »To je zelo gnil argument, samo denar hočete izpuliti od mene!« Prav zaradi intuitivne privlačnosti takšnega odziva so v antiki sklep razumeli kot argument *proti* nauku o usodi. Reakcija se mi zdi povsem upravičena, ampak *logik* jo mora podpreti z analizo napačnega koraka. Konstruktivna dilema je veljavna deduktivna oblika, premise niso očitno zmotne, zakaj je tedaj sklep nesprejemljiv? Problem je zastavljen kot *paradoks*, v

[4] »Either E will occur or it will not. If E does not occur, it will not occur even if you do A. If E does occur, it will occur even if you omit to do A. If E does not occur, A, if you did it, will have been fruitless. If E does occur, A, if you did it, will have been redundant« (Dummett, 1993: 354).

oznaki Sainsburyja (2009: 1): navidez nesprejemljivi sklep sledi po navidez sprejemljivem sklepanju iz navidez sprejemljivih premis. Nečemu se moramo odpovedati: morda sklep ni zares nesprejemljiv, morda je nekaj narobe s sklepanjem ali pa s premisami.

Prikažimo logično zgradbo bolj natančno. Oprl se bom na Stalnakerja (prim. 1975: 150), pri katerem se ključni sklepalni premiki povezani s pogojnim dokazom in pogojniki (‘D’ – »deklica bo dobila davico,« ‘C’ – »deklico cepimo,« navpična črta označuje pogojni dokaz).

1. D ali ne-D. Tavtologija

2. | D hipoteza za pogojni dokaz

3. | Če C, potem D. iz 2, logični zakon »implikacije«

4. | C je neučinkovito. iz 3

5. Če D, potem je C neučinkovito. 2–4, pogojni dokaz

6. | ne-D hipoteza za pogojni dokaz

7. | Če ne-C, potem ne-D. iz 6, logični zakon »implikacije«

8. | C je nepotrebno. iz 7

9. Če ne-D, potem je C nepotrebno. 6–8, pogojni dokaz

10. C je neučinkovito ali nepotrebno. 1, 5, 9, konstruktivna dilema

11. C je nesmiselno. iz 10

Začnimo z (1), ki je logični zakon *izključene tretje možnosti*, in prav Dummett je dokazoval, da mora antirealist zavrniti zakon *bivalence* (»vsaka propozicija je bodisi resnična bodisi neresnična«) in z njim vred zakon izključene tretje možnosti (Dummett, 1991: 9). Zanimivo pa je, da se Dummett, ki sicer niha glede (anti)realizma o preteklosti, v primeru lenega argumenta izogne vsaki razpravi o tem logičnem zakonu: »gotovo lahko fatalistov argument razkrinkamo kot sofizem, ne da bi zabredli v globoke vode metafizike ali osnov logike« (Dummett, 1993: 354). Pri tem navaja Quinea, ki govori o »obupni skrajnosti, da bi resno vzeli Aristotelovo fantazijo, po kateri resnica trditve ‘Resnično je, da *p* ali *q*’ ni zadostni razlog za ‘Resnično je, da *p* ali resnično je, da *q*’« (Quine, 1966:21).

Gre za znamenito razpravo o fatalizmu, do katerega naj bi vodila resnica propozicij o prihodnosti. Aristotelov primer je stavek »Ali jutri bo pomorska bitka ali pa jutri pomorske bitke ne bo«. Ker sedaj (še) ni ničesar v svetu, kar bi naredilo ta stavek za resničen ali neresničen, načelo bivalence (resnično je, *da p* ali resnično je, da *ne-p)* ne drži. Toda celotna disjunkcija »Ali jutri bo pomorska bitka ali pa jutri pomorske bitke ne bo« je kljub temu logično resnična zgolj zaradi svoje oblike. V splošnem iz zakona izključene tretje možnosti še ne sledi načelo bivalence. Problem resnice, časa in nujnosti je seveda povezan s

fatalizmom, ampak na drugačen način kot leni argument, zato se tu lahko strinjamo z Dummettom in privzamemo, da prvi korak ni vprašljiv.

Tudi razpravi o naslednjem koraku (od 2 do 3 in 6 do 7), ki temelji na »Filonovem« zakonu: »P, torej, če Q, potem P,« se Dummett izogne. Zadovolji se s tem, da je sklepanje veljavno v mnogih interpretacijah besedice »če«. Seveda pa je to samo vrh ledene gore: gre za znane paradokse materialne implikacije. Že stoiki so vedeli, da je veljavno nenavadno sklepanje: »Če si živ, potem, če si mrtev, si živ.« Kadar je konsekvens (»živ si«) resničen, lahko v implikaciji nastopa poljuben antecedens (»mrtev si«), pa bo celoten pogojnik, razumljen kot *materialna implikacija*, takoj resničen. Ali lahko razumemo pogojnike, ki nastopajo v dokazu, kot materialne implikacije? To je eno od ključnih vprašanj za razrešitev »paradoksa,« ampak Dummett (1964: 346) pravi, da ne bo izgubljal časa z iskanjem prave interpretacije pogojnikov v naravnem jeziku. Kot bomo videli Dummett vseskozi beži pred pogojniki, ampak nikdar zares ne pobegne.

Za Dummetta sta ključna naslednja koraka (od 3 do 4 in potem 5, ter od 7 do 8 in potem 9). Ker se mu zdi jasno, da je, ob predpostavki, da deklica zboli za davico, cepljenje zares neučinkovito, v (5) ne dvomi. Drugače pa je z drugim pogojnim dokazom (od 6 do 8), s tezo, da je cepljenje, ob predpostavki, da deklica davice ne dobi, *nepotrebno*. Ali iz resnice (7), *»Četudi* ne-C, *še vedno* ne-D«, zares sledi, da je cepljenje nepotrebno (8), ker ne gre za učinkovit način preprečitve bolezni? (Pozor na besedico *četudi*!). Za Dummetta je C učinkovit način za preprečitev D, kadar sta resnična pogojnika *učinkovitosti*: (i) Če C, potem ne-D; (ii) Če ne-C, potem D. Teza o nepotrebnosti cepljenja bi zanj sledila samo, če bi resnica (7) izključevala resnico obeh pogojnikov (i) in (ii). No, zlahka se prepričamo, da je, vsaj za materialno implikacijo, resnica pogojnika (7) – »če ne-C, potem ne-D« – združljiva z resnico *obeh* pogojnikov (i) in (ii) – v primeru, ko je C res, D pa ne. Torej (7) *ne* zanika tako definirane učinkovitosti cepljenja.

V kasnejšem članku Dummett (1993: 355) eksplicitno opiše to rešitev s pomočjo pogojnikov v veznem naklonu (angl. *subjunctive*, uporabljal jih je že v izvornem članku). Samo če bi veljalo, da *bi* brez cepljenja deklica še vedno ne obolela, *bi* bilo cepljenje nepotrebno. Fatalist sicer ima, če združi (6) in (7): »Če ne-D, potem, četudi ne-C, še vedno ne-D«, ampak iz tega ne sledi nepotrebnost cepljenja. Samo če bi imel pogojnik v veznem naklonu, »Če ne-D, potem četudi ne *bi* bilo res, da C, še vedno ne *bi* bilo res, da D«, bi lahko sklepal na nepotrebnost cepljenja. To je razvidno tudi iz Dummettove rešitve: resnica (materialnega) pogojnika »če ne-C, potem ne-D« je zares združljiva z resnico pogojnika »če ne-C, potem D«. Toda resnica *protidejstvenega* pogojnika »če ne bi bilo res, da C, potem ne bi bilo res, da D«, ni združljiva z resnico protidejstvenika »če ne bi bilo res, da C, potem bi bilo res, da D«. Po standardni definiciji je protidejstvenik A > B resničen na svetu *i*, če in samo če je B resničen v vseh A-svetovih, ki so najbolj podobni svetu *i*. Če je v najbližjih svetovih, v katerih ne cepimo deklice, res, da ne oboli za davico, potem ne more biti obenem res, da v najbližjih svetovih, v katerih deklico ne cepimo, oboli za davico. Fatalist

ima v svojem argumentu za nepotrebnost cepljenja samo »četudi ne C, še vedno ne-D«, potreboval pa bi neko močnejšo, *modalno* zvezo med odsotnostjo cepljenja (previdnostnih ukrepov) in nenastopom bolezni.

Značilen je nastop besedice »četudi« v Dummettovi formulaciji lenega argumenta. Mislim, da z njo neopazno zamenjamo smisel pogojnika. Recimo, da sprejmemo: »Če si živ, potem, če si mrtev, tedaj si živ«. Ampak že sklepanje »Če si živ, potem, *četudi* si mrtev, si *še vedno* živ«, bi težko sprejeli. Sploh pa ne »Če si živ, potem, *četudi* bi bil mrtev, bi še bil *vedno* živ«. Protidejstvenike (pogojnik v veznem naklonu, ki ima neresničen antecedens), v katerih je konsekvens resničen, je Goodman (1947) poimenoval poldejstveniki [*semifactual*]. Chisholm (1949: 492) je zagovarjal ekvivalenco med »Četudi bi bilo res, da A, bi še vedno bilo res, da B« in »Ni resnično: če bi bilo res, da A, bi bilo res, da ne-B«. Denimo, »Četudi bi spal celo jutro, bi še vedno bil utrujen« zanj pomeni: »Ni resnično, da če bi spal celo jutro, potem ne bi bil utrujen«. Goodman ločuje med praktičnim in dobesednim pomenom poldejstvenikov. Dobesedno vzeto sta »Četudi A, še vedno ne-B« in »Če bi bilo resnično A, bi bilo resnično B«, kontrarna, ne pa protislovna, saj sta lahko oba neresnična. Praktično pa za Goodmana velja Chisholmova ekvivalenca – protidejstvenik zatrjuje, poldejstvenik pa zanika obstoj zveze med antecedensom in konsekvensom (Goodman, 1947: 115). Dejstvo, da razumemo »če bi bilo resnično, da A (ne pa ne-A, kot se je dejansko zgodilo), tedaj bi bilo resnično, da B (kar se je dejansko zgodilo)« kot »Četudi A, še vedno B« govori o »naravnem« razumevanju protidejstvenikov: pričakujemo, da bo imela sprememba od A do ne-A nek učinek. Ker je B dejansko resničen in ostane resničen, se zdi, da ta sprememba nima vpliva, zato »Četudi A, še vedno B« razumemo kot zanikanje zveze ali odvisnosti B od A.

Toda fatalist v svojem argumentu sploh *nima* pogojnikov v veznem naklonu (če bi .... potem bi ...), zato tudi ne more imeti poldejstvenika, ki bi zanikal modalno zvezo med cepljenjem in preprečitvijo bolezni. »Četudi« je retorično učinkovit, ni pa logično podprt. Vrsta razpravljavcev ugotavlja, da se fatalistova zmota skriva v razumevanju *moči* dvojnega pogojnika: »Če ne-D, če ne-C, potem ne-D«.

V naši formalizaciji bi to bilo (7’):

…

6. | ne-D hipoteza za pogojni dokaz

7. | Če ne-C, potem ne-D. iz 6, logični zakon »implikacije«

7’ | Če ne-D, potem, *četudi* ne-C, ne-D. 6, 7, logični zakon

8. | C je nepotrebno. iz 7’

9. Če ne-D, potem je C nepotrebno. 6–8, pogojni dokaz

Kako razumeti (7’)? Za protidejstvenik »Če bi bilo res, da A, potem bi bilo res,

da B« bom uporabljal oznako »A > B,« za materialni pogojnik »Če A, potem B« pa oznako »A → B« (‘¬’ je negacija). Možnosti so:

i. ¬D → (¬C → ¬D)

ii. ¬D → (¬C > ¬D)

iii. ¬D > (¬C > ¬D)

Zgradba je za nekatere še bolj kompleksna, Buller (1995: 118) tako meni, da bi fatalist moral imeti kot premiso: »Če ne-D, potem za *vsak* možni ukrep X, če bi izvedel ali pa ne bi izvedel X, bi še vedno bilo res, da ne-D«. Navadni *materialni* pogojnik (i), ki ga dobimo s pogojnim dokazom v vrstici (7), ne more utemeljiti nepotrebnosti C. Močnejši pogojnik bi to lahko utemeljil, ampak fatalist ni pokazal, da ob predpostavki, da pri deklici, ki ni cepljena in *dejansko* ne oboli za davico, odsotnost te bolezni nastopi tudi v relevantnih *možnih* svetovih, kjer ni cepljena. Rekli bi lahko, da ima fatalist samo indikativ (pogojnik v povednem naklonu), ki ga dobi »na prazno,« glede na čudne lastnosti materialne implikacije. Nima pa konjunktiva (pogojnika v veznem naklonu), ki bi zares utemeljeval neko odvisnost med stanji stvari v svetu. Prav protidejstveniki v veznem naklonu pa izražajo metafizične odvisnosti in zveze v svetu, saj povedo, kaj v svetu bi se spremenilo, če bi antecedens bil resničen. In fatalist ni utemeljil odsotnosti vzročne zveze med cepljenjem in odsotnostjo bolezni (ali *bi* ne-D zares nastopil tudi brez C?).

3.

Toda Dummett s svojo rešitvijo ni bil zadovoljen. V njej nastopajo pogojniki v veznem naklonu, ki so zanj izvor »zmešnjav« [angl. *morass*]. Njihov pomen se mu zdi obskuren, celo smisel rabe teh pogojnikov naj bi bil nejasen (Dummett, 1993: 355). Nenavadno, saj danes celo raziskave psihologov kažejo na tesno zvezo med razmišljanjem o vzročnih zvezah in protidejstveniki! Dummett si zdaj želi zavrnitve fatalistovega sklepanja brez zanj spornih pogojnikov, zato fatalistovo razmišljanje tolmači kot vprašanje, ali je *racionalno* upoštevati previdnostne ukrepe oz. ali je v splošnem racionalno poskušati storiti nekaj, kar bi nekaj preprečilo ali do česa privedlo. Zdi se mi, da s tem Dummettova prvotna rešitev začenja izgubljati svojo vrednost. Dummetta zdaj zanima, ali nas lahko fatalist s svojim argumentom *prepriča*, da preventivni ukrepi niso racionalni. Toda ali nismo že od vsega začetka vedeli, da to ni res? Spomnimo se impulzivne reakcije in antičnega razumevanja lenega argumenta kot *reductio* stoiškega nauka o usodi. Zanimivi so prav logični in semantični »triki,« ki dajejo sklepanju videz racionalnosti!

Dummettov novi predlog temelji na precej finem in ne preveč razumljivem razlikovanju med nepotrebnostjo [angl. *redundant*] in odvečnostjo [angl. *superfluous*] nekega dejanja. Dejanje C je nepotrebno za ne-D, samo če velja pogojnik v veznem naklonu: »četudi bi bilo res, da ne-C, bi še vedno bilo res, da ne-D«. To je bila prvotna rešitev (takega pogojnika fatalist nima), ki jo zdaj

želimo izboljšati. Fatalist po mnenju Dummetta trdi, da nikdar ni razumno storiti česarkoli, da bi s tem kasneje nekaj nezaželenega preprečili ali nekaj zaželenega dosegli. In kdaj imamo razumen motiv, da storimo C z namenom preprečitve D? Kadar imamo *dobro osnovo*, da sprejmemo C kot zadostni in neodvečni pogoj za ne-D. Dobro osnovo za zadostnost C pa imamo, če imamo dobre razloge za trditev: »Če storimo C, tedaj nastopi ne-D«. In C ni odvečni pogoj za ne-D, če nimamo osnove za trditev, da če ne storimo C, tedaj ne nastopi D. Še več, C je neodvečni pogoj za ne-D, kadar imamo *osnovo* za trditev: »Če ne storimo C, tedaj nastopi D«.

Toda spet smo končali pri pogojniku, rešili smo se pogojnika *nepotrebnosti'*. »četudi bi bilo res, da ne-C, *bi* še vedno bilo res, da ne-D« in problem zvedli na osnovo za trditev pogojnika *odvečnosti*: »če ne-C, potem ne-D«. V čem je tu kak napredek? Fatalist naj bi zdaj trdil, da je C *odvečni* pogoj za ne-D, ker je resničen pogojnik (7) »Četudi ne-C, še vedno ne-D«. Ampak to po Dummettu ne sledi, ker:

> Za razliko od podobnega pojma nepotrebnosti, ki smo ga obravnavali prej, ni na voljo nobenega pojasnila za ta, kdaj je neko dejanje odvečno. Pojasnila ni na voljo, ker ga ne moremo podati brez pogojnikov v veznem naklonu, ki pa se jim zdaj izogibamo ... V sedanjem kontekstu sploh ne potrebujemo pojma *odvečnosti* dejanja, zadošča že pojem razumnega obravnavanja dejanja kot odvečnega. Razlika je v tem, da je vprašanje, ali lahko neko dejanje tako obravnavamo ali ne, vprašanje tega, kar vemo, ali za kar imamo razloge za verjetje: zatorej noben argument, ki dokazuje, da je dejanje dejansko odvečno, če določen pogojnik, za katerega nimamo nobene osnove, da bi ga imeli za resničnega, dejansko drži [angl. *holds good*], ne more biti prepričljiv (Dummett, 1993: 357).

Daljši odlomek, ki lepo ilustrira Dummettov stil pisanja,[5] naj bi pomenil ključno in dokončno zavrnitev lenega argumenta. Spomnimo se napovedi (Dummett, 1993:355):

> Raje bi imel čistejšo rešitev. Ta, ki sem jo pravkar podal, je neurejena [angl. *messy*], ker uporablja pogojnike v veznem naklonu.

Poskušajmo dešifrirati, v čem je nova zavrnitev – na prvi pogled se zdi kvečjemu *manj* razumljiva od prvotne. Sprva naj bi fatalist sklepal na *nepotrebnost* cepljenja: »četudi bi bilo res, da ne-C, bi še vedno bilo res, da ne-D«, toda tega protidejstvenika fatalist ni utemeljil. To se je zdelo prepričljivo, ampak protidejstveniki so zdaj »mora«. Po novem je fatalistova zmota v tem, da iz (7) »Četudi ne-C, še vedno ne-D«, sklepa na odvečnost cepljenja. Dummett mu ugovarja, da iz indikativa (7) odvečnost ne sledi. Zakaj ne, iz česa pa potem sledi? Odlomek nakazuje, da bi vsako pojasnilo temeljilo na protidejstvenikih, ki pa se jim ravno želimo izogniti. Toda ali iz nemožnosti *pojasnila* X (odvečnost)

[5] Leta 1993 je Dummett izdal knjigo *Grammar and Style for Examination Candidates and Others*, kar so primerjali s tem, da bi Hun Atila napisal knjigo o lepem obnašanju.

na način Y (protidejstveniki) zares sledi odsotnost X? Zdi se mi, da Dummettovemu protiargumentu nekaj zmanjka, saj fatalist lahko reče:

> Menim, da odvečnost sledi iz indikativa, ti praviš, da ne. Od mene zahtevaš, da pojasnim odvečnost brez protidejstvenikov, in praviš, da je to nemogoče. Ampak dokler je ti nekako ne pojasniš, je še vedno mogoče, da imam prav jaz! Predpostavimo, da deklica ne bo obolela, obenem pa se še nisem odločil, ali naj deklico cepim ali ne. Ali ni, ob tej predpostavki, *razumno* sklepati, da je cepljenje povsem odveč? Tvoja prvotna rešitev je bila, da je C učinkovit način za preprečitev D, kadar sta resnična pogojnika *učinkovitosti*: (i) Če C, potem ne-D; (ii) Če ne-C, potem D. Ampak, kot praviš, je »če ne-C, potem ne-D« združljivo z resnico teh dveh pogojnikov. Nezdružljiva sta pogojnika v veznem naklonu »če ne-C, potem bi bilo res, da ne-D« in pa »če ne-C, potem bi bilo res, da D«, ki pa jih jaz v svojem argumentu nisem utemeljil. Prav. Toda nezdružljiva sta že *indikativa*: »če ne-C, potem ne-D« in »če ne-C, potem D« (prim. Gibbard, 1981). Prvega *imam* (7), zato sklepam, da drugi ni resničen. In če ni, potem lahko zanikam potrebnost cepljenja!

Dummett na koncu navedenega odlomka odgovarja, da nimamo osnove za trditev »Četudi ne storimo C, še vedno nastopi ne-D«. To rešitev poznamo od prej, ko fatalist ni imel osnove za trditev »četudi *bi* bilo res, da ne-C, *bi* še vedno bilo res, da ne-D«. Glede na zvezo med protidejstveniki in vzročnostjo je bil ta ugovor smiseln. Zdaj imamo samo še indikativ, v katerem ne gre več za zveze in odvisnosti v svetu. Indikativi izražajo značilnosti našega (spo)znanja, povezani so z racionalno spremembo prepričanj – v kaj naj bomo prepričani, ko izvemo tisto, kar sporoča antecedent. Toda, kaj točno pomeni, da tak pogojnik velja [angl. *holds good*]? Kaj pomeni imeti osnovo, da sprejmeš nek pogojnik kot resničen? Eden od standardnih predlogov je, da je pogojnik »Če A, potem B« *zatrdljiv*, kadar je pogojna verjetnost B glede na A visoka. Ampak ali lahko o tem sploh tvorimo kako *objektivno* pravilno mnenje?

Gibbard (1981: 231–232) je predstavil argument, ki temu oporeka. Obstajajo primeri, ko ima ena oseba povsem ustrezne razloge za sprejetje pogojnika »Če A, potem B« in zavrnitev »Če A, potem ne-B,« druga oseba pa ima povsem ustrezne razloge za sprejetje pogojnika »Če A, potem ne-B« in zavrnitev »Če A, potem B«. Ampak sprejeti »Če A, potem B« pomeni zavrniti »Če A, potem ne-B«, v tem se večina teoretikov strinja z našim fatalistom. Eddingtonova (1997: 107) navaja svoj primer: vohuna X in Y vsak iz svojega opazovališča skrivaj nadzirata sobo, v kateri so šef mafije in trije gangsterji, A, B in C. Vohun X vidi, da je B zapustil sobo (Y tega iz svoje perspektive ne vidi). Vohun Y pa vidi, da je C zapustil sobo (tega pa X iz svoje perspektive ne vidi). Oba slišita, kako je šef nekomu, ki je v sobi, dal določena navodila. X poroča: »Če šef navodil ni dal A-ju, potem jih je C-ju (ne pa B-ju)«. Y pa pravi: »Če šef navodil ni dal A-ju, potem jih je B-ju (ne pa C-ju)«. Nihče od njiju ne razmišlja zmotno in oba imata prav – kako je to lahko racionalno? Sodbi ne moreta biti hkrati resnični in ne

hkrati zmotni, a obenem ne moremo reči, da je samo ena zmotna. Ne gre za to, da bi dodatna informacija vsakemu od njiju omogočila pravilno razmišljanje. Če bi imela več informacij, bi vedela, da je A dobil navodila, in sploh ne bi potrebovala pogojnikov (prim. Bennett, 1988: 520)!

Mislim, da so indikativ in njegovi pogoji zatrdljivosti enako, če ne še bolj zapleteni kot pogoji resnice protidejstvenikov. Pravzaprav je metafizična *odvisnost*, ki jo izražajo protidejstveniki, bolj razumljiva kot pa osnova za trditev pogojnikov v povednem naklonu. Dummett ta problem enostavno pomete pod preprogo, zato njegova končna rešitev ni dosti boljša od očetove impulzivne reakcije. Dummettova zavrnitev lenega argumenta se zdaj zvede na:

> Sklep iz pogojnika »Četudi ne-C, še vedno ne-D« na odvečnost C je neutemeljen. Kajti glede na to, kar vemo ali za kar imamo kako osnovo za ustrezno prepričanje, ni nobenega razloga, da bi razumno obravnavali C kot odvečen preventivni ukrep.

Toda to smo vendar vedeli že od vsega začetka! Spomnimo se očetove impulzivne reakcije na zdravnikovo različico argumenta: »To je zelo gnil argument, samo denar hočete izpuliti od mene!« Uporabil bi lahko tudi bolj prefinjen angleški stil:

> Glede na vse, kar vem ali za kar imam kako osnovo za ustrezno prepričanje, ni nobenega razloga, da bi razumno obravnaval cepljenje proti davici kot odvečni preventivni ukrep. Dozdeva se mi, da želite samo neupravičeno prevaliti breme kritja stroškov cepljenja na starše ali skrbnike otrok.

Impulzivni odziv je bolj učinkovit (sploh če je podkrepljen z ustrezno gestikulacijo in tonom), ampak vsebina povedanega je skoraj enaka. Zakaj smo se sploh morali prebijati skozi razne pogojnike in logične analize? Jasno je, da ker sklepa ne sprejemamo, fatalist ni podal razumnih razlogov za odvečnost cepljenja. Ampak v čem je *videz* racionalnosti? Logično zanimiv se mi zdi prav argument in njegova varljiva logika, ne pa izpodbijanje teze, da nikdar nimamo nobenega motiva za katerokoli dejanje, neresnica te teze je bila vendar jasna že v antiki!

Dummett kot svoj *credo* navaja naslednjo oznako filozofije:

> Filozofija ne poskuša odkriti novih resnic o svetu, ampak si prizadeva za jasen vpogled v to, kar o svetu že vemo in smo prepričani. Uspeh tega podvzetja je odvisen od bolj eksplicitnega dojetja zgradbe naših misli; to pa je odvisno od odkritja, kako sistematično pojasniti delovanje jezika, medija, v katerem izražamo svoje misli.[6]

Ampak kako se v sistematičnem pojasnilu delovanja jezika in misli sploh lahko *izognemo* pogojnikom? Tega so se zavedali že prvi raziskovalci logične zgradbe

[6] Prim. http://www.pyke-eye.com/view/phil_I_20.html (pridobljeno 3. 5. 2013), objavljeno v Pyke, S. 1995. *Philosophers*. London: Zelda Cheatle Press.

naših misli. William in Martha Kneale v svojem znanem delu (Kneale, 1962: 128) omenjata vroče razprave med megariki, zaradi katerih je pesnik Callimachus (4. – 3. stol. pr. n. št.) skoval epigram: »Celo krokarji na strehi krakajo o naravi pogojnikov«. Dummett sicer beži pred pogojniki, ampak nikdar zares ne pobegne.

4.

Na tem mestu bi temo lahko naravno pripeljali do njenega konca – Dummettova prvotna rešitev (1964) je povsem zadovoljiva, fatalist nima pogojnika, ki bi utemeljeval odsotnost vzročne zveze med cepljenjem in odsotnostjo bolezni. Korak (8) ne sledi iz (7), zato argument ni veljaven. Ampak spomnimo se zadnjega odgovora fatalista:

> Menim, da odvečnost cepljenja sledi že iz *indikativa*. Predpostavimo, da deklica ne bo obolela, obenem pa se še nisem odločil, ali naj deklico cepim ali ne. Ali ni, ob tej predpostavki, *razumno* sklepati, da je cepljenje povsem odveč? Nezdružljiva sta že *indikativa*. »če ne-C, potem ne-D« in »če ne-C, potem D«. Prvega *imam* (7), zato sklepam, da drugi ni resničen. In če ni, potem lahko zanikam potrebnost cepljenja!

Ali lahko temu zares *oporekamo*? Gre za sklepanje:

6. | ne-D hipoteza za pogojni dokaz

7. | Če ne-C, potem ne-D. iz 6, logični zakon »implikacije«

8. | C je nepotrebno. iz 7

9. Če ne-D, potem je C nepotrebno. 6–8, pogojni dokaz

Denimo, da razmišljam o nakupu delnic sklada Balkan 21. Bančni uslužbenec mi obljublja 8% donos. V tem primeru sem, če jih kupim, za 8% v dobičku. Torej se nakup izplača. Ali:

6’. | Delnice imajo 8% dobiček. hipoteza za pog. dokaz

7’. | Če delnice kupim, potem imajo 8% dobiček. iz 6

8’. | Nakup se izplača. iz 7

9’. Če imajo delnice 8% dobiček, potem se nakup izplača. 6’–8’, pogojni dokaz

V kontekstu 8% dobička se nakup zares izplača, (9’) se zdi resnično – ampak, kako lahko tedaj kritiziramo fatalistov sklep (9), po katerem je cepljenje nepotrebno?

Zanimivo rešitev ponuja Stalnaker (1975). Sklepanje od (6) do (9) je zanj zares *razumno*, v kontekstu, v katerem lahko trdimo ali predpostavimo premiso (6) kot ustrezno in presojamo o preventivnih ukrepih (*odprto* je, ali je antecedens pogojnika resničen), je sklep poddokaza (9) sprejemljiv. V kontekstu cl, v katerem predpostavim, da bo deklica obolela in je odprto, kaj bom storil, je

povsem razumno sprejeti, da je cepljenje neučinkovito. V kontekstu c2, v katerem predpostavim, da deklica ne bo zbolela in je še odprto, kaj bom storil, pa je razumno sprejeti, da je cepljenje nepotrebno. Toda končni sklep, ki naj bi sledil po konstruktivni dilemi (»C je neučinkovito ali nepotrebno«) ni niti razumen niti veljaven. Nobenega konteksta c3 ni, v katerem bi sprejeli oba poddokaza in izpeljali ta končni sklep. Zato nismo upravičeni sklepov poddokazov »izvoziti« iz njunih kontekstov in ju sestaviti skupaj v sklep konstruktivne dileme.

Fatalistova zmota, po Stalnakerju, je v zamenjavi razumnosti z veljavnostjo. V čem je razlika? Veljavnost (sledenje) je klasični semantični pojem – nemogoče je, da bi bile *propozicije* resnične, sklep pa ne. Razumno sklepanje pa je pragmatična relacija, ki povezuje jezikovna dejanja, ne pa propozicij. Sklepanje iz zaporedja *trditev* ali *domnev* [angl. *suppositions*] je *razumno*, samo če je v vsakem kontekstu, v katerem lahko premise ustrezno zatrdimo ali predpostavimo, nemogoče, da bi kdorkoli sprejel premise, ne da bi se zavezal k sklepu. Kontekst tvori skupno vedenje, predpostavke, privzetja vseh sodelujočih v pogovoru. Lahko ga predstavimo kot množico svetov, ki niso izločeni s splošnimi predpostavkami konverzacije. Kadar pa B *sledi* iz A, takrat je vsak svet, v katerem je resnično A, tak, da je v njem resnično tudi B. S tem so zajeti *vsi* konteksti, zato iz veljavnosti sledi razumnost, obratno pa ne velja! Sklepanje »B, torej če A, potem B« je lahko *razumno*: če je B združljiv z danim kontekstom (ali del tega konteksta) in je tudi A združljiv s kontekstom. Ampak sklepanje ni *veljavno*: denimo, da je res, da B, toda obstaja nek svet *izven* konteksta trditve B, v katerem je resnično A, B pa ne. In prav to se zgodi v koraku od (6) do (9): ob predpostavki, da deklica ne oboli, je razumno sprejeti tudi: »če deklice ne cepimo, potem ne dobi davice (in je cepljenje nepotrebno)«. Ampak to ne velja izven tega konteksta in še manj v *vsakem* kontekstu! Ali shematsko:

1. D ali ne-D.

Tavtologija, ki je resnična v vsakem kontekstu.

5. Če D, potem C je neučinkovito.

Razumno znotraj konteksta c1 (predpostavke D).

9. Če ne-D, potem je C nepotrebno.

Razumno znotraj konteksta c2 (predpostavke ne-D).

10. C je neučinkovito ali C je nepotrebno.

Sklep *ne* sledi iz 1, 5 in 9. Sklep ni razumen glede na 1, 5 in 9.

Sklepanje v poddokazu (2) – (5) je razumno, ni pa veljavno. Sklepanje v poddokazu (6) – (9) je prav tako razumno, ni pa veljavno. Končni sklep (10) pa je nerazumen in neveljaven. Če bi bila oba poddokaza veljavna, bi sledil tudi končni sklep konstruktivne dileme, ampak poddokaza sta samo razumna. In iz tega, da je vsak od poddokazov razumen v svojem kontekstu, ne sledi, da je

celoten dokaz razumen. V kontekstu cl je cepljenje neučinkovito, v kontekstu c2 je cepljenje nepotrebno, v nobenem kontekstu pa ni nepotrebno ali neučinkovito.

V obeh razrešitvah lenega dokaza nastopajo »kontekstualni« premiki. V izvornem članku Dummett fatalistu dopušča, da iz »Deklica ne zboli« izpelje »Če deklice ne cepimo, deklica ne zboli«. Ampak smisel, v katerem je to sklepanje veljavno, ne dopušča, da iz »Če deklice ne cepimo, deklica ne zboli« izpeljemo »Cepljenje je neučinkovito za preprečitev bolezni« (prim. Dummett, 1964: 341). Toda Dummett se izogne razpravi o teh različnih smislih in skuša, po mojem mnenju neuspešno, problem razumevanja pogojnikov povsem zaobiti. Stalnaker diagnosticira kontekstualne premike med poddokazoma in končnim sklepom. Pravzaprav Dummettovo formulacijo lenega argumenta izkorišča prav kot ilustracijo svojega formalnega aparata. V ozadju se skriva kompleksna teorija o pogojnikih, kontekstualizmu, semantiki in formalni pragmatiki .... ki bi zaslužila posebno obravnavo.

Naj omenim le, da moč Stalnakerjeve teorije vidim tudi v tem, da zajame in pojasni uspeh Dummettove prvotne rešitve. Spomnimo se – fatalist po Dummettu nima protidejstvenika, ki bi *zanikal* odvisnost ozdravitve od cepljenja in trdil: »četudi *bi* bilo res, da ne-C, *bi* še vedno bilo res, da ne-D«. Po Stalnakerju v presoji o resnici protidejstvenika nastopajo najbližji svetovi, v katerih je antecedens resničen, toda ti svetovi *niso* del privzetega konteksta. Vzemimo znani par:

A Če Oswald ni ubil Kennedyja, potem ga je nekdo drug.

B Če Oswald ne bi ubil Kennedyja, potem bi ga nekdo drug.

(A) je resnično – antecedens v indikativu zadeva samo žive možnosti *znotraj* danega konteksta *c* in predpostavke konverzacije v tem primeru vključujejo Kennedyjevo smrt, ki jo ohranjamo v presoji o resnici (A). In če je Kennedy umrl v atentatu, pa ga ni ubil Oswald, potem ga je očitno nekdo drug. Pogojnik (B) pa ni resničen, antecedens v protidejstveniku zadeva hipotetične svetove, ki so zunaj privzetega konteksta (če ga ne *bi* ubil Oswald, ga morda ne bi nihče drug). Zdaj zapuščamo znano in privzeto in se sprašujemo, kaj bi bilo, če bi bile predpostavke drugačne, in (B) se zato ne zdi resnično. Indikativ je *ustrezen* v kontekstu, ki je združljiv z antecedensom, protidejstvenik pa moramo izraziti v veznem naklonu, ki naznači, da zapuščamo dani kontekst privzetij in predpostavk. Fatalist ima »Če ne-C, potem ne-D« znotraj konteksta cl (recimo, da ne-D), nima pa »Če bi bilo res, da ne-C, potem bi bilo res, da ne-D«, saj, ko opustimo privzetje, da deklica ne oboli, pogojnik ni resničen. Ker je veljavnost resnica v *vseh* kontekstih, tako pri protidejstvenikih kot pri veljavnosti zapuščamo dani kontekst privzetij.

