ANαliZA

časopis za kritično misel

Letnik 18, številka 1-2, 2014 · ISSN 1408-2969

STROJI, ZAVEST IN FENOMENOLOGIJA

Janez Bregant

Stroji in zavest: problem takšnosti

Filozofska fakulteta Univerze v Mariboru

Citiranje: Janez Bregant, »Stroji in zavest: problem takšnosti«, *Analiza: časopis za kritično misel* 18, št. 1-2 (2014): 45–74.

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/deed.sl.

Izdajatelj: Društvo za analitično filozofijo in filozofijo znanosti, Ljubljana.

Elektronska različica je bila leta 2026 rekonstruirana iz izvirnih stavnih datotek in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

Janez Bregant
Filozofska fakulteta Univerze v Mariboru

STROJI IN ZAVEST: PROBLEM TAKŠNOSTI

Članek se ukvarja z vprašanjem, ali je fenomenalna zavest, tj. 'kako je biti', nekaj, kar je rezervirano izključno za človeka, ali pa nekaj, kar bi lahko v enem izmed možnih svetov imel tudi stroj. Iz različnih problemov zavesti izlušči tri, ki se najpogosteje povezujejo s tako imenovanim problemom takšnosti in tvorijo težek problem zavesti, tj. vprašanje, kako so lahko takšnosti fizične. Na poti do takšnega odgovora nanj, ki bi ohranil fizikalizem, stoji *argument iz znanja*, ki dokazuje, da takšnosti niso fizične, filozofi pa si glede tega, kako ga zavrniti, da ohranimo resničnost fizikalizma, niso edini. Članek s pomočjo enostavne dileme namiguje na to, da za ugotovitev tega, ali lahko obstajajo tudi zavestni stroji, takšnega soglasja sploh ne potrebujemo, saj narava takšnosti pri tem, ali lahko pripišemo fenomenalno zavest tudi, npr. robotom, nima nobene vloge.

Ključne besede: fenomenalna zavest, takšnosti, 'kako je biti', argument iz znanja, obrnjen spekter.

0 Uvod

> Zavest je bila dolgo časa izločena iz možganskih raziskav in to preprosto zato, ker je znanstveniki niso znali pojasniti. Princip, ki so se ga držali, se je glasil »Če ne veš, kako bi nekaj razložil, se delaj, kot da to ne obstaja«. Še v 80. letih 20. stoletja so nekateri znanstveniki o njej govorili omalovažujoče, npr. Stuart Sutherland, ki je bil takrat direktor *Centra za raziskovanje percepcije in kognicije na Univerzi Sussex*, je imel o njej takšno mnenje: »Kolikor je meni znano, ni bilo o zavesti napisanega nič takega, kar bi bilo vredno papirja, na katerem je bilo objavljeno« (Copeland, 1993: 163–164). Danes je stvar drugačna, raziskovanje zavesti ni več zanemarjeno, na empirični ravni poteka znotraj nevroznanosti, na teoretični znotraj filozofije, obe ravni njenega raziskovanja pa se trudi pokriti psihologija. Freud je nekoč pripomnil: »O tem, kaj mislimo z zavestjo, ne rabimo razpravljati: to je več kot jasno« (Freud, 1932: 76–77).

Danes je več kot jasno tudi to, da se je hudo zmotil. Pojem zavesti namreč označuje več različnih pojavov, ki so med seboj le ohlapno povezani. Ker namen tega članka ni razpravljati o definicijah zavesti, ne glede na to, kako različne so lahko,[1] se bomo osredotočili zgolj na tisto, za katero se zdi najmanj verjetno, da bi jo lahko pripisali tudi strojem. Gre za tako imenovano *fenomenalno zavest*, ki jo običajno opisujemo z besedno zvezo 'kako je biti' in v slovenščini imenujemo *takšnost*, medtem ko je v tuji literaturi za njeno označitev uveljavljen latinski izraz *quale* (mn. *qualia*). V tem pomenu ima bitje zavest, ko njegove čutne interakcije z okoljem vsebujejo tudi kvalitativna subjektivna izkustva tega, kar zaznava, npr. občutke, kako je biti žejen, kako je biti bolan, kako je biti zaljubljen itd. Fenomenalna zavest vključuje naslednje:

(i) zaznavno izkustvo: *videti* rdeč paradižnik, *slišati* violino, *vohati* gorgonzolo, *okušati* česen, *občutiti* hrapavost itd.;

(ii) telesne občutke: bolečino, srbenje, lakoto, žejo, vročino itd.;

(iii) čustva: strah, veselje, ljubezen, obžalovanje, sovraštvo itd.;

(iv) razpoloženje: srečo, potrtost, mirnost, dolgočasnost, napetost itd.