Trdi fatalizem je za Dummetta *filozofsko* zanimiv, ker je zanj zmoten na enak način, kot je zmoten retro argument proti možnosti vplivanja na preteklost. Dummettov končni obračun s fatalistom, ki temelji na tem, kaj je razumno verjeti, se mi zdi ali prazen (od vsega začetka smo vedeli, da je fatalizem

nerazumen) ali pa spet pristane v starodavnih »zmešnjavah« pogojnikov. Trdi fatalizem ni filozofsko vznemirljiv, ker bi z dokazi resno utemeljeval držo vdanosti v usodo pred neogibno prihodnostjo. Takšna drža je nesmiselna, kot so vedeli že v antiki. Logično zanimiva je paradoksalna zgradba lenega argumenta (sprejemljive premise in sklepanje, ne pa sklep), saj se izkaže, da potrebujemo presenetljivo zapleten logiški aparat za njegovo razrešitev. Sili nas v razmislek o naravi pogojnikov, česar so se zavedali že prvi raziskovalci logične zgradbe naših misli, za katere se temi fatalizma in pogojnikov prepletata. In *ta* razprava se tudi po dva tisoč štiristo letih še ni končala.

DUMMETT, FATALISM AND CONDITIONALS

In “Bringing about the past” (1964) and “Causal Loops” (1986) Dummett defends the possibility of backwards causation. Argument against such a possibility (there is an intrinsic absurdity in doing something in order that something else should happen *earlier* in time) is convincing but fallacious. The old “lazy” argument (there is an intrinsic absurdity in doing something in order that something else should *subsequently* happen) has the same form and is also fallacious. Fatalist starts with: “Either E will occur or it will not. If E does not occur, it will not occur even if you do A. If E does occur, it will occur even if you omit to do A.” But for Dummett (1964) on any sense of “if” on which this inference is valid, it is impermissible to pass from “If E does occur, it will occur even if you omit to do A” to “If E does occur, A, if you did it, will have been redundant.” A is redundant, when E occurs, only if E would have occurred even if A had not been performed. The indicative conditional, being of a kind whose truth is guaranteed by the truth of its consequent, will not support the corresponding subjunctive conditional. Dummett (1986) is no longer satisfied with his earlier solution – it is supposed to be messy because it appeals to subjunctive conditionals. Fatalist is wrong in arguing that we can never have a reasonable motive for doing anything. The notion of having a reasonable motive is now explained in terms of indicative conditionals (“We have a ground for asserting that, if we do not do A, E will not occur.”) But Dummett (1986)

is either saying that fatalism is not grounded (ancient authors already knew that – the lazy argument was traditionally understood as a *reductio* of stoic doctrine of fate!) or else introducing, without analysis, indicative conditionals, which are also a »morass.« At the end I sketch and defend a solution, based on a complex apparatus of formal pragmatics, proposed by Stalnaker.

Keywords: Dummett, fatalism, lazy argument, conditionals, Stalnaker.

Literatura

Baumeister, R., Crescioni, W. in Alquist, J. (2011). »Free Will as Advanced Action Control for Human Social Life and Culture«. *Neuroethics*, 4, str. 1–11.

Bennett, J. (1988). »Farewell to the Phlogiston Theory of Conditionals«. *Mind*, 97, str. 509–527.

Berčič, B. (2002). »Fatalizem«. *Analiza*, 6, št. 3, str. 3–22.

Bernstein, M. (2002). »Fatalism«. V *The Oxford Handbook of Free Will*, Oxford: Oxford University Press, str. 65–81.

Bobzien, S. (1998). *Determinism and Freedom in Stoic Philosophy*. Oxford: Oxford University Press.

Brennan, T. (2005). *The Stoic Life*. Oxford: Clarendon Press.

Buller, J. D. (1995). »On the 'Standard' Argument for Fatalism«. *Philosophical Papers*, 24, str. 111–125.

Chisholm, R. (1949). »The Contrary-to-Fact-Conditional«. V Feigl H., Sellars W. (ur.), *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, str. 482–497.

Dummett, M. (1964). »Bringing about the Past«. *Philosophical Review*, 73, str. 338–359.

Dummett, M. (1991). *The Logical Basis of Metaphysics*. London: Duckworth.

Dummett, M. (1993). »Causal Loops«. V *The Seas of Language*, Oxford: Clarendon Press, str. 349–375. Prva objava v Flood, R. in Lockwood, M. (ur.) (1986). *The Nature of Time*. Oxford: Basil Blackwell, str. 135–169.

Dummett, M. (2004). *Truth and the Past*. New York: Columbia University Press.

Dummett, M. (2006). *Thought and Reality*. Oxford: Oxford University Press.

Eddington, D. (1997). »Truth, Objectivity, Counterfactuals and Gibbard«. *Mind*, 106, str. 107–116.

Gibbard, A. (1981). »Two Recent Theories of Conditionals«. V Harper,

Stalnaker, Pearce (ur.), *Ifs*, Dordrecht: Reidel, str. 211–247.

Goodman, N. (1947). »The Problem of Counterfactual Conditionals«. *Journal of Philosophy*, 44, str. 113–128.

Kneale, W. in M. (1962). *The Development of Logic*. Oxford: Clarendon Press.

Price, H. (1996). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. New York: Oxford University Press.

Quine, W. v. O. (1966). »On a Supposed Antinomy«. V *The Ways of Paradox and Other Essays*, Cambridge, Mass.: Harvard University Press, str. 19–21.

Sainsbury, M. (2009). *Paradoxes*. Cambridge: Cambridge University Press.

Stalnaker, R. (1991). »Indicative Conditionals«. V Jackson, F. (ur.), *Conditionals*, Oxford: Basil Blackwell, str. 136–154.

Taylor, R. (2007). »Usoda«. V Šuster, D. (ur.), *O svobodni volji*, Maribor: Aristej, str. 15–31.

Maja Malec
CEU v Budimpešti

KRIPKE IN DUMMETT O MOŽNOSTI OBSTOJA SAMOROGOV

Kripke glede možnega obstoja samorogov zagovarja dve tezi: metafizično tezo, da nobene protidejstvene situacije ne moremo primerno opisati tako, da bi v njej obstajali samorogi, in epistemološko tezo, da arheološko odkritje, da so obstajale živali z vsemi lastnostmi, ki jih pripisujemo samorogom v ustreznem mitu, ne bi bilo samo po sebi dokaz, da so samorogi obstajali. Pri tem izhaja iz svoje semantične teorije izrazov za naravne vrste in svoje vzročne teorije reference. Dummett nasprotuje obema tezama, sicer nekoliko drugače formuliranima, in dokazuje, da sta v nasprotju z našo jezikovno prakso. Metafizične teze poleg tega ni mogoče konsistentno zagovarjati niti znotraj Kripkejeve teorije, sprejetje možnosti obstoja samorogov pa razkrije, da je pravi modalni sistem za metafizično modalnost S4, in ne intuitivno sprejemljivejši S5. Poskusila bom ugotoviti, kateri od njiju ima prav, Kripke ali Dummett?

Ključne besede: Kripke, Dummett, samorogi, naravne vrste, togi označevalci, vzročna teorija reference, modalnost

1. Uvod

V *Imenovanju in nujnosti* (Kripke, 2000) Kripke kritizira opisno teorijo imen, ki jo poimenuje Frege-Russellova teorija, pri čemer nasprotuje tako zamisli, da je imenu pridružen opis z njim sinonimen in mu daje pomen (teorija pomena), kakor tudi zamisli, da opis določi individuum, na katerega se ime nanaša (teorija reference). Po njegovem se lastna imena neposredno, neodvisno od kakega pridruženega opisa, nanašajo na imenovano. Referenca imena je prvotno določena tako, da se na stvar pokaže ali pa poda edinstveno določujoči opis (krst), nato pa se ta prenaša od enega do drugega govorca, pri čemer ima vsak namen ime uporabiti z isto referenco kot ta, od kogar ga je prevzel (vzročna ali zgodovinska teorija reference). Oseba lahko referencialno uspešno uporablja ime, četudi se pravzaprav ne spomni, od koga ga je prevzela ali pa ga povezuje z

opisom, ki napačno opiše poimenovani predmet, saj je bistvena povezanost z drugimi govorci. Teoriji se tudi razlikujeta v analizi modalnih stavkov. Po Kripkeju so namreč imena togi označevalci, ki označujejo isti predmet v vseh možnih svetovih, v katerih ta obstaja (drugače pa nič). To pomeni, da je referent imena, ki nastopa v določenem modalnem stavku, isti kot v dejanskem svetu in da sprašujemo o njem, tj. *tem* predmetu, ali bi to, kar je o njem rečeno, bilo res v neki protidejstveni situaciji ali ne, pri čemer ni treba, da je v tej situaciji o njem resnično tisto, kar mu v dejanskem svetu pripisuje opis. Nasprotno pa je v Frege-Russellovi teoriji, kjer pot do referenta vodi preko opisa, referent imena v določeni protidejstveni situaciji tisti predmet, ki edinstveno izpolni imenu pridruženi opis, to pa ni nujno isti predmet, ki ga izpolni v dejanskem svetu. Semantični teoriji potemtakem dajeta različne modalne rezultate.

To se pokaže tudi pri izrazih za naravne vrste, ki se po Kripkeju obnašajo podobno kakor lastna imena. »Tiger« je ime vrste, ki je določena na podlagi paradigmatičnih primerkov, s katerimi smo v vzročnem stiku. Ti posedujejo pojavne lastnosti, ki nastopajo v slovarski definiciji tigra, vendar pa te niso nujni in zadostni pogoj za članstvo v vrsti, temveč enaka notranja zgradba, ki jo empirično odkrijemo s preiskavo paradigmatskih primerkov. Ker je v protidejstvenih situacijah referent izraza enak dejanskemu, imajo tigri tudi v teh situacijah enako notranjo zgradbo kot v dejanskem svetu, lahko pa imajo drugačne pojavne lastnosti. Po Fregeju, Russellu, pa tudi Millu je obseg izraza v nemodalnih in modalnih kontekstih določen na podlagi slovarske definicije, ki podaja značilne pojavne lastnosti tigra. Zato so te nujni in zadostni pogoj za članstvo v vrsti, ne pa enaka notranja zgradba. Teoriji pa se ne razlikujeta le v modalnih rezultatih, ko je denimo stavek »Tigri bi lahko bili zeleni« po Kripkejevi teoriji resničen in po opisni teoriji neresničen, temveč gre prav tako za različni vrsti modalnosti. Po opisni teoriji obsege izrazov določajo naši običajni pojmi ali slovarske definicije in gre torej za pojmovno ali jezikovno utemeljeno modalnost. Po Kripkejevi teoriji obseg izrazov določa znanost, ki odkrije bistveno notranjo zgradbo dejanskih primerkov te vrste, zato gre za objektivno, tj. metafizično modalnost.

Razliko med obema teorijama lepo ilustrira primer samorogov. Po Kripkeju samorogi tvorijo mitološko naravno vrsto, ki pa ni dejanska, tj. nima primerkov v dejanskem svetu. Je njihov neobstoj zgolj kontingentno dejstvo in bi v nekoliko drugačnih okoliščinah lahko obstajali ali je nujno dejstvo, kakor je denimo neobstoj okroglega kvadrata? Naša intuicija govori v prid možnega obstoja samorogov, ki ji pa Kripke znotraj okvirja svoje semantične teorije ne more pritrditi. V tem prispevku bom najprej predstavila Kripkejevo stališče, tako metafizični in epistemološki vidik, ki ga najdemo v *Imenovanju in nujnosti* (Kripke, 2000),[1] »Vacuous Names and Fictional Entities« (Kripke, 2011)[2] in

[1] Citiram slovenski prevod samostojne knjižne izdaje *Naming and Necessity* iz leta 1980, ki je pravzaprav prepis treh predavanj, ki jih je Kripke imel januarja leta 1970 na Univerzi Princeton, z obsežnejšim predgovorom in dodatki.

[2] To je nekoliko popravljen prepis prispevka, ki ga je predstavil na konferenci »Language, Intentionality, and Translation Theory« na Univerzi v Connecticutu marca leta 1973.

Reference and Existence (Kripke, 2013).[3] Nato bom predstavila Dummettov protiargument iz »Could There Be Unicorns?« (Dummett, 1993). Dummett je sicer privrženec standardne semantične teorije, a v tem članku poskuša pokazati, da je zanikanje možnosti obstoja samorogov nemogoče konsistentno zagovarjati znotraj Kripkejeve teorije. Na koncu bom predstavila še mojo oceno o tem, kdo od njiju ima prav.

2. Kripkejev metafizični in epistemološki argument

Na začetku prvega od treh predavanj, ki tvorijo *Imenovanje in nujnost* (Kripke, 2000), Kripke svoje stališče o nemožnosti obstoja samorogov omeni kot eno od tistih njegovih stališč, ki se zdijo vsem očitno napačna. Kripke se sicer strinja s tistimi, ki pravijo, da so nekateri predikati prazni zgolj kontingentno in ne po nujnosti, ne strinja pa se, da je tak primer samorog:

> »Pravi se, da bi samorogi, čeprav vsi vemo, da jih ni, seveda *lahko* obstajali. Samorogi *bi* v določenih okoliščinah *obstajali*. To je primer nečesa, za kar mislim, da ne drži. Po mojem mnenju resnica morda ne bi smela biti podana z besedami, da je nujno, da samorogi ne obstajajo, temveč da preprosto ne moremo reči, v kakšnih okoliščinah bi samorogi lahko obstajali. Menim tudi, da čeprav bi arheologi ali geologi jutri odkrili fosile, ki bi ustrezali vsemu, kar iz mita o samorogu vemo o njih, to še vedno ne bi dokazalo, da so samorogi obstajali« (Kripke, 2000: 24).

Zakaj torej samorogi ne bi mogli obstajati oziroma, kot se v nadaljevanju natančneje izrazi Kripke, zanje ni mogoče reči, v kakšnih okoliščinah bi lahko obstajali? Nekaj o tem zvemo iz dodatka na koncu knjige, kjer Kripke pojasni, da pravzaprav zagovarja dve tezi – metafizično in epistemološko. Metafizična teza je, »da nobene protidejstvene situacije ne moremo primerno opisati, tako da bi v njej obstajali samorogi« (prav tam: 116), epistemološka pa, »da arheološko odkritje, da so obstajale živali z vsemi lastnostmi, ki jih pripisujemo samorogom v ustreznem mitu, ne bi bilo samo po sebi dokaz, da so samorogi obstajali« (prav tam).

Kripke izhaja iz predpostavke, da samorogi tvorijo mitološko, tj. hipotetično naravno vrsto, in se pri dokazovanju sklicuje na svoje ugotovitve o semantiki izrazov za naravne vrste.[4] Najprimernejša je primerjava s poimenovanjem dejanske živalske vrste, recimo tigrov. V *Slovarju slovenskega knjižnega jezika* je podana naslednja definicija besede »tiger«: »velika rumenkasto ali rdečkasto rjava zver s črnimi progami«. Definicija, kot vidimo, našteva značilne pojavne lastnosti tigrov, a kot prepričljivo dokazuje Kripke, ni tiger vsaka žival, ki tej

[3] Popravljen prepis z dodanimi opombami predavanj, ki jih je Kripke imel leta 1973 na Oxfordski univerzi. V zadnjih dveh delih mu samorogi ne služijo zgolj kot zanimiva ilustracija njegove semantične teorije izrazov za naravne vrste in z njo povezanega novega razumevanja nujnosti in apriornosti, temveč spregovori tudi o ontološkem statusu, ki ga imajo kot mitološki liki. O tej sicer zanimivi temi v tem prispevku ne bom govorila.

[4] Kot rečeno, Kripke obravnava izraze za biološke vrste (»tiger«, »limona«), substance (»voda«, »zlato«) in fizikalne količine (»toplota«, »svetloba«) po istem kopitu kot lastna imena. Več o Kripkejevi teoriji reference in njegovi kritiki stare, t. i. »Frege-Russellove« teorije najdete v Šuster, 2000.

definiciji ustreza, in četudi neka žival definiciji ne ustreza, je lahko še vedno tiger. Potrditev zadnje trditve bi bil denimo beli tiger, za katerega, čeprav ni rumenkasto ali rdečkasto rjav s črnimi progami, še vedno menimo, da je tiger.[5] Na podlagi značilnih pojavnih lastnostih namreč zgolj identificiramo domnevne člane vrste, da bi pa res pripadali tej vrsti, morajo imeti enako notranjo zgradbo, ki jo empirično odkrijejo biologi.[6] Zato tudi žival, ki bi povsem ustrezala slovarski definiciji tigra, ne bi bila tiger, če bi imela drugačno notranjo zgradbo od dejanskih tigrov, s katerimi smo v vzročnem stiku. Torej, če bi odkrili žival, katere zunanji videz bi popolnoma ustrezal slovarski definiciji tigra, a bi s preiskavo ugotovili, da ima plazilsko zgradbo, zanjo ne bi rekli, da pripada vrsti tiger.

Samorogi so tako kot tigri živalska vrsta, zato veljajo pri aplikaciji izraza »samorog« ista pravila kot za izraz »tiger«. V *SSKJ* najdemo sledečo definicijo besede »samorog«: »konju podobna bajeslovna žival s čarovnim rogom na čelu«. Če odmislimo mitološke pritikline, potem definicija podaja značilne pojavne lastnosti samorogov – konju podobna žival z enim rogom na čelu –, nič pa ne pove o notranji zgradbi samorogov. A če ne poznamo notranje zgradbe vrste, potem je ne moremo edinstveno določiti. V primeru tigrov smo preučili paradigmatične primerke, s katerimi smo v vzročnem stiku, in tako empirično določili njihovo skupno notranjo zgradbo, a samorogi dejansko ne obstajajo, zato tega ne moremo storiti. Možno je seveda, da bi obstajale živali, ki so videti kot samorog iz mita, a kaj ko je prav tako možno, da se med seboj razlikujejo po notranji zgradbi – ene so sesalci, druge plazilci, spet tretje dvoživke. Za katero od njih bi lahko rekli, da so člani vrste samorogov? Ker nam mit ne posreduje podatkov o njihovi notranji zgradbi, Kripke meni, da za nobeno od teh možnih vrst ni mogoče reči, da je vrsta samorogov (Kripke, 2000: 116, Kripke, 2011: 65–66 in Kripke, 2013: 43–47).

Trditev, ki jo Kripke poimenuje metafizična teza, torej ni, da samorogi sami po sebi ne bi mogli obstajati. Potem bi šlo, kot se izrazi, za »nemožnost običajne vrste, kakor je denimo nujen neobstoj okroglih kvadratov« (Kripke, 2011: 69). Samorogi nam povzročajo težave, ker mit premalo pove o njih, da bi jih lahko določili kot edinstveno vrsto, in zato pravzaprav sploh ne moremo reči, kdaj bi nekaj bilo samorog. Njihov ustrezni zunanji videz namreč ni ne zadosten ne nujen pogoj za pripadnost vrsti – ta je določena z notranjo zgradbo. Te ne poznamo, saj mit o tem molči, samorogi pa v dejanskem svetu ne obstajajo, zato je ne morejo empirično odkriti biologi. Izraz »samorog« zato nima popolnoma določenega pomena in ni kriterija na podlagi katerega bi posamezne živali identificirali kot člane te vrste. Ker pa je po Kripkeju izraz za naravno vrsto togi označevalec, ki v vseh možnih svetovih označuje isto vrsto kot v dejanskem svetu,[7] pravzaprav nimamo ničesar oprijemljivega, o čemer bi odločali, ali bi lahko obstajalo ali ne. Zaradi tega sklep ni, da samorogi ne morejo obstajati, saj

[5] 25. 4. 2005 je denimo *Delo* poročalo, da se je na jugovzhodu Španije skotil beli, modrooki tiger brez prog.
[6] Več o Kripkejevem zagovoru esencializma v Miščević, 2002.
[7] O tem, da je v definicijo togega označevalca že vgrajena modalno stabilna referenca, glej tudi Šuster, 2002.

to predpostavlja, da smo imeli nek določen predmet in po tehtanju vseh možnih situacij prišli do sklepa, da ne obstaja v nobeni od njih, temveč, da za nobeno situacijo ne moremo reči, da v njej obstajajo samorogi. Ne vemo namreč dovolj, da bi sploh identificirali referenta izraza v dejanskem svetu, ki je potem ta, o katerem tehtamo, če bi lahko bilo o njem nekaj res oziroma če bi lahko obstajal v neki protidejstveni situaciji.

Pri dokazovanju metafizične teze se Kripke opre na pogoje, ki določajo aplikacijo izrazov za naravne vrste: izraz za določeno vrsto lahko uporabimo le za člane te vrste, ki jih identificiramo na podlagi njihove notranje zgradbe. Tu vzročna teorija reference nima bistvene vloge, kakor pravilno ugotovi Dummett (Dummett, 1993: 332), kajti temeljno vprašanje je, ali je kriterij za aplikacijo izraza ustrezen zunanji videz organizma ali ustrezna notranja zgradba. In ker je pravi kriterij notranja zgradba članov te vrste, bi lahko neko protidejstveno situacijo opisali kot situacijo, v kateri obstajajo samorogi, samo če bi res (tj. v dejanskem svetu) obstajali samorogi, s katerimi smo v *vzročnem* stiku, in bi biologi lahko odkrili njihovo notranjo zgradbo oziroma če bi bila ta natančno *opisana* v samem mitu (Kripke, 2011:67).

Druga teza, ki jo Kripke postavi o možnem obstoju samorogov, je epistemološka, in odgovarja na vprašanje, ali bi lahko odkrili, da so samorogi dejansko obstajali oziroma da dejansko obstajajo. Pri dokazovanju te teze se Kripke opre na vzročno teorijo reference. Predpostavimo, da arheologi odkrijejo fosile, ki imajo vse lastnosti, ki jih mit pripisuje samorogom. Bi to bilo samo po sebi dokaz, da so samorogi obstajali? Po vzročni teoriji reference ime referira na neki predmet zato, ker ima ta določeno vlogo v vzročnem pojasnilu rabe imena. Na začetku imamo neposredni stik med imenovalcem in imenovanim: imenovalec običajno poimenuje predmet tako, da pokaže nanj in pove njegovo ime. Nato se ime prenaša po verigi, pri čemer ima vsak uporabnik namen ime uporabiti tako, da se nanaša na to, kar je bilo referent predhodne rabe. Zato za določitev referenta imena ni dovolj (in ne nujno), da predmet izpolni imenu pridruženi opis, temveč mora obstajati vzročna veriga med uvedbo imena in kasnejšimi uporabami. Enako velja za izraze za naravne vrste. Prvi imenovalec je vrsto identificiral s pomočjo paradigmatskih primerkov: »Tiger je ta vrsta živali«. Pri vsaki nadaljnji rabi izraza pa ima vsak govorec namen tega uporabiti zgolj za člane te iste vrste, tj. za živali, ki imajo enako notranjo zgradbo kot ti dejanski primerki (s katerimi je bil v vzročnem stiku prvi imenovalec).

Nasploh lahko torej rečemo, da pri določitvi referenta oziroma obsega določenega izraza pomembno vlogo igra način, na katerega je bil izraz prvotno uveden, in vzročna oziroma zgodovinska povezava med njegovo uvedbo in poznejšo rabo. Zato samo odkritje živali z lastnostmi, ki jih pripisujemo mitološkim samorogom, ne zadošča za ugotovitev, da so samorogi obstajali. Morali bi tudi ugotoviti, trdi Kripke, da je mit zgodovinsko povezan s temi živalmi, tj. da mit govori o njih. Odkriti bi torej morali, da so samorogi v mitu opisani tako, kakor so, zato, ker so ljudje takrat, ko je nastal mit, bili zgodovinsko

in dejansko povezani s temi živalmi. Če pa nasprotno ugotovimo, da je mit popolna izmišljotina, je dejanski obstoj živali, ki so videti kot samorog iz mita, golo naključje in ni res, da so mitološki samorogi nekoč obstajali (Kripke, 2000: 116–117, Kripke, 2011: 66–67 in Kripke, 2013: 48–50).

Kripke to ugotovitev podkrepi s primerom Sherlocka Holmesa. Četudi bi odkrili, da je v resnici obstajal detektiv, ki je počel vse to, kar počne Sherlock Holmes v romanih Conana Doyla, to še ne bi pomenilo, da je ta človek Sherlock Holmes. Če je namreč Conan Doyle pisal izmišljotino in ne resnično zgodbo o tem človeku, gre v tem primeru le za naključno podobnost literarnega lika resničnemu človeku (Kripke, 2000: 117).

3. Dummettova kritika

Kripke torej trdi, da nobene protidejstvene situacije ne moremo primerno opisati tako, da bi v njej obstajali samorogi, in da odkritje fosilov živali z vsemi pojavnimi lastnostmi mitološkega samoroga samo po sebi ne dokazuje, da so samorogi obstajali. Dummett v članku »Could There Be Unicorns?« (Dummett, 1993) nasprotno trdi, da bi samorogi lahko obstajali in da če bi odkrili dejanske živali, ki bi bile videti kot mitološki samorogi, bi jih upravičeno imenovali samorogi.

Najprej se loti Kripkejevega epistemološkega argumenta, ki je, kot smo videli, bistveno odvisen od vzročne teorije reference. Da bi dejanske živali, ki so videti kot samorogi, upravičeno proglasili za samoroge, moramo dokazati, da mit govori prav o njih, da torej med mitom in temi živalmi obstaja zgodovinska povezava. Če je mit popolna izmišljotina, potem dejanske živali ne morejo biti samorogi, tako kot Sherlock Holmes kot leposlovni lik ne bi mogel biti resnična oseba. Če bi obstajale živali, ki bile videti kot mitološki samorogi, ali pa oseba z junaštvi Sherlocka Holmesa, bi to namreč bilo zgolj naključje, pa naj se zdi še tako neverjetno. Dummett se strinja, da imena leposlovnega lika ne moremo z istim smislom uporabiti za resnično osebo, prav tako kot ne moremo o resnični osebi domnevati, da bi lahko bila zgolj izmišljeni lik (Dummett, 1993: 334). Vendar pa se primer Sherlocka Holmesa razlikuje od primera samorogov, saj o prvem vemo, da je plod domišljije Conana Doyla, medtem ko »avtorji« mitov o samorogih niso znani, zato ne vemo, ali so si samoroge izmislili ali pa gre za bolj ali manj ustrezen opis resničnih živali. Dummett tako spomni, da so v srednjem veku številni verjeli, da samorogi res obstajajo (prav tam: 337). Sama bi dodala še, da so samorogi pravzaprav na zahodu prvič omenjeni v naravnih filozofijah antičnih piscev, katerih namen je bil prikaz stvarnosti in ne pisanje fikcije. Ti so poročali, da samorogi živijo na koncu sveta, v Indiji.[8]

[8] V *Encyclopedia Mythica* piše, da jih je prvi omenil grški »zgodovinar« Ktesias leta 398 pr. n. št.. Opiral naj bi se na zgodbe popotnikov in v opisu samoroga združil nekatere lastnosti indijskega nosoroga, himalajske antilope in divjega osla. Bili naj bi »divji osli, ki so tako veliki kot konji oziroma še večji. Njihov trup je bel, glava temno rdeča in oči temno modre. Na čelu imajo en sam rog, ki je približno pol metra dolga rem V Encyclopedia Mythica *Online*, http://www.pantheon.org/articles/u/unicorn.html, pridobljeno 10.8.2013).

Za samoroge torej ne moremo s tako gotovostjo kot za Sherlocka Holmesa trditi, da so izmišljeni liki, tj. zgolj plod človeške domišljije, in da je beseda »samorog« ime izmišljene naravne vrste, s katero ne moremo označiti obstoječih živali. Če bi danes potemtakem odkrili fosile živali, ki bi bili podobni samorogom, ne bi mogli kar takoj zavreči možnosti, da lik samoroga dejansko temelji na teh živalih in da miti, čeprav jim napačno pripisujejo čudežno moč, v osnovi govorijo o njih. Možnost, da bi odkrili take živali, se zdi dokaj neverjetna, ampak če bi jih, potem se zdi bolj verjetno, da med njimi in miti obstaja povezava, kot pa da gre zgolj za naključno ujemanje. Podobno bi bilo v primeru Sherlocka Holmesa: če ne bi bilo znano, ali je Conan Doyle pisal zgodbice ali opisoval dogodivščine resničnega detektiva, in bi odkrili, da je taka oseba res obstajala, potem bi predvidevali, da je pisal o resnični osebi, saj je naključno ujemanje bolj neverjetna od obeh možnosti.

Dummett gre še korak naprej in trdi, da če ne vemo, da je bil izraz vpeljan kot ime za izmišljeno naravno vrsto, je razumno domnevati, da je od takrat prišlo do nenamerne spremembe smisla, in da izraza danes ne dojemamo kot intrinzično izmišljenega, temveč ga uporabljamo glede na njegovo opisno vsebino (Dummett, 1993: 338). Po njegovem naša jezikovna praksa takšno spremembo v primeru besede »samorog« tudi potrjuje in da bi s tem imenom poimenovali živali, ki bi bile podobne mitološkim samorogom, ne da bi sploh predpostavljali, da so »avtorji« mitov pisali o njih.

Toda, nadaljuje Dummett, dejstvo, da samorogi lahko obstajajo, opravičuje prehod k trditvi, da čeprav samorogi dejansko ne obstajajo, bi lahko. Kripke sicer razlikuje med epistemsko in metafizično modalnostjo, zato ta prehod od »lahko« v »bi lahko« ni samoumeven. To razlikovanje lahko pojasnimo na primeru identitetnega stavka »Večernica je Danica« (Kripke, 2000: 78–79, 105–107). Ta je metafizično nujno resničen, saj obe imeni v dejanskem svetu označujeta isti predmet, tj. planet Venero, kar pomeni, da kot toga označevalca označujeta ta isti predmet tudi v vsakem možnem svetu, v katerem obstaja, drugače pa ničesar. Ker pa je referent imena »Večernica« prvotno določen na podlagi opisa »najsvetlejše nebesno telo na nočnem nebu«, ki je zgolj kontingentna oznaka Venere, si lahko zamislimo situacijo, v kateri je ta opis ne izbere, temveč neko drugo nebesno telo, ki najsvetleje sije na nočnem nebu. Recimo, da se ime »Danica« še vedno nanaša na Venero, ker je »najsvetlejše nebesno telo na jutranjem nebu«. V tej situaciji bi bila resnična izjava »Večernica ni Danica«, kajti imeni bi označevali dve različni nebesni telesi. Takšna situacija je epistemsko možna in lahko bi se izkazalo, da je opis dejanskega sveta. Toda empirična raziskava razkrije, da ni in da se v dejanskem svetu obe imeni nanašata na isto nebesno telo, tj. Venero. Situacija zato ni metafizično možna.

Je torej obstoj samorogov zgolj epistemsko možen? Razlika med epistemsko in modalno možnostjo nastane, pravi Dummett, ker lahko določeno stvar identificiramo s pomočjo njenih naključnih lastnostih in ne zgolj bistvenih. Tako je bistveno za Večernico, da je istovetna z Danico, toda referenta imena

»Večernica« ne določimo na podlagi te bistvene lastnosti, temveč na podlagi naključne lastnosti »najsvetlejše telesno nebo na večernem nebu«. Zato se nam zdi, da bi Večernica lahko bila različna od Danice, čeprav to ni tako. Za naravne vrste imamo podoben model: obseg izraza za vrsto določimo na podlagi določenih pojavnih lastnostih nekaj živali, ki pa v resnici niso ne nujne ne zadostne, da bi žival pripadala tej vrsti, kajti bistvena za članstvo v vrsti je določena notranja zgradba. Tako se nam torej zdi, da bi tiger lahko imel drugačno notranjo zgradbo, kot jo ima, čeprav to ni tako (Dummett, 1993: 339),

Dummett se sicer ne strinja s Kripkejem, da je pripadnost živalski vrsti določena na podlagi notranje zgradbe, saj je ta po njegovem le zanesljiv pokazatelj pravega kriterija pripadnosti vrsti, tj. skupnega izvora oziroma evolucijske zgodovine. Tako denimo pravi, da »četudi bi bila na Marsu bitja, ki bi bila zemeljskim tigrom podobna po zunanjem videzu *in* po notranji zgradbi, ta ne bi bila tigri, če ne bi izvirala iz zemeljskih tigrov« (Dummett, 1993: 332). Sam Kripke v kasnejšem članku priznava, da je najbrž bil nekoliko površen pri obravnavi živalskih naravnih vrst in da je uporabljal izraz »vrsta« dokaj neformalno. Kot dodatni kriterij za članstvo v vrsti pa omeni prav mesto na evolucijskem drevesu (Kripke, 2011: 67), zato lahko brez posledic omenjeni nujni pogoj za članstvo v vrsti dopolnimo s skupnim izvorom. Za sam argument sicer ni toliko pomembno, ali je nujni pogoj za članstvo v vrsti enaka notranja zgradba ali evolucijska zgodovina, temveč predvsem to. da ne zadošča, da žival ustreza slovarski definiciji, v kateri so naštete značilne pojavne lastnosti, s tem pa se strinjata tako Kripke kakor Dummett.

O epistemski možnosti, ki se razlikuje od metafizične, bi torej lahko govorili, če bi imeli opravka s sodbo, da bi samorog lahko imel drugačno notranjo zgradbo od dejanske, a to ni predmet naše razprave. Vprašanje, ki nas zanima, je, ali so kriteriji pri aplikaciji izraza »samorog« v hipotetičnih kontekstih drugačni od kriterijev v dejanskih kontekstih. Zanima nas torej pod katerimi pogoji, v skladu z našo obstoječo jezikovno prakso, bi bilo prav določeno žival imenovati »samorog« in Dummett meni, da so kriteriji isti v dejanskih in hipotetičnih kontekstih. Predpostavimo, da nek predmet običajno identificiramo na podlagi neke naključne lastnosti, in si zamislimo situacijo, v kateri predmet te identifikacijske lastnosti nima. Navkljub temu še vedno vemo, za kateri predmet gre, ker je del našega obstoječega razumevanja izraza, s katerim se nanj nanašamo, da ima svoje bistvene lastnosti. Ni sicer treba, da vemo natančno katere, saj zadostuje, da vemo za katere vrste stvar gre in posledično tudi kateri kriterij identitete velja zanje. Tako denimo vemo, da »Večernica« referira na nebesno telo, četudi ne vemo, ali je planet ali zvezda. Podobno posamezne živali identificiramo kot tigre na podlagi njihove zunanjosti, a vemo tudi, da je »tiger« izraz za živalsko vrsto in da je kriterij za članstvo v vrsti določena notranja zgradba. Tako dopuščamo, da smo žival pravilno identificirali samo, če ima isto notranjo zgradbo kot druge živali, ki smo jih enako identificirali. A na tej prvi stopnji ni potrebno vedeti, kakšna pravzaprav je ta skupna notranja zgradba, ki jo lahko šele naknadno (empirično) odkrijemo. In zato tudi vemo, da bodo v

hipotetičnih situacijah tigri živali iste vrste kot te, ki jih dejansko identificiramo na podlagi zunanjega videza. Tako v dejanskih kot hipotetičnih situacijah torej velja isti kriterij identitete. Posledično so za Dummetta vprašanja o t. i. »metafizični možnosti« istovetna »vprašanjem o naših obstoječih kriterijih za določanje reference ali aplikacijo« izrazov (Dummett, 1993: 341).

Glede samorogov se nam torej zastavljata dve vprašanji. Prvo vprašanje je: »Ali lahko obstajajo samorogi?«, oziroma natančneje: »Upoštevaje naše vedenje o tem, kakšen je svet, ali bi lahko obstajale živali, ki bi jih, če bi vedeli zanje, upravičeno, tj. v skladu z načeli, ki določajo našo sedanjo rabo te besede, imenovali 'samorogi'?« Drugo vprašanje je: »Ali bi lahko obstajali samorogi?«, oziroma natančneje: »Ko si svet predstavljamo drugače, kakor vemo, da je, ali bi ga lahko konsistentno opisali, kakor da vsebuje samoroge, se pri tem držali dejanskega sedanjega smisla te besede in bi ta tudi zagotavljal razumljivost opisa?« Odgovor na obe vprašanji je odvisen od pogojev, ki jih dejanska jezikovna praksa postavlja za identifikacijo neke živali kot samoroga, in ali so ti, glede na to, kakšen je po našem vedenju svet, lahko izpolnjeni. Gre pa za izpolnjevanje istih semantičnih pogojev: pogoj, ki določa, kdaj lahko v hipotetičnih situacijah neko stvar upravičeno imenujemo »samorog«, je namreč isti pogoj, ki določa, kdaj lahko neko stvar, ki jo dejansko srečamo, upravičeno imenujemo »samorog« (Dummett, 1993: 341). Dummett zato meni, da se Kripke ne more sklicevati na razliko med epistemsko in metafizično možnostjo in tako preprečiti prehoda od »Samorogi lahko obstajajo« k »Samorogi bi lahko obstajali«. Razlikovanje med metafizično in epistemsko modalnostjo torej pri vprašanju obstoja samorogov ne igra nobene vloge.

Če je bila v epistemološkem argumentu bistvena zahteva, da med sedanjo rabo besede »samorog« in njeno vpeljavo obstaja zgodovinska povezava, pa Kripke pri dokazovanju metafizične teze izhaja iz pogojev, ki določajo aplikacijo izrazov za naravne vrste: z izrazom za določeno vrsto lahko poimenujemo le člane te vrste, ki jih definiramo na podlagi njihove notranje zgradbe. Ker samorogi dejansko ne obstajajo, mit pa opiše le njihove pojavne lastnosti, nič pa ne pove o njihovi notranji zgradbi, ne moremo reči, v kakšnih okoliščinah bi samorogi lahko obstajali. Oziroma, kot Dummett poenostavi Kripkejevo metafizično tezo: če samorogi ne obstajajo v dejanskem svetu, potem ne bi mogli obstajati.

Dummett meni, da je metafizični argument nekoherenten. Ključnega pomena za argument je, ali dejansko obstajajo živali, ki so podobne samorogom iz mitov. Če obstajajo, potem te živali *so* samorogi, če pa ne, potem samorogi ne morejo obstajati, saj ni določenega pogoja, ki ga mora žival izpolniti, da bi bila samorog. Po Dummettu pa Kripke hoče zagovarjati močnejšo trditev, kakor jo izraža pogojnik v povednem naklonu: »Če obstajajo živali, podobne samorogom, te so samorogi«. Kajti naj bo še tako gotov, da dejansko ne obstajajo samorogom podobne živali, še vedno želi reči, da če bi obstajale, bi jih upravičeno imenovali »samorogi« (pogojnik v veznem naklonu). A obojega ne moremo konsistentno trditi, pravi Dummett, »ker govoriti o tem, kako bi stvari bile v dejanskem svetu

v nekih hipotetičnih situacijah, ni nič drugega, kakor govoriti o tem, kako stvari so v določenih možnih svetovih« (Dummett, 1993: 343).

Po Dummettu torej Kripke zagovarja nezdružljivi tezi. Prvič, »Možno je, da obstajajo samorogi«, ker če bi v dejanskem svetu obstajale samorogom podobne živali, bi te bile samorogi. In da bi v dejanskem svetu lahko obstajale samorogom podobne živali, je Dummett pokazal, ko je upravičil prehod od »Samorogi *lahko* obstajajo« v »Samorogi *bi lahko* obstajali«. Drugič, »Ni možno, da obstajajo samorogi«, ker če v dejanskem svetu ni takih živali in vemo, da jih ni, potem jih ni niti v zgolj možnih svetovih, saj ni določen pogoj, ki bi ga posamezna žival morala izpolniti, da bi bila samorog v nekem možnem svetu.

Dummett zato meni, da Kripke mora sprejeti trditev, da bi samorogi lahko obstajali. A ker teh v dejanskem svetu ni in ker mit ne uspe določiti za kakšno vrsto gre, se znajdemo v težavah, kot pravilno ugotavlja Kripke, saj bi samorogi lahko bili denimo iz reda lihoprstih kopitarjev (red *Perissodactyla*), v katerega sodijo konji, ali pa iz reda sodoprstih kopitarjev (red *Artiodactyla*), v katerega sodijo jeleni. V jeziku možnih svetov bi potemtakem imeli dejanski svet w, v katerem ni samorogov, svet u, v katerem so samorogi iz reda sodoprstih kopitarjev, in svet v, v katerem so samorogi iz reda lihoprstih kopitarjev. Možnost obstoja obeh vrst samorogov izhaja iz dejstva, da po Dummettu za identifikacijo živali kot članice določene vrste ni potrebno poznati za to vrsto značilno anatomsko zgradbo – zadošča poznavanje značilnih pojavnih lastnosti in spoznanje, da morajo imeti določeno, vsem članom vrste skupno anatomsko zgradbo, ki pa se jo določi šele naknadno, v vsakem možnem svetu posebej. V sistemu S5, za katerega je značilna ekvivalenčna relacija, seveda te modalne sodbe ne morejo biti vse resnične. V svetu u so samorogi nujno iz reda sodoprstih kopitarjev, v svetu v pa nujno iz reda lihoprstih kopitarjev. Zato svet v ni relativno možen glede na svet *u* in obratno. V svetu w je možno, da so samorogi iz reda lihoprstih kopitarjev, zato w ne more biti relativno možen glede na *u*, v katerem so samorogi nujno iz reda sodoprstih kopitarjev, medtem ko *u* je relativno možen glede na w. Enako velja za odnos med w in v. Iz tega razberemo, da relacija relativne možnosti ni simetrična, je pa refleksivna in tranzitivna, kar je značilno za sistem S4 (Dummett, 1993: 346–347). Če ima potemtakem Dummett prav in bi samorogi res lahko obstajali, metafizični nujnosti ne ustreza SS, temveč zgolj S4, in metafizično nujne resnice ni mogoče definirati kot tiste, ki je resnična v vseh možnih svetovih. A prav ta definicija povezuje analizo modalnih pojmov s pomočjo možnih svetov z našim običajnim razumevanjem modalnosti.

4. Kdo ima prav – Kripke ali Dummett?

Najprej naj opozorim, da je razprava o možnosti obstoja samorogov smiselna le, če sprejmemo, da so samorogi biološko možne živali in da tvorijo posebno naravno vrsto. To seveda ni samoumevno. Mitološka bitja so običajno t. i. mešana bitja, ki sestojijo iz različnih živalskih delov. Egipčanska sfinga je denimo pošast s človeško glavo in levjim telesom, kentavri pa imajo človeško

poprsje in konjski trup. Samorog je največkrat opisan kot konju podobna žival z enim čarovnim rogom na čelu, a včasih tudi kot žival, katere zadek je jelenji, rep levji, trup in glava pa konjska.[9] Ali bi taka bitja sploh lahko obstajala v resničnem svetu? Prav tako naj bi samorog imel čudežne, tj. nadnaravne moči, sposoben naj bi bil denimo očistiti zastrupljeno vodo tako, da vanjo pomoči svoj rog. In četudi morda sprejmemo, da samorogi so biološko možna bitja, lahko vseeno zanikamo, da tvorijo posebno naravno vrsto. V mitih tega podatka gotovo ne najdemo. Kripke in Dummett sicer oba privzameta, da so biološko možni in naravna vrsta, katere slovarska definicija naj bi bila konju podobna žival z enim rogom na čelu; čudežne lastnosti preprosto odmislita.

Kdo ima torej prav – Kripke ali Dummett? Bi samorogi lahko obstajali? Dummett najprej poskuša dokazati, da samorogi lahko obstajajo, pri čemer je njegova glavna poanta, da danes ne vemo več, kako je bil vpeljan izraz »samorog«, tj. ali so z njim poimenovali izmišljena ali dejansko obstoječa bitja. Danes zato pri uporabi izraza izhajamo iz njegove opisne vsebine in bi se ob odkritju fosilov, ki bi ustrezali tej vsebini, odzvali tako, da bi jih imenovali »samorogi«. Strinjam se z Dummettom, da bi se v tem primeru večina govorcev odzvala na tak način. Gotovo bi kdo pripomnil: »Zanimivo, samorogi so res obstajali, jaz pa sem mislil, da so čisto izmišljotina«. Ampak ta odziv po mojem ravno dokazuje, da izraz razumemo kot ime za izmišljena bitja in da nas odkritje dejanskih živali, ki so videti kot mitološki samorogi, pripravi do razmišljanja, da so morda »avtorji« mitov vendarle opisovali dejansko obstoječe živali, ki so jim zgolj dodali nekaj čudežnih lastnosti. Nekaj podobnega bi se zgodilo v primeru, da bi konji že zdavnaj izumrli in bi o njih vedeli le iz srednjeveških zgodb o vitezih in njihovih junaštvih. Dokler ne bi našli fosilov konjev, bi lahko mislili, da so le plod avtorjeve domišljije in ne opis resničnih živali. Prav tako ni res, da bi se pri uporabi izraza »samorog« ravnali le po pridruženem opisu, saj če bi znanstveniki hkrati odkrili, da so te živali izumrle že zdavnaj, tj. dolgo pred nastankom mitov, ne bi bilo tako samoumevno, da jih lahko imenujemo »samorogi«. To je še bolj očitno, če si zamislimo, da bi odkrili samorogom podobne živali, za katere bi se izkazalo, da so se razvile šele po nastanku mitov. Teh gotovo ne bi imenovali »samorogi« in bi se najbrž čudili nad naključjem, da je narava ustvarila nekaj tako podobnega mitološkim samorogom.

Ne zdi se mi torej, da naša jezikovna praksa podpira Dummettov sklep. Z izrazom »samorog« ne bi označili kar vsako žival, ki izpolni njegovo slovarsko definicijo, temveč je naš namen, da ga uporabimo tako, da bo imel istega referenta kot pri predhodni rabi. Izvira iz mita in imenuje izmišljeno naravno vrsto, zato nima referenta v dejanskem svetu. Možnost, da samorogi dejansko obstajajo, smo pripravljeni dopustiti zgolj zato, ker ne vemo, kaj so »avtorji« mitov imeli v mislih. Najverjetneje so si samoroge izmislili in potem gre za izmišljena bitja, bi pa odkritje konjem podobnih živali z enim rogom na čelu zelo povečalo verjetnost, da so pisali o resničnih živalih, ki so jim zgolj pripisali

[9] Tak opis najdemo v *Miti in legende* (1998).

določene čudežne lastnosti. V tem primeru bi lahko pritrdili, da so »avtorji« mitov pravzaprav pisali o teh živalih in da so ves ta čas bile referent izraza »samorog«. To bi bilo podobno primeru preroka Jone, o katerem v Bibliji zvemo, da ga je Bog poslal pridigat v Ninive in da ga je pogoltnila velika riba. Čeprav po mnenju zgodovinarjev nič od tega ni res, pa naj bi Jona res živel, zato navkljub napačnem opisu, ime označuje resničnega človeka. Vzročna veriga med vpeljavo izraza in današnjo rabo obstaja. Čeprav mislim, da Dummettu ne uspe dokazati, da bi vsako žival, ki bi bila podobna mitološkim samorogom, upravičeno imenovali »samorog« in da naša jezikovna praksa ne podpira Kripkejeve zahteve po vzročni oziroma zgodovinski povezavi med vpeljavo izraza in sledečimi rabami, pa ima prav, da samorogi lahko obstajajo. A temu ne oporeka niti Kripke, ki Dummettov pogoj, da dejansko obstajajo živali, ki so videti kot samorogi iz mitov, dopolni s pogojem, da med njimi in mitom obstaja zgodovinska povezava. Tega scenarija prav nič ne preprečuje, saj bi vsako empirično odkritje lahko bilo tudi drugačno in bi se torej lahko izkazalo, da živimo v svetu, v katerem dejansko obstajajo mitološkim samorogom podobne živali in da mit dejansko govori o njih. Torej je res, da bi dejanski svet lahko bil tak da bi v njem obstajali samorogi.

Videli smo, da pri metafizičnem argumentu igra bistveno vlogo to, da so izrazi za naravne vrste togi označevalci, ki v vsakem možnem svetu označujejo isto stvar kot v dejanskem svetu. Ker izraz »samorog« ne označuje ničesar v dejanskem svetu, ne označuje ničesar niti v nobenem drugem možnem svetu. Dummett ima zato prav, da je za argument ključnega pomena, ali dejansko obstajajo živali, ki so podobne samorogom iz mitov, Kripke bi še dodal pogoj, da miti res govorijo o njih. Če take živali obstajajo, potem lahko seveda ugotovimo, ali imajo (večina) enako notranjo zgradbo in potemtakem edinstveno določimo vrsto samorogov. In ker samorogi obstajajo, bi lahko tudi obstajali, in lahko ugotavljamo, kaj bi lahko oziroma bi nujno bilo res o živalih, ki pripadajo *tej* vrsti. Če pa take živali ne obstajajo, potem ni načina, da bi edinstveno določili vrsto samorogov, saj ne moremo določiti notranje zgradbe, ki je bistvena za članstvo v tej vrsti. Zaradi tega pa tudi ni načina, na katerega bi lahko v zgolj hipotetičnih situacijah ugotavljali, ali določene živali pripadajo *tej* vrsti samorogov, saj ne vemo, kaj naj bi *ta* vrsta bila. Če smo povsem natančni, obstaja še en način, na katerega bi lahko bila edinstveno določena vrsta: če bi mit opisal notranjo zgradbo, ki jo morajo imeti živali, da bi bile samorogi, potem bi samorogi lahko obstajali, čeprav v dejanskem svetu ni nobene take živali. Ampak ostanimo pri osnovnem modelu, saj mit tega podatka ne posreduje.

Torej, naš svet je tak, da samorogi ne obstajajo, zato velja, da nobene protidejstvene situacije ni mogoče opisati tako, da bi v njej obstajali samorogi. Dummett poenostavi, da je sklep, da samorogi ne bi mogli obstajati. To pa je v protislovju s trditvijo, da če bi obstajale mitološkim samorogom podobne živali, bi jih upravičeno imenovali »samorogi«, ki jo po Dummettu Kripke prav tako mora sprejeti. Kot smo videli, je dokazoval, da iz »Samorogi lahko obstajajo« sledi »Samorogi bi lahko obstajali«, saj so pogoji, ki jih dejanska praksa

postavlja za identifikacijo neke živali kot samoroga, enaki v dejanskih in hipotetičnih situacijah. Posamezne živali identificiramo kot samoroge na podlagi njihove zunanjosti, a vemo tudi, da je »samorog« izraz za živalsko vrsto in da morajo torej te živali imeti tudi enako notranjo zgradbo, ki jo sicer lahko naknadno ugotovimo. Prav določena notranja zgradba je namreč kriterij za članstvo v vrsti tako v dejanskih kot tudi v hipotetičnih situacijah. To drži, a na kakšen način dejstvo, da so kriteriji za aplikacijo izraza »samorog« enaki v dejanskih in hipotetičnih kontekstih, podpira prehod od »Samorogi lahko obstajajo« k »Samorogi bi lahko obstajali«? Tu gre vendarle za dva hipotetična konteksta in ne za en dejanski in en hipotetični kontekst in ključno vprašanje je, ali gre v obeh kontekstih za iste vrste možnost.

Dummettov odgovor je, da resničnost obeh stavkov temelji na izpolnjevanju istega semantičnega pogoja, ki določa, kdaj neko stvar upravičeno imenujemo »samorog«, zato naj bi šlo za isto vrsto možnosti. A med njima obstaja razlika. Ko sprašujemo, ali lahko obstajajo samorogi, drži, da bomo za samoroge šteli le konjem podobne živali z rogom na čelu, ki imajo vsi enako notranjo zgradbo, tehtamo pa situacije, kjer je vse tako, kakor v našem svetu, odmislimo le empirično evidenco, ki je pokazala, da takih živali ni. Dejansko torej presojamo možnosti, v katerih bi uporabljali drugačen jezik, saj bi izraz »samorog« označeval edinstveno živalsko vrsto. Podobno ime »Večernica« lahko označuje neko drugo nebesno telo, če odmislimo dejstvo, da je dejansko ime planeta Venera. V tem primeru odmislimo empirični podatek, da je Venera najsvetlejše nebesno telo na večernem nebu, in si zamislimo, da ta pogoj izpolnjuje denimo Mars. V primeru samorogov pa odmislimo empirični podatek, da samorogi ne obstajajo, in si zamislimo take živali. Dejansko ugotavljamo, kaj je možno na ravni smisla, torej če odmislimo, na kaj se lastna imena in izrazi za naravne vrste kot togi označevalci dejansko nanašajo. Pri metafizični modalnosti pa ne gre za raven jezika, temveč za objektivno raven. Sprašujemo se, kaj je možno ali nujno glede tega dejanskega sveta in teh dejanskih predmetov, ali bi Večernica lahko bila Mars in ne, ali bi v nekih drugih okoliščinah ime »Večernica« lahko označevalo planet Mars. Bi Sherlock Holmes lahko obstajal glede na to, da je v dejanskem svetu ta leposlovni lik plod domišljije Conana Doyla, in ne, ali bi lahko v nekih drugih okoliščinah to ime označevalo neko dejansko osebo. Ali bi samorogi lahko obstajali glede na to, da jih v dejanskem svetu ni, in ne, ali bi v nekih drugih okoliščinah izraz »samorog« lahko označeval dejansko obstoječe živali te vrste.

Gre torej za dve vrsti možnosti. Ko Dummett pravi, da samorogi lahko obstajajo, pravzaprav postavlja semantično trditev. Če bi bil naš svet tak, da bi obstajale konjem podobne živali z enim rogom na čelu, bi izraz »samorog« označeval edinstveno vrsto. A to ni tako, v dejanskem svetu samorogi ne obstajajo in v našem jeziku »samorog« ne označuje edinstvene vrste. Podobno kot če bi rekli, da je Večernica lahko Mars. Če bi bil naš svet tak, da bi bil Mars najsvetlejše nebesno telo na večernem nebu, bi ime »Večernica« označevalo planet Mars. A to ni tako, v našem svetu ta opis edinstveno izpolnjuje Venera, zato je Večernica

Venera in ne bi mogla biti Mars. Tu ne gre za ugotavljanje, na kaj bi se imena in izrazi za naravne vrste nanašali, če bi bil dejanski svet drugačen od tega, kakršen je, temveč za metafizično možnost, kaj bi, glede na to, kakšen naš svet je, lahko bilo res o tem svetu in ne o nekem poljubnem svetu.

Torej ne drži, da je Kripkejev metafizičen argument nekoherenten. Kripke lahko trdi, da če samorogi dejansko ne obstajajo, potem ne morejo obstajati oziroma da nobene protidejstvene situacije ne moremo ustrezno opisati tako, da bi v njej obstajali samorogi, in obenem sprejme trditev: »Če bi dejansko obstajale mitološkim samorogom podobne živali, bi jih upravičeno imenovali ‘samorogi’«. Zaradi tega se ne bo zapletel v protislovje, saj iz omenjenega pogojnika ne sledi, da samorogi lahko obstajajo. Ne smemo namreč mešati semantične in metafizične ravni, ravni jezika in ravni predmetov. Podobno bi potem namreč iz pogojnika »Če bi v dejanskem svetu Mars bil najsvetlejše nebesno telo na večernem nebu, bi ga upravičeno imenovali ‘Večernica’« morali sklepati, da je Večernica lahko Mars. Temu nasprotuje tudi Dummett, ki je, kot smo videli, poskušal pokazati, da se v primeru samorogov ne da sklicevati na razlikovanje med epistemsko in metafizično možnostjo, kot se to da v primeru lastnih imen. Zakaj ne? Za kaj sploh gre pri epistemski možnosti? Posamezna dejstva o svetu, ki jih odkrijemo empirično, bi lahko bila tudi drugačna, in posledično bi imena in predikati imeli druge referente oziroma obsege. Najsvetlejše nebesno telo na večernem nebu bi tako lahko bil Mars in ime »Večernica« bi referirala nanj in ne na Venero, in v tem svetu bi bile metafizično nujne ali možne druge stvari kot v našem svetu. To se razlikuje od primera samorogov v toliko, da pri slednjem ne gre za to, da ima ime drugega referenta, temveč da ga sploh ima, saj je v našem svetu »biti samorog« predikat, ki ima prazen obseg, a načeloma gre še vedno za isto stvar.

Tudi v primeru samorogov torej lahko razlikujemo med metafizično in epistemsko možnostjo, pri čemer gre pri epistemski možnosti pravzaprav za to, da ostanemo na ravni smisla in ne upoštevamo, na kaj se izrazi našega jezika dejansko nanašajo. Dummett pravi, da v primeru izrazov za naravne vrste razlike ni mogoče vzpostaviti, ker se že v osnovi ne zanašamo zgolj na slovarsko definicijo, temveč vemo, da morajo primerki imeti tudi enako notranjo zgradbo, ki pa jo lahko točno določimo šele naknadno. Iz njegove predstavitve modela možnih svetov, ki naj bi dokazoval, da relacija relativne možnosti ni simetrična, je očitno, da po njegovem to enako notranjo zgradbo, ki je nujni pogoj za članstvo v vrsti, naknadno določimo v vsakem možnem svetu posebej in da se torej lahko od enega do drugega sveta razlikuje. To pa je v nasprotju z modalno stabilnostjo reference, ki je vgrajena v definicijo togega označevalca, kar pomeni, da Dummett dejansko ne ostaja znotraj parametrov Kripkejeve teorije.

Dummettu torej ni uspelo dokazati, da iz Kripkejevih lastnih predpostavk izhaja, da bi samorogi lahko obstajali in da je posledično za metafizično modalnost pravi sistem modalne logike S4 in ne S5.

KRIPKE AND DUMMETT ON THE POSSIBILITY OF THE EXISTENCE OF UNICORNS

Concerning the possible existence of unicorns Kripke defends two theses: a metaphysical thesis that no counterfactual situation is properly describable as one in which there would have been unicorns; and an epistemological thesis that an archeological discovery of animals with all the features attributed to unicorns in the appropriate myth would not in itself constitute proof that there were unicorns. In defending these theses, he relies on his semantic theory of natural kind terms and his causal theory of reference. Dummett rejects both, although slightly differently formulated, and argues that they are conflicting with our linguistic practice. Additionally, the metaphysical thesis cannot even be consistently defended within Kripke's theory, while the acceptance of the possibility that unicorns exist reveals S4 and not the intuitively more acceptable S5 as the right modal system for metaphysical modality. I will try to determine who of the two is right – Kripke or Dummett?

Keywords: Kripke, Dummett, unicorns, natural kinds, rigid designators, causal theory of reference, modality

Literatura

Dummett, M. (1993). »Could There Be Unicorns?«. V Dummett, M., *The Seas of Language*, Oxford: Clarendon Press, str. 328–348.

Kripke, S. (2013). *Reference and Existence: The John Locke Lectures*. New York: Oxford University Press.

Kripke, S. (2011). »Vacuous Names and Fictional Entities«. V Kripke, S., *Philosophical Troubles: Collected Papers, Volume I*, New York: Oxford University Press, str. 52–74.

Kripke, S. (2000). *Imenovanje in nujnost*. Prevedla Bojan Vučko in Marjan

Šimenc. Ljubljana: Krtina.

Miti in legende: Ilustrirana enciklopedija (1998). Prevedel Uroš Kalčič. Ljubljana: Mladinska knjiga.

Miščević, N. (2002). »Kripkejev zagovor esencializma«. *Analiza*, 6, str. 77–83.

Šuster, D. (2002). »O nujnosti *a posteriori*«. *Analiza*, 6, str. 84–103.

Šuster, D. (2000). »Lastno ime je togi označevalec«. V Kripke, S. (2000), *Imenovanje in nujnost*, Ljubljana: Krtina, str. 123–142.

»Unicorn« (2013). V *Encyclopedia Mythica*, http://www.pantheon.org/articles/u/unicorn.html.



Funkcionalizem in vprašanje umetne inteligence

Janez Bregant
Filozofska fakulteta Univerze v Mariboru

FUNKCIONALIZEM, RAČUNALNIKI IN »KITAJSKA SOBA«

V članku je predstavljena zveza med funkcionalizmom, ki mentalna stanja obravnava kot funkcionalna stanja, računalniki, ki so po njegovi računski različici sposobni misliti, in argumentom iz »kitajske sobe«, ki naj bi enkrat za vselej pokazal, da stroji nikoli ne bodo inteligentni. Prvi del prikazuje osnovno idejo in motivacijo za funkcionalizem, skupaj z glavnimi ugovori proti njemu, drugi del članka miselni eksperiment s »kitajsko sobo«, ki formuliran kot argument zavrača idejo o mislečih strojih na podlagi tega, da manipulacija s formalnimi simboli še ni dovolj za razumevanje njihove vsebine, tretji pa povzema glavne ugovore argumentu iz »kitajske sobe« in pri odgovoru nanje prinaša nekaj novih idej, ki zagovarjajo prepričanje, da so inteligentni računalniki kljub vsemu načeloma možni.

Ključne besede: »kitajska soba«, umetna inteligenca, simbolna hipoteza, jezik misli, računska teorija duha.

1 Funkcionalizem

> V začetku 60. let 20. stol. je Putnam (1960) oblikoval teorijo duha, ki je dobila ime *funkcionalizem*, njene začetke pa lahko najdemo že pri Aristotelu (Block, 1980b/2007: 147, 158). Ta je dušo opredelil kot skupek posebnih sposobnosti živega organizma, s katerimi usmerjamo in uravnavamo življenjske funkcije z namenom, da preživimo in napredujemo kot živo, dejavno, zaznavajoče in razumno bitje:[1] »Zmožnosti, o katerih smo govorili, so: prehranjevalna, želelna ali nagonska, zaznavna, gibalna po prostoru ter razumska. Rastlinam pripada zgolj prehranjevalna zmožnost, drugim živim bitjem pa je na razpolago tako ta kot zaznavna zmožnost. Če pa jim pripada zaznavnost, jim pripada tudi želelna ali nagonska zmožnost /.../ Marsikaterim živalim pa poleg teh zmožnosti pripada še zmožnost gibanja po prostoru, drugim pa tudi razumska zmožnost ter um, tako

[1] Aristotel je npr. trdil, da imajo stvari svoj namen – telos – po svojem bistvu.

recimo ljudem /…/« (Aristotel, 1993: 123–125). Danes bi lahko funkcionalizem na kratko definirali kot filozofsko teorijo, ki mentalna stanja označuje v pojmih njihovih vzročnih vlog, kaj to pomeni, pa si bomo ogledali v nadaljevanju.

Strojni funkcionalizem

Putnam je kot osnovo za funkcionalizem predlagal analogijo med človeško duševnostjo in t.i. Turingovim strojem. To je enostavni abstraktni sistem za dokazovanje teoretičnih predpostavk, ki si ga je leta 1936 izmislil Alan Turing,[2] sestavljen pa je iz gibljive *glave* in neskončnega *traku*. Glava se pomika preko polj, na katere je razdeljen trak in katera so zasedena s simboli (vsako polje z enim), ki so predvideni na končnem seznamu znakov. Delovanje stroja določa pet elementov/parametrov: (i) simbol, ki ga zazna (*z*), (ii) notranje stanje, v katerem se v danem trenutku nahaja (*s*), (iii) naslednji simbol v istem polju (*z + 1*), (iv) naslednje stanje (*s + 1*) in (v) naslednje polje/smer gibanja (*p*). Prva dva elementa se nanašata na sedanje obnašanje, ostali trije pa na obnašanje v naslednjem trenutku. V splošnem lahko glede na to stroj naredi naslednje: (a) ostane v stanju, v katerem je ali preide v neko drugo stanje, (b) pusti na traku zapisan isti simbol, ali ga izbriše, ali ga izbriše in vpiše novega ter (c) premakne se po traku levo ali desno ali ostane na istem mestu. Kje se ustavi, kaj s poljem stori in kam se giblje, je vnaprej določeno s pravili, ki imajo obliko pogojnika naslednje vrste: če je stroj v stanju *s* in dobi vnos *z*, proizvede naslednji iznos *z + 1* in gre v naslednje stanje *s + 1*.

Takšen niz pogojnikov je pogosto formuliran kot *strojna tabela*, sistem, ki je tako opisan, pa imenujemo avtomat: v bistvu je vsak avtomat zgolj ena vrsta Turingova stroja, kar lahko ponazorimo z enostavnim primerom stroja za coca-colo (Block, 1980b/2007, Podor, 1981). Predpostavimo, da je njegova naloga, da izvrže pločevinko coca-cole, ko prejme 1 evro. Vanj lahko vržemo kovanec za 50 centov in za 1 evro. Avtomat se lahko nahaja v dveh stanjih, *s*, in *s 2*, ki sta določeni glede na vnos, tj. kateri kovanec prejme. Možnosti z vidika obeh stanj so potem naslednje:

> (s_1): (i) če je avtomat v s_1 in prejme 50 centov, ne izvrže ničesar, zgolj preide v s_2, (ii) če pa je avtomat v s_1 in prejme 1 evro, izvrže coca-colo in ostane v s_1;
>
> (s_2): (i) če je avtomat v s_2 in prejme 50 centov, izvrže coca-colo in preide v s_1, (ii) če pa je avtomat v s_2 in prejme 1 evro, izvrže coca-colo in 50 centov ter preide v s_1.

Shematsko lahko to ponazorimo z naslednjo tabelo:

[2] Za podrobnosti *glej* Turing (1936).

vnos/stanje	s_1	s_2
50 centov	ni iznosa pojdi v s_2	izvrži coca-colo pojdi v s_1
1 evro	izvrži coca-colo ostani v s_1	izvrži coca-colo in 50 centov pojdi v s_1

Putnam bistvo takšnih avtomatov izrazi takole: »Ko je Turingov stroj opisan s pomočjo 'strojne tabele', je opisan kot nekaj, kar ima trak, tiskalnik, 'skener' (to mogoče ne rabi biti nič drugega kot točka stroja, ki je v kateremkoli danem trenutku poravnana zgolj z enim kvadratom na traku) in končni niz (A, B, C itd.) 'stanj'. /.../ Razen tega je opisan samo še z vnaprej določenimi pravili, ki opredeljujejo vrstni red, po katerem si stanja sledijo in kaj je kje natisnjeno« (Putnam, 1960: 158–159). Pomembno je, da ostane v takšnem opisu Turingovega stroja vprašanje narave njegovih stanj, tj. njegove notranje zgradbe, povsem odprto oziroma nedoločeno: »Zlasti pa 'logični opis' Turingovega stroja ne vključuje nobene specifikacije fizične narave teh 'stanj' – ali celo fizične narave celotnega stroja« (Putnam, 1960: 159). Njegova stanja so definirana v odnosu do simbolov na traku (vnosov) in v odnosu do drugih stanj, o tem, kako se med seboj razlikujejo glede fizične strukture, pa ni govora, realizira jih lahko karkoli, tako nekaj fizičnega kot tudi nefizičnega.[3] Poudariti je treba, da opisujejo strojne tabele takšne vrste operacije *deterministčnega avtomata*, kjer program za vsako stanje in niz vnosov določi naslednje stanje sistema in iznos, ki ga proizvede. Danes se večini avtorjev to ne zdi več pravi model za opredelitev človeškega duha, kar je hitro ugotovil tudi Putnam, ki je nekaj let kasneje razvil idejo, da smo ljudje *probabilistični avtomati*, kjer program za vsako stanje in niz vnosov določa zgolj verjetnost naslednjega stanja sistema in iznosa, ki ga proizvede (Putnam, 1967/1 980a).

Kakorkoli, Putnam je bil prvi, ki je značilnosti delovanja Turingovega stroja prenesel na področje našega mentalnega življenja. Tudi psihologijo zanima zgolj funkcioniranje organizma v različnih okoliščinah, ne pa to, iz česa je narejen. Pri svojem početju povsem abstrahira od fizičnih lastnosti stanj, skozi katere bitje prehaja, ko opravlja različne operacije. Glede na to, da obstajajo številni materialni in nematerialni sistemi, ki so funkcionalno izomorfni, ne moremo pričakovati redukcije funkcionalne vloge na zgolj enega izmed njih. Isti psihološki opis lahko zadovoljijo z vidika zgradbe povsem različni organizmi, tudi takšni brez nevronov, zaradi česar psihologije ne moremo skrčiti na nevrofiziologijo, saj je po svojem namenu splošnejša: nanaša se na vse funkcionalno izomorfne sisteme ne glede na razlike v njihovi zgradbi (Sesardić, 1984: 86).

[3] S tem se funkcionalizem nasploh izogne t.i. *argumentu iz večvrstne realizacije*, ki preprečuje identifikacijo mentalnih stanj s fizičnimi stanji (Putnam, 1967/1980a) in *onemogoča identitetno teorijo duha*, ki mentalna stanja enači z možganskimi (Smart, 1959/2007).

Tako med Turingovim strojem in človeško duševnostjo obstaja analogija, saj so strojeve strukturalne lastnosti (hardver) podobne človeškim nevrofiziološkim, v tem primeru gre za opis obnašanja v fizičnih pojmih (v slovarju kemije in biologije), strojeve funkcionalne lastnosti (softver) pa našim mentalnim stanjem in njihovi z zakoni določeni organizaciji v smislu vrstnega reda pojavljanja, v tem primeru gre za opis obnašanja v psihičnih pojmih (v slovarju psihologije). »Zanimiva je ugotovitev, da tako kot obstajata dva možna opisa obnašanja Turingovega stroja – inženirjev strukturalni načrt in logikova 'strojna tabela' – obstajata tudi dva možna opisa človeške psihologije. Behavioristični pristop /.../, katerega cilj je priskrbeti popoln fizikalistični opis človeškega obnašanja /.../ To je skladno z inženirjevim ali fizikovim opisom fizično realiziranega Turingovega stroja. Toda ravno tako bi bilo možno najti bolj abstrakten opis človeških mentalnih procesov v pojmih 'mentalnih stanj' (z nedoločeno fizično realizacijo, če sploh) in 'vtisov' (ki imajo na strojevem traku vlogo simbolov) – opis, ki bi specificiral zakone, ki bi nadzorovali vrstni red, po katerem si stanja sledijo /.../ Ta opis bi bil analogen 'strojni tabeli'/.../« (Putnam, 1960: 160–161). Takšen funkcionalizem je dobil ime *strojni funkcionalizem*, v svoji najbolj preprosti različici pa trdi, da je vsaka vrsta mentalnega stanja identična enemu izmed stanj v strojni tabeli. V našem primeru s coca-colo (če pustimo ob strani dejstvo, da si stroj ne more ničesar želeti) tako velja, da je s_1 želja po 50 centih, s_2 pa želja po 1 evru, res pa je, da opis človeških mentalnih stanj v pojmih Turingovega stroja zahteva kompleksnejše strojne tabele.