Ni pa to, kako je biti' bistven del vsake zavesti, ni ga zaslediti pri tako imenovani *intencionalni zavesti*, tj. pri prepričanjih, mišljenju, željah, namerah ipd., saj npr. misli za to, da bi obstajala, ne rabim čutiti.[2]

Vprašanje, na katerega bomo skušali najti odgovor, je, kdo ali kaj ima lahko fenomenalno zavest, z drugimi besedami, ali je to nekaj, kar je rezervirano izključno za človeka, ali pa nekaj, kar bi lahko v enem izmed možnih svetov, če že ne v dejanskem, imel tudi stroj. Kljub temu, da se zdi slednje za večino ljudi malodane heretično, saj stvaritev iz plastike, silikona, železa itd. ne more imeti podobnega kakovostnega izkustva, ko zagleda npr. rdeč paradižnik, kot ga imamo mi, bomo navijali za to, da lahko fenomenalno zavest pripišemo komurkoli ali čemurkoli, r bi bilo zgrajeno na ustrezen način in tako tudi računalnikom. Ker moramo tudi oje obravnavati kot zavestne sisteme, če so takšnosti fizične – v nasprotnem imeni bi za »merjenje« istih pojavov uporabljali različne standarde, zaradi česar bi nas lahko nekdo obtožil *šovinizma vrst*, ki favorizira ljudi pred stroji, ko za to ni nobene osnove -, bi bil za naš uspeh dovolj že prepričljiv argument, ki bi dokazal, da so takšnosti resnično fizične. V prvem delu članka je tako predstavljena kratka novejša zgodovina zveze med zavestjo in znanostjo, v drugem pa problemi zavesti, kot jih je zbral in uredil Tye (1995),

[1] Npr. (i) zavest kot sposobnost zaznavati svet s pomočjo čutnih organov in sposobnost razmišljati, tehtati, soditi, predpostavljati, načrtovati ipd., (ii) zavest kot posedovanje nekega kontrolnega sistema *X* višjega reda, ki je zmožen nadzorovati večino kognitivnih stanj in procesov, pri čemer sam ni zavesten, ter tiste, ki izpolnijo osnovne pogoje," naredi"' zavestne, to pa se zgodi takrat, ko se jih zavem, (iii) zavest kot očitno stanje biti buden, (iv) zavest kot občutljivost za določeno razsežnost človekovega obstoja (npr. politična zavest), (v) zavest kot samozavedanje ali »jaz«, (vi) zavest kot metaforični »notranji prostor« itd. (Copeland, 1993: 164)

[2] Tukaj sicer med filozofi ni enotnega mnenja, nekateri menijo, da obstaja tudi fenomenalna intencionalnost, *glej* npr. Kriegel (2013), sami pa se bomo držali bolj tradicionalnega pogleda. (1980), Block s

izmed katerih predvsem trije tvorijo tako imenovani *problem takšnosti*, ki je osrednja tema naše razprave. Tretji del članka opisuje dva znana argumenta proti fizikalistični rešitvi problema takšnosti, ki bi omogočala pripisovanje zavesti tudi strojem, »kako je biti netopir« in »Marijino sobo«. V četrtem so predstavljene različne bolj ali manj uspešne obrambe fizikalizma, ki tako v večji ali manjši meri odpirajo prostor, v katerem bi lahko obstajali tudi zavestni računalniki, v petem, sklepnem delu članka, pa enostavna dilema, ki jasno zagovarja pripisovanje takšnosti tudi npr. robotom, in to ne glede na to, ali je zavest fizična ali nefizična.

1 Zavest in znanost

Idejo o zavestnih strojih temelji na funkcionalističnem prepričanju, da ni pomembna zgradba bitja, ampak opravljanje funkcije, ki jo mentalni pojavi imajo. Po njem niso mentalna stanja nič drugega kot notranja/fizična stanja s tipičnimi vzroki in tipičnimi učinki. Ker lahko isti psihološki opis z vidika zgradbe zadovoljijo povsem različni organizmi, tudi takšni brez nevronov, se zdi, da je pot do zavestnih strojev odprta. O tem, ali je funkcionalizem sprejemljiva teorija, smo pisali v eni izmed prejšnjih številk *Analize*,[3] kjer smo povzeli glavne argumente, s katerimi ga njegovi nasprotniki zavračajo, omeniti pa velja, da funkcionalizem kljub temu, da mentalna stanja identificira z njihovimi vzročno-funkcionalnimi vlogami, ohranja *fizikalizem*, ki v svoji najsplošnejši obliki pravi, da je vse, kar obstaja, sestavljeno iz fizičnih delcev.[4]

Nedvoumno je, da znanost navija za fizikalizem,[5] vprašanje pa je, ali lahko zavest zadovoljivo pojasni. Problem zanjo predstavlja pojasnitev tega, *od kod pridejo občutki*, kar Chalmers (1996) imenuje *težak problem zavesti*, Levine (1983) pa *pojasnjevalna vrzel*. Težava je, da nam nobeno raziskovanje korelacij med možgani, tj. proženjem nevronov in občutki, npr. bolečino, ne da odgovora na vprašanje, *'kako je biti'*. Znanost lahko pojasni funkcionalne vloge različnih mentalnih stanj in najde fizične mehanizme, ki jih realizirajo, zdi pa se, da to ni dovolj; obstaja še nekaj, kar moramo pojasniti. Namreč, potem ko vemo vse o sinaptičnih povezavah med nevroni, se še vedno sprašujemo: »Zakaj to občutimo tako, kot občutimo«? Težak problem zavesti se je pojavil v drugi polovici 20. stol, kot posledica hitrega napredka znanosti, ki je grozil, da bo zavesti vzel obstoj in jo

tako izključil iz realnosti. Ko smo svet napolnili s silami, atomi in molekulami, v njem naenkrat ni bilo več prostora za drugačna in ločena zavestna stanja.

Pred 20. stol, so tudi znanstveniki menili, da obstaja neodvisna zavest, ki je

[3] *Glej* Bregant (2013).

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

[5] Najbolj znan nevroznanstvenik, ki je zagovarjal (interakcijski) dualizem, je bil Eccles (1994).

ločena od kakršnekoli snovi, še več, zavest je bila tista, ki je bila temeljna, ne pa snov, kot se razlaga danes. Ni jih bilo malo, ki so trdili, da »Kamni, palice in ostale fizične stvari nimajo nobene druge realnosti razen tiste, ki jo dobijo s čutnim zavedanjem«, kar pomeni, da materialni svet sploh ni obstajal (Papineau, Selina, 2000). Ideja je čudna, le kdo bi danes to trdil, ne da pa se je tako lahko zavrniti: idealist seveda dovoli, da vidimo kamen, zanika le, da je vzrok za to vidno izkustvo neka materialna substanca. In kako lahko to zavrnemo glede na to, da bo vsak naš naslednji dokaz za nasprotno spet zgolj vidno izkustvo, tj. predstava predmeta, nikoli pa neposreden materialni predmet (substanca)? Najbolj znan zagovornik idealizma je bil prej ko slej Berkeley (1710/1976), razvil pa ga je kot kritiko *reprezentacionalizma*, ki ga je v približno istem času zagovarjal Locke (1689/2012). Berkeley je napadel Lockovo razlikovanje med primarnimi in sekundarnimi kvalitetami, ki ga v duhu ene izmed trditev reprezentacionalizma, tj. da so nam pri spoznavanju predmeta neposredno dostopne zgolj njegove reprezentacije/predstave, sam predmet pa ne, ki jo imenujemo *epistemološki antirealizem*, Locke ne more upravičiti. Berkeleyjeva ideja je bila, da razlike med enimi in drugimi ne moremo dokazati, ko pa so vse predstave, ki jih imamo, oblikovane v duhu, saj so zunanji predmeti zaznani zgolj posredno, zaradi česar so izključno čutne predstave in nič drugega.

V 20. stol, so se psihologi, gnani s prepričanjem, da znanstvena psihologija ne more temeljiti na introspekciji subjektivnega izkustva, tj. posameznikovi oceni lastnih občutkov, ampak na eksperimentalnem raziskovanju obnašanja, že obrnili proti idealizmu. V filozofiji je v tem času idealizem zagovarjal Russell, na drugi strani pa je imel hudega nasprotnika v Wittgensteinu, ki je menil, da govorjenje o mentalnih dogodkih ne more biti stvar zasebne introspekcije, saj potem sploh ne bi vedeli, o čem govorimo. V podporo tej trditvi je navajal miselni eksperiment, ki ga imenujemo »hrošč v škatli«. Zamislimo si ljudi, ki imajo škatle, v njej pa (domnevno) hrošče, no, vsaj obnašajo (govorijo) se tako, kot bi jih imeli. Težava je, da če vsebina njihovega govora (zavestnih mentalnih stanj) ni objektivno preverljiva, nikoli ne moremo vedeti, kaj imajo v škatlah – lahko res hrošče, lahko kaj drugega, lahko pa ničesar (Wittgenstein, 1953/1977). Enako naj bi veljalo tudi za pripisovanje zavesti: sistem, organizem ali bitje ima zavest, če se obnaša na ustrezen (določen, tipičen) način, npr. če je žejen, potem pije. Ker takšna teorija duha ne ponuja odgovora na to, kaj so mentalna stanja, ampak, kako jih analiziramo, jo imenujemo (analitični ali) *logični behaviorizem*. Danes je njegova ideja o tem, da lahko nekaj o mentalnih pojavih vemo zgolj na osnovi javno opazljivega in preverljivega obnašanja, preživeta.[6] Sta pa logični in v psihologiji metodološki behaviorizem vsak s svojimi pogoji za pripisovanje zavesti tlakovala pot, ki je vodila do končne zavrnitve tradicionalne slike zavesti, po kateri je ta bila prej ločena od snovi.