Metafizični funkcionalizem

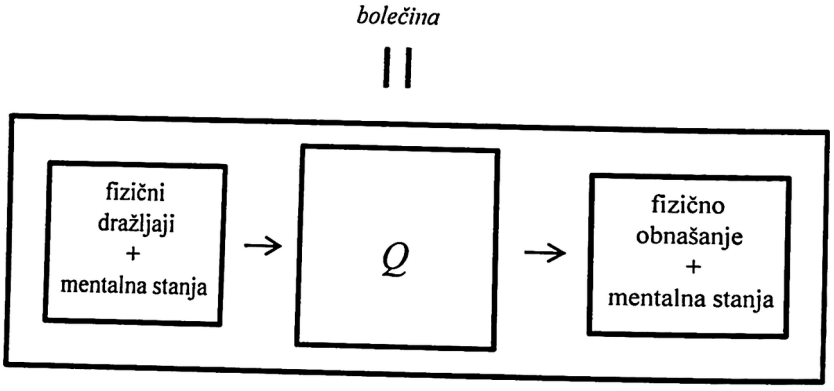
Kljub temu, da ima strojni funkcionalizem prednost pred *behaviorizmom*,[4] saj so stanja strojnih tabel opisana v pojmih relacij do vnosov in iznosov ter v pojmih relacij do trenutnega stanja strojne tabele, npr. dežnik vzamem (*iznos*), če sem prepričan, da bo deževalo (*vnos*) in če želim ostati suh (*trenutno stanje*), se je večina funkcionalistov od njega kmalu poslovila, saj je prevladalo prepričanje, da ne zagotavlja splošnega opisa tega, kaj funkcionalizem pomeni (Block, 1980b/2007). Ker smo ljudje lahko v istem trenutku v več kot zgolj enem mentalnem stanju oziroma lahko v istem trenutku opravljamo več kot zgolj eno mentalno operacijo, se zdi trditev, da smo pri proizvodnji naših duševnih stanj podobni avtomatu za coca-colo, neverjetna (Block in Fodor, 1972). Tako se je kmalu zatem uveljavil pojem funkcionalizma, ki mentalnega življenja ni definiral zgolj v pojmih determiniranih stanj iz strojne tabele, ampak tudi v pojmih vzročnih zvez med čutnimi vnosi in behaviorističnimi iznosi, ki ga lahko imenujemo *metafizični funkcionalizem*.[5] Armstrong je o strojnem funkcionalizmu denimo zapisal: »Splošna ideja je primerjava duha z računalnikom in mentalnih stanj z računalniškimi stanji v smislu njihovega označevanja v pojmih programa softvera. Ena stvar glede te ideje pa je povsem jasna: 'program, ki teče v naših

[4] Za behaviorizem in njegovo kritiko *glej* npr. Putnam (1963/1980a)
[5] Za ostale različice funkcionalizma *glej* Block (1980b/2007).

možganih, bi moral biti drugačen od kateregakoli programa, ki ga lahko napišemo zdaj in v bližnji prihodnosti« (Armstrong, 1999: 105). Kljub temu pa je ideja, da lahko notranja stanja organizma opišemo z njihovimi s pravili določenimi relacijami do vnosa, iznosa in trenutnih notranjih stanj, ostala bistvo vseh funkcionalističnih različic.

Podobno kot strojni funkcionalizem tudi metafizični funkcionalizem išče odgovor na vprašanje, kaj je narava mentalnih stanj, pri čemer pa jih ne enači zgolj z ozkim strojnim stanjem, ampak s širšo *vzročno-funkcionalno vlogo*, zaradi česar bolečina ni več strojno, ampak vzročno stanje: »Bolečina je identificirana z abstraktno vzročno lastnostjo, ki je povezana z dejanskim svetom samo preko svojih posrednih in neposrednih zvez z vnosi in iznosi« (Block, 1980b/2007: 154). To, kar naredi bolečino za bolečino, je določena vzročno-funkcionalna vloga, ki jo ima, ta pa je definirana v pojmih tipičnih vzrokov in tipičnih učinkov ter drugih stanj, ki so v vzročni interakciji in se nahajajo med njima. Ponazorimo to z naslednjim primerom. Predpostavimo, da ureznina na roki in prepričanje, da smo jo staknili, povzročita določeno stanje (fizično ali psihično) in hkrati s tem celo vrsto vzročnih interakcij med njim in drugimi stanji (imenujmo to »črna škatla« – *Q*), to pa povzroči glasen krik in željo po prenehanju. Mentalni pojem, ki ga v primeru, ko se to zgodi, uporabimo za opis tega, kar občutimo, je seveda bolečina. Fizično informacijo in mentalno stanje, ki povzročita verižno vzročno reakcijo v »črni škatli«, sta vnos, druga mentalna stanja in fizično obnašanje, ki so njena posledica, pa iznos. Tako ni bolečina (po tej preprosti teoriji) nič drugega kot lastnost, ki jo nekdo ima, ko je v stanju, ki je povzročeno z urezom z nožem in ki povzroča glasen krik in željo po prenehanju. »Tako bi lahko teorija bolečine, ki bi jo ponudil metafizični funkcionalist, bolečino označevala delno s tem, da je povzročena s poškodbo tkiva, delno s tem, da povzroči željo, da bi se je rešili, in delno s tem, da povzroči dejanje, namenjeno temu, da se poškodovani del telesa loči od tistega, za kar se misli, da je bolečino povzročalo« (Block, 1980b/2007: 149).

To lahko ponazorimo z naslednjo splošnejšo shemo:



Kljub temu, da metafizični funkcionalizem mentalna stanja identificira z

njihovimi vzročno-funkcionalnimi vlogami, se zdi, da ohranja *fizikalizem*, ki v svoji najsplošnejši obliki pravi, da je vse, kar obstaja, sestavljeno iz fizičnih delcev, saj zavrača zgolj trditev, da so mentalna stanja *identična* s fizičnimi.[6] »Funkcionalisti so lahko fizikalisti s tem, ko dopuščajo, da so vse entitete (stvari, stanja, dogodki itn.), ki obstajajo, fizične entitete, pri čemer zanikajo zgolj to, da je tisto, kar povezuje posamezne vrste stvari, fizična lastnost« (Block, 1980b/2007: 149).

Računsko-reprezentacijski funkcionalizem ali računska teorija duha[7]

Posebno vrsto funkcionalizma, »ki je postavil temelje klasični kognitivni znanosti in simbolnemu modeliranju kot empirični teoriji /…/«. (Markič, 2010: 34), pa predstavlja različica funkcionalizma, ki primerja psihološko razlago tega, kako je lahko misel proizvod fizičnih predmetov, npr. naših možganov, z računalniškim programom.[8] Njeno bistvo je priskrbeti/odkriti funkcionalno analizo mehanskih procesov, na katere lahko skrčimo naše mentalne operacije, ker pa ta običajno vsebuje algoritemski zapis, tj. zaporedje korakov računanja, ki je podoben temu, kako deluje Turingov stroj, je tudi ta vrsta funkcionalizma z njim povezana. Funkcionalizem v tem smislu je v grobem kombinacija reprezentacijske teorije, ki predstavlja svet preko *jezika misli* in računalniške analogije, po kateri so psihični procesi obravnavani kot računanje s takšnimi reprezentacijami, zaradi česar ga imenujemo *računsko-reprezentacijski funkcionalizem* ali *računska teorija duha*. Block ga rahlo ironično opiše takole: »Kakršno koli skrivnostnost naj bi naše mentalno življenje sprva imelo, jo funkcionalna analiza mentalnih procesov razčleni do te mere, da se zdi, kot da je sestavljena iz računskih operacij, ki so tako primitivne kot operacije digitalnega računalnika – procesi, ki so tako neumni, da psihološke razlage, ki se na njih sklicujejo, niti slučajno niso krožne« (Block, 1980b: 147).

V resnici je računska teorija duha sestavljena iz treh sicer povezanih delov: (i) *jezika misli*, (ii) *strukturne občutljivosti sistemskih procesov* in (iii) *reprezentacijske teorije duha. (i)* Jezik misli ali »mentalese« je leta 1975 predstavil Fodor v svoji knjigi *The Language of Thought*, gre pa za trditev, da so mentalne predstave zgrajene na isti način kot jezik, kar se včasih opisuje s frazo »misli so stavki v glavi«. Jezikovna struktura, ki je za opis misli primerna, pa vsebuje dva elementa, *kombinatorno sintakso* in *kompozicionalno semantiko* (Fodor in Pylyshyn, 1988). Za njuno ponazoritev vzemimo jezik stavčne logike, ki s simboli in vezniki tipično predstavlja stavke naravnega jezika in predpostavimo, da *p* označuje stavek »Peter je suh«, *s* stavek »Sara je zabavna«, *m* pa stavek »Marko je debel«. Vsak izmed simbolov je *atomarna predstava*

[6] Za argumente v prid trditvi, da iz resničnosti funkcionalizma sledi neresničnost fizikalizma *glej* Block (1980c), Block in Fodor (1972/1980), Putnam (1966/1975).

[7] Uporablja se tudi izraz *računski funkcionalizem*.

[8] Trditev, da so človeški možgani zelo podobni računalnikom in da se lahko z njihovim preučevanjem dokopljemo do pomembnih ugotovitev glede našega duha, sicer imenujemo *računalniška metafora*. (Miščević in Smokrović, 2001: 9)

določenega naravnega stavka, če pa jih združimo v stavek »Peter je suh« in »Sara je pametna« ali »Marko je debel« in to označimo kot (p a s) v *m*, dobimo njegovo *sestavljeno predstavo*. V tem primeru lahko rečemo, da jezik stavčne logike (in katerikoli drug jezik) uporablja kombinatorno semantiko, saj za označevanje uporablja atomarne in sestavljene predstave, kar je nujni in zadostni pogoj, ki ga mora jezik izpolniti, da ima to lastnost.

Razen tega imajo predstave v jeziku stavčne logike tudi pomen (vsebino), tj. lahko so resnične ali neresnične, kakšne so, pa je odvisno od resničnostne vrednosti sintaktičnih delov, njihove razvrstitve in veznikov med njimi. Če kateregakoli izmed naštetih elementov spremenimo, se spremeni tudi resničnostna vrednost celotne predstave/izraza, zaradi česar pravimo, da vsebuje jezik stavčne logike (in katerikoli drug jezik) tudi kompozicionalno semantiko, »ki za vsako možno kombinacijo določa, kako pomeni sintaktičnih delov sestavljenega izraza prispevajo k pomenu celote« (Markič, 2010: 36).

(ii) Strukturno občutljivost sistemskih procesov, ki je drugi del računske teorije duha, Fodor in Pylyshyn (1988) opredelita kot zmožnost sistema, da spremembe mentalnih stanja definira z načeli, ki so nad strukturo mentalnih predstav. Tega pa je sposoben zgolj v primeru, ko uporablja računske procese, ki niso občutljivi na vsebino, ampak samo na sintakso: modus ponens ($p \rightarrow q$, p, $\therefore q$) je npr. takšna računska operacija, pri kateri se možnost izpeljave iz predstave p -> q skupaj s predstavo p do predstave q, ne nanaša niti na pomen obeh propozicij niti na pomen veznika. »Z vidika formalnega procesa je pri uporabi *modusa ponensa* pomembno le to, da ima p v drugi premisi isto obliko kot p na levi strani prve premise in da ima q v sklepu isto obliko kot q na desni strani prve premise« (Markič, 2010: 37). To sistemu daje prostor, da spremembe mentalnih stanj npr. definira v pojmih vzročno-funkcionalnih vlog, ki so implementirane z njegovo sintakso ter tako realizira mentalnost/inteligentnost ne glede na svojo naravo.

(iii) Reprezentacijska teorija duha pa je sestavljena iz dveh predpostavk. Prva trdi, da lahko *intencionalna stanja*, tj. stanja, ki se na nekaj nanašajo (predmet ali situacijo) ali so na nekaj usmerjena, kar imenujemo njihova *vsebina*, zaradi česar so stanja *»o* nečem«, opredelimo s t.i. *propozicionalnimi naravnanostmi*, izrazom, ki se nanaša na povedi, v katerih v glavnem stavku tipično nastopajo glagoli verjeti, želeti ali upati, njim pa sledi predmetni odvisnik z veznikom ‘da’, npr. »Želim, da bi nehalo deževati«. Tako so propozicionalne naravnanosti sestavljene iz relacije med subjektom, tj. tistim, ki si »želi« (ima to naravnanost) in propozicijo (»da bi nehalo deževati«), ki je identična s vsebino našega mentalnega stanja. Ali: »Za katerikoli organizem *O* in katerokoli naravnanost do propozicije *P* obstajata (‘računska’/‘funkcionalna’) relacija *R* in mentalna reprezentacija *MP*, ki sta takšni, da *MP* pomeni *P* in da *O* ima *A*, če in samo če, je *R* med *0* in *MP«* (Podor, 1987:17).

Poleg te trditve, ki se nanaša na *naravo propozicionalnih naravnanosti*, pa reprezentacijska teorija duha vsebuje še eno, ki se dotika *narave mentalnih*

procesov, ki so v njej opisani kot vzročna zaporedja primerkov mentalnih predstav. Zaradi tega je sosledje misli v bistvu vzročno zaporedje primerkov mentalnih predstav, primerki mentalnih predstav pa izražajo propozicije, ki so objekti misli. Npr. »Misliti 'Deževalo bo. Torej grem noter.' je uprimeriti mentalno predstavo, ki pomeni *Grem noter*, povzročeno na določen način z uprimerjanjem mentalne predstave, ki pomeni *Deževalo bo«* (Podor, 1987: 17).

Turing je v svojem članku *Stroji, ki računajo in inteligenca* iz leta 1950 pompozno trdil, da v primeru, ko se »verbalno« obnašanje stroja, ki poteka preko ekrana in tipkovnice, ne razlikuje od verbalnega obnašanja človeka, ko odgovarjata na ista vprašanja, stroj, ki je tega sposoben, *misli*.[9] No, računska teorija duha je v resnici obrnjena Turingova trditev, da je, če mu to uspe, računalnik duh, pravi namreč, da je mišljenje računski proces oziroma da je duh računalnik. Kakorkoli, njena razlaga tega, kako lahko možgani proizvedejo misel, ki črpa svoj navdih iz stavčne logike (teorije dokaza) in Turingove ideje, da so formalni postopki lahko mehanizirani, je še danes osrednja tema kognitivne znanosti in temelj umetne inteligence, ki pa ima tako svoje zagovornike kot tudi nasprotnike.

Ugovori funkcionalizmu

Vrnimo se k najbolj splošni trditvi/obliki funkcionalizma, ki pravi, da so mentalna stanja sestavljena iz dispozicije delovati na določen način in drugih mentalnih stanj, kar je posledica določenih čutnih in mentalnih vnosov (Block, 1980c: 268), ter si oglejmo nekaj težav, ki ga spremljajo. Ugovore, ki izhajajo iz njegove predpostavke, da ima mentalno življenje vse, kar izpolni ali realizira ustrezno vzročno-funkcionalno vlogo, lahko razvrstimo v dve skupini: (i) v prvo spadajo tisti, ki pravijo, da iz tega sledi podobnost sistemov, za katere niti v sanjah ne bi rekli, da imajo skupnega karkoli bistvenega za njihovo naravo, (ii) v drugo pa spadajo tisti, ki trdijo, da iz tega sledi priznavanje zavesti sistemom, katere niti slučajno ne dojemamo kot mentalne, kar kaže na to, da se pri funkcionalizaciji človeške duševnosti izgubi ena od njenih bistvenih značilnosti.

(i)

Ugovor iz funkcionalne izomorfnosti

Funkcionalizem predpostavlja, da sta dva stroja izomorfna takrat, kadar izpolnita isti funkcionalni opis ne glede na to, v kolikšni meri se razlikujeta glede svojih strukturalnih lastnosti. To pa vodi do govorjenja o funkcionalni izomorfnosti sistemov, ki si z vidika funkcioniranja še zdaleč niso podobni, saj je treba pri njihovem definiranju upoštevati tudi opis na nižjih ravneh abstrakcije (Kalke, 1969). Vzemimo mačko in mišolovko. Če ju opišemo dovolj abstraktno, obe realizirata isto funkcionalno shemo, in sicer *P* (pojav miši – vnos), *x* (notranje stanje – »črna škatla«), *Q* (lovljenje miši – iznos). Če pa ju opišemo bolj

[9] Za razpravo o tem *glej* Bregant (2011, 2012).

podrobno, tj. osredotočimo se na to, kaj se dogaja znotraj mačke in kaj znotraj mišolovke, takoj ugotovimo, da sta povsem različni. Sklep je, da sta dva sistema lahko funkcionalno izomorfna zgolj na visoki ravni abstrakcije, pri podrobnejšem opisu njune funkcionalne organizacije, pa se izkažeta za bistveno drugačna: očitno je, da je funkcionalna shema pri živih organizmih bolj zapletena, kot pa pri neživih artefaktih.

Poanta Kalkejevega argumenta zoper funkcionalizem je tako naslednja: bolj, ko je nek stroj abstraktno opisan, tj. več kot je stvari, ki se ne upoštevajo, večja je verjetnost, da bomo našli stroj, ki mu je funkcionalno podoben; in manj, ko je nek stroj abstraktno opisan, tj. manj kot je stvari, ki jih šteje za nepomembne, manjša je verjetnost, da obstaja stroj, ki mu je funkcionalno izomorfen. In govoriti, da je psihologija danes neodvisna znanost, ker pojasnjuje naše obnašanje s pomočjo prikazanih funkcionalnih shem, že dolgo ni več resnično, zadnje čase se vse bolj opira na strukturalne lastnosti sistemov, tj. na nevrofiziološko organizacijo naših možganov (Churchland, 2011), ki pa jo funkcionalizem povsem zanemarja.

Ugovor iz splošnih in abstraktnih sistemov

Funkcionalizem trdi, da so mentalna stanja definirana s svojim kontekstom in da spadajo v isto naravno vrsto izključno na osnovi mesta, ki ga imajo v funkcionalni shemi, ne glede na razlike v svojih fizičnih značilnostih. To pa pripelje do tega, da klasificiramo kot pojave iste vrste tudi pojave, ki nimajo med sabo skupnega ničesar razen, da so primerki iste funkcionalne sheme. Vzemimo funkcionalni model (Sesardić, 1984), ki ga lahko uprimerita tako mehanski kot tudi elektronski sistem in ki opisuje oscilacije, v enem primeru mehanske, v drugem pa elektronske: $A\ (d^2x/dt^2) + B\ (dx/dt) + Cx = D$. Skladno s kontekstom simbolom v tej enačbi pripišemo različne pomene, oba sistema pa se s funkcionalnega vidika kljub temu obnašata izomorfno, zaradi česar se lahko mehanski in elektronski parametri prikažejo z istim funkcionalnim modelom. Predpostavimo, da imajo simboli v teh dveh sistemih naslednje pomene:

	A	B	C	D	x
1. sistem (mehanski):	masa	koeficient trenja	togost	sila	razdalja
2. sistem (elektronski):	električni tok	električni upor	kapacitivnost $^{-1}$	električni potencial	naboj

Problem je, da bi glede na funkcionalistični princip, ki razvršča stanja v vrste zgolj na osnovi položaja, ki ga v abstraktnih shemah zasedajo, v prikazani funkcionalni shemi v isto vrsto spadala npr. razdalja in naboj, česar pa noben resen teoretik ne bi zastopal, saj si nista podobna v nobeni bistveni podrobnosti, na osnovi katere bi ju lahko razvrstili v isto vrsto. Da bi ju tja uvrstili, bi o obeh morale obstajate zanimive posplošitve, ampak o razdalji in naboju jih ni, edino, kar imata skupnega, je, da uprimerjata isti funkcionalni model, da pa je to že dovolj za to, da sta tudi stanji iste naravne vrste, je, milo rečeno, prehuda

poenostavitev. To se zgodi zato, pravi Gendron (1971), ker funkcionalizem svoj model predstavlja kot *splošen*, kot sta npr. Newtonova mehanika ali Maxwellova elektromagnetika, ki opredeljujeta obnašanje posameznih primerkov in vnaprej določata, katere vrste so, v resnici pa je *abstrakten*, kot je npr. McCullochov in Pittsov logični model nevrona, ki pa ne pojasnjujeta obnašanja posameznih realizatorjev in ne določata, v katero vrsto spadajo.

Z vidika psihologije kot posebne znanosti definiranje človeške duševnosti v pojmih vzročno-funkcionalnih vlog tako ni zanimivo, saj ne vodi do nobenih posplošitev o pojavih, kar bi nam zagotavljalo klasifikacije v vrste in prispevek k razumevanju sveta, ampak gre zgolj za računanje s simboli, kar pa je domena matematike, ki ji tako jemlje njeno neodvisnost.

(ii)

Ugovor iz zombijev[10]

Obstaja več definicij zavesti, najbolj zanimiva in za funkcionalizem najbolj problematična pa je tista, ki pravi, da ima bitje zavest, ko njegove čutne interakcije z okoljem vsebujejo tudi kvalitativna subjektivna izkustva tega, kar zaznava, tj. *takšnosti* ali *kvalije*. Gre za tipične fenomenalne značilnosti občutkov, ki se običajno opisujejo kot načini čutenja, pojasnjujejo pa s pomočjo stavkov ‘kako je biti’, npr. kako je videti rdeč paradižnik ali kako je čutiti bolečino v kolenu. V zvezi z njimi je problematično predvsem to, kako naj bodo opisane, da bodo ustrezno predstavljene drugim, saj se predpostavlja, da to, ‘kako je biti’, ve samo tisti, ki to izkusi.

Kaj imajo torej vsi primerki bolečine skupnega, da so vsi tudi primerki bolečine? Funkcionalistov odgovor na to vprašanje je jasen, in sicer tipične vzroke, tj. fizične dražljaje ali mentalna stanja, npr. poškodbo ali travmo, in tipične učinke, npr. kričanje in željo po prenehanju. Vendar se zdi, da mentalni pojavi, za katere so značilne takšnosti, *fenomenalni pojavi*, vsebujejo še nekaj več od golega opisa v pojmih njihovih vzročno-funkcionalnih vlog. Kar imajo vsi primerki bolečine skupnega, da so vsi tudi primerki bolečine, je, da vsi bolijo. Kaj to pomeni? To pomeni, da takšnosti ne moremo opisati v funkcionalnih pojmih, pravi Jackson (1995). Pri tem se sklicuje na primer z Marijo, ki ve o izkustvu rdečega vse, kar se da o njem vedeti na osnovi funkcionalnih informacij, pa o izkustvu rdečega kljub temu nima popolnega znanja. Manjka ji takšnost rdečega, ki pa jo dobi šele takrat, ko sama vidi nekaj rdečega. Podobno velja tudi za bolečino. Opis bolečine kot vzburjenosti živčnih vlaken, ki povzroča kričanje in željo po prenehanju, nam o tem, kako je čutiti bolečino, ne pove ničesar. O izkustvu bolečine tako nimamo popolnega znanja, saj nam manjka njena takšnost, ki pa jo dobimo šele takrat, ko nas nekdo npr. udari (Bregant, 2004: 218).

Funkcionalistove muke pa lahko še zaostrimo z naslednjim miselnim

[10] Imenujemo ga lahko tudi *problem dvojnikov*.

eksperimentov, ki ima svoje zametke že pri Descartesu (1637/2007), in njegovem zanikanju, da bi lahko bili avtomati kadarkoli zavestni, razvil pa se je v 70. letih 20. stol. kot orožje za boj proti fizikalizmu nasploh.[11] Predstavljajte si bitje, ki je molekulski duplikat vas samih, tj. s fizičnega vidika je z vami povsem identično, nima pa nobene zavesti oziroma občutkov. Gre za t.i. filozofskega zombija, ki se razlikujejo od filmskega zombija: filmski je neroden, nad svojim obnašanjem nima nadzora, prehranjuje se z ljudmi in ni nujno brez takšnosti, filozofski pa se od nas razlikuje zgolj v tem, da nima fenomenalnega izkustva, kar pomeni, da se tudi obnaša na povsem enak način kot njegov fizični izvirnik. Npr. ko sediva jaz in moj zombi na plaži, se obnašava povsem enako, oba pijeva, kot da bi bila žejna, oba naročiva vino, kot da bi marala njegov okus, oba uživava v razgledu, kot da bi imela vidno izkustvo itn. Razlika je zgolj v tem, da jaz resnično vse to izkušam, moja fizična kopija pa ne, čeprav se v naju, ker sva postavljena v isto okolje, odvijajo enaki fiziološki procesi. Katerikoli dražljaj, ki se pojavi kot vnos, bo moja kopija obdelala na enak način kot jaz, kar pomeni, da bo tudi njeno obnašanje kot iznos enako kot moje. Skratka, vse pri njem je enako kot pri meni razen, da zanj ne obstaja to ‘kako je biti’ v nekem stanju.

Zamisljivost zombijev je tako za funkcionalizem problematična, ker zombiji, tako kot funkcionalizem to zahteva, opravljajo funkcionalno vlogo, opisano v pojmih tipičnih vzrokov in učinkov, npr. ko so žejni, pijejo, ne da bi sploh imeli kakršnokoli mentalno stanje; z drugimi besedami, identifikaciji fenomenalnih pojavov z njihovimi vzročnimi vlogami se izmuzne kvalitativna značilnost, ki jo tipično imajo, zaradi česar funkcionalizem *odsotnosti takšnosti*, kot se ta problem tudi opisuje, ne more pojasniti. Seveda takšni zombiji niso dejansko ali *nomološko* možni, tj. v aktualnem svetu z znanimi naravnimi zakoni jih ni, so pa vsaj zamisljivi in tako logično ali *metafizično* možni.[12]

Ugovor iz obrnjenega spektra

Podobno težavo predstavlja za vse, ki pravijo, da je naš duh zgolj funkcionalna organizacija naših možganov, še en miselni eksperiment, katerega začetki segajo k Locku (1689/2012), orodje za zavračanje funkcionalizma pa je prav tako postal v 70. letih 20. stol.[13] Zamislite si Janeza, ki vidi rdeče, ko vi vidite zeleno in ki vidi zeleno, ko vi vidite rdeče, tj. ko gleda rdeč paradižnik je za njega ‘kako je biti’ tako kot za vas, ko gledate zeleno travo, in ko gleda zeleno travo je za njega ‘kako je biti’ tako kot za vas, ko gledate rdeč paradižnik. To imenujemo *obrnjeni barvni spekter* ali *obrnjena takšnost* (Block in Podor, 1972), zanj pa je značilno, da ga ima Janez od rojstva, da se dogaja v isti jezikovni skupnosti in da se Janez obnaša na enak način kot vi. Z drugimi besedami, Janez in vi sta fenomenalno različna, tj. vajini subjektivni kvalitativni izkustvi sta obrnjeni, vendar

[11] *Glej* npr. Shoemaker (1975), Block (1980c), Chalmers (1996).
[12] Vsi teoretiki se s tem ne strinjajo, za nezamisljivost zombijev *glej* npr. Harnad (1995), Tye (2006), Dennet (1991/2012).
[13] *Glej* npr. Block in Fodor (1972), Block (1980c, 1990), Shoemaker (1982).

funkcionalno enaka, tj. na vprašanje, kakšne barve je paradižnik, oba odgovorita, da rdeče, pa tudi z vrta oba vedno prineseta le zrel paradižnik. To je t.i. klasični/tradicionalni/običajni argument iz obrnjenega spektra, ki za funkcionalista predstavlja težavo, ker isti vzrok v vzročni specifikaciji vidnega izkustva povzroči isti učinek, čeprav imata Janez in vi različni takšnosti, kar kaže na to, da zavestna stanja niso del vzročno-funkcionalnih vlog oziroma, da funkcionalizem obrnjene takšnosti ne more pojasniti. Podobno kot zombiji je tudi obrnjen spekter zgolj metafizično možen, tj. mogoče si ga je zamisliti in to enostavno, ne glede na to, kako težko bi ga bilo v praksi v našem svetu izvesti, tj. da bi postal tudi nomološko možen.

Razen tega bolj lockovsko obarvanega obrnjenega spektra pa obstaja še en, ki si ga je izmislil Block (1990) in ga imenoval *obrnjena Zemlja*. Zamislite si drugo Zemljo, kjer je nebo resnično rumeno, zrel paradižnik zelen in trava resnično rdeča. Njeni prebivalci imajo tako drugačno fenomenalno izkustvo, ko pogledajo v nebo, jedo paradižnik in kosijo travo, kot njihovi sorodniki na normalni Zemlji. Razen tega so tudi vsebine vzročno-funkcionalnih vlog prebivalcev te druge Zemlje obrnjene: mislijo, da je nebo rumeno, paradižnik zelen in trava rdeča. Vse ostalo je na obrnjeni Zemlji prav takšno kot na normalni. Sedaj pa predpostavite, da vas neko noč znanstveniki ugrabijo, opremijo z lečami, ki izničijo ali nevtralizirajo učinek obrnjenega spektra in z normalne Zemlje pošljejo na obrnjeno. Zaradi leč, ki jih nosite, bodo vaše takšnosti enake kot doma, nebo boste videli modro, zrel paradižnik rdeč in travo zeleno, z vidika ‘kako je biti’ za vas ne bo nobene spremembe, v bistvu boste mislili, da ste še vedno doma. Toda ko bo minilo dovolj časa in ko boste posvojili njihovo poimenovaje ter se zlili z njihovo kulturo in fizičnim okoljem, se bodo vsebine vaših vzročno-funkcionalnih vlog začele ujemati z njihovimi, tj. tudi vi boste mislili, da je nebo rumeno, paradižnik zelen in trava rdeča. Skratka, ko se boste dovolj integrirali, se boste obnašali enako kot oni, s čimer se bo vaša funkcionalna vloga v primerjavi z vami, kot ste bili prej, spremenila, čeprav bodo subjektivna izkustva (še vedno nosite leče) ostala ista, zaradi česar gre za, recimo, *nasprotni obrnjeni spekter*. Vendar tudi ta enako kot prejšnji pokaže, da »’kako je biti’ ne more temeljiti na funkcionalnih značilnosti vaših mentalnih stanj« (Tye, 1995: 29).

Kakorkoli, funkcionalistična ideja, da se danes pojem mentalnih stanj nanaša na formalne vzorce, ne pa na njihove materialne elemente, je postala osnova za govorjenje in ustvarjanje inteligentnih strojev, o obstoju katerih pa si vsi teoretiki še zdaleč niso edini, kar si bomo ogledali v nadaljevanju.

2 »Kitajska soba«

V 90. letih 20. stol. je Edward Fredkin, ki je bil takrat direktor laboratorija za umetno inteligenco (UI) na MIT-ju, menil, da so se v zgodovini zgodili trije pomembni dogodki. Prvi je bil nastanek vesolja, drugi nastanek življenja. Tretji, za katerega je trdil, da je enako pomemben, pa je bil nastanek umetne inteligence

(Copeland, 1993: 1). S pojmom ‘inteligenca’ mislimo na izvajanje umskih/spoznavnih operacij, s pojmom ‘umetna’ pa na to, da jih ne izvaja človek, ampak stroj/sistem (tj. nekaj, kar ni rezultat narave, ampak človekovega ali strojevega dela). Odgovora na to, kaj ga je pripeljalo do tega, da je nastanek umetne inteligence postavil ob bok nastanku vesolja in življenja, ni težko najti, saj gre za obliko življenja, ki je povsem drugačna od običajne in s potencialom, ki si ga težko predstavljamo.

Umetno inteligenco lahko najenostavneje definiramo kot znanost o izdelavi mislečih strojev, ki so sposobni reproducirati različne zmožnosti našega duha. Način, kako jim to uspe, ni pomemben, zaradi česar je za to početje zelo prikladna osnovna funkcionalistična ideja o tem, da so mentalna stanja stanja Turingovega stroja. Ker si ni težko zamisliti, da se bodo stroji razvijali, s tem postajali pametnejši in nekega dne samostojno načrtovali ter izdelovali druge stroje, je povsem na mestu vprašanje, kaj se bo zgodilo s svetom, v katerem ima prvenstvo človek kot najbolj inteligentno bitje? Strah, da bo to prevlado, če že ne bo prišlo do uničenja sveta, kot ga poznamo iz filmov, izgubil, izhaja iz tega, da si je težko zamisliti stroj, ki je pametnejši od najpametnejšega človeka, pa kljub temu dela zgolj tisto, kar od njega zahtevamo. Lažje si je predstavljati situacijo, kjer se takšen stroj poniža do te mere, da se z nami igra, pogovarja in druži, podobno, kot se mi obnašamo do naših domačih živali, ki nam krajšajo čas.

Iz tega pa izhaja kopica filozofskih vprašanj, izmed katerih so najbolj pomenljiva naslednja: (i) je misel biološki pojav in tako umetnim strojem nedosegljiva, podobno, kot to velja za fotosintezo (produkcija kisika) ali laktacijo (izločanje mleka), ali pa sta mišljenje in zaznavanje bolj podobna letenju, nečemu, kar lahko delajo tako biološki organizmi kot tudi umetni sistemi; (ii) ali bi lahko postal računalnik kadarkoli pametnejši od ljudi, ki so ga programirali; (iii) ali je lahko stroj v svojih odločitvah svoboden, tj. ali so njegova dejanja posledica njegove lastne volje; (iv) ali bi računalnik lahko imel zavest, tj. ali bi lahko imel subjektivna kvalitativna izkustva ‘kako je biti’; (v) ali nismo tudi mi zgolj neke vrste stroji (Copeland, 1993: 1–2).

V nadaljevanju se bomo osredotočili zgolj na prvo, pri čemer si bomo ogledali slavni Searlov argument iz »kitajske sobe«, ki so ga leta 1980 ob izidu njegovega članka *Duhovi, možgani in programi*, nasprotniki inteligentnih strojev označevali kot »popolno zmagoslavje«, njihovi zagovorniki pa kot nekaj »preprosto smešnega«. Pri tem ni odveč omeniti, da so od rojstva umetne inteligence,[14] leta 1956 na slavni darthmouthski konferenci, mislečim strojem vedno bili bolj naklonjeni računalničarji (programerji), kot pa filozofi. Turing je npr. že leta 1950 v svojem članku *Stroji, ki računajo in umetna inteligenca* menil, da bodo danes inteligentni stroji nekaj običajnega.[15] Ali lahko stroj misli? To vprašanje se ni prvič pojavilo šele v 20. stol., ko so bili računalniki zreli za masovno

[14] Prvič naj bi se ideja o umetni inteligenci sicer pojavila v McCulloch in Pitts (1943).
[15] Za to *glej* Bregant (2012).

proizvodnjo, ampak že daljnega leta 1637, ko je Descartes trdil, da noben stroj nikoli ne bo sposoben simbolne manipulacije, ki bi mu omogočala tvorjenje smiselnih stavkov, na osnovi česar bi mu lahko pripisali mišljenje (Bregant, 2010:55). Šele po stoletjih čakanja s prihodom prvih elektronskih računalnikov v 50. letih 20. stol. je ideja o mislečem stroju dobila nov zagon. Od takrat je sicer doživela vzpone in padce, hiter napredek informacijskih tehnologij pa jo še vedno, mogoče celo bolj kot kdajkoli prej, dela živo.