Kakorkoli, to se še ne bi zgodilo tako hitro, če ne bi k temu pripomogel nagel

[6] Za njegovo kritiko *glej* npr. Putnam (1963/1980).

razvoj nevrofiziologije v zadnjih 100 letih. Načrtala je nevronsko mrežo in analizirala mehanizme, ki so dogovorni za aktivnost nevronov, zaradi česar vemo o kemiji sinaptičnih povezav, tj. molekulah nevrotransmiterjev, s katerimi živčne celice komunicirajo med sabo, danes več kot kdajkoli prej. Seveda to ni končni dokaz za to, da v zvezi z zavestjo ne delujejo neke posebne mentalne sile, je pa to sedaj vsekakor manj verjetno (Papineau, Selina, 2000).

2 Številni problemi zavesti

V luči prepričanja, da je najboljša ontološka teorija v tem trenutku fizikalizem, Tye (1995) vsem njegovim zagovornikom v brk vrže deset problemov, katere morajo rešiti, da zavest ne bi ostala zunaj njega in plavala na nebu brez materialnega sidra. V nadaljevanju bomo vse probleme na kratko predstavili, osredotočili pa se bomo na zadnje tri, ki podrobneje razlagajo in izražajo težak problem zavesti, s skupnim nazivom pa jih bomo od sedaj naprej preprosto imenovali *problem takšnosti*.

(1) Problem dvojnikov

Predpostavimo, da je najboljša fizikalistična teorija duha funkcionalizem, kar je z našega stališča, če zagovarjamo zavestne stroje, tudi najbolj smiselno, saj jim je ta, kot smo že omenili, najbolj naklonjena, in si oglejmo, s katero težavo se srečamo že takoj na začetku.

Vprašajmo se, kaj imajo vsi primerki bolečine skupnega, da so sploh 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, tj. *fenomenalni pojavi*, vsebujejo še nekaj več od golega opisa v pojmih njihovih vzročno-funkcionalnih vlog. To prikazuje naslednji miselni eksperiment, ki ima svoje zametke že pri Descartesu (1637/2007) in njegovem zanikanju tega, da bi lahko bili avtomati kadarkoli zavestni, razvil pa se je v 70. letih 20. stol, kot orožje za boj proti fizikalizmu nasploh.[7]

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 tako imenovanega 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 njegovi fizični izvirniki. Npr. ko sediva jaz in moj zombi v jazz klubu, se obnašava povsem enako, oba pijeva, kot da bi bila žejna, oba naročiva pivo, kot

[7] Shoemaker (1975), Block (1980), Chalmers (1996).

da bi marala njegov okus, oba uživava v glasbi, kot da bi imela slišno izkustvo itd. 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 fizični 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, ki jih je uvedel Kripke (1972/2000), 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 imenuje, 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,[8] kar pa je dovolj, da dodobra zamajejo, če že ne porušijo, navidezno trdne temelje funkcionalizma.

Fizikalist ima kot rešitev na voljo vsaj dva odgovora, oba pa zanikata možnost zombijev: (i) maksimalni funkcionalni dvojniki – bitja, ki so enaka v smislu vseh notranjih in zunanjih vzročnih interakcij (vključno z okoljem in zgodovino) – *so* vedno enaki tudi v fenomenalnem smislu: če je mikrofizična identiteta v tem smislu nujni in zadostni pogoj za izkustveno identiteto, potem bodo vse moje fizične opije imele tudi fenomenalno zavest;

(ii) gre za dve različni zamišljanji dvojnika, med katerima lahko preskakujemo, ne pa za eno: zombija si lahko predstavljamo od zunaj (P), tj. da se obnaša povsem enako kot izvirnik (med njima je funkcionalna izomorfnost), lahko pa si ga predstavljam tudi od znotraj (0, tj. da ne čuti absolutno ničesar, ne moremo pa si tega dvojega predstavljati skupaj, saj je sklep, da zamisliti si prvo in zamisliti si drugo *zadostuje* za zamisliti si prvo z drugim, modalna zmota naslednjega tipa (Tye, 1995: 200): Možno, da P.