Od začetka obstajata v splošnem dve glavni pojmovanji/razumevanji umetne inteligence. Po prvem je njen cilj zgraditi stroj, ki ima vse običajne (človeške) psihične lastnosti, kot so mišljenje, razumevanje in sklepanje, še več, celo zavest, čustva in občutke, skratka, *ustrezno programiran računalnik bi resnično imel duh* v *istem smislu kot ga imate vi ali jaz*, kar imenujemo *močna umetna inteligenca*. Po drugem, manj skrajnem, pa je njen cilj zgraditi stroj, ki lahko opravi enake stvari kot človek, s to razliko, da človeku, ko jih naredi, pripišemo psihične lastnosti, stroju pa ne, skratka, *programirani računalniki zgolj simulirajo mišljenje*,[16] kar imenujemo *šibka umetna inteligenca*. Searlov argument iz »kitajske sobe« je uperjen proti močni umetni inteligenci, katere osrednjo trditev sam opredeli takole: »[Tukaj] računalnik ni le orodje za raziskavo duha; tu je ustrezno programiran računalnik *sam* duh, v tem smislu, da za računalnike, ki jih ustrezno programiramo, lahko rečemo, da dejansko razumejo in imajo še druga spoznavna stanja« (Searle, 1980/1990:361).

Vprašanje je, ali sploh obstaja kdo, ki je to pripravljen zagovarjati? Dvom v to ni na mestu, saj se zdi, da to sprejemajo vsi ustanovitelji umetne inteligence vključno s Turingom,[17] kasneje pa tudi npr. Newell (1980), Dennett (1985), Kurzweil (1998) itd. Zdi se, da »je *kdorkoli*, ki *resno* opisuje elektronski digitalni računalnik v psiholoških pojmih ali predlaga reprodukcijo psihičnih pojavov v takšnih računalnikih, zavezan močni UI« (Preston, 2002: 16). Enako naj bi veljalo tudi za Schanka in Abelsona, ki sta v 70. letih 20. stol. na Yaleu razvila program z imenom *Sam*, ki je bil sposoben brati med vrsticami, tj. bil je sposoben v zgodbi zapolniti luknje oziroma prepoznati manjkajoče podrobnosti. To se zdi nujno za *razumevanje* kateregakoli teksta in predstavlja nekaj, s čimer ljudje, npr. v novinarskih člankih, kjer se to običajno dogaja zaradi varčevanja s prostorom, nimamo težav. Na vprašanje znamo odgovoriti pravilno tudi takrat, ko odgovor v zgodbi ni izrecno naveden. Enako sposobnost je imel tudi Sam, ki je imel vgrajene različne scenarije/besedila za različne situacije, t.i. *skripte*. Z njimi je razširil pomanjkljivo zgodbo in potem brez težav zapolnil manjkajoče podrobnosti v njej, kar mu je omogočalo, da je pravilno odgovoril na vsa vprašanja v zvezi z njo in tako prestal preizkus razumevanja. Z drugimi besedami, stroj, na katerem bi tekel takšen program, bi *razumel* zgodbo na enak

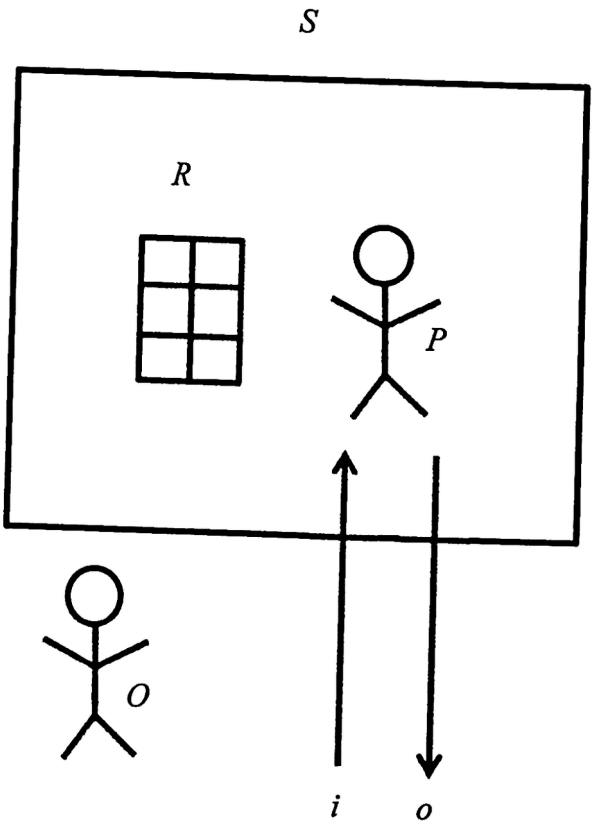
[16] Ta simulacija je sicer koristna za spoznavo tega, kako duh deluje, vendar iz tega še ne sledi, da je tudi sam računalnik duh.

[17] Turing sicer ne velja za očeta umetne inteligence, ampak za začetnika filozofije umetne inteligence (Copeland, 1993).

način, kot jo razumemo sami, zaradi česar bi morali razumevanje pripisati obema ali nikomur.[18] Searlov argument iz »kitajske sobe« naj bi pokazal, da je takšno prepričanje zmotno.

V svoji izvirni različici poteka njegov miselni eksperiment takole. Zamislite si Searla, ki ne obvlada nobenega drugega jezika razen angleščine, zaprtega v sobi z odprtino, skozi katero dobi tri svežnje kitajskega besedila: prvi predstavlja t.i. *skripte*, tj. scenarije za različne situacije, drugi predstavlja *zgodbo*, vključuje pa še pravila v angleščini, ki Searlu omogočajo, da poveže formalne simbole iz skriptov s temi iz zgodbe, pri čemer 'formalen' pomeni, da jih lahko spozna izključno po njihovih oblikah, tretji pa predstavlja *vprašanja*, na katera mora odgovoriti in prav tako vključuje pravila v angleščini, ki pa mu tukaj omogočajo, da poveže formalne simbole iz skriptov in zgodbe s temi iz vprašanj. Torej, Searle ima v sobi na voljo tri svežnje kitajskih pismenk, tj. skripte, zgodbo in vprašanja ter pravila za njihovo povezovanje. Ljudje zunaj sobe sedaj od njega pričakujejo, da jim skozi odprtino posreduje odgovore na vprašanja, ki jih je dobil, zamislite pa si, da se Searle v rokovanju s simboli sčasoma tako izuri, da se ob odličnih programih, ki jih pri tem uporablja, njegovi odgovori v kitajščini na dana vprašanja za nekoga, ki stoji zunaj sobe, ne razlikujejo od odgovorov rojenih Kitajcev. Slikovno lahko to prikažemo takole (Bringsjord in Noel, 2002:148): S

[18] Vprašanje je, ali so res filozofi poklicani za to, da o tem sodijo? Res je sicer, da nimajo vpogleda v tehnične podrobnosti računalnikov in ne vedo, česa vsega so sposobni in kdaj, kljub temu pa lahko ocenjujejo, kateri je najboljši način za opisovanje psihičnih sposobnosti v ne-tehničnih pojmih in z miselnimi eksperimenti zavračajo tiste, ki se izkažejo za neprimerne (Preston, 2002: 16).



Slika 1: *S* označuje sistem, *R* pravila, *P* osebo, tj. Searla, *O* opazovalce, tj. rojene Kitajce, *i* vnose in *o* iznose.

Vprašanje je, ali Searle, ki deluje kot računalniški program, tj. izvaja operacije na formalno opredeljenih elementih oziroma proizvaja kitajske odgovore tako, da uporablja neraztolmačene formalne simbole, razume kitajsko? Odgovor je slej ko prej *ne*. Čeprav se njegovi vnosi in iznosi ne razlikujejo od tistih rojenega Kitajca in čeprav lahko izvede katerikoli naložen program, je in bo med njima vedno velika razlika: Searle ne razume, rojeni Kitajec pa razume. »Iz istega razloga tudi Schankov računalnik ne razume ničesar iz nobene zgodbe, pa najsi bo ta v angleščini, kitajščini ali kateremkoli drugem jeziku, saj je v kitajskem primeru računalnik jaz, v primerih pa, ko ni jaz, računalnik nima ničesar več, kot imam jaz v primeru, ko ne razumem ničesar« (Searle, 1980/1990: 363–364). (To, mimogrede, predstavlja ugovor funkcionalizmu, saj imamo v tem primeru dva sistema, Searla in rojenega Kitajca, ki sta funkcionalno enaka, pa kljub vsemu psihološko različna, tj. eden razume, drugi pa ne, kar je s funkcionalizmom, ki pravi, da iz funkcionalnega izomorfizma sledi tudi psihološki, nezdružljivo: opravljati prave funkcije, vsaj kadar so določene zgolj z vnosi in iznosi, še ni dovolj za posedovanje mentalnih stanj).

Skratka, Searlov miselni eksperiment pokaže, da obvladovanje slovnice/sintakse,

tj. poznavanje niza pravil za simbolno manipulacijo, še ne pomeni obvladovanja pomena/semantike, tj. poznavanje tega, kar simboli dejansko pomenijo. To pa je v nasprotju z močno umetno inteligenco, ki pravi, da ustrezno programirani računalniki tudi razumejo, zaradi česar jih lahko prištevamo med inteligentne stroje, ki so sposobni izvajati tudi druge kognitivne operacije in imajo poleg intencionalnih stanj tudi fenomenalna: »/.../ *v* dobesednem pomenu programirani računalnik razume toliko, kot razumeta avto in seštevalni stroj, to pa je natanko nič. Računalnikovo razumevanje ni niti delno niti nepopolno (kot je moje razumevanje nemščine); je povsem nično« (Searle, 1980/1990: 366).

3 Ugovori argumentu iz »kitajske sobe«

Ko želimo govoriti o »kitajski sobi« kot o argumentu, takoj naletimo na dve težavi, prva je, da Searle svojega miselnega eksperimenta nikoli ne zapiše v obliki argumenta, druga pa, da zaradi tega ni povsem jasno, proti čemu je uperjen. Če bi »kitajsko sobo« skušali predstaviti kot argument, bi ta verjetno bil videti takole (Preston, 2002: 26):

1. Oseba v sobi ima dostop zgolj do formalnih, sintaktičnih značilnosti simbolov, ki so ji posredovani.

2. Da bi razumela kitajske iznose, bi morala imeti dostop do semantičnih značilnosti njihovih vnosov.

3. Noben niz formalnih ali sintaktičnih načel ni zadosten za razumevanje.

/. 4. Oseba v sobi ne razume kitajskih vnosov.[19]

Problem je, da takšen sklep, ker se nanaša zgolj na osebo v sobi, ne zavrača močne umetne inteligence, saj ta ne trdi, da razumejo vsi deli ustrezno programiranega računalnika, ampak takšen stroj kot celota, zaradi česar zagreši *zmoto sestave*, kar bomo podrobneje predstavili v nadaljevanju.

Kakorkoli, glavna tarča argumenta iz »kitajske sobe« je nedvomno močna umetna inteligenca,[20] vprašanje je samo, v kakšnem smislu. Searle namreč ne zavrača splošne ideje, da lahko stroji mislijo, saj je prepričan, da smo tudi ljudje zgolj neke vrste stroji: »Ali lahko stroj misli? Odgovor je očitno – da. Mi sami smo natanko takšni stroji« (Searle, 1980/1990: 375). Razen tega meni, da lahko mislijo celo umetno narejeni stroji, tj. takšni, ki jih je naredila človeška roka, če bi bili z vidika živčne zgradbe dovolj podobni nam. Sprejema celo predstavo, da lahko mislijo tudi računalniki, saj smo tudi mi samo neke vrste računalniki: »/…/

[19] Kasneje Searle (npr. 1997) svoj argument iz »kitajske sobe«, t.i. »brutalno preprost argument«, predstavlja takole:

1. Programi so povsem formalni (sintaktični).
2. Duh (vsaj človeški) ima semantiko, tj. mentalno vsebino.
3. Sintaksa sama po sebi ni niti ista kot semantika niti ni zanjo zadostna.

∴ 4. Programi sami po sebi niso niti bistveni niti zadostni za duh.

[20] Razen tega je uperjen še proti Turingovemu testu (*glej* npr. Harnad, 2002, Coulter in Sharrock, 2002) in kot smo že na kratko omenili, tudi funkcionalizmu (*glej* npr. Rey, 2002, Block, 2002).

ali lahko misli digitalni računalnik? Če z ‘digitalnim računalnikom’ mislimo karkoli, kar lahko opišemo kot izvedbo računalniškega programa, potem je odgovor seveda spet pritrdilen, saj smo tudi mi sami ‘realizacija’ neke vrste računalniškega programa in mi že lahko mislimo« (Searle, 1980/1990: 375).

Tisto, čemur Searle nasprotuje, je, da ima lahko nekaj psihične lastnosti *izključno* zaradi tega, ker je programirano na ustrezen način oziroma, da je biti sposoben izpeljati ali izvesti primeren program, zadostni pogoj/razlog za pripisovanje inteligence. »Toda ali lahko nekaj misli, razumeva in tako naprej *že samo* zaradi tega, ker je računalnik s pravo vrsto programa? Ali je lahko izvedba programa, seveda če je ta pravi, že sama po sebi zadosten pogoj za razumevanje? To pa, mislim, je tisto vprašanje, ki si ga je treba zastaviti, čeprav ga običajno mešajo s prejšnjimi vprašanji, odgovor nanj pa je – ne« (Searle, 1980/1990: 375). Razlog za njegovo nestrinjanje s tem smo že omenili, gre pa za to, da formalni program nima nobene intencionalnosti, je vsebinsko prazen in brez smisla, zaradi česar sistemu ne more zagotoviti razumevanja, če ga ta ni imel že prej: »Intencionalnost, kakršno naj bi imeli računalniki, obstaja samo v glavah tistih, ki jih programirajo, in tistih, ki jih uporabljajo, tistih, ki vnašajo input in tistih, ki tolmačijo output« (Searle, 1980/1990:376).

Sistemski ugovor

Čeprav Searle »kitajske sobe« kot argumenta ni zapisal v prej predstavljeni obliki, pa nekateri avtorji menijo, da v miselnem eksperimentu oseba v sobi ni analogna računalniku (Copeland, 1993, Block, 2002, Haugeland, 2002), kot Searle običajno trdi, ampak zgolj *procesorju*,[21] tj. osrednjemu delu računalnika, ki obdeluje (procesira) podatke ter nadzoruje in upravlja ostale enote. Če to drži, argument iz »kitajske sobe« zgreši svojo tarčo, saj se trditve močne umetne inteligence nanašajo na računalnik kot celoto, ne pa samo na enega izmed njegovih delov, če pa vztrajamo, da se nanaša na celoten sistem, zagrešimo, kot smo že omenili, zmoto sestave, kar se lepo vidi iz njegovega standardnega zapisa:

1. Nobena simbolna manipulacija, ki jo oseba v sobi opravi, ji ne bo pomagala, da bi razumela kitajski vnos.

/. 2. Nobena simbolna manipulacija, ki jo oseba v sobi opravi, ne bo širšemu sistemu, katerega del je, pomagala, da bi razumel kitajski vnos (Copeland, 1993: 125).

Ker v sklepu nastopa pojem ‘sistem’, o katerem ni v premisah ne duha ne sluha, je argument očitno neveljaven, tj. iz premis ta sklep preprosto ne sledi. »Logični odgovor«, kot sistemski ugovor imenuje Copeland (1993), Searlu očita osredotočenje na napačnega akterja, vprašanje ni, ali razume procesor, ampak, ali razume računalnik. Sistemski ugovor Searle povzame kot »Res je sicer, da oseba,

[21] Centralno procesna enota (CPE).

ki je zaklenjena v sobi, ne razume zgodbe, vendar je res tudi to, da je sama le del celotnega sistema, ta sistem pa zgodbo razume« (Searle, 1980/1990: 366). Vprašanje je, ali zagovorniki sistemskega ugovora to tudi resnično trdijo?

Preden na to odgovorimo, si oglejmo Searlovo zavrnitev sistemskega ugovora. Vsebuje tri pripombe, vsaka izmed njih pa naj bi dokazovala, da ne drži, (i) Pr... pravi, da nima sistem za razumevanje kitajskih simbolov na voljo nič takšnega, česar ne bi imela na voljo tudi oseba v sobi, tj. drugi deli sistema – knjige, papir, svinčniki itd., ki jih pri računanju uporablja, ne prispevajo k tolmačenju formalnih simbolov nič novega, zaradi česar bi celotni sistem razumel, njegov del (procesor) pa ne. (ii) Druga trdi, da je od sistema nesmiselno pričakovati, da se bo samo zaradi konjunkcije osebe in ostalih delov »kitajske sobe« (pravil, svinčnikov, papirja itd.) iz nečesa, kar ne razume, takoj prelevil v nekaj, kar razume. »Šlo naj bi za to, da oseba sicer resda ne razume kitajščine, vendar naj bi jo nekako razumela *zveza* te osebe in kosov papirja. Le stežka si predstavljam, da se lahko ta misel komurkoli, ki ni v primežu ideologije, sploh zdi verjetna« (Searle, 1980/1990: 366). (iii) Tretja pa pravi, da vodi sprejetje sistemskega ugovora v absurd, saj bi očitno nementalni deli sistema, kot npr. ledvice, jetra in želodec, naenkrat postali mentalni: res je, da obstaja (funkcionalna) raven opisa mojih jeter, kjer te obdelujejo vnosne podatke in izvajajo iznosne operacije, vendar to še ne pomeni, da karkoli razumejo.

Searle (1993) sicer meni, da je najmočnejši dokaz za to, da je sistemski ugovor ničvreden, prva, vprašanje pa je, ali je prepričljiv sploh katerakoli izmed naštetih. Copeland (1993) pravi, da je tarča Searlovih odgovorov nekaj, česar sistemski ugovor sploh ne trdi, in to je, da del sistema sicer ne razume, razume pa sistem kot celota. Poanta »logičnega odgovora« je zgolj ta, da je argument iz »kitajske sobe« neveljaven, tj. da iz dane premise ne sledi navedeni sklep. O njegovi resničnosti ne pove ničesar, še najmanj to, kar mu pripisuje Searle, zaradi česar njegovi odgovori niso uperjeni proti zagovornikom sistemskega ugovora, ampak proti nekomu drugemu. V tej luči je tudi njegova opazka o krožnosti »logičnega odgovora« neumestna, saj bi to morda držalo samo v primeru, če bi Searlu, ki skuša dokazati, da delujoči program ni sposoben razumeti, ugovarjali s tem, da oseba v sobi sicer ne razume kitajsko, razume pa sistem, katerega del je (kar Searle sicer neupravičeno domneva), saj bi tako res že na začetku predpostavili tisto, kar bi morali šele dokazati. Ker pa sistemski ugovor tega ne sprejema kot izhodišče, ni jasno, kako naj bi navedene tri pripombe predstavljale orožje za njegovo zavrnitev.

Poleg tega pa se zdi, da lahko Searla obtožimo tudi *zmote delitve*, ki se kaže v njegovi zadnji pripombi na »logični odgovor«. Če je bilo prej sporno pripisovanje nerazumevanja celoti na osnovi pripisovanja nerazumevanja njenemu delu, pa je sedaj zgrešena njegova trditev, da vodi sistemski ugovor v absurden nastanek mentalnih jeter, ledvic ali želodca. Iz tega, da je sistem mentalen, namreč še ne sledi, da so mentalni tudi vsi njegovi deli, podobno kot ne velja za vsako rimsko legijo, da je bila nepremagljiva, če je bila nepremagljiva

rimska armada. Ker je celota več kot zgolj seštevek njenih delov, je možno, da ostanejo »notranji organi« sistema nementalni, njihova specifična organizacija pa zagotavlja celoti novo lastnost, npr. mentalnost, ki je sami nimajo. V končni fazi med nami in računalniki z vidika osnovnih sestavnih delov, tj. kemičnih elementov, ni razlik, da se po prevladujočem mnenju v nas pojavi duh, v njih pa ne, mora to očitno biti posledica razlike v urejenosti med nami in njimi, saj bi v nasprotnem primeru bili primorani razmišljati o obstoju neke posebne substance v slogu Descartesovih nematerialnih duhov, ki so se po nekem čudnem naključju (ali morda božji intervenciji) znašli v človeških telesih, ne pa tudi v računalniških ohišjih.[22]

Ugovor z robotom

Kaj pa, če »kitajska soba« ne razume, ker gre preprosto za sistem (ki v slogu enega izmed ugovorov Turingovemu testu, in sicer *ugovora iz čutil*, ki pravi, da se ta omejuje zgolj na strojevo sposobnost dajanja verbalnih odgovorov, ne pa tudi na njegovo zmožnost povezati besede s stvarmi, ki jih označujejo), ki ni sposoben proizvesti vzročnih relacij med simboli in stvarmi, na katere se nanašajo. Če pa bi npr. Schankov računalnik dali v robota, kjer bi predstavljal njegove ‘možgane’, bi se takšno bitje odzivalo na dražljaje, odločalo glede na značilnosti okolja in obnašalo skladno z njimi, tj. opremljeno s kamero, ki zaznava zunanji svet, bi delovalo po načelu vzroka in učinka, zaradi česar bi mu lahko pripisali tudi pravo razumevanje. Searle ugovor z robotom povzame takole: »Recimo, da damo v računalnik robota, in ta računalnik ne bo le sprejemal formalnih simbolov skoz input in jih nato oddajal skoz output, temveč bo dejansko vodil robota, tako da bo ta opravljal funkcije, ki bodo zelo podobne zaznavanju, hoji, pomikanju, zabijanju žebljev, jedenju, pitju – kar pač že hočete. Ta robot bi recimo na sebi imel pritrjeno televizijsko kamero, s katero bi lahko videl, imel bi roke in noge, s katerimi bi lahko ‘nastopal’, in vse to bi nadzorovali njegovi računalniški ‘možgani’. Tak robot bi imel, za razliko od Schankovega računalnika, pravo zmožnost razumevanja in druga mentalna stanja« (Searle, 1980/1990: 369).

Searle meni, da je ugovor z robotom kontraproduktiven, saj s tem, ko od sistema za razumevanje poleg sintakse zahteva še vzročne relacije z zunanjim svetom, priznava, da sintaksa sama po sebi še ni dovolj za mentalno vsebino in tako postavlja pod vprašaj močno inteligenco: »Prvo, kar opazimo pri odgovoru z robotom, je, da molče priznava, da spoznavanje ni zgolj stvar manipulacije s formalnimi simboli, saj dodaja kup vzročnih zvez z zunanjim svetom /.../« (Searle, 1980/1990: 369). Komentatorji menijo (Preston, 2002), da Searle ugovoru z robotom ni naklonjen že zato, ker temelji na redukciji semantike na (vzročne) naravne pojave, ki so brez vsakršne intencionalnosti, kar pa zanj, ki je zagovornik t.i. *internalistične teorije semantike*, ni sprejemljivo, saj naj bi se tako

[22] Za Searlovo možno rešitev iz zagate *glej* Preston, 2002: 30, vprašanje pa ni zgolj, ali bi se z njo strinjal, ampak tudi, ali bi navedeno sploh (upravičeno ali neupravičeno) sprejel kot problem.

izgubila *subjektivnost*, kije bistvo mentalne vsebine.

Razen tega Searle pravi, da »takšne ‘zaznavne’ in ‘motorne’ zmožnosti v ničemer ne povečujejo razumevanja oziroma intencionalnosti prvotnega Schankovega programa« (Searle, 1980/1990: 369–370). Ne glede na to, kako bogato opremimo našega robota, in ne glede na to, kako zahteven Turingov test bo potem lahko opravil, še vedno bo zgolj manipuliral s simboli (noben še tako sofisticiran tehnološki dodatek ga ne bo pripeljal do razumevanja), zaradi česar opravljen Turingov test nikoli ne more biti zagotovilo za to, da nekaj ima duševnost.

Harnad (1993) na to odgovarja, da kakor hitro v igro pripeljemo tako zaznavno in motorno razvitega robota, argument iz »kitajske sobe« pade, saj moramo v ten primeru *klasični Turingov test, v* katerem računalnik zgolj v grobem in iz druge roke opisuje svoje občutke, ko npr. igra tenis (ki ga v resnici seveda ne igra), zamenjati z *absolutnim Turingovim testom*, kjer jih opisuje iz prve roke (zdaj pa tenis resnično igra). To pa vodi Searla v naslednjo (konstruktivno) dilemo (po analogiji je tukaj razumevanje zamenjano z videnjem), ki je kombinacija sistemskega ugovora in ugovora z robotom, z rogovoma, ki zanj nista sprejemljiva:

1. Ali robot (prej Searle) zgolj sprejema iznos optičnega aparata (prej iznos možganske kalkulacije) ali pa robot dejansko gleda, npr. kip Bude.

2. Če robot zgolj sprejema iznos optičnega aparata, potem ne implementira celotnega sistema (ampak le njegov del), torej velja sistemski ugovor. 3. Če robot dejansko gleda kip Bude, potem implementira celoten sistem (ne zgolj njegovega dela), torej resnično vidi.

/. 4. Ali velja sistemski ugovor ali pa robot resnično vidi (Harnad, 1993: 17).

Dejstvo je, da mora Searle takšen modificiran ugovor z robotom zavrniti, saj v nasprotnem primeru njegov argument iz »kitajske sobe« ne deluje več: prvi rog ga sili v ubadanje z že predstavljenim »logičnim odgovorom«, ki mu očita zmoto sestave in zmoto delitve, drugi rog pa po analogiji s takšnim sofisticiranim računalnikom v kombinaciji z robotom pripisuje dejansko razumevanje, kar je hotel ovreči.[23]

Ugovor z možganskim simulatorjem

Zamislimo si računalnik, ki je s pomočjo več programov, ki delujejo vzporedno, sposoben simulirati možgane rojenih Kitajcev, ko obdelujejo formalne elemente kitajskih vnosov in iznosov: na ravni nevronov in sinaps, ki pri reševanju zahtevanih formalnih nalog tvorijo dobro delujočo mrežo, med računalniškim programom in programom možganov rojenih Kitajcev ni nobene razlike. Ugovor z možganskim simulatorjem Searle povzame kot »Ta stroj sprejema kot input kitajske zgodbe in vprašanja o njih, simulira formalno strukturo dejanskih

[23] Za zagovor argumenta iz »kitajske sobe« proti Harnadovi konstruktivni dilemi *glej* Bringsjord in Noel (2002).

kitajskih možganov, ko obdelujejo te zgodbe in oddaja kot output kitajske odgovore. Zamišljamo si lahko celo, da stroj ne deluje samo z enim serijskim programom, temveč s celim nizom programov paralelno, tako kot domnevno delujejo tudi človeški možgani, ko obdelujejo materni jezik« (Searle, 1980/1990: 370). Vprašanje je, ali ne bi morali v tem primeru reči, da računalnik razume, saj bi v nasprotnem primeru morali odreči razumevanje tudi ljudem?

Searlov odgovor je, kajpak nikalen, začne pa ga z naslednjo ironično pripombo: »/.../ mislil sem /.../, da je poglavitna misel UI ta, da nam ni treba vedeti, kako delujejo možgani, če hočemo zvedeti, kako deluje duh /…/ Če bi morali vedeti, kako delujejo možgani, da bi razvili UI, se pač ne bi ukvarjali z UI« (Searle, 1980/1990: 370). Motivacija za takšno razmišljanje se skriva v tem, da je bil v času nastanka argumenta iz »kitajske sobe« prevladujoč pristop na področju umetne inteligence t.i. *simbolna hipoteza*, ki pravi, da naš silni duh ni nič več kot univerzalni sistem simbolov, zaradi česar je tudi človeško mišljenje zgolj manipulacija z njimi.

Kakor hitro pa v igro pripeljemo antropomorfistično zahtevo po poznavanju zgradbe in delovanja človeških možganov za to, da bi stroj mislil, se oddaljimo od ideje močne umetne inteligence, po kateri je *»/.../* duh nasproti možganom toliko, kot je program nasproti hardveru, zaradi česar lahko razumemo duha, ne da bi se poglabljali v nevrofiziologijo« (Searle, 1980/1990: 371).

Ugovor z možganskim simulatorjem pa z vidika konstrukcije mislečega računalnika po njegovem mnenju na neuspeh obsodi tudi takrat manjšinski pristop k umetni inteligenci – *nevronske mreže*, ki trdi, da ima ustrezno sestavljena (iz enot, ki predstavljajo nevrone, in povezav med njimi, ki predstavljajo sinapse) in vrhunsko izurjena konekcionistična mreža duh. Zagovorniki tega t.i. *konekcionizma* predpostavljajo, da ima tako zgrajen sistem pravo duševnost izključno zaradi primerne (našim možganom podobne) konfiguracije in izvežbanosti nevronske mreže, kar pa za izdelavo inteligentnega stroja spet zahteva antropomorfizem in tako spodbija zamisel močne inteligence, ki v duhu računskega funkcionalizma pravi, da *»/.../* obstaja raven mentalnih operacij, ki sestojijo iz računskih procesov nad formalnimi elementi, ki sestavljajo bistvo mentalnega in se jih da uresničiti v najrazličnejših možganskih procesih, prav tako kot lahko vsakteri računalniški program uresničnimo v različnih računalniških hardverih */.../«* (Searle, 1980/1990:371).

Za primer, da bi se konekcionisti iz Searlovega primeža skušali izviti z očitkom, da njihovo pojmovanje močne inteligence tega, kar je navedeno v citatu, ne zahteva, saj pravi zgolj, da »ima ustrezno sestavljena in izurjena nevronska mreža duh izključno zaradi svoje konfiguracije in izvežbanosti« (Preston, 2002: 33), pa Searle (1990) razvije t.i. *miselni eksperiment iz »kitajske telovadnice«*. Zamislite si telovadnico namesto sobe, v kateri imamo veliko samo angleško govorečih ljudi, ki izvajajo iste operacije kot nevroni in sinapse v konekcionistični mreži. Rezultat je na koncu enak kot v sobi, kjer s kitajskimi simboli s pomočjo knjig s

pravili manipulira zgolj en človek. Nihče ne govori kitajsko, čeprav se zunanjim opazovalcem tako zdi in ne obstaja noben način, da bi se celota naučila pomena kitajskih besed (Searle, 1990: 22). Če prej ni razumel en človek, sedaj ne razume več ljudi, in če prej ni razumel sistem z enim procesorjem, ki je svoje operacije izvajal zaporedno, sedaj ne razume sistem z več procesorji, ki svoje operacije izvajajo vzporedno; druge razlike ni.

Copeland se s tem ne strinja in pravi, da je »sistemski ugovor« dovolj za zavrnitev »kitajske telovadnice«: posamezni igralci v mreži sicer res ne razumejo kitajsko, vendar iz tega še ne sledi, da kitajsko ne razume celoten sistem. »Napaka, ki je posledica sklepanja iz dela na celoto, je tukaj celo bolj očitna kot v izvirni različici argumenta iz kitajske sobe« (Copeland, 2002: 116).

Ugovor »drugih duhov«

Vprašajmo se, kako sploh vemo, da npr. naš prijatelj Janez, s katerim redno razpravljamo o aktualnih družbenih, gospodarskih in kulturnih problemih, karkoli od tega, o čemer se z njim pogovarjamo, sploh razume? Zdi se, da razen njegovih primernih in smiselnih verbalnih odgovorov na vprašanja, ki se v danem kontekstu porajajo, nimamo na voljo nobenega drugega dokaza, da to, o čemer teče beseda, razume. Če pa to drži, bi morali tudi računalnik, ki se je v danem okolju sposoben obnašati na ustrezen način, bodisi verbalno bodisi fizično, obravnavati kot bitje ali sistem, ki to, kar je vzrok za njegovo takšno ali drugačno delovanje, razume. V nasprotnem primeru bi nas lahko nekdo pri »merjenju« duševnosti obtožil uporabe različnih vatlov za različne organizme oziroma t.i. šovinizma vrst, ki favorizira ljudi pred stroji, ko za to ni nobene osnove. Searle ugovor drugih duhov povzame takole: »Kako pa veste, da drugi ljudje zares razumejo kitajsko ali karkoli drugega? Samo po njihovem vedenju. Računalnik lahko ravno tako uspešno opravi vedenjske teste kot oni (v načelu); če torej pripisujemo drugim ljudem spoznavnost, jo morate načeloma pripisati tudi računalnikom« (Searle, 1980/1990:373).

Zanimivo je, da z njim opravi na kratko, meni namreč, da »zgreši poanto«, saj pri argumentu iz »kitajske sobe« ne gre za to, *kako* vemo, da imajo druga bitja duh, ampak zato, *kaj* jim pripisujemo, ko pravimo, da ga imajo. »Ost dokazovanja je v tem, da to ne morejo biti zgolj računski procesi in njihov output, saj lahko računski procesi in njihov output obstajajo brez kognitivnega stanja« (Searle, 1980/1990:373).

Zdi se mi, da si ugovor drugih duhov kljub vsemu zasluži več pozornosti, posebej v luči (i) Searlovega ostrega razlikovanje med sintakso in semantiko in (ii) njegove splošne teorije duha. Searle Turingu očita, da se ni nikoli vprašal, *kaj* je tisto, zaradi česar ima neka naprava mentalno vsebino, tj. semantične značilnosti. S tem je zanemaril glavno razliko med strojem in človekom – slednji je po njegovem mnenju pač edini organizem, ki ima dostop tako do sintakse kot tudi do semantik, in hkrati s tem tudi glavni razlog, zakaj mi razumemo ali imamo

intencionalnost, računalniki pa ne. V svojih odgovorih na kritike argumenta iz »kitajske sobe«, kot smo videli, vedno poudarja, da niti celoten sistem niti sofisticiran robot niti možganski simulator nimajo na voljo ničesar, da bi simbolom pripisali pomene, edino, kar delajo je, da s pomeni, ki jih simboli imajo, manipulirajo. Še tako izbrana naprava (od znotraj) simbolom ne more dati pomena, to lahko stori zgolj nekdo od zunaj, človek, ki pa še zdaleč ne rabi biti tako sofisticiran. To, da sintaksa ni zadostna za semantiko, je za Searla zelo *preprosta logična resnica* (Searle, 1992: 200).

Ljudje pa se zavedamo, da v duhu obstaja vsebina, ki nam omogoča vednost o tem, da so naše misli o npr. o predmetih ali simbolih, ki nekaj predstavljajo in imajo pomen in da je to več od zgolj manipulacije s formalnimi simboli. Vprašanje je, zakaj mi to vemo, stroji pa ne? Odgovor na to se skriva v njegovi teoriji duha, ki pravi, da so mentalni pojavi *biološki* pojavi, kar bolj konkretno pomeni, npr., daje vse, kar je zavestno, v resnici biološko, ki jo imenuje *biološki naturalizem*.[24] Natančneje pravi, da so *možgani oboje, vzrok in realizator zavesti*, podobno, kot je voda oboje, vzrok in realizator tekočnosti. Razlikuje med *mikrolastnostmi* sistema, tj. njegovimi sestavnimi deli, kisikom in vodikom ter relacijami med njima pri vodi oziroma nevroni in povezavami med njimi pri možganih, ter *makrolastnostmi* sistema, tj. njegovimi vidnimi lastnostmi, npr. tekočnostjo pri vodi oziroma zavestjo pri možganih. Poanta je, da mikrolastnosti povzročajo makrolastnosti, kisik in vodik ter odnos med njima tekočnost vode oziroma nevroni in interakcije med njimi zavest možganov, pri čemer sistem, voda oziroma možgani, predstavljata okvir, v katerem se to realizira. To vodi do tega, da Searle zavest pripisuje možganom – sistemu (makroraven), ki v emergentističnem slogu vznikne iz nevronov – sestavnih delov (mikroraven), ne pripisuje pa je tudi njim samim (kot je dokazoval, da mora sprejeti zagovornik sistemskega ugovora, če predpostavi, da je sistem zavesten, saj so potem zavestni tudi njegovi deli, npr. jetra, ledvice in želodec, kar pa je, kot pravi, zmotno). Skratka, ljudje se zavedamo mentalne vsebine, ki nam omogoča napredovanje od sintakse do semantike zato, ker imamo ustrezno zgrajene biološke možgane, ki mentalno vsebino hkrati povzročajo in realizirajo, stroji pa ne.

Warwick (2002) dokazuje, da nas argument iz »kitajske sobe« vodi v omenjeni šovinizem vrst, in da pade, če nam uspe pokazati, da ima lahko tudi stroj, računalnik ali robot zavest, pri čemer za izhodišče uporabi Searlovo trditev, da je zavest lastnost možganov. Če to drži, potem jo težko odrečemo umetnim silikonskim možganom ali umetnim nevronskim možganom; smiselno pa je predpostaviti, da gre za različne vrste zavesti, saj je vedno realizirana v različnih vrstah možganov. Neumestno pa je trditi, da se lahko zavest pojavi zgolj v bioloških sistemih, saj je po definiciji tega, kaj je življenje, kmalu jasno, da pogoje zanj, ki segajo od samoorganizacije, neodvisnosti in odgovornosti do reprodukcije, evolucije in metabolizma, izpolnjujejo tudi številni stroji (Warwick,

[24] Za kritiko tega stališča *glej* npr. Penrose (2002).

2002). Zavest, ki jo ima v mislih, je *sposobnost opravljanja takšnih in drugačnih operacij*, v primeru z roboti *Sedem palčkov* je to npr. gibanje naprej brez zadevanja predmetov iz okolice. Po analogiji s človekom, kjer nadzor njegovega telesa s strani človeških možganov pomeni uprimeritev človeške zavesti, v primeru stroja nadzor njegovega telesa s strani strojnih možganov pomeni uprimeritev strojne zavesti. Poanta njegovega odgovora je tako, da so tudi stroji (v okviru svoje vrste) zavestni in sposobni učenja, čeprav nimajo bioloških možganov, ampak umetne, zaradi česar je možno razumevanje (preskok od sintakse do semantike) načeloma pripisati tudi njim.

Kakorkoli, Searle s svojim miselnim eksperimentom iz »kitajske sobe« namiguje na to, da vprašanje, ali stroj lahko misli, ni empirično, ampak miselno, zaradi česar nam pri iskanju odgovora nanj noben tehnološki napredek, pa naj bo še tako velik, ne pomaga. Kljub dosežkom umetne inteligence (včeraj, danes ali jutri) lahko z argumentom iz »kitajske sobe« takoj dokažemo, da inteligentni stroji ne glede na to, kako bodo realizirani, nikoli ne bodo stvarnost. Videli pa smo, da vsaj sistemski ugovor z opozarjanjem na njegovo logično neveljavnost in ugovor drugih duhov v kombinaciji s kritiko biološkega naturalizma, to napoved postavljata pod vprašaj.

FUNCTIONALISM, COMPUTERS, AND »CHINESE ROOM«

The article introduces a relation among functionalism which considers mental states as functional states, computers that are according to its computational version able to think and the Chinese Room Argument which is supposed to prove, once and for all, that machines will never be intelligent. The first part shows the fundamental idea and motivation for functionalism together with the main objections to it, the second part of the article represents the Chinese Room Thought Experiment which, formulated as an argument, rejects the idea of thinking machines on the grounds that a manipulation with symbols is not enough for understanding their contents, and the third part summarizes the key objections to the Chinese Room Argument as well as principle answers to them bringing about thereby some new ideas in order to defend the belief that intelligent computers are, in principle, possible.

Keywords: the Chinese Room Argument, artificial intelligence, the symbol system

hypothesis, the language of thought hypothesis, computation-representation functionalism.

Literatura

Armstrong, D. M. (1999). *The Mind-Body Problem*. Boulder: Westview Press.

Aristotel (1993). *O duši*. Ljubljana: Slovenska matica.

Block, N. (ur.) (1980a). *Readings in Philosophy of Psychology, vol. 1*. Cambridge: Harvard University Press.

Block, N. (1980b/2007). »Uvod: Kaj je funkcionalizem?«. V Markič, O. in Bregant, J. (ur.), *Narava mentalnih pojavov*, Maribor: Aristej, str. 147–166.

Block, N. (1980c). »Troubles with Functionalism«. V Block, N. (ur.), *Readings in Philosophy of Psychology, vol. 1*, Cambridge: Harvard University Press, str. 268–305.

Block, N. (1990). »Inverted Earth«. *Philosophical Perspectives*, 4, str. 53–79.

Block, N. (2002). »Searle's Arguments against Cognitive Science«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 70–79.

Block, N. in Fodor, J. (1972/1980). »What psychological states are not«. V Block, N. (ur.), *Readings in Philosophy of Psychology, vol. 1*, Cambridge: Harvard University Press, str. 237–250.

Bregant, J. (2004). *Misel kot vzrok*. Maribor: Pedagoška fakulteta.

Bregant, J. (2010). »Ali lahko stroj misli?«, *Analiza*, 4, str. 55–72.

Bregant, J. (2012). »Turing se je motil«, *Analiza*, 1/2, str. 5–31.

Bringsjord, S. in Noel, R. (2002). »Real Robots and the Missing Thought-Experiment in the Chinese Room Dialectic. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 144–166.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Churchland, P. (2011). *Braintrust*. Princeton: Princeton University Press.

Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.

Copeland, J. (2002). »The Chinese Room from a Logical Point of View«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 109–122.

Coulter, J. in Sharrock W. (2002). »The Hinterland of the Chinese Room«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 181–200.

Dennett, D. C. (1985). »Can Machines Think«. V Shafto, M. (ur.), *How we Know*, San Francisco: Harper & Row, str. 121–145.

Dennett, D. C. (1991/2012). *Pojasnjena zavest*. Ljubljana: Krtina.

Descartes, R. (1637/2007). *Razprava o metodi*. Ljubljana: ZRC SAZU.

Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Fodor, J. (1981). »The Mind-Body Problem«, *Scientific American*, 244, str. 124–132.

Fodor, J. (1987). *Psychosemantics*. Cambridge: The MIT Press.

Fodor, J. A. in Pylyshyn, Z. W. (1988). »Connectionism and Cognitive Architecture: A Critical Analysis«, *Cognition*, 28, str. 3–71.

Gendron, B. (1971). »On the Relation of Neurological and Psychological Theories. A Critique of the Hardware Thesis«. V Buck, R. C. in Cohen, R. S. (ur.), *In Memory of Rudolf Carnap*, Dordrecht: Reidel, str. 483–495.

Harnad, S. (1993). »Grounding Symbols in the Analog World with Neural Nets: A Hybrid Model«, *Think*, 2, str. 12–20.

Harnad, S. (1995). »Why and How We Are Not Zombies«. *Journal of Consciousness Studies*, 1, str. 164–167.

Harnad, S. (2002). »Minds, Machines, and Searle2: What's Right and Wrong aboout the Chinese Room Argument«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 294–307.

Haugeland, J. (2002). »Syntax, Semantics, Physics«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 379–392.

Jackson, F. (1995). »What Mary Didn't Know«. V Moser P. K. in Trout, J. D. (ur.), *Contemporary Materialism*, London: Routledge, str. 180–189.

Kalke, W. (1969). »What is Wrong with Fodor and Putnam's Functionalism«. *Nous*, 3, str. 83–94.

Locke, J. (1689/2012). *An Essay Concerning Human Understanding*. Verzija je dosegljiva na www.earlymoderntexts.com.

Kurzweil, R. (1998). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking.

Markič, O. (2010). *Kognitivna znanost*. Maribor: Aristej.

Miščević, N. (2001). »Uvod«. V Miščević, N. in Smokrović, N. (ur.), *Računala, mozak in ljudski um*, Rijeka: Izdavački centar Rijeka, str. 9–42.

Newell, A. (1980). »Physical Symbol Systems«. *Cognitive Science*, 4, str. 135–183.

McCulloch, W. in Pitts, W. (1943). »A Logical Calculs of the Ideas Immanent in Nervous Activity«, *Bulletin of Mathematical Biophysics*, 5, str. 115–133.

Preston, J. (2002). »Introduction«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 1–50.

Putnam, H. (1960). »Minds and Machines«. V Hook, S. (ur.), *Dimensions of Mind*, New York: New York University Press, str. 148–179.

Putnam, H. (1966/1975). »The mental life of some machines«. V *Mind, Language and Reality*, Cambridge: Cambridge Unviersity Press, str. 408–428.

Putnam, H. (1963/1980a). »Brains and Behaviour«. V Block, N. (ur.), *Readings in Philosophy of Psychology, vol. 1*, Cambridge: Harvard University Press, str. 24–36.

Rey, G. (2002). »Searle's Misunderstanding of Functionalism and Strong AI«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 201–225.

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, str. 20–25.

Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge. The MIT Press.

Searle, J. (1993). »The Failures of Computationalism«. *Think*, 2, str. 68–71.

Searle, J. (1997). *The Mystery of Consciousness*. New York: A New York Review Book.

Sesardić, N. (1984). *Fizikalizam*. Novi Beograd: Istraživačko-izdavački centar SSO Srbije.

Shoemaker, S. (1975). »Functionalism and Qualia«. *Philosophical Studies*, 27, str. 291–315.

Shoemaker, S. (1982). »The Inverted Spectrum«. *Journal of Philosophy*, 79, str. 357–381

Smart, J. J. C. (1959/2007). »Občutki in možganski procesi«. V Markič, O. in Bregant, J. (ur.), *Narava mentalnih pojavov*, Maribor: Aristej, str. 79–93.

Turing, A. (1936). »On Computable Numbers, with an Application to the *Entscheidungsproblem*«, *Proceedings of the London Mathematical Society*, 42, str. 230–265.

Turing, A. (1990). »Stroji, ki računajo, in inteligenca«. V Hofstadter, D. R. in Dennet, D. C. (ur.), *Oko duha*, Ljubljana: Mladinska knjiga, str. 61–74.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge: The MIT Press.

Tye, M. (2006). »Absent Qualia and the Mind-Body Problem«. *Philosophical*

Review, 115, str. 139–168.

Warwick, K. (2002). »Alien Encounters«. V Preston, J. in Bishop, M. (ur.), *Views into the Chinese Room*, str. 308–318.



Iz zgodovine filozofije

Peter Lukan
Filozofska fakulteta Univerze v Ljubljani

ARISTOTEL IN POJEM NAKLJUČJA

Aristotel za naključje uporablja dva pojma in sicer *týche* in *taùtómaton*, od katerih se prvi nanaša na človeški svet, drugi pa na naravna dogajanja. Bistvena je njegova oznaka naključja kot sekundarnega vzroka, s čimer po eni strani kaže na človeško dojemanje in način govora o njem kot o vzroku, po drugi pa na rezultat Aristotelove analize, ki pokaže, da gre tu v resnici za nesubstancialni vzrok. Naključje kot *dogodek* je postavljeno v teleološki kontekst – gre za pomenljiv redek dogodek, ki ga povzroči naključni vzrok in kateremu umanjka smotrni vzrok. Izvor naključja je vzpostavljen tudi v odnosu do pojma pogojne nujnosti, to je tipa nujnosti, ki opisuje dogajanja v sublunarnem svetu in ki dopušča tudi redke izjeme pri naravnih dogajanjih. Problematična točka so za Aristotela spački in nespolno razmnoževanje, saj ta dva pojava kažeta na potrebo po uvedbi pojma naključne substance, kar nasprotuje Aristotelovi lastni definiciji naključja kot nesubstancialnega vzroka. V članku odprte probleme na koncu tolmačimo v perspektivi (statistične) pogostosti/redkosti pojavov.

Ključne besede: naključje, Aristotel, teleologija, vzroki, statistika.

Pri razdelavi Aristotelovega pojma naključja se bom opiral na monografijo Johna Dudleyja *Aristotle's Concept of Chance* (2012), ki je domnevno prvo izčrpno delo na to temo. Ta pojem je v evropski miselni tradiciji prisoten že od samih začetkov evropske filozofske misli, z njim so operirali že misleci, kot so Demokrit, Empedokles, Heraklit, Anaksagora, vendar je bil Aristotel poleg Demokrita edini, ki je podal filozofsko *razlago* tega pojma. Še Aristotelov učitelj Platon je pojem uporabljal brez posebne pojasnitve. Aristotelova motivacija za podrobnejše ukvarjanje s pojmom naključja naj bi izhajala iz njegove želje po zavrnitvi Demokritovega pripisovanja kozmičnega reda naključju (Dudley, str. 2012:2). Prvotna motivacija za ukvarjanje s tem pojmom pri njem spada torej v območje kozmologije oziroma kozmogonije. Glavno refleksijo v zvezi s pojmom naključja v kozmološkem kontekstu poda v *Fiziki*. Aristotel sicer uporablja ta pojem tudi v svojih bioloških delih v glavnem zato, da pojasni izvor nekaterih

življenjskih oblik. V svojih etičnih delih obravnava, kakšno vlogo ima naključje pri sreči v življenju. Dudley meni, da je Aristotelovo pojmovanje naključja sploh temeljnega pomena za razumevanje tako njegove filozofije narave kot etike.

Po kratki terminološki razjasnitvi in predstavitvi pojma naključja pri mislecih pred Aristotelom se bomo osredotočili na njegovo pojmovanje naključnega vzroka, torej na opredelitev, kakšne vrste vzrok je sploh naključje. Nadalje nas bo posebej zanimalo, v kakšnem odnosu je naključje do pojma smotra oziroma teleološkega mišljenja, ter razmerje med pojmoma naključja in nujnosti, pri razumevanju katerega sta posebno zanimiva primera spačkov in spontane generacije organizmov. Navedli bomo še Aristotelov pogled na morebitno »znanost o naključju« ter opredelili mesto, ki ga ima naključje v njegovem pojmovanju sreče v etičnih delih.

Pojma *týche* in *taùtómaton*

Aristotel svojo osrednjo doktrino naključja posreduje v *Fiziki II 4–6*. Dva glavna izraza v zvezi z naključjem sta *týche* in *taùtómaton*, ki ju slovenski prevajalec Valentin Kalan prevaja kot naključje in samodejnost. Angleški prevajalci so izraza večinoma prevedli kot »luck« in »chance« oziroma se pojem *taùtómaton* alternativno prevaja tudi kot »spontaneity« ter »the automatic«. Ta izraz je sprva pomenil »samo od sebe« (tako kot na primer padajo lasje, teče reka, itd.), vendar se Aristotel odmakne od tega prvotnega pomena in pri njem kasneje pomeni naključje.

Izraz *týche* Aristotel rabi skoraj izključno v povezavi s človeškimi dejanji, kar angleški prevod »luck« dobro povzema, slovenski prevod naključje pa nekoliko manj, medtem ko se izraz *taùtómaton* uporablja tudi za dogajanja v naravi in je zato širši ter vključuje tudi pomensko polje *týche*. Slovenski izraz naključje se velikokrat uporablja tako v kontekstu ljudi kot v kontekstu drugih naravnih dogajanj (npr. v stavku »strela je po naključju zadela drevo in povzročila požar«). Da bi ti dve rabi bolje razločili, bomo uporabljali izraza naključje in sreča oziroma nesreča v smislu dogodka (tako kot rečemo npr. »imel sem srečo, da sem dobil glavni zadetek«), saj ima pojem sreče očitno antropocentrično konotacijo, to pa je ravno razlikovalna lastnost pojma *týche* nasproti pojmu *taùtómaton*. V slovenščini lahko (ne)srečni dogodek dobro razločujemo od pojma sreče v smislu duševnega blagostanja s pomočjo fraz »*imeti* (ne)srečo« in »*biti* (ne)srečen«.

Dudley z natančno analizo pokaže, da se izraz *týche* občasno pojavlja tudi kot splošni pojem za naključje, zato je pri prevajanju in tolmačenju potrebno paziti na kontekst. Ko se *týche* nanaša na človeška dejanja, je torej boljši prevod sreča, ko pa je izraz uporabljen v najbolj splošnem smislu, je boljši prevod zanj naključje. Potrebna previdnost pri tolmačenju izraza *týche* je posledica tega, da je pojem naključja pri Aristotelu doživel manjši razvoj, tako da je prvotni splošni

pojem *týche* zamenjala dvojica pojmov *týche* in *taùtómaton*[1]. *V* odlomkih, kjer torej nastopata oba izraza, ima pojem *týche* pomen naključja v človeškem svetu oziroma (ne)srečnega dogodka, na kratko sreče ali nesreče.

Aristotel na nekaterih mestih v smislu naključnega dogodka uporabi tudi izraz *sýmptóma*, ki pomeni sovpadanje (angleški prevod je »coincidence«)[2]. V slovenščini ima izraz sovpadanje šibkejšo konotacijo naključja (na primer v stavku »moj prihod je ravno sovpadel z njegovim, tako da se nisva iskala«). V angleščini pa tudi na primer v italijanščini (izraz *coincidenza*) je pomensko povezava med sovpadanjem in naključjem bolj izrazito vidna v vsakdanjem govoru. To dejstvo zrcali eno od pomembnih lastnosti naključja, namreč to, da gre pri naključju za sovpadanje kakšnih pomembnih okoliščin.[3] Izraz *sýmptóma* ima pri Aristotelu lahko vzročne konotacije ali pa ne, vedno pa se zgodi naključno.

Pojem naključja pri drugih grških mislecih

Ena bistvenih značilnosti pojma naključja pri predsokratikih je bila, da se ta ni pomensko izključeval s pojmom nujnosti.[4] Nujnost je pri njih bila postavljena v sam začetek obstoja sveta, Aristotel pa jo je prvi uporabljal tudi pri opisu naravnega dogajanja. To je po Dudleyjevem mnenju posredno povezano tudi z dejstvom, da so predsokratiki poznali samo snovni in dejavni vzrok, ne pa tudi oblikovnega in smotrnega,[5] o čemer Aristotel razpravlja na več mestih (npr. v *Metafiziki A)*. Za predsokratike je naključje nekaj, kar je kontingentno dano in je kot tako hkrati nujno in naključno, v smislu, da je neka danost pač nujna, kakšna je ta danost, pa je kontingentno. Za takšne dogodke bi torej lahko rekli, da so nujni *po danosti* in kontingenti *po pojavnosti*. Tako opredeljen pojem naključja je precej vsebinsko prazen oziroma nevtralen, v današnji znanstveni rabi je na primer pojem naključnega dogodka vsebinsko bolj opredeljen.

---

[1] Aristotel naj bi po Dudleyjevem prepričanju najprej napisal *Fiziko II*, 4–6 samo z rabo izraza *týche*, kasneje naj bi napisal *Fiziko II*, 4, ki vsebuje razlikovanje med *týche* in *taùtómaton*, naknadno pa še revidiral *Fiziko II*, 4–5, da bi bolje vključevala doktrino iz *Fizike II*, 6. Po mnenju Dudleyja je bila distinkcija med *týche* in *taùtómaton* uvedena precej pozno, Aristotel jo je uporabil le še na nekaterih mestih v *Metafiziki*. Na splošno sicer Aristotel ni veliko spreminjal svojih pogledov, piše Dudley, stališče o razvoju njegove misli, ki ga gojijo nekateri raziskovalci, je zanj pretirano. Dudley pravi zgolj, da obstajajo nekateri pojmi, ki so pri tem izjeme, vendar je potrebno njihov domnevni razvoj dobro argumentirati. Pojem naključja je po njegovo ena izmed teh izjem.

[2] Dudley navaja naslednja mesta: *Fizika II 8*, 199 a 1–5; *Retorika I 9*, 1367 b 25; *Politika II 21*, 1274 a 12; *Politika V 4*, 1304 a 1 ter *Politika V 6*, 1306 b 6–12.

[3] Carl Gustav Jung je v tem pomenu uporabljal svoj neologizem *sinhroniciteta*, ki ga je natančneje razložil v svojem delu *Synchronicity: An Acausal Connecting Principle* (1952). Sinhroniciteto opredeli kot časovno sovpadanje dveh dogodkov, ki sta povezana po pomenu, oziroma kot »nevzročni princip povezovanja«.

[4] Ravno tako se pojem naključja ni nujno bil s pojmom racionalnega mišljenja. Demokritovo pojmovanje naključja je skoraj v celoti znano iz kritik in premišljanj mislecev za njim, ki so imeli dostop do njegovih del, vendar v knjigi *Fragmenti predsokratikov* (ur. Gorazd Kocijančič) najdemo tudi kakšen ohranjen fragment, ki se šteje za pristnega, v katerem je naključje eksplicitno omenjeno. Demokrit pravi: »Ljudje so iz naključja napravili sveto podobo kot izgovor za lastno nepreudarnost. Naključje se namreč le redko bojuje z razumnostjo /…/« (Epikur, *Sent*. 16, v: *Fragmentipredsokratikov*, str. 911).

[5] Slovenski prevajalec Kalan uporablja za Aristotelove štiri vzroke naslednje izraze: snovni, gibalni, oblikovni in dejavni vzrok in jih tukaj prevzemamo.

Kot že omenjeno, so grški misleci pred Aristotelom z izjemo Demokrita operirali s pojmom naključja na nereflektiran način. Empedokles je verjel, da je splošni okvir poteka zgodovine vnaprej določen, medtem ko so naključni samo individualni dogodki, in sicer zato, ker se elementi srečujejo in mešajo na nepredvidljiv način (Dudley, 2012: 159–160). V grobem takšno razumevanje sveta sprejema tudi Aristotel. Platon je o naključju menil, da je božanskega izvora, čeprav je na nekaterih mestih ta izraz uporabljal v enakem pomenu kot kasneje Aristotel, torej kot vzrok brez božanskega pridiha (Dudley, 2012: 157–158). Pojem naključnega pred Aristotelom torej večinoma pomeni kontingentno, v primeru nastanka kozmosa pa hkrati tudi nujno. Stoiki so kasneje pojem naključja reducirali na status človeškega pomanjkljivega vedenja, tisto človeku nepoznano pa pripisali božji previdnosti. S tem so eliminirali nevarnost, ki jo je za doseganje človeške sreče predstavljal neracionalni pojem naključja v smislu kontingence.

Demokrit je imel za spoznanje bolj dodelano pojmovanje naključja. V njegovi teoriji atomizma je po Aristotelovem mnenju naključje pripisano zgolj in samo vrtinčenju atomov,[6] čeprav pri Demokritu to v resnici pomeni, da za vrtinčenje ni vzroka, saj njegov pojem naključja pomeni odsotnost vsakršnega vzroka (Dudley, 2012: 144–153). Pri Demokritu se nujnost in naključje tako kot kasneje pri Aristotelu že izključujeta, razlika med njima pa je v tem, da vrtinci nastanejo po naključju, delujejo pa po nujnosti. Drugače rečeno, vrtinci nastanejo brez vzroka, so nepovzročeni, delujejo pa po poti vzrokov. Tako je Demokrit naključje kot vzrok (čeprav ravno to ni) reduciral na primarni vzrok nastanka vesolja, dogajanje v njem pa se odvija po nujnosti. Približno takšno je tudi stališče moderne kozmološke znanosti.

Naključje kot sekundarni vzrok

Aristotel se z naključjem ukvarja kot del obravnave vzrokov in sicer predvsem v *Fiziki II*. Pojem naključja je pri njem ukoreninjen v teorijo vzrokov, ker pri tej analizi Aristotel za izhodišče refleksije vzame pogovorno rabo besede in pojmovanje modrih (tj. filozofov pred njim), ki so imeli naključje za *substancialni*[7] vzrok. Zanj je nasploh značilno, da kot izhodišče svojih refleksij večkrat vzame splošno mnenje in ga potem dialektično razdela.[8] Aristotel skozi razlago pojem naključja desubstantivira, saj ugotovi, da gre pri naključju za vzrok *dogodka* in ne substance. Ko govori o tem, da od stvari, ki nastanejo, nekatere nastanejo z namenom, druge pa ne (*Fizika II 5*, 196 b 17–18), je

[6] Ohranjen je tudi pristni fragment, ki omenja vrtince na začetku nastanka sveta: »Od vesolja se je ločil vrtinec vsakovrstnih oblik« (Simplicius, *Phys*. 327, 24, v: *Fragmenti predsokratikov*, str. 929). Simplicius o njih zapiše. da je Demokrit s to izjavo želel reči, da vrtinec nastane samodejno in naključno, torej brez vzroka.

[7] Izraz substanca Kalan prevaja kot bitnost. V pridevniški obliki se ustrezni izraz glasi bitnosten. Sam uporabljam v tekstu pridevniški izraz substancialen, kjer je govora o substancah pa uporabljam tudi samostalniški izraz bitnost.

[8] O tem pišejo tako Valentin Kalan v spremni študiji k Aristotelovi *Fiziki* in *Metafiziki* kot tudi John Dudley in M. R. Johnson v svojih monografijah.

potrebno izraz »stvari« razumeti v smislu dogodkov, pravi Dudley. Ta opredelitev je torej nekoliko bolj fenomenološko obarvana.

Aristotel večkrat pravi, da je o vzrokih mogoče govoriti na več načinov (npr. *Fizika II 3*, 195 a 29–31). Njegov najbolj znani način govora o vzrokih so seveda štirje vzroki (snovni, gibalni, oblikovni, smotrni), ki so v bistvu oblike izrekanja o nekem izvornem vzroku na četvero načinov. Aristotel na več mestih navede tri *počela* vzrokov, ki so narava, umetnost, naključje (*Metafizika Z 7*, 1032 a 12–13), na treh drugih mestih pa dodaja še srečo (*Metafizika Λ 3*, 1070 a 6–7; *Metafizika K 8*, 1065 b 3; *Fizika II 6*, 198 a 10), pri čemer na dveh mestih pove, da sta *týche* in *taùtómaton* ontološko sekundarna glede na naravo in um oziroma umetnost. Čeprav je naključje vzrok oziroma »vzrok po naključnosti« [*aítion katá symbebekós*] (*Fizika II 5*, 197 a 5–6, 33), pri tem ne gre za nobenega od Aristotelovih štirih *vrst* vzrokov, temveč za sekundarni tip vzroka, to je takšnega, ki ni vzrok substanc, temveč dogodkov. Naključje ni enostaven vzrok, torej ni substancialni vzrok, temveč je postavljeno *nasproti* le-tem (*Fizika II*, 196 b 26 in 198 a 9). Substancialna oziroma primarna vzroka sta lahko samo narava in um (*phýsis* in *nôus*), ta dva delujeta preko štirih znanih tipov vzrokov. Tako ne moremo imeti na primer naključnih gibalnih vzrokov ali naključnih smotrnih vzrokov. Aristotel to razloži s primerom gradnje hiše (*Fizika II 5*, 196 b 25–28): gibalni vzrok hiše je ta, ki zna graditi (graditelj), naključni vzrok hiše pa bel človek ali glasbenik, saj sta to slednje le naključni lastnosti primarnega vzroka. Naključni vzrok je lastnost kot pritiklina, je dodaten podrejeni, vendar pomembni vzrok, ki si zasluži svoj posebni status: »naključje ni za Aristotela niti substancialna realnost niti povsem subjektivni pojem, temveč ima pristni in pomenljivi status naključnega vzroka« (Dudley, 2012: 13).

Pri naključnem *dogodku* gre vedno za *sovpadanje dveh* vzrokov, od katerih je eden substancialen drugi pa naključen glede na prvega. Na pomemben način lahko opredelimo razliko med substancialnimi in naključnimi vzroki v njihovem odnosu do kategorij.[9] Substancialni vzroki povzročajo postopno spreminjanje in sicer glede na eno od naslednjih štirih kategorij: bitnost, kakšnost, kolikšnost in prostorsko relacijo; za spremembo po bitnosti se uporablja izraz *génesis*, za ostale tri vrste spreminjanja pa izraz *kínesis*. Naključni vzrok lahko po drugi strani pripada katerikoli kategoriji in preide v bivanje v *trenutku*, se tako rekoč *zgodi*, in sicer v substancialnem vzroku, zaradi česar naključni vzroki lahko sploh delujejo kot vzroki.[10] Substancialni vzrok je lahko odločitev nekoga ali pa narava

[9] Kategorije (*kategórema*) so Aristotelova določila bivajočega in predstavljajo pomemben osrednji del njegove misli. Kot navaja Reale, Aristotel bivajoče obravnava na štiri načine: bivajoče za sebe (to je kot bitnost, včasih pa tudi v smislu kategorij), bivajoče kot možnosti in udejanjenosti, bivajoče kot resnično in bivajoče v pomenu pridevka (naključno bivajoče). Pri tem Reale poudarja, da se vsi ostali načini razumevanja in opredeljevanja bivajočega naslanjajo na opredelitve bivajočega s kategorijami (Reale, 2002: 31–313).

[10] Reale piše o razmerju med opredeljevanjem bivajočega kot naključnega in bivajočega po sebi, tj. na način kategorij, takole: »pridevno bivajoče je *prigodni dogodek, ki se uresniči v skladu z različnimi (nujnimi) liki kategorij*« (Reale, 2002: 313). Pridevno oziroma naključno bivajoče je torej v odnosu do kategorij v smislu dogodkovne izpolnitve le-teh, pri čemer so kategorije nujne, pridevno bivajoče pa je slučajno. Spremembe po nekaterih od prej navedenih kategorij (bitnost, prostorska relacija, kolikost, kakšnost) so postopne in imajo v

nekoga. Vse omenjeno ponazarja Aristotelov primer nekoga, ki se odloči izkopati luknjo na vrtu, pri tem pa odkrije zaklad (*Metafizika Δ 30*, 1025 a 15–19). V tem primeru nas veriga nujnih vzrokov, ki je pripeljala do odkritja zaklada, privede samo do naključnega vzroka pojavitve zaklada v zemlji, ta se ni pojavil po nobeni nujnosti.

Še en primer, s katerim Aristotel ponazori razliko med substancialnim in naključnim vzrokom, je primer človeka, ki gre iz hiše do vodnjaka piti vodo, ker je poprej pojedel preveč začinjeno hrano, zunaj pa naleti na bandite, ki ga ubijejo; Aristotel ga omenja na več mestih (npr. *Metafizika E 3*, 1027 b 6–8). Po Aristotelu ni pravilno reči, da so vzrok smrti banditi, čeprav v vsakdanjem jeziku tako govorimo, saj je srečanje z banditi nepovzročena akcidenca oziroma naključni vzrok, ne pa substancialni. Na bandite je nesrečnež zgolj naletel, pojavili so se v nekem trenutku, lahko bi rekli na način sovpadanja. Tudi zaužitje pekoče hrane ni pravi vzrok za smrt, temveč zgolj naključen, čeprav je ta dogodek sprožil nujno vzročno verigo dogodkov, ki so privedli do smrti. Pravi, tj. substancialni vzrok je tukaj racionalna odločitev, ki je sledila zaužitju pekoče hrane: »Če se naknadno vprašamo, zakaj je človek *nujno* šel skozi različne faze, ki so pripeljale do njegove smrti, je odgovor ta, da je veriga gibalnih vzrokov podrejena smotrnemu vzroku, ki nalaga racionalne odločitve. Razlog, zakaj je človek nujno moral ravnati tako, kot je, je ta, da je človeški in se od njega pričakuje, da bo ravnal racionalno, to je tako, kot smotrni vzrok (njegovo dobro) od njega *zahteva*« (Dudley, 2012: 306). Tudi iz tega primera je razvidno, da je naključni vzrok v podrejenem statusu nasproti substancialnemu.

Opis naključnega dogodka z naključnimi vzroki pokaže tudi razliko med pojmoma *týche* in *taùtómaton*, ki smo jo že omenili. Substancialni vzrok vsega, kar je *taùtómaton*, je zunanji, medtem ko je pri *týche* substancialni vzrok notranji (namreč odločitev), tako kot pri naravi: »vzrok samodejnosti je namreč zunaj, medtem ko je vzrok narave znotraj« (*Fizika II 6*, 197 b 34). Aristotelov primer za *taùtómaton* je recimo padec trinožnika na tla na takšen način, da se lahko nekdo nanj usede (*Fizika II 6*, 197 b 16–17). Nezanemarljiv problem pri takšnem razumevanju pojma *taùtómaton* predstavljajo tako imenovane pošasti in spontana (nespolna) generacija organizmov, ker gre tukaj za substance, ki so po naravi, ta pa po Aristotelu deluje po notranji nujnosti, kot notranji vzrok (o tem kasneje). S tema dvema primeroma ima Aristotel nekaj težav, ker se ne ujemata dobro z njegovo definicijo naključja, ki se posredno ali neposredno navezuje pretežno na človeški svet odločitev in dejanj.

Povezava med naključjem in smotrom

Sprva Aristotel pojem naključja v *Fiziki* opredeli tako, da se nanaša samo na nenavadne dogodke, kasneje pa se osredotoči na povezavo med naključjem in

tem smislu substancialne vzroke, te kategorije so tudi pomembnejše za opredelitev bitnosti od ostalih.

smotrom (Dudley, 2012: 22), ko stvari deli na tiste, ki se zgodijo nujno (enostavno), tiste, ki se zgodijo večinoma, in tiste, ki se ne zgodijo niti nujno niti večinoma (*Fizika II 5*, 196 b 15–18; *O nastanku živali IV 4*, 770 b 9–10). V slednjem primeru gre za slučajne dogodke, ki nimajo substancialnih vzrokov, te imajo po njegovem mnenju samo prvi dve vrsti stvari, medtem ko imajo redki dogodki vedno le slučajne vzroke, ki le deležijo na substancialnih. Vendar kljub temu Aristotel pravi, da tudi redki dogodki niso nujno brez pomena, temveč jim ga včasih lahko pripišemo – ravno takšne dogodke poimenuje naključne. Naključni vzrok, kot smo ga opredelili v prejšnjem podpoglavju, ne pripelje vedno do naključnega dogodka, kajti da je naključni vzrok hkrati tudi vzrok naključnega dogodka, mo ta dogodek imeti nek pomen. Po Aristotelovo rečeno: ta dogodek mora pripadati območju stvari, ki so zavoljo nečesa: »Torej je jasno, da je slučajnost vzrok po naključnosti na področju stvari 'zaradi nečesa', ki so po izbiri« (*Fizika II*, 5 197 a 5–6). V tem smislu to, da glasbenik zgradi hišo, ni nobeno naključje, človek kot glasbenik je zgolj naključni vzrok zgraditve hiše; da bi tu šlo za naključje, bi ta dogodek moral imeti poseben pomen. Seveda pa je naključni dogodek vedno povzročen od naključnega vzroka.