2. Možno, da *Q*.

3. Možno, da P in *Q*.

(2) Problem obrnjenega spektra

Drug miselni eksperiment, ki prav tako namiguje na to, da mentalni pojavi, za katere so tipične takšnosti, vsebujejo še nekaj več od golega opisa v pojmih njihovih vzročno-funkcionalnih vlog, s čimer podvomi o funkcionalizmu, sega k

[8] Vsi teoretiki se s tem ne strinjajo, za nezamisljivost zombijev *glej* npr. Harnad (1995), Tye (2006), Dennet (1991/2012).

Locku (1689/2012), orodje za zavračanje funkcionalizma pa je prav tako postal v 70. letih 20. stol.[9]

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 Fodor, 1972/1980), 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 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 tako imenovani klasični/tradicionalni/običajni argument iz obrnjenega spektra,[10] 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 izvesti, s čimer bi postal tudi nomološko možen.

Fizikalist se lahko iz primeža obrnjenega spektra izvije s pomočjo sklicevanja na razliko v *funkcionalni organizaciji duha* med normalnim in »obrnjenim« spoznavalcem, zaradi česar bi bila predpostavka o funkcionalni izomornosti, ki jo vsebuje hipoteza o obrnjenem spektru, neresnična. Funkcionalizem sicer res predpostavlja, da sta dva stroja izomorfna takrat, kadar izpolnita isti funkcionalni opis ne glede na to, v kolikšni meri se razlikujeta z vidika svojih strukturalnih lastnosti, ker pa to vodi do govorjenja o funkcionalni izomorfnosti sistemov, ki si glede funkcioniranja še zdaleč niso podobni, npr. mačka in mišolovka, ki sta lahko funkcionalno izomorfna zgolj na visoki ravni abstrakcije, pri podrobnejšem opisu njune funkcionalne organizacije (nižja raven abstrakcije) pa se izkažeta za bistveno drugačna in ker znotraj psihologije pojasnjevanje našega obnašanja izključno s pomočjo prikazanih funkcionalnih shem že dolgo ni več resnično, se zadnje čase vse bolj opira na strukturalne lastnosti sistemov, tj. na nevrofiziološko organizacijo naših možganov (Churchland, 2011), pri tem nima smisla vztrajati.

Funkcionalna organizacija duha očitno vsebuje vsaj dve ravni opisa, splošnejšo, po kateri sta si normalni in »obrnjeni« spoznavalec resnično podobna, npr. zunanje obnašanje, in konkretnejšo, ki se povezuje z notranjimi/strukturalnimi/fizičnimi lastnostmi organizmov, po kateri sta si

[9] Glej npr. Block in Fodor (1972/1980), Block (1980, 1990), Shoemaker (1982).

[10] Block (1990) si je izmislil drugačen obrnjeni spekter, vendar z enako uničujočimi posledicami za funkcionalizem, ki ga je imenoval *obrnjena Zemlja*.

različna, npr. nevrofiziološka zgradba. Zato je govoriti o absolutni funkcionalni izomornosti oziroma funkcionalni identiteti utopija, kljub temu, da sta si res podobna v obnašanju (višja raven abstrakcije), imata še vedno povsem različno nevrofiziologijo (nižja raven abstrakcije), in ravno to je tisto, zaradi česar se fenomenalno gledano razlikujeta, funkcionalno pa ne.

Poudariti velja, da se s takšnim odgovorom funkcionalist morebiti izogne problemu obrnjenega spektra, če pa se za predpostavko o dveh enako pomembnih delih funkcionalnih shem vendarle odloči, korenito poseže v eno izmed temeljnih načel svoje teorije, tj. zanemarjanje pomembnosti fizične zgradbe sistemov pri realizaciji zavestnih stanj, ki sega do tega, da ima lahko duha praktično karkoli, če je le ustrezno organizirano.

(3) Problem fenomenalne vzročnosti

> Če se kaj v fizičnem svetu ne zdi sporno, je to to, da fenomenalna stanja, tj. subjektivne lastnosti našega izkustva, vplivajo na naše obnašanje. Kričimo, ker nas nekaj boli, sopihamo, ker smo utrujeni, vzdihujemo, ker smo npr. žalostni, kupujemo, ker nekaj potrebujemo, pijemo, ker smo žejni itd. Z drugimi besedami, za vsak opisani učinek obstaja mentalni vzrok in niti slučajno ne dvomimo, da so v tem smislu mentalna stanja vzročno učinkovita. Prav tako pa ne moremo ali ne želimo zanikati, da obstaja za vsak omenjeni učinek tudi popolna fizična razlaga, čeprav vseh podrobnosti še ne poznamo: vzrok za npr. kričanje žensk, obsojenih na grmado zaradi čarovništva, je v končni fazi sprožitev motornih nevronov z vsemi svojimi kemičnimi posebnostmi. Kljub temu, da vse podrobnosti še niso znane, pa fizična razlaga, kot vidimo, ne vsebuje nobenih občutkov, čustev, razpoloženja ali česa podobnega, tj. ne vključuje nobenih subjektivnih zaznavnih lastnosti.