Za opredelitev naključnega dogodka sta torej pomembna dva kriterija – redkost in pomenljivost dogodka. Povezava med naključjem in namenom je, da »naključje nima nobenega smotra, vendar pripada območju tistega, kar je zavoljo nečesa, to je tistega, kar bi *lahko* bilo narejeno zavoljo nečesa, vendar *ni* bilo« (Dudley, 2012: 25). Razlika med naključjem v človeškem svetu (torej naključjem v smislu sreče) in naključjem v naravnih dogajanjih je ta, da je pri prvem poleg pogojev o redkosti in pomenljivosti dogodka potrebno tudi, da bi dogodek načeloma lahko spadal v območje izbire in bi bil izbire vreden (ustrezen angleški izraz je *choiceworthiness*). V primeru ljudi je torej potreben še pogoj, da gre za nekaj, do česar imajo ljudje pozitiven ali negativen vrednostni odnos (Dudley, 2012: 47). Pojem *taùtómaton*, kot smo ga že opredelili, tako vedno vključuje *týche*, saj je širši. Izraz *taùtómaton* se lahko uporablja tako za področje narave (*phýsis*) kot mišljenja (nus), medtem ko se *týche* uporablja zgolj na območju mišljenja (Dudley, 2012: 57). Tako imajo (ne)srečo oziroma *týche* lahko le bitja, ki imajo možnost svobodne izbire – po Aristotelu v to skupino spadajo odrasle osebe, živali in otroci pa ne (*Fizika II 6*, 197 b 18–22). Če na primer mesečnik zadane z glavo v strop (kar je redkost, pomenljivo pa zato, ker se pri tem lahko zbudi), gre po Aristotelovem mnenju za *taùtómaton* in ne za *týche*, ker zadevanje z glavo v strop načeloma ne spada v območje odločanja. Pri tem je potrebno biti previden, saj če rečemo, da nek dogodek spada v območje odločanja, je s tem mišljeno racionalno odločanje, pri katerem za nas ugoden dogodek izberemo, neugodnega pa zavrnemo. Na ta način so seveda zajeti tako primeri, ko imamo srečo, kot primeri, ko imamo nesrečo – slednje bi ob možnosti izbire zavrnili.

Meja med *taùtómaton* in *týche* je tako po Aristotelu zmožnost odločanja za izpolnitev nekega dogodka. Menim, da je meja med obema območjema precej

zabrisana in premična, je stvar perspektive, saj je veliko stvari mogoče šele napraviti za predmet odločitve, čeprav v nekem poprejšnjem času to ni bilo možno, recimo v primeru tehnologije. Za najbolj uničujoč meteorski dež v pisani zgodovini, ki je padel na Zemljo leta 1490 n. š. in izbrisal celo mesto na Kitajskem,[11] bi rekli, da se je zgodil *taùtómaton*, samodejno in naključno ravno nad nekim naseljenim območjem, saj je šlo za dogodek, o katerem se niti načeloma ni mogoče odločati. Po drugi strani bo najbrž nekoč v prihodnosti ta dogodek spadal v območje *týche*, saj se tehnologija preusmerjanja nebesnih teles, ki bi lahko usodno trčila z Zemljo, intenzivno razvija. Če bi do trka le prišlo, bi lahko rekli, da je šlo za *týche*, torej nesrečo v tem smislu, da bi »izšolani preusmerjevalci nebesnih teles« lahko ukrepali drugače. Šlo bi za klasično nesrečo tipa prometne nesreče, kjer gre za načelno možnost drugačne vožnje od tiste, ki je do nesreče privedla. Meja med *týche* in *taùtómaton* torej temelji na podobni osnovi kot meja med naravo in človeško tehnologijo ali kulturo.

Izraz *taùtómaton* Aristotel primerja tudi z besedo *máten*, ki pomeni zaman (*Fizika II 6*, 197 b 18–34). Med njima napravi vzporejanje, recimo s primerom kamna, ki pade na nekoga in ga ubije (*Fizika II 6*, 197 b 30–32). To se lahko zgodi po naključju, brez namena, vendar bi se lahko zgodilo tudi z namenom, tako da bi nekdo vrgel kamen na nekoga; v tem primeru bi šlo za namen metalca kamna, ne pa za kak višji namen. Kar se zgodi po naključju, je za Aristotela samo po sebi *máten*, torej zaman, brez namena oziroma smotra, tega mu je mogoče podeliti le skozi »antropocentrična očala«.

Za Aristotela lahko imajo smoter samo substance in človeške odločitve, naključni dogodki sami po sebi pa ne. Substancialen vzrok za *týche* je, kot že omenjeno, človeška odločitev oziroma natančneje duša, kar ponazarja primer s človekom, ki gre do vodnjaka potešit žejo, a ga ubijejo razbojniki. Odločitev je dejavni vzrok, ki pa sam po sebi še ne *pojasni* naključnega dogodka. To je mogoče storiti šele tedaj, ko naključje uvrstimo v območje doseganja ciljev, kot je storil Aristotel. Naključni dogodki se torej ne zgodijo z namenom, drži pa, da imajo za človeka poseben *pomen* – to spada v njihovo definicijo, sicer teh dogodkov sploh ne opazimo. Za to, da imajo nekateri dogodki poseben pomen, je nujno potreben razum oziroma po Aristotelu duša ter njena zmožnost odločanja in refleksije. »Brez pričakovanja intelekta ni [substancialnega] vzroka (za izkustvo) sovpadanj [*coincidences*], ki se dogajajo naključno, ne da bi bila opažena« (Dudley, 2012:35). Naključni dogodek kot sovpadanje mora tako, ker sam po sebi nima pomena, pripadati območju človeških namenov, ciljev. Lahko bi rekli tudi, da je vzrok za doživljanje (ne)srečnih naključij človeška dejavnost postavljanja ciljev, odločanje pa je seveda vezano nanjo. V Aristotelovem jeziku gre tukaj za dva vidika istega dogajanja, ki ga enkrat gledamo skozi gibalni vzrok (odločitev),

[11] Podatek najdemo na primer v knjigi Johna S. Lewisa *Comet and Asteroid Impact Hazards on a Populated Earth*, str. 12 (Academic Press, San Diego, 2000). V njej avtor piše tudi o začetkih preučevanja in popisovanja nebesnih teles z namenom preusmerjanja objektov, ki bi lahko povzročili usodne trke z Zemljo.

drugič pa skozi smotrni vzrok (zadani cilj).

Veliko razlagalcev Aristotela je spregledalo, da naključnim dogodkom samim zase ne pripisuje posebnega pomena, piše Dudley. Ravno tako je veliko njegovih tolmačev izpodbijalo, da naj bi Aristotel upošteval tako srečna kot nesrečna naključja, temveč naj bi razlagal samo prva, vendar to ne drži. Nesporazum je bil ravno v tem, da je bilo mišljeno, da nesrečen dogodek ne more spadati v območje namenov, saj nobeden ne bi izbral nečesa, kar je zanj neugodno, vendar na osnovni ravni ne gre za to, ali bi človek izbral takšno neugodno možnost, temveč za to da tudi nesrečen dogodek predstavlja pomenljiv dogodek v odnosu do doseganja nekega cilja. S tem se uvršča na področje namenov – nesrečen dogodek je namreč takšen redek dogodek, ki preprečuje ali povsem onemogoči nek namen in je torej v negativni povezavi z doseganjem cilja. V primeru možnosti odločanja bi takšen dogodek bil vreden zavrnitve, ne pa nevtralne drže.

Eminentno področje doseganja namenov oziroma smotrov je *téchne*, to je področje veščin in umetnosti. Naključje in veščina se medsebojno izključujeta, saj pri veščini dosegamo nek smoter, namreč kvaliteten izdelek. To, kar človek običajno ustvari z umetnostjo, se redko zgodi samo od sebe po naravni poti, torej po naključju (recimo da se skala oblikuje tako, da ima približno obliko stola). V naravnem dogajanju le *redko* nastane to, kar umetnost proizvaja *običajno* (na to nasprotje v regularnosti med človeškim svetom in naravnimi dogajanji bomo v nadaljevanju še večkrat opozorili). Po drugi strani pa se tudi pri človeških izdelkih znotraj neke obrti kdaj reče, »da se je nek izdelek posebej posrečil«, kar je spet dojeto kot redek dogodek, ne kot pravilo, ki ga v tem primeru predstavlja neka veščina z osrednjim smotrom izdelave.

Med veščinami in odločanjem obstaja določena vzporednica, ki jo vidi tudi Aristotel, ko pravi: »Razred stvari, ki dopuščajo variacijo [v okviru nekega smotra] vključuje tako izdelane stvari (*poietón*) kot izvedena dejanja (*praktón*)« (*Evdemova etika*, 4, 1140 a 1–2; citat preveden po Dudley, 2012: 41). S tem je mišljeno, da nek smoter, tako v primeru odločanja kot v primeru izdelovanja artefaktov, lahko dosežemo na več načinov, med katerimi izbiramo. Svobodna volja je v primeru tehnike implicitna veščini, zaradi doseganja določenega namena je omejena, saj je podrejena doseganju določene *pravilnosti* oziroma *regularnosti*. Ko neki izdelovalec uspe doseči določeno želeno stopnjo regularnosti, se reče, da je izkušen.

Aristotel razpravlja o primeru ozdravitve bolnika, kateri botruje toplota (*Metafizika Z 7*, 1032 b 21–26). Ozdravitev se lahko zgodi zaradi veščine, lahko pa tudi po naključju. V prvem primeru posreduje zdravnik z veščino zdravljenja, v drugem primeru pa gre lahko za to, da je neki prijatelj bolniku svetoval, naj se greje ali pa se je enostavno zgodilo, da se je vreme segrelo in s tem povzročilo ozdravitev. V zadnjih dveh primerih gre za naključje, le da gre v prvem od njiju po Aristotelovo za *týche*, v drugem pa za *taùtómaton*. Kot smo že omenili,

imamo pri naključnem dogodku lahko neskončno možnih naključnih vzrokov, ti so po obliki lahko gibalni ali snovni, vendar le omejeno število oblikovnih oziroma smotrnih vzrokov. Če v navedenem primeru rečemo, da je vzrok ozdravitve toplota, smo s tem navedli le gibalni vzrok, do katerega je lahko prišlo na zelo mnogo mogočih načinov (po poti več možnih kavzalnih verig). Če opredelimo še oblikovni vzrok pa bolje *razumemo*, kako je do ozdravitve prišlo. Pri ozdravitvi zaradi zdravnika gre tako poleg toplote kot gibalnega vzroka še za oblikovni vzrok, namreč za idejo zdravja, pravi Aristotel.[12] Če pa je bolnik pozdravljen po naključju, gre samo za gibalni vzrok, medtem ko oblikovni manjka. V primeru dobrohotnega prijatelja gre za *týche*, v smislu, da je bil nasvet posrečen, ker je privedel do ozdravitve, vendar je še vedno šlo za naključen vzrok. V primeru otoplitve vremena pa gre za *taùtómaton*, saj otoplitev vremena ne more spadati na področje človeškega odločanja.[13]

Nujnost in naključje

Pojem nujnosti (*ànánke*) je po mnenju mnogih raziskovalcev pri Aristotelu eden osrednjih. Omenili smo že, da je Aristotel prvi mislec, za katerega je nujnost popolno nasprotje naključja, zato je za boljše razumevanje njegovega pojma naključja potrebno preučiti tudi pojem nujnosti. Kot sam pravi, naključni vzroki in počela očitno morajo obstajati, sicer bi bile vse stvari iz nujnosti (*Metafizika E 2*, 1027 a 11–12).

Za Aristotela je nujno »to, kar ne more ne biti« (*Metafizika Γ 4*, 1006 b 32; *Metafizika Δ 5*, 1015 a 33–36). Kljub tej tako kategorični opredelitvi, ki jo Dudley ponovno tolmači zgolj kot izhodišče za refleksijo, pa Aristotel v mnogih drugih tekstih[14] napravi pomembno distinkcijo med absolutno in hipotetično oziroma pogojno nujnostjo (*tò d' áplôs ánagkaîon* in *tà d' eks ýpothéseos ánagkaîon*), s čimer uvede tako rekoč nižjo stopnjo nujnosti. Prvi tip nujnosti najdemo na področju večnega, recimo v matematiki ali pri nebesnih telesih. Tako matematični objekti, ki imajo nujne lastnosti (npr. trikotnik, pri katerem je vsota notranjih kotov nujno enaka 180°), kot nebesna telesa in njihova gibanja obstajajo absolutno nujno.

V sublunarnem svetu, kjer so štirje elementi (za razliko od supralunarnega, kjer je samo eter), pa je večno samo spreminjanje in zato je to edino, kar je tam absolutno nujno. Ta nujnost spreminjanja se pri Aristotelu, kar zadeva živa bitja,

[12] Pri tem ni odveč poudariti, da Aristotel idej ne pojmuje v Platonovem smislu kot splošnih in neodvisnih bitnosti, temveč ideje in oblike nasploh (*eidos*) pojmuje kot specifične v nasprotju s splošnimi ter kot vezane na konkretne pojave (na snov) v nasprotju z neodvisnimi (pri tem so izjema le povsem netelesne bitnosti. ki pa ne bivajo na Zemlji). O tem več Reale, 2002: 304–307.

[13] Pri tem izpuščam primer, ko človek lahko vpliva na naravo v tolikšni meri, da lahko lokalno spreminja fizikalne pogoje (npr. hišno ogrevanje). To vprašanje po mojem mnenju spada v področje človeške kulturalizacije oziroma tehnizacije sveta.

[14] Zelo pomenljivo se zdi, da Aristotel pojem pogojne nujnosti opredeli v glavnem v delih, ki se nanašajo na biološki svet: *O delih živali* (1, 1, 639 b 23–26; IV, 2, 677 a 15–19), O *nastanku živali* (V, 3, 782 a 22–24; IV, 8, 776 b 31–33), *O snu in bdenju* (455 b 26–28); mesta so navedena po Dudleyju.

nanaša samo na vrste, ne pa na posamezna bitja, podobno kot je razmišljal Empedokles. Absolutna nujnost velja tudi za elemente, ki se gibljejo glede na svojo naravo; Aristotel govori s tem v zvezi o materialni nujnosti. Tu je mišljena recimo nujnost padanja kamna, če mu tega kaj ne preprečuje; kamen namreč sestavlja osnovni element zemlje, ki teži k svojemu naravnemu mestu. Nujnost padanja kamna je iste vrste kot nujnost kroženja nebesnih teles (s stališča moderne znanosti oboje povzroča gravitacija). Po drugi strani nujnost medsebojnega pretvarjanja elementov, na primer zraka v vodo, kot se dogaja pri nastanku dežja (in obratno pri izhlapevanju), ni iste vrste kot nujnost krožnega gibanja nebesnih teles, temveč je tej samo podobna, je njeno posnemanje. Stvari, ki obstajajo po absolutni nujnosti, obstajajo vedno na isti način (za razliko od stvari, ki obstajajo večinoma ali pa niti vedno niti večinoma). Vidimo lahko, da se pojma absolutne in pogojne nujnosti razlikujeta tudi po regularnosti dogajanja, saj pogojna nujnost povzema ravno to, da se dogajanja na Zemlji nekoliko pogosteje ne izidejo po pričakovanjih, ker lahko kaj poseže vmes. To razlikovanje bi iz modernega stališča lahko opredelili na statističen način, z večjo ali manjšo pogostostjo nekega pričakovanega dogodka.

Hipotetična oziroma pogojna nujnost ni vrsta nujnosti, ki bi temeljila na naravi snovi, temveč je to nujnost *sredstev*, ko je enkrat dan smotrni ali oblikovni vzrok (Dudley, 2012: 108). Ta tip nujnosti se nanaša na vse naravne substance, ki so v procesu postajanja, in tudi na človeške artefakte, ki so v procesu izdelave. To ne pomeni, da pogojna nujnost nima nikakršne veze s snovjo, ki je del nekega živega bitja ali izdelka. Pri pogojni nujnosti v bistvu govorimo o nujnosti dogajanja, do katerega pride tudi zaradi narave snovi, iz katere je neko bitje oziroma izdelek, vendar le ob pogoju, da je dan tudi določen oblikovni ali smotrni vzrok.

Če nekoliko obrnemo perspektivo in o nujnosti govorimo s stališča štirih vzrokov, lahko rečemo, da po absolutni nujnosti delujeta samo snovni in gibalni vzrok, medtem ko oblikovni in smotrni vzrok delujeta le po pogojni nujnosti (O *nastanku živali II* 1, 731 b 20–24). To je tako, ker smotrni vzroki ne privedejo vedno do izpolnitve smotra. Zdravljenje na primer ne uspe vedno, živa bitja pa se kdaj razvijejo v popačene primerke svoje vrste. Ta in podobna dogajanja kažejo na to, da gre pri smotrnih vzrokih za stvari, ki nastanejo večinoma, nikakor ne pa ne nujno (če pri tem izvzamemo nebesne sfere, kjer se dogajanje odvija po absolutni nujnosti, tam pa vlada eter). Stvarem, ki so večinoma na nek način, torej pritiče razlaga s pogojno nujnostjo. Med začetno snovjo, ki ima pri Aristotelu vlogo potencialnosti, v kateri se lahko razvije določena oblika, in smotrnim vzrokom v takšnih primerih ni logično nujne povezave, vseeno pa je mogoče sestaviti silogizem, ki vključuje smoter, sredstvo in prejemnika koristi, ne pa tudi snovi.[15]

[15] O tem piše M. R. Johnson v svoji monografiji *Aristotle on Teleology*, str. 52–55. V zvezi s teleološkimi vprašanji se je veliko kritikov ukvarjalo s tem, da bi sestavili silogizme v zvezi s smotrnimi vzroki, ki bi

Aristotel je pri predhodnikih v zvezi s pojmovanjem nujnosti kritiziral rabo pojma absolutne nujnosti, ne tudi pogojne nujnosti (Dudley, 2012: 135–153). Narava se v osnovi spreminja oziroma razvija po pogojni nujnosti: ob danem znanem smotru, ki deluje kot nekakšno negibno gibalo,[16] spremembe narave sledijo po pogojni nujnosti. Za razliko od živih bitij se elementi, kot smo že omenili, gibljejo po absolutni nujnosti, saj Aristotel zanje ne poda smotrnega vzroka.[17] Elementi se gibljejo po notranjem vzroku, tako kot naravne stvari, vendar se za razliko od njih ne morejo sami ustavljati, lahko jih ustavi le nekaj drugega. Po Dudleyju to lahko razložimo le s hilozoizmom, to je s stališčem, da snov vsebuje *princip* duše.

Razlaga narave in oblik, ki jih v njej najdemo, je tista, ki poleg snovnega in gibalnega vzroka zahteva še oblikovni in smotrni vzrok. Pri tem so to za Aristotela komplementarni in ne izključujoči načini razlage. V okviru iste razlage dogajanja Aristotel lahko poda na primer tako snovni kot smotrni vzrok, recimo ko razlaga, zakaj nekaterim rogatim živalim odpadejo rogovi (*O delih živali III 2*, 663 b 12–24). Ravno tako sta komplementarna pojma absolutne in pogojne nujnosti, ki tudi lahko nastopata v isti razlagi. To drugače rečeno pomeni, da so pri istem naravnem pojavu nekateri vidiki tega pojava bolj nujni kot drugi, pri čemer za oboje želimo razlago.

Aristotel s pomočjo pojma nujnosti razloži tudi, od kod izvor naključja. Stvari, ki bivajo po nujnosti, »ne morejo biti na drug način, nekatere bivajoče stvari pa ne bivajo iz nujnosti niti vedno, temveč večinoma: to je počelo in to je vzrok biti naključja; kar namreč pač biva, a ni niti vedno niti večinoma, to pravimo, da je pripetljaj« (*Metafizika E 2*, 1026 b 29). Prostor za naključje med bivajočimi stvarmi torej obstaja, izvor ima v dejstvu, da se nekatere stvari dogajajo zgolj večinoma in ne vedno, vezano je torej na dokaj veliko mero regularnosti, ki pa nujno pozna izjeme. Substancialni vzrok za to je snov: »torej bo snov (*hýle*), na način, kakor poleg dogajanja večinoma omogoča še drugačno dogajanje, vzrok naključja« (*Metafizika E 2*, 1027 a 13). Pri Aristotelu je snov tista, ki lahko prepreči uresničenje smotra, saj deluje po absolutni nujnosti, nujnost pa je lahko

podkrepili razumevanje Aristotelovega smotrnega vzroka kot racionalno utemeljenega. Aristotel sam je namreč, kljub temu, da je pravzaprav zasnoval metodo silogističnega mišljenja, tega presenetljivo redko uporabljal v svojih delih. Kot ugotavlja tudi Hankinson, Aristotel v svojih naravoslovnih delih pogosto samo s pogostostjo pojavljanja utemeljuje, zakaj je nekaj takšno in takšno. To je tudi dovolj za priklic smotrnih vzrokov. »Pogosto te [njegove formulacije] niso niti univerzalne, kaj šele nujne: večina, vendar ne vsi moški imajo brade *[Posteriorna analitika II 12, 96* a 8–11]. V resnici se vsi naravni dogodki (za razliko od slučajnih) zgodijo bodisi vedno bodisi večinoma, kar jih po Aristotelu dela dovzetne za teleološke oziroma smotrne razlage. Vendar medtem ko so stvari v supralunarnem svetu nespremenljive, stalne in absolutno nujne, tiste v sublunarnem svetu, ki je domena aristotelske fizike, niso; tukaj splošnosti držijo le v večini primerov« (Hankinson, 1995: 113).

[16] Tukaj ni mišljeno Prvo Gibalo oziroma Negibno Gibalo, ampak naravne substance, torej živali in rastline, ki so posnemanja Negibnega Gibala. M. R. Johnson (Johnson, 2005: 249) piše, da so za Aristotela dokaz za to, da lahko neka substanca začne svoje lastno gibanje, ravno živali in ljudje, in sicer zaradi svoje duše.

[17] Johnson (Johnson, 2005: 13–136) temu oporeka in trdi, da je cilj elementov gibanje k njihovemu naravnemu mestu in da je njihovo premikanje mogoče razlagati teleološko. Tudi če se do neke mere lahko strinjamo s slednjim, pa Johnson sam nakaže, da so elementi precej specifičen tip substanc v primerjavi recimo z nebesnimi telesi, ki so iz etra (in so za Aristotela tudi živa bitja), ali z zemeljskimi bitji.

tudi prisila (*Metafizika Δ 5*, 1015 a 17–24).

To je potrditev, da so naključni dogodki na nek način povezani s snovjo, ki sicer deluje po absolutni nujnosti. Kontingentne lastnosti pri organizmih, kot je recimo barva oči, nastopajo po absolutni in ne pogojni nujnosti. Tisti vidiki snovnega vzroka, ki ne prispevajo k oblikovnemu ali smotrnemu vzroku, so naključni, ve dar to ne pomeni, da je *vzrok zanje* naključen, temveč deluje po absolutni nujnosti. Rečemo lahko, da je naključni *vzrok* v resnici vedno materialni ali pa dejavni vzrok, ki je absolutno nujen, kar pa je pri njem posebnega je to, da sopovzroči nek naključni *dogodek*, zaradi česar šele *retrospektivno* dobi status naključnega vzroka.

Pojem pogojne nujnosti nam pomaga dobiti dodaten uvid ravno v povezavo med naključnim vzrokom in preostalimi štirimi substancialnimi vzroki, ker vzpostavlja polje vzročne obravnave *pogostih* pojavov, s čimer pa potihem vključuje tudi *redke*. Po mojem mnenju je pogojna nujnost tisti pojem, ki najbolje opisuje odnose med snovnim in gibalnim vzrokom ter oblikovnim in smotrnim vzrokom. Snovni in dejavni vzrok sta s tem pojmom v svojih neštetih možnih delovanjih na nek način zamejena z verjetnim *izidom*, ki je v smislu razlage njun predpogoj, saj je izid ravno to, kar je potrebno razložiti.

Spački in spontana generacija

Povedali smo, da imamo pri delovanju narave ali uma opravka z oblikovnim in smotrnim vzrokom, ki sta povezana z gibalnim in snovnim vzrokom na pogojen način. Snov bi se lahko obnašala tudi drugače, vendar se vede tako, da doseže nek naravni ali umski smoter. V primeru naključnega dogodka, ko se zgodi nekaj, kar se običajno dogaja po veščini (recimo kamen v obliki preprostega stola), je odsotnost oblikovnega vzroka neproblematična (karikirano: narava nima v mislih stola), saj lahko to proizvede tudi »slep« gibalni vzrok. Tak primer je tudi prej omenjena ozdravitev, za katero je potrebna toplota, ta pa po naključju pride naravnost od sonca (lahko bi prišla od kod drugod). Ta način razlage s substancialnimi vzroki za Aristotela predstavlja problem pri stvareh, ki nastajajo po naravi, saj narava zanj vedno že pomeni notranji princip, ki deluje z nekim smotrom. Takšna sta že večkrat omenjena primera spačkov ter spontane generacije organizmov.

O spačkih Aristotel najpodrobneje piše v *O nastanku živali IV 3–4*. Sem uvršča potomce, ki niso podobni staršem, dvojčke pri živalih, ki imajo običajno samo po enega potomca, živali z manjkajočim ali dodatnim udom ali udom nepravilne oblike, tumorje in podobno. Njegova razlaga spačkov se giblje znotraj področja smotrnih vzrokov, spački so spodleteli primeri udejanjanja naravnih smotrov. Z vidika nujnosti pri spačkih odpove nujnost smotrnega vzroka, saj niso nujni po motni, nastanejo zgolj po poti »slepe« absolutne nujnosti snovnega vzroka. S em ni mišljeno, da so spački absolutno nujni, temveč da se pri njih snov vseeno vede

po neki nujnosti, kar pa ne privede do običajnega izida, ker nekaj to preprečuje. Lahko bi rekli, da pogoji za doseganje smotra v tem primeru niso izpolnjeni, čeprav je smoter obstajal. Vzrok za spačke Aristotel deloma pripisuje semenu, ki je gibalni vzrok in ga ima za moški princip, deloma jajčecu, ki je snovni vzrok in ga ima za ženski princip.[18] Vsekakor oba vzroka delujeta po absolutni nujnosti. V primeru spačkov naj bi šlo za to, da gibalnemu vzroku ne uspe ukrotiti snovnega vzroka (O *nastanku živali IV 3*, 767 b 13–15).

Kot smo že omenili, stvari, ki so *večinoma* na nek način, niso nujne zato, ker jim to preprečuje snov, ki deluje po absolutni nujnosti. Snov je tisti krivec, da oblikovni vzrok ne uspe vedno doseči cilja, ki je postavljen po naravi; le v supralunarnem imamo absolutno nujnost, ker tam ni snovi, temveč samo eter. Primer spačkov je z vidika naključja tako zanimiv, ker je naključje v naravnih stvareh (*taùtómaton*) povzročeno od zunanjega vzroka, pri spačkih pa gre za naravne substance, kar pomeni, da je na delu narava kot *notranji* princip, ki po Aristotelu mora vsebovati smotrni vzrok. Dudley (Dudley, 2012: 126) spačke kategorizira kot naključne *substance* za razliko od naključnih *dogodkov*, kamor Aristotel sicer uvršča naključje. Pri spačkih gre torej za pristno naključje na ravni bitnosti, kar je pri Aristotelu nekaj nepričakovanega.

Če imajo spački načeloma oblikovni vzrok in pri njih zato ni takšnega problema, predstavlja Aristotelu pravi izziv za razlago mehanizem tako imenovane spontane generacije organizmov (*génesis aùtómatos*). Ta pojav najbolje opredeli v *O nastanku živali III*. Pod tem pojmom razume to, da se nekatera živa bitja množijo kar brez semena, torej nespolno. Glavni problem pri tem je, kako razložiti spontano generacijo s teorijo štirih vzrokov, saj mora tu oblikovni vzrok iskati v sami snovi, kar je zanj problematična situacija. Po teoriji vzrokov ta pojav ni naraven, po izkustveni plati pa je premalo redek, da bi ga imeli za naključje.

Dudley tudi spontano generacijo organizmov uvrsti med naključne substance. Splošna razlaga njihovega nastanka, ki jo poda Dudley, je, da nastanejo, ko je oblikovni vzrok »pretvorjen« v gibalnega in konkurira s kakšno nenavadno kategorijo, rezultat pa ni dosežen cilj kot smotrni vzrok, čeprav bi načeloma lahko bil. Aristotelovo teorijo spontane generacije je po njegovem mnenju mogoče razumeti vzporedno s teorijo produkcije. Glavna razlika med njima je, da pri spontani generaciji ni oblikovnega vzroka, ki ga mora proizvajalec imeti v mislih, pri obeh pa imamo gibalni vzrok (pri spontani generaciji naj bi to bila svetloba, kije pove kasnejšega oblikovnega vzroka). Ko so spontani organizmi ustvarjeni, so d narave, čeprav niso perfektni v tem smislu, da niso podobni staršem, zato so vsi deloma nenaravni, saj je bistvo naravne produkcije, da iz enakega nastane enako. Vzrok spontane generacije je pri Aristotelu tako v osnovi hilozoizem, to je prepričanje, da je snov prežeta s principom duše. To lahko zaznamo na primer v naslednjem odlomku: »naravne stvaritve, ki morejo nastati

[18] Dudley (Dudley, 2012: 170–172) sicer navaja, da se je Aristotelova razlaga nastanka spačkov sicer nekoliko spreminjala, to je domnevno zadnja različica, pri kateri je obstal.

po samodejnosti, prav kakor v umetnosti, so tiste, katerih snov se more sama od sebe gibati po tistem [tj. takšnem] gibanju, ki ga povzroča seme, medtem ko bitja, ki take snovi nimajo, ne morejo nastajati kako drugače kakor iz samih roditeljev« (*Metafizika Z 9*, 1034 b 5). Gre torej za takšno snov, ki se zmore spreminjati na način, ki ga povzroča seme, kar pomeni tako, kot se razvijajo zarodki.

Skupno obema primeroma je to, da gre tako pri spačkih kot pri spontani generaciji za nekaj, kar doživi (notranjo) intervencijo snovi in se ne razvija izključno po naravi. Ena glavnih razlik med spački in spontano generacijo ter *taùtómaton* je ta, da prvo dvoje v nekem smislu le nastane po naravi, torej iz *notranjega* vzgiba, natančneje s prevlado snovnega vzroka nad smotrnim, medtem ko pojem *taùtómaton* vključuje *zunanjo* intervencijo snovnega vzroka. Ni pa nepomembna tudi naslednja razlika: Aristotel ne more ravnodušno opraviti s spački in spontano generacijo ravno zato, ker ta dva pojava izkazujeta neko (čeprav majhno ali zelo majhno) pogostost pojavljanja in se ravno po tem pomembno razlikujeta od pojma naključnega dogodka, ki je resnično redek, predvsem pa se ne pripeti na živem bitju samem. Ta dvoumni položaj naključnih substanc Aristotelu ne da miru. Zanje nekje tudi pravi, da se zgodijo v nasprotju s tem, kar je v naravi običajno, ne pa tudi v nasprotju z naravo v celoti, pošastno pa te bitnosti delujejo zato, ker so v nasprotju s splošnim pravilom (*O nastanku živali IV 4*, 770 b 9–11; 772 a 36–37). Aristotel pravi, da se v tem primeru nekaj zgodi v nasprotju z nekim partikularnim redom in da gre za naravno deformacijo, *ànapéria phýsike* (*O nastanku živali IV 4*, 770 b 9–11).

V *O nastanku živali* Aristotel pravi, da se spački ne pojavljajo naključno, glede istega vprašanja pri spontanem generaciji organizmov pa se ne opredeli, dejstvo pa je, da jo je klasificiral kot *taùtómaton (Fizika II, 6*, 197 b 32–37). Dudley Aristotelovo težavo opredeli takole:

> »[S]pačke in spontano generirana bitja, ki so naključne substance, lahko le s težavo uskladimo z razlago naključnih dogodkov v *Fiziki II, 4–6*. Težave so v tem, da spački in spontana generacija niso tako zelo neobičajni ali spontani kot naključni dogodki. Neobičajna kategorija v primeru spačkov in tako način produkcije oblikovnega vzroka kot neobičajna kategorija v primeru spontane generacije se ravno tako ne ujemajo dobro s teorijo naključnega dogodka« (Dudley, 2012: 373).

Nezmožnost znanstvenega preučevanja naključja

Kot smo že povedali, je za naključne dogodke značilna njihova pomenljivost. Pri tem pogoju se tudi pri naključnih dogodkih, ki se dogajajo v naravi, torej *taùtómaton*. tako rekoč ne moremo izogniti antropocentričnemu elementu, čeprav ne gre za področje človeškega odločanja. Po razpravljanju o naravi in izvoru naključja v *Metafiziki E 3* Aristotel tako za bivajoče na način resnice kot za naključno bivajoče pravi, da ne obstajata zunaj mišljenja:

> »[Kjer pa sta spletanje in ločevanje v razumu, ne pa v stvareh, je na ta način bivajoče drugačno od na odločilen način bivajočih stvari /.../; vzrok naključja je namreč nedoločen, medtem ko je vzrok drugega neko stanje mišljenja, tako da se oba nanašata na preostali rod bivajočega in ne razodevata neke narave bivajočega, bivajočo zunaj [mišljenja] – zaradi tega naj ju torej pustimo ob strani, treba pa je raziskovati vzroke in počela bivajočega samega, kolikor je bivajoče« (*Metafizika E 4*, 1027 b 30).

Ker meni, da je to tako, obe področji izvzame iz nadaljnjega raziskovanja bitnosti, nadalje raziskuje samo še bivajoče na način potencialnosti in aktualnosti ter bivajoče kot takšno, opredeljeno s kategorijami in podvrženo substancialni vzročnosti. Tudi Reale (Reale, 2002: 308) piše, da imajo ostali trije načini bivajočega svoje ustreznice tudi v nebivajočem, le nebivajočega po pridevku (slučajnega nebivajočega) Aristotel ne pozna, ker je zanj pridevno (slučajno) bivajoče »nekaj, kar je zelo blizu nebivajočemu« (*Metafizika*, E 2, 1026b21). Zato se zdi smiselno trditi. da ima naključje zanj zelo šibek ontološki status oziroma ima bolj epistemološki status.

O bivajočem po naključju ni nobenega spregledovanja ali ugotovitev, »nobeni znanosti namreč naključje samo ni na skrbi (*èpimelés*), niti delovanjski, niti proizvajalni niti pozorno premišljevalni« (*Metafizika E 2*, 1026b 5–6). O naključno bivajočih stvareh oziroma o naključju po samem sebi po Aristotelu sploh ne obstaja nobena umetnost oziroma veščina (*Metafizika E 2*, 1027 a 7–9). Povsem izrecno tudi zatrdi, da znanosti o neobičajnih dogodkih ni, pri čemer seveda misli na znanost z demonstracijo: »Ne more biti znanosti (*èpistéme*), ki uporablja demonstracijo (*àpódeiksis*) o tem, kar se dogaja po naključju *(týche)« (Posteriorna analitika 1* 30, 87 b 19–27).