Vprašanje je, kako lahko potem ob dveh konkurenčnih vzročnih razlagah, mentalni in fizični, zavestna stanja vplivajo na to, kar v našem življenju počnemo ali kako lahko fenomenalni značaj izkustva bolečine vpliva na to, da vzamemo aspirin? Povedano drugače, če obstaja popolna fizična razlaga tega, zakaj se naša telesa premikajo tako, kot se, zakaj sploh rabimo in kako sploh bi subjektivna fenomenalna stanja karkoli povzročala? To imenujemo *problem -vzročne Izključitve*. Pri odgovoru nanj se znajdemo v škripcih: na eni strani se fizične razlage ne moremo kar tako odkrižati, če ne želimo priti navzkriž z dognanji znanosti in dopustiti obstoj nematerialnih »kartezijanskih duš«, na drugi strani pa se ne moremo otresti naših intuicij o mentalnih vzrokih naših fizičnih dejanj, saj nismo pripravljeni govoriti, da npr. pijemo, ker so se sprožili ustrezni nevroni. Rešitev je najti ustrezno *relacijo* med obema vzrokoma, ki bi ohranila tako mentalni realizem, tj. zavestna stanja kot vzročno učinkovita, kot fizični

monizem, tj. obstoj zgolj materialne substance.[11]

Priljubljen fizikalistični predlog, ki naj bi to zagotovil, ponuja kot takšno relacijo *supervenienco*,[12] ki jo je v sodobni filozofiji duha prvi omenil Davidson (1970/1995), trdi pa, da so zgoraj ležeče mentalne lastnosti *določene z ali odvisne od* spodaj ležečih fizičnih lastnosti v smislu, da če organizem realizira katero izmed fizičnih lastnosti, ki recimo predstavljajo popolno bazo žeje, potem realizira tudi to mentalno lastnost, in obratno, če organizem ne realizira nobene izmed lastnosti, ki predstavljajo popolno fizično bazo žeje, ne realizira niti te mentalne lastnosti. Supervenienca vsak mentalni dogodek poveže s fizičnim prostorom tako, da mu priskrbi njegovo bazo; vsak fizični dogodek iz nje pa je nujen in zadosten za pojav mentalnega življenja. S tem prepreči, da bi obstajalo še eno ontološko neodvisno področje, ki bi imelo vzročni vpliv na dogodke iz fizičnega sveta, hkrati pa s tem, ko trdi, da je duh odvisen od telesa, ponudi dobro razlago tega, kako sta mentalni in fizični vzrok istega dogodka povezana. Rekli bi lahko, da zgoraj ležeča zavestna stanja s pomočjo supervenience vzročno moč preprosto podedujejo od spodaj ležečih fizičnih stanj, ki so njihovi realizatorji, res pa je, da se vsi filozofi s tem, da to predstavlja rešitev problema vzročne izključitve niti slučajno ne strinjajo.[13]

(4) Problem prepustnosti

V kakšnem smislu so zavestna stanja prepustna? Vzemimo naslednji primer. Osredotočimo se na moder kvadrat. Intuitivno se brez težav zavemo njegove kvadratosti in modrosti kot nečesa, kar je zunaj v svetu. Gre za lastnosti površine tega predmeta, ki predstavljata *vsebino* našega zavedanja. Zdaj pa se obrnimo navznoter in se skušajmo zavesti golega subjektivnega izkustva modrega kvadrata – neke njegove notranje lastnosti, ki ga razlikuje od ostalih – ne da bi se hkrati zavedli njegovih zunanjih lastnosti (kvadratosti in modrosti) (Tye, 1995: 30).

Zdi se, da to ni mogoče; bolj ko se trudimo, bolj se zavest vrača k izkustvu kvadratosti in modrosti kot skupne lastnosti, uprimerjene na zunanjem predmetu. Ko se obrnemo navznoter, da bi imeli dostop do golega subjektivnega izkustva, tj. nečesa, kar se dogaja znotraj nas, vedno končamo pri osredotočanju na to, kar se dogaja zunaj nas (pri kvadratosti in modrosti), tj. pri njegovi vsebini oziroma, o čem je. Vprašanje je, kako je lahko fenomenalna zavest prepustna na opisan način, tj. kako je možno, da se ne moremo zavedati nobene druge lastnosti našega fenomenalnega izkustva – neke njegove notranje lastnosti, ki bi bila neodvisna od vsega drugega – ampak vedno pristanemo pri tem, o čem je?

Možen odgovor na to vprašanje je naslednji: subjektivno izkustvo predstavlja ve

[11] Za mentalno vzročnost glej npr. Heil in Mele (1995).
[12] Zanjo *glej* npr. Kim (1993).
[13] Za ugovor *glej* npr. Kim (2001).

dno zgolj svojo vsebino ali predmet, ne pa neke svoje notranje lastnosti, zaradi tega, ker je fenomenalni značaj takšnih stanj – 'kako je biti' – identičen z njihovo intencionalno vsebino, tj. občutek kvadratosti in modrosti je npr. identičen s tem, da nekaj vidimo kvadratno in modro (Tye, 1995: 136). (5) Problem superslepote

(5) Problem superslepote

Tako imenovana superslepota je pojav, pri katerem so ljudje zaradi poškodovanih možganov slepi za del vidnega polja, kljub temu pa so pod določenimi pogoji na osnovi informacij, ki jih ta del dobi, sposobni posredovati pravilne izjave o svetu, npr. glede prisotnosti, barve, položaja itd.: pravilno, recimo, ugotovijo, da je prisoten *A* in ne *B*, določijo barvo, pa tudi oko potuje enako kot pri normalnih ljudeh, ko nekaj »zagledajo«. Zdi se, da lahko takšni ljudje, če se dovolj izurijo, kar je logično možno, funkcionirajo kot normalni ljudje ne glede na to, da je del njihovega vidnega polja slep, kar pomeni, da se v njih, ko je ta del »aktiven«, ne dogaja nič fenomenalno izkustvenega. To pa kaže s stališča fizikalizma na dvoje nasprotujočih si dejstev: ali ni zavest s takšnostmi nič posebnega, saj med normalnim in superslepim spoznavalcem ni razlike (kar govori v prid fizikalizmu) ali pa ima funkcionalist star problem v nov preobleki.

Prva možnost izraža problem superslepote, ki je vprašanje, kakšna je razlika med superslepim človekom, ki brez vidnega izkustva verjame, da je npr. *A* prisoten, in normalnim človekom, ki dejansko vidi (izkuša), da je *A* prisoten? Zagovornik fenomenalne zavesti lahko podre analogijo med njima tako, da se sklicuje na štiri vidike, v katerih se razlikujeta, s čimer ohrani posebnost takšnosti in ponovno podvomi o fizikalizmu. Ti so naslednji:

(i) vsebina: izkustvo predstavlja trdno lastnost sveta, prepričanje pa ne, saj je potrebno še nekaj drugega, nepojmovnega, da nekaj dokončno verjamemo;

(ii) mehanizem: izkustvo predstavlja svet s pomočjo matric in mrež, prepričanje pa preko besed in stavkov;

(iii) vloga spomina: izkustvo ne temelji na shranjenih predstavah, prepričanje pa za prepričevanje v resničnost nečesa nima druge izbire;

(iv) kognitivni sistem: izkustvo je medsebojno usklajeno s pomočjo kognitivnega aparata, prepričanje pa je že del tega aparata (Tye, 1995: 142–143). Druga možnost fizikalizmu ni naklonjena, saj obudi vse tegobe, ki jih ima funkcionalizem s problemom dvojnika. Logična možnost obstoja superslepih spoznavalcev, ki se ne razlikujejo od normalnih ljudi, je zanj problematična, saj podobno kot zombiji tudi oni opravljajo funkcionalno vlogo, opisano v pojmih tipičnih vzrokov in učinkov, npr. ko jih vprašamo, kakšne barve je banana pred njimi, postavimo pa jo v njihovo slepo vidno polje, odgovorijo, da rumene, ne da bi sploh imeli kakršnokoli vidno izkustvo, kar pomeni, da se identifikaciji fenomenalnih pojavov z njihovimi vzročnimi vlogami izmuzne kvalitativna značilnost, ki jo po prevladujočem mnenju takšni pojavi tipično imajo, zaradi

česar funkcionalizem odsotnosti takšnosti pri superslepih ljudeh ne more razložiti.

(6) Problem lokacije

Predpostavimo, da imamo v nogi občutek nečesa bolečega, tj. v njej čutimo bolečino, pravimo pa, da nas boli noga. Ampak v resnici v naši nogi ni nobene bolečine, pa naj se nam to zdi še tako očitno, saj je bolečina tako imenovani *mentalni predmet*, katerega obstoj je odvisen od ustrezne živčne aktivnosti in predstavlja vsebino našega občutka.[14] To pomeni, da je naš občutek bolečine zmoten, ker nam pravi, da je v naši nogi bolečina, v resnici pa je tam niti slučajno ni. Vprašanje je, ali fenomenalno izkustvo resnično predstavlja svet na napačen način? (Tye, 1995: 31).