V *Metafiziki E 6* pri obravnavi neobičajnih dogodkov Aristotel razloži, zakaj o njih ne more biti znanosti – namreč deduktivne znanosti.[19] Znanost se nanaša samo na nespremenljivo in običajno, le pri teh pojavih lahko pridemo do vednosti (*Metafizika E 2*, 1027 a 20–24). Pri takšnih pojavih človek lahko predvidi substancialne vzroke, njihovo vzročno verigo. Za naključne dogodke trdi, da to ni stvar naše nevednosti, ampak da so inherentno nepredvidljivi.[20] Za dan naključni dogodek na področju človeškega delovanja lahko imamo več primarnih vzrokov, tako rekoč nedoločeno število le-teh. Eden od razlogov za inherentno nepredvidljivost naključnih dogodkov je svoboda izbire primarnega vzroka, ki je rezultat mišljenja. Drugi razlog je to, da imajo različni dogodki različno težo za posameznega svobodnega individuuma, kar je povezano z razmišljanjem o izidu

[19] O tem razpravlja tudi v *Metafiziki K 8*, 1064 b 32 – 1065 a 6 in *Metafiziki Δ 30*. S tem seveda ne trdimo, da naključni dogodki ne obstajajo na pristen način, vendar je ta način za Aristotela nekako v nasprotju s pojmom bitnosti, zaradi česar je imel tudi prej omenjene težave z naključnimi substancami, kot jih je poimenoval Dudley.

[20] Podobna diskusija se je vnela konec 19. in v začetku 20. stoletja v zvezi s statističnimi zakoni. Klasično pojmovanje, ki se je ujemalo z determinizmom, je bilo, da je statistika odraz našega pomanjkljivega vedenja o pojavih. Kvantna mehanika je kasneje v znanost uvedla stališče, da so dogodki na atomskem nivoju dejansko inherentno nepredvidljivi, kar je formalno povzelo Heisenbergovo načelo nedoločenosti.

izbire. Tretji razlog je ta, da gre pri naključnih dogodkih po definiciji za neobičajne, redke dogodke, aristotelska znanost pa se ukvarja z dogodki, ki se dogajajo vedno ali večinoma. Kar smo povedali, velja za (ne)srečne dogodke (*týche*), za naključne dogodke (*taùtómaton*) pa je največji problem zgolj v njihovi redkosti.

Tudi v odnosu do vprašanja morebitne znanosti o naključju Dudley izpostavi problem spačkov in spontane generacije: »Položaj je takšen, da Aristotel pojavljanja spačkov in spontane generacije ni zvedel na znanost, kljub temu da hkrati več ne ustrezajo poljubnosti naključja (če so v resnici sploh kdaj). Regularnost pojavljanja [neke vrste] */…/ je* ravno tisti argument, ki ga je Aristotel uporabil za zavrnitev neteleološke razlage biološkega razvoja, kot ga podajata Empedokles in Demokrit« (Dudley, 2012: 194).

Aristotelov pojem sreče

S pojmom naključja je po Aristotelovo do neke mere povezan tudi pojem sreče kot duševnega blagostanja. Pojem naključja v njegovih etičnih delih se jasno povezuje z vprašanjem človeške sreče. Na podlagi splošno razširjenega mnenja Aristotel želi oblikovati stališče glede tega, do kolikšne mere je naključje v smislu kontingence izvor sreče. O tem najobsežneje razpravlja v *Nikomahovi etiki*, *Evdemovi etiki* in *Maximi moralii*. Izvora sreče kot duševnega blagostanja sta za Aristotela kontemplacija in praktična modrost oziroma etično udejstvovanje, pri čemer je prvo večji izvor sreče kot drugo. V tem najbolj splošnem smislu človeška sreča za Aristotela ni stvar naključja. Vendar ta splošni odgovor še ne vključuje razdelave odnosa med srečo in naključjem. Za to je potrebno podrobneje razdelati Aristotelovo teorijo sreče.

Aristotel loči *izvore* sreče, ki so locirani v človeku, od *zunanjih pogojev* sreče. Poleg tega Dudley izpostavlja, da pri njem obstajajo tudi implicitne stopnje sreče (Dudley, 2012: 219–220) in da je sreča, o kateri se Aristotel večinoma sprašuje, *absolutna* sreča. Za to, da bi dosegli absolutno srečo, so sicer potrebni tako izvori sreče kot tudi izpolnjeni zunanji pogoji, tako da človek, ki ne živi v srečnih okoliščinah, ne more biti absolutno srečen (*Nikomahova etika I 8*, 1099 a 31–32; *Maxima moralia II 8*, 1206 b 30–34). Ravno pri zunanjih pogojih oziroma okoliščinah nastopi pojem naključja. Vzrok za zunanje okoliščine je po Aristotelu namreč naključje in sicer v takšnem pomenu, kot ga v njegovih fizikalnih spisih ne najdemo, namreč v predsokratskem pomenu kontingentno danega (Dudley, 2012: 203, 12. opomba). To je po njegovo tako, ker o okoliščinah, v katere se rodimo, ne moremo odločati.

Kljub temu Aristotel priznava, da obstajajo ljudje, za katere se dozdeva, da imajo pogosteje srečo kot drugi (npr. *Evdemova etika VIII 2*, 1247 a 11). V tem primeru je to mišljeno v smislu srečnih dogodkov, ne v smislu duševnega blagostanja, čeprav gre ravno za to, da prvo s svojo pogostostjo vpliva tudi na drugo. Za take

ljudi, pri katerih se izkazuje določena mera pogostosti srečnih dogodkov, kdaj rečemo, da so »pravi srečneži« ali pa, da jih »spremlja sreča«.

V *Evdemovi etiki* in *Maximi moralii* se Aristotel sprašuje po izvoru takšne trajne pogostosti srečnih dogodkov (*eùtychía*). Najprej pomisli, da bi vzrok za to lahko bila nekakšna srečna narava posameznika, naključje namreč ne more biti, saj je po definiciji nekaj redkega (tu lahko vidimo še eno mesto več, kjer se nakazuje Aristotelovo povezovanje naravnega s pogostim in naključnega z redkim). To možnost potem zavrne in se osredotoči na človekova dejanja. V tem pogledu razločuje stvari, ki jih naredimo iz nagiba in odločitve, ter tiste, ki jih ne storimo na tak način, temveč nezavedno. Človek ima v nekem konkretnem primeru srečo (doživi srečno naključje), če nekaj naredi brez razmišljanja in ravno tako uspe. Tukaj je narava posameznika deloma odgovorna, a sreče ne moremo povsem zvesti nanjo, pravi Aristotel. Če človek sprejema zavestne odločitve in pri tem dosega neko regularnost srečnih dogodkov, pa Aristotel pravi, da gre za neko kombinacijo narave in človekove želje po bogu kot smotrnem vzroku, kar Dudley označi kot intuicijo o dobrem (Dudley, 2012: 247), čeprav seveda Aristotel ne uporablja tega izraza, ampak piše, da gre za »nerazumski nagib k dobrim stvarem« (*àneu lógou òrmén tàgathá*) (*Maxima moralia II 8*, 1207 a 35–1207 b 5).

V *Evdemovi etiki* tako ločuje dve vrsti sreče v smislu dogodkov. Tista božjega tipa, pri kateri ne ravnamo po zavestni odločitvi, temveč bolj po intuiciji, je bolj kontinuirana. Tista, pri kateri ravnamo brez odločanja (na primer pri metanju kocke), pa po definiciji ni pogosta. Hkrati loči tudi dve vrsti vzrokov srečnih dogodkov, ki pa niso substancialni, temveč naključni vzroki. V primeru srečnega meta kock gre za naključno srečo, ki je nestalna in je enostavno dana, kot na primer rojstvo, ni je mogoče razložiti. V drugem primeru gre za to, da je eden od vzrokov nenamerna odločitev *porine)*, ki je analogna odločitvi na območju smotrnih vzrokov. To slednje je bolj sreča v pravem pomenu besede, pravi Aristotel. Prava zunanja sreča je torej tista, ki temelji na *òrmé*, na lastnem nezavednem ravnanju. S tem je Aristotel razvil tudi takšen pojem sreče, ki ima neko stalnost pri prinašanju zunanjega blagostanja, ki je pogoj absolutne sreče.

Naključje v smislu kontingentnosti je torej le eden od vzrokov za srečne dogodki drugi je intuicija v okviru smotrnosti božjega vzroka. To lahko imamo za Aristotelov nauk kontinuirane sreče, ki je hkrati njegova teorija intuicije.[21] Aristotel torej priznava neko stopnjo korelacije med značajem oziroma intuicijo in srečo v smislu srečnih naključnih dogodkov in s tem naš vpliv na zunanje pogoje sreče kot duševnega blagostanja. Vendar to, da se nekdo rodi z naravo, ki je dovzetna za smotrno kavzalnost boga, je stvar naključja v smislu kontingence danega, tako da nimajo vsi možnosti tega vpliva (Dudley, 2012: 267).

[21] Te v *Nikomahovi etiki* še ni razvil, doktrino kontinuirane sreče oziroma intuicije najdemo le v *Evdemovi etiki* in *Maksimi moralii*.

Vprašanje, s katerim se Aristotel še ukvarja, je kdaj lahko na podlagi dolgoročne sreče v zunanjih okoliščinah rečemo, da je nekdo absolutno srečen. V splošnem rečeno je zanj absolutno srečen nekdo, ki je srečen vse življenje (*bíos téleios*). Glavni problem je tu, kaj pomeni vse življenje, oziroma to, po kolikšnem času lahko za nekoga rečemo, da je imel srečno življenje. Ali lahko to rečemo šele po smrti, retrospektivno? Iz odlomkov v *Nikomahovi etiki* sta po Dudleyju mogoči dve interpretaciji, in sicer da gre za celo življenje do smrti ali pa da gre za neko nekoliko krajše vendar dovolj dolgo obdobje (to je mogoče sklepati zgolj na podlagi *Nikomahove etike I, 10*). Dudley obema branjema priznava težo in pravi, da Aristotel uporablja izraz »absolutno srečen« v obeh smislih, saj se v njegovem diskurzu mešata pojma absolutne sreče v danem trenutku in absolutne sreče v celotnem življenju. Dudley razišče, kolikšna bi bila ocena starosti človeka, da zanj lahko rečemo, da je absolutno srečen. Na podlagi teksta *Nikomahove etike* pride do ocene okrog 50 let. Otroci ne morejo biti absolutno srečni, ker še nimajo praktičnih in intelektualnih vrlin, odraščajoči tudi ne, saj naj bi po Aristotelu človek dosegel intelektualni višek okrog 50. leta. To je glavni razlog za to Dudleyjevo oceno, ki je Aristotel sicer ne poda eksplicitno.

Srečni dogodki, ki so izvor zunanjega blagostanja (ne pa absolutne sreče), spadajo pod *týche*, vendar v smislu kontingentnosti danega, izvori zunanjega blagostanja, ki je zunanji pogoj absolutne sreče, so narava, intuicija in naključje kot (slučajna) sreča, izvora absolutne sreče pa sta intelektualna vrlina in praktična vrlina. Aristotel sprejema kombinacijo naključja (v smislu kontingence) in smotrne kavzalnosti negibnega gibala kot izvor zunanjega blagostanja in potrebni pogoj za absolutno srečo »resničnih srečnežev«. »Naključje je gospodar zunanjih dobrin, za katere Aristotel meni, da so nepogrešljiv pogoj sreče. Naključje je tudi gospodar kvalitete narave, s katero se vsak posameznik rodi in zato določa, ali bo naša narava odprta za smotrno kavzalnost boga. /.../ Končno je naključje tudi gospodar zunanjega uspeha, ki je spremljevalna posledica vrlih dejanj in je potreben za popolno srečo« (Dudley, 2012: 267).

Aristotelova pojma naključja in pogojne nujnosti v perspektivi statistike in verjetnosti

Na vprašanje naključja se pri Aristotelu navezuje še več sorodnih vprašanj, od katerih se nekatera tičejo povsem človeškega sveta vključno z etiko, druga pa tudi vprašanj o naravi. Kljub temu najdemo pojem naključja predvsem na področju praktično-delovanjskih (etičnih) in proizvajalno-storitvenih znanosti, medtem ko ima Aristotel na področju teoretičnih znanosti s pojmom naključja težave – po eni strani ga hoče iz preučevanja izključiti, po drugi pa se mu vrača v podobi spačkov in spontano generiranih organizmov. Pri tem je po eni strani v svojih delih o etiki ohranil predsokratski pomen naključja kot čiste kontingence danega, v delih o naravi pa je razvil nov pojem naključja kot sovpadanja, kar bi

po vrsti lahko označili za trdi in mehki pojem naključja. Glavni pojmi, ki so pri njem povezani s pojmom naključja, so vzrok, teleologija, nujnost, (ne)naravno, odločanje, sreča.

Mehkejši pojem naključja je pri Aristotelu tako po eni strani razumljen kot *vrsta vzroka*, čeprav le vzroka sekundarne vrste, ki je v določenem odnosu do substancialnega vzroka. Po drugi strani je naključje razumljeno kot *dogodek*, ki spada v območje stvari, ki so zavoljo nečesa. Pri tem naključni dogodek vključuje pojem naključnega vzroka, slednji je del prvega. To dihotomnost lahko razumemo tako, da upoštevamo, da Aristotel svojo refleksijo začenja pri mnenju večine, ki naključje pojmuje kot vzrok, kasneje pa ugotavlja, da je pri tem zelo pomemben dogodek sovpadanja kakšnih pomembnih okoliščin. Zato sam sprejemam kot njegovo pojmovanje naključja v bistvu pojem naključnega dogodka, ki v povezavi s človeškim delovanjem pomeni (ne)srečen dogodek Naključni dogodek vedno povzročata dva vzroka, namreč naključni in substancialni, ki je vedno narava ali um. Naključni vzrok deleži na substancialnem in nenadoma povzroči pomenljivo sovpadanje, povzročen je po absolutni nujnosti in je v osnovi snovni vzrok, morda bi lahko bil tudi gibalni. V resnici gre tako za *sovpadanje*, za srečanje dveh vzročnih verig, od katerih je ena absolutno nujna, druga pa pogojno nujna v smislu prej opisne pogojne nujnosti, po kateri delujejo oblikovni in smotrni vzroki. Narava kot vzrok vedno pomeni neko regularnost in s tem priskrbi kontekst pravilnosti, v katerem je neko konkretno sovpadanje redek dogodek. Um, če ga razumemo kot racionalnega, ravno tako priskrbi kontekst pravilnosti nekega dogajanja in je s tem substancialni vzrok, v tem kontekstu se potem lahko zgodi za človeka srečno ali nesrečno redko sovpadanje. V tem vidimo, da je mehki pojem naključja neizogibno povezan z neko regularnostjo in zato s človeškimi *pričakovanji*. Pri naključnih substancah Aristotela moti prav to, da pri njih snovni vzrok tako zelo spremeni prvotno bitnost, da gre pravzaprav malodane že za drugo bitnost. Posebej pri spontani generaciji organizmov je težko videti, kaj je tukaj narava kot vzrok in kaj je po naključju, dokler jih ne obravnavamo bolj kot snovi in ne toliko kot organizme.

Vsekakor je Aristotel naključje po malem dojemal tudi kot posebno vrsto vzroka, medtem ko so ga njegovi predhodniki dojemali kot zgolj danost, kontingentno in torej prej kot odsotnost vzroka. Zato gre pri Demokritu, ki je trdil, da je red svetu nastal po naključju kot snovnem vzroku, za to, da je takšno pripisovan posledica tega, da išče smisel oziroma razlago za nekaj, kar razume kot smiseln namreč red v vesolju. Njegov pojem naključja pri razlagi nastanka sveta vključuje zgolj čudenje nad tem, da obstajamo, razlagalno pa je prazen (Dudley, 2012: 353). Zanikanje nefizikalnega (božanskega) vzroka stvarjenja sveta je odziv na psihološko nepričakovan red na nebu, naključje je tu torej v smislu čudenja, izrednega, pomenljivega. Za Aristotela je red deloma nujno in deloma običajno stanje univerzuma, zato ga je pripisal naravi (naravno je med drugim to, kar se

dogaja vedno ali večinoma) oziroma smotrnemu vzroku Negibnega gibala, ki pa ni dejavni urejevalec (ni gibalni vzrok), ne posega aktivno v dogajanja v svetu. Za zavrnitev Demokritovega pojmovanja uporabi Aristotel nov, dodelan mehki pojem naključja, ker hoče razložiti naše čudenje nad redom, tega pa trdi pojem naključja ne pojasnjuje.

Sam menim, da bi oba pojma naključja lahko bila kompatibilna (tj. del iste slike), kar bi lahko ponazoril z verjetnostnim opisom. Trdemu pojmu naključja kot kontingentne danosti po mojem ustreza pojem prostora vseh možnih izidov (t.i. fazni prostor), mehkemu pojmu pa nekateri dogodki v tem prostoru možnosti, ki imajo majhno verjetnost in veliko pripisano vrednost. Takšno opredelitev bi bilo mogoče celo formalno artikulirati v okviru statistično-verjetnostnega računa in se v svetu igralništva praktično ujema z glavnim dobitkom v neki igri.

Aristotelov mehki pojem naključja je možen le v teleološkem kontekstu. To lahko trdimo, ker opredelitev mehkega pojma naključja strukturno vsebuje tudi substancialni vzrok, na katerem naključni vzrok zgolj deleži, in sicer bodisi naravo bodisi um, ki po Aristotelu oba delujeta tudi preko smotrov. Pri tem je pomembno poudariti, da ima naključni dogodek *negativni* odnos do teleologije v tem smislu, da gre pri naravnih stvareh za *neizpolnitev smotra*, pri človeških stvareh pa za *nenačrtovano izpolnitev smotra*. Poleg tega je pojem naključja močno vezan na pojem neregularnosti oziroma redkosti, kar je z modernega stališča povsem statistični pojem. »Pojem naključja je torej odvisen od Aristotelove distinkcije med običajnim in izjemnim. Ravno tako je odvisen od teleološkega konteksta človekovega zasledovanja ciljev, saj so vsi izjemni dogodki, ki jih imenujemo naključne, rezultat sovpadanja nepričakovane okoliščine z vzrokom na sebi, ki poskuša doseči smoter« (Dudley, 2012: 367). Z vidika bitnosti gledano mehki pojem naključja temelji tudi na obstoju duše oziroma je z njo povezan, saj morajo substancialni vzroki naključnih dogodkov imeti neko vrsto duše, ki jo Aristotel priznava tudi živalim in rastlinam (Dudley, 2012: 351). Pri naravnih elementih in spontani generaciji organizmov zagato rešuje princip duše oziroma hilozoizem.

Tudi razliko v pojmovanju med absolutno in pogojno nujnostjo je po mojem mnenju mogoče razumeti skozi perspektivo regularnosti. Kar se dogaja vedno, se dogaja po absolutni nujnosti, kar se dogaja običajno, pa po pogojni nujnosti. Razlog, da se slednje ne dogaja vedno, je v tem, da lahko pride do preprečitve. Pojem pogojne nujnosti Aristotel uporablja predvsem v svetu organizmov. Stvari, ki so še redkejše, se ne dogajajo po nujnosti, temveč po naključju, pri čemer je dejstvo, da je veliko takšnih redkih dogodkov, ki jih sploh ne opazimo, ker so povsem nezanimivi in si ne zaslužijo imena naključje (Aristotel da primer nekoga, ki se ravno umiva, ko se zgodi sončni mrk). V takšnem pojmovanju je po mojem mnenju nakazano, da se pri razlagi dogajanja v zvezi z organizmi pojem nujnosti nekako mehča, to mehčanje pa gre v smeri zmanjševanja regularnosti nekega pojava.

Vsakič, ko Aristotel razlaga nekaj, kar se dogaja vedno, uporabi absolutno nujnost in s tem snovni in gibalni vzrok. Ko razlaga nekaj, kar se dogaja večinoma, potrebuje oblikovni in smotrni vzrok Ko razlaga nekaj, ker se zgodi redko, uporabi razlago, ki ima še vedno odnos do smotrnega in oblikovnega vzroka, vendar negativen. Razlage s smotrnim vzrokom nastopajo tako pri absolutno nujnih dogodkih, ko gre za nebesne pojave, kot pri običajnih dogodkih, ko gre za naravne pojave na Zemlji, med redkimi dogodki pa imajo samo naključni dogodki odnos do nje, drugi redki dogodki pa ne.

Z absolutno nujnostjo *ne moremo* razložiti forme oziroma smotrov. Pri stvareh, ki so običajno na neki način, je krivec za to, da niso nujno vedno na ta določen način, snov, ki deluje po absolutni nujnosti, snov je ta, ki občasno prepreči udejanjenje naravnega smotra. Vendar tu govorimo vedno o kompleksnih organizmih. V tem smislu imajo kompleksni organizmi neko neregularnost razvoja po absolutni nujnosti, neregularnost je pri njih absolutna nujna (kot da bi rekli: ni pravila brez izjeme). Aristotel tudi meni, da če izjem ne bi bilo, potem bi se vse dogajalo po absolutni nujnosti, kar pa očitno ne drži.

Velik problem predstavljajo spački in spontano generirani organizmi, ker ne spadajo med naključne dogodke, ampak naključne bitnosti. Ta razkorak bi lahko omehčal fenomenološki opis, ki je načeloma bolj desubstantiviran. Ko jih Aristotel poskuša razlagati, pri njih pride do interakcije med tipi substancialnih vzrokov, to pa je velik problem, saj načeloma v razlagah tipi vzrokov ne konkurirajo med seboj. Primer spontane generacije kaže na bližino teh organizmov elementom, ko je stvar snovno manj kompleksna. V tem je nekoliko podobna rasti kristalov, ki niso živa bitja.

V Aristotelovih delih naj bi približno ena petina navedenih organizmov bila takšnih, ki se razmnožujejo spontano (Dudley, 2012: 184). Statistično gledano gre za vprašanje, ali je to dovolj pogosto, da se lahko šteje za naravno. Tudi pri spačkih Aristotel pravi, da jih to, da morda nastajajo z večjo regularnostjo, dela manj spačke (*O generacij živali IV 4*, 15–17). Opažam, da takšna zmerna pogostost, ki ni niti večinoma niti izjemoma (recimo nekje okrog 20 %), Aristotela večkrat zmoti, ker s kategorijo tovrstne pogostosti (manj pogosto za razliko od redko) ne operira. V tem smislu mu manjka kvantitativne subtilnosti, ki jo premore statistika. Najbolj ga moti to, da snovi sami po sebi v glavnem ne priznava regularnosti, ampak jo priznava samo naravi, zato je v dvomu, ali je spontana generacija naravna. Dudleyjev končni odgovor je, da je Aristotelu blizu hilozoizem, kjer gre za princip duše v obliki *težnje snovi*, pri kateri ne potrebujemo smotrnega vzroka. Nekaj podobnega pravi Johnson za gibanja elementov, ko zagovarja tezo, da so štirje osnovni elementi bitnosti, ker imajo svoje težnje, čeprav priznava, da glede tega problema med preučevalci Aristotela ni konsenza (Johnson, 2005: 131–159). Pri elementih gotovo ne gre za paradigmatične primere bitnosti, kar so sicer živa bitja, temveč spet za nekaj, kar z določeno regularnostjo obnašanja izkazuje princip duše.

Pri opredelitvi regularnosti srečnih dogodkov kot potrebnega pogoja dolgoročne absolutne sreče smemo videti obrise tega, kar se v statistiki imenuje časovno povprečje in dolgoročna konvergenca. Absolutna sreča je takšna, da mora trajati skozi celo življenje, le redko se lahko zgodi, da je nekdo absolutno srečen, nato ni več, potem pa postane spet, pravi Aristotel. Če je kdo absolutno srečen, potem običajno ostane tako vse življenje, kljub posameznim nesrečnim dogodkom. Za to, da rečemo, da je kdo srečno živel, je potrebno, da če je človek doživel velike nesreče, od tedaj mine veliko časa in da se ta doživetja uravnovesijo; obdobje velike nesreče mora biti omiljeno z daljšim obdobjem sreče. Razlikovanje med obdobjem celotnega življenja in »dovolj dolgega« življenja, ki ga razbere Dudley, je analogno matematičnemu problemu konvergence relativnih frekvenc pri verjetnosti oziroma statistiki (t.i. centralnemu limitnemu izreku). Statistično rečeno gre za to, da se verjetnost za to, da se trenutno povprečje neke količine skazi, s časom manjša, doseženo povprečje pa je vedno bolj reprezentativno. Tudi v statistiki gre za vprašanje v smislu »kdaj smo že dovolj blizu pravi vrednosti, da jo lahko razglasimo za pravo vrednost«. Pomembna in v oči bodeča lastnost pri tem pa je, da se sreče ne da kvantificirati, ampak le kvalificirati (opredeliti po stopnjah). Vendar Aristotel tudi svoje fizike ne kvantificira, kar je ena ključnih metodoloških razlik med njegovim opisnim in modernim preučevanjem narave.

Ravno v Aristotelovih poskusih razlage možnih izvorov dolgoročne sreče s pojmom srečne narave nekega človeka (čeprav je to razlago opustil) je očitno, da pojem narave v veliki meri temelji na statistični pravilnosti, saj ta pojem poskuša razložiti ravno to pravilnost. Kasneje to zamenja pojem intuicije, katere smotrni vzrok je bog. Tudi ta razlaga je v bistvu razlaga regularnosti srečnih izidov, ravno ti so znak intuicije, tako jo prepoznamo. Ta individuacija intuicije z regularnostjo srečnih naključij je v jedru statistična, njena opredelitev pa se oblikuje v jeziku vzrokov. Aristotelova opredelitev intuicije je poleg pojmov narave in uma še ena opredelitev več v zvezi s področjem doseganja smotrov, ki se opira na *regularnost pojavov*. Podobno, vendar v zrcalno nasprotnem smislu pa se njegov (mehki) pojem naključja opira na *neregularnost* oziroma redkost pojavov ter ima s področjem smotrov *negativno* povezavo.

ARISTOTLE AND THE CONCEPT OF CHANCE

Aristotle uses two terms for chance, namely *týche* and *taùtómaton*, the first relating to the human world and the second to natural events. His key definition of chance as a secondary cause on one hand points to the human understanding of

chance as a cause while on the other hand it points to the result of Aristotle's analysis of chance as a non-substantial cause. Chance events are explicated in a teleological context – they are rare meaningful events which are caused by chance causes and lack a final cause. The source of chance is explicated in relation to the concept of hypothetical necessity, a kind of necessity for describing events in the sublunary world and that leaves room for rare exceptions in natural events. A problematic point for Aristotle are monsters and spontaneous generation because these two phenomena show a need for the introduction of the concept of chance substance, which contradicts Aristotle's own definition of chance as non-substantial cause. We end the article by addressing the open problems from the perspective of (statistical) (in)frequency of phenomena.

Keywords: chance, Aristotle, teleology, causes, statistics.

Literatura

Aristotel (2004). *Fizika: knjige 1, 2, 3, 4*. Ljubljana: Slovenska matica.

Aristotel (1999). *Metafizika*. Ljubljana: Založba ZRC, ZRC SAZU.

Aristotel (2000). *Generation of Animals*. Cambridge, MA, London: Harvard University Press.

Aristotel (2002). *Nikomahova etika*. Ljubljana: Slovenska matica.

Aristotel (1981). *The Eudemian Ethics*. Cambridge, MA: Harvard University Press; London: W. Heinemann.

Aristotel (1997). *Posterior Analytics*. Cambridge, MA [etc.]: Harvard University Press.

Dudley, John (2012). *Aristotle's Concept of Chance*. Albany: State University of New York Press.

Hankinson, R. J. (1995). »Philosophy of Science«. V Barnes, J. (ur.), *The Cambridge Companion to Aristotle*, Cambridge [etc.]: Cambridge University Press, str. 109–139.

Johnson, Monte R. (2005). *Aristotle on Teleology*. Oxford: Oxford University Press.

Kocijančič, Gorazd (ur.) (2012). *Fragmenti predsokratikov*. Ljubljana:

Študentska založba.

Reale, Giovanni (2002). *Zgodovina antične filozofije – II. Zvezek: Platon in Aristotel* (9. izdaja). Ljubljana: Studia humanitatis, Ljubljana (prva izdaja 1975–1980).

Navodilo avtorjem

Prispevke pošljite v izpisu z dvojnim razmakom na naslov uredništva. Priložiti je treba povzetek (v slovenščini in v enem izmed svetovnih jezikov – angleščini, nemščini, francoščini), ki povzema glavne poudarke dela. Povzetku v tujem jeziku je treba dati naslov v tem jeziku. Povzetek ne sme presegati 150 besed. Na koncu povzetka navedite do 5 ključnih besed (deskriptorjev). Na primer – *Ključne besede*: filozofija jezika, metafora, analogija ... oziroma *Keywords*: philosophy of language, metaphor, analogy ...

Prispevki naj ne presegajo obsega ene in pol avtorske pole (45.000 znakov) vključno z vsemi opombami. Prispevki naj bodo notranje razčlenjeni, torej razdeljeni na razdelke, in opremljeni – če je mogoče – z mednaslovi. V besedilu dosledno uporabljajte dvojne narekovaje (na primer: pri navajanju naslovov člankov, citiranju besed ali stavkov in kadar navajate strokovne in posebne izraze). Izjema so citati znotraj citatov. Pri teh uporabite enojne oklepaje. Naslove knjig, periodike in tuje besede je treba pisati *ležeče*. Primeri: *Kritika praktičnega uma*, *Anthropos*, *a priori*.

Reference se tiskajo v besedilu (“author-date” sistem). Oblikovane morajo biti takole: (avtorjev priimek, letnica: str. ali pogl.); na primer (Davidson, 1967: 312) ali (Blackburn, 1988: III). Na koncu poslanega prispevka morate priložiti popoln, po abecednem redu urejen bibliografski opis citiranih virov. Primeri takih opisov:

Davidson, D. (1967). “Truth and Meaning”. *Synthese*, 17, str. 304-323.

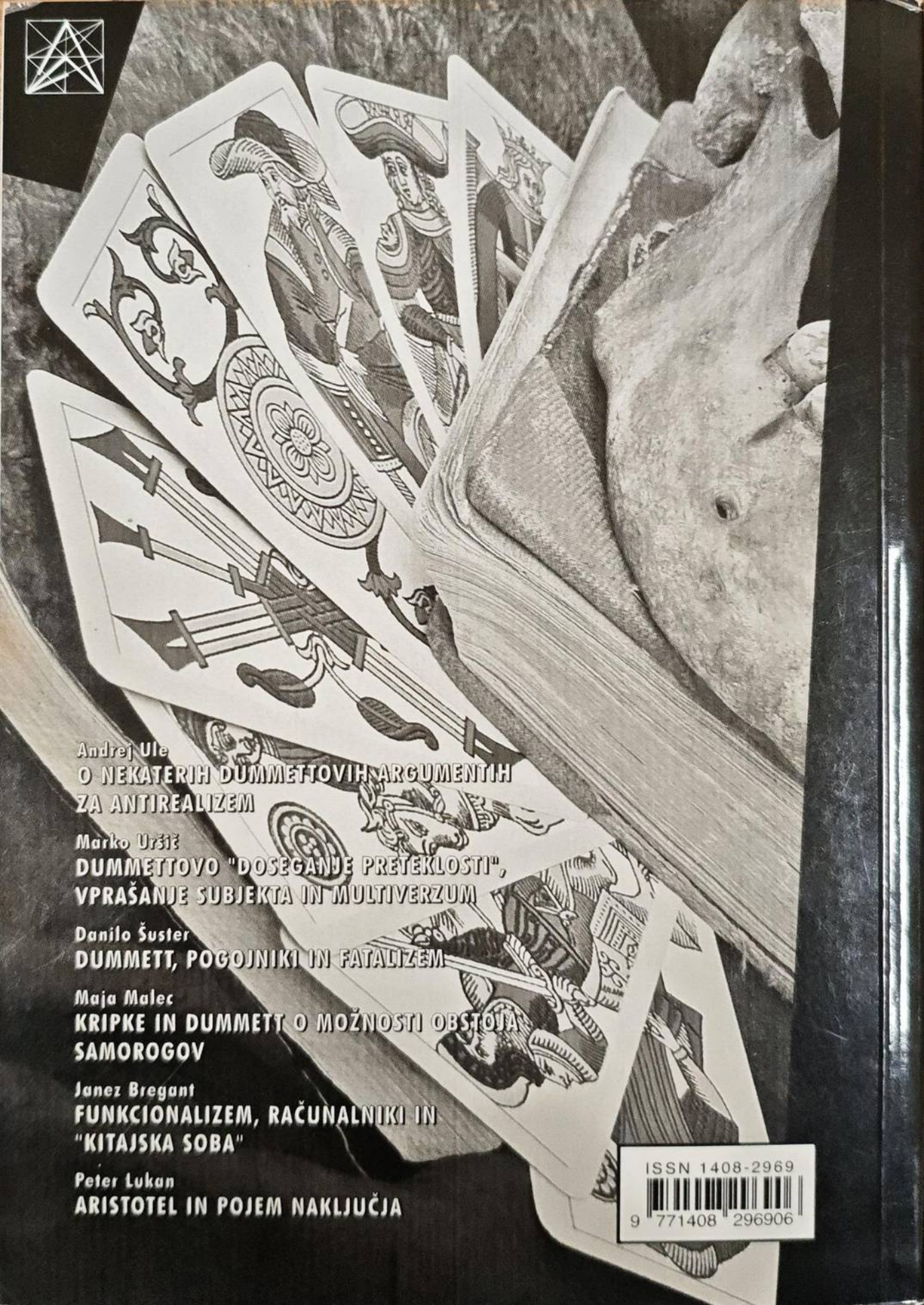
Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davidson, D. (1984). “Truth to the Facts”. V Davidson, D., *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, str. 37-54.

Davidson, D. (1997). “Radical Interpretation”. V Baillie, J. (ur.), *Contemporary Analytic Philosophy*, New Jersey: Prentice Hall, str. 383-393.

Pod črto navajajte *samo* opombe. V besedilu je treba opombe označiti z dvignjenimi indeksi[1]. Ne uporabljajte opomb za navedbo reference.

Poleg izpisa teksta priložite tudi tekst na 3.5” disketi. Tekst naj bo napisan na IBM združljivem računalniku v urejevalniku besedil Word ali WordPerfect, po možnosti v okolju Windows 95. Poleg datoteke, ki je bila shranjena v Word ali WordPerfect formatu, shranite isto datoteko še v ASCII formatu. Poskrbite, da se bosta besedili na disketi in izpisu natančno ujemali. Nenaročenih rokopisov ne vračamo.



Andrej Ule
O NEKATERIH DUMMETTOVIH ARGUMENTIH ZA ANTIREALIZEM

Marko Uršič
DUMMETTOVO "DOSEGANJE PRETEKLOSTI", VPRAŠANJE SUBJEKTA IN MULTIVERZUM

Danilo Šuster
DUMMETT, POGOJNIKI IN FATALIZEM

Maja Malec
KRIPKE IN DUMMETT O MOŽNOSTI OBSTOJA SAMOROGOV

Janez Bregant
FUNKCIONALIZEM, RAČUNALNIKI IN "KITAJSKA SOBA"

Peter Lukan
ARISTOTEL IN POJEM NAKLJUČJA

ISSN 1408-2969
9 771408 296906