Eden izmed fizikalističnih odgovorov na to je, da ima v kontekstu, v katerem nastopajo zavestna stanja, predlog 'v' drugačen pomen, kot pa v normalnem kontekstu, kjer je razumljen prostorsko, zaradi česar lahko govorimo, da bolečina *je* v moji nogi, čeprav je tam *prostorsko* gledano ni.[15] Block (1983) to zagovarja z naslednjim argumentom:

1. Čutim bolečino v prstu.
2. Moj prst je v mojih ustih.
3. Čutim bolečino v svojih ustih.

Če razumemo V v obeh premisah prostorsko, sklep sledi in argument je veljaven, če pa 'v' v njiju razumemo v običajnem pomenu, sklep ne sledi in argument je neveljaven. To kaže na to, da je V v prvi premisi in sklepu razumljen drugače *Vprašanje je, kako)* kot v drugi premisi, kjer je razumljen prostorsko. Tako naše izkustvo sveta ne predstavlja na nepravilen način, ko govorimo, da je bolečina v moji nogi, saj s tem ne mislimo, da je tam locirana, ampak hočemo povedati zgolj, da je prišlo na tistem mestu do telesne motnje, ki pa so jo zaznali in obdelali možgani.[16]

(7) Problem tujega uda

V primeru »tujega uda« gre za neke vrste psihično motnjo, pri kateri ljudje poročajo, da občutijo svoj ud, kot da sploh ni njihov. Znan je primer moža, ki je prišel v bolnico na preizkuse in je moral nekaj časa ostati tam. Potem, ko ga je medicinska sestra pustila v postelji, mu je nekako uspelo, da je iz nje padel in pristal na tleh. Ko je opisoval, kaj se mu je zgodilo, je dejal, da je, ko se je zbudil, v svoji postelji našel nogo nekoga drugega, pravo, lepo oblikovano stvar, ki jo je

[14] Z vidika rešitve problema prepustnosti bi lahko sedaj rekli, da je mentalni predmet identičen s fenomenalnim občutkom.

[15] V tem primeru se poraja vprašanje, kaj pa potem npr. V, ko se nanaša na izkustvo, sploh pomeni?

[16] Za ugovor *glej* npr. Tye (1995).

lahko prijel v roko. Ker pa je bila čudna in hladna, jo je vrgel ven, ob metu noge iz postelje pa je na tleh nekako pristal tudi sam. In zdaj je noga, čeprav to težko verjame, pripeta na njegovo telo. Ko so mu povedali, da je želel od telesa odtrgati svojo nogo, je bil presenečen in zmeden, saj se mu ni zdela prava in resnična, ni se mu zdela del njegovega telesa (Tye, 1995: 33).

Poanta je, da ljudje s takšno težavo ud lahko identificirajo kot npr. nogo, problem je, da ga ne čutijo kot svojega oziroma se jim zdi, da gre za tujo nogo. Tako se zdi, da značilnost zavestnih stanj ni zgolj to, da npr. občutimo bolečino v nogi, ampak tudi to, da je bolečina v *naši* nogi. Problem tujega uda tako zastavlja vprašanje, kako je lahko subjekt občutka tudi sam vključen v občutek oziroma, kako lahko občutimo bolečino v svoji nogi?

Možen odgovor na to je naslednji: bolečino v svoji nogi občutimo takrat, ko imamo celostno čutno predstavo o svojem telesu, ki dano nogo povezuje z ostalim jazom in s čutno predstavo poškodbe tkiva v delu te noge. Z drugimi besedami, neprestano smo del občutkov, ki potekajo v ozadju našega telesa in nam omogočajo, da imamo o njem celostno predstavo, kar nam zagotavlja, da posamezna fenomenalna stanja občutimo kot svoja.

(8) Problem takšnosti (problem lastništva + problem subjektivnega vidika + problem mehanizma)

(a) Problem lastništva

Predpostavimo, da poleti sedimo na terasi in v svoji roki naenkrat začutimo pekočo bolečino – sosedova čebela je skrenila s svoje običajne poti in nas zamenjala za slasten cvetni nektar.

Recimo, da se enako zgodi našemu prijatelju, ki na terasi poseda z nami, s to razliko, daje njega pičila druga čebela, tudi on pa je na roki začutil pekočo bolečino. Vprašanje je, ali lahko on občuti našo bolečino, mi pa njegovo? Odgovor je seveda ne, le kako bi lahko kdorkoli izmed nas občutil bolečino kogarkoli drugega. Bolečina, ki jo občutimo, je izključno naša, imeti je ne more nihče drug, zaradi česar pravimo, da je *zasebna*, tj. do nje ima dostop zgolj njen lastnik. Res je, da naš prijatelj verjetno svojo bolečino občuti tako, kot mi našo, toda to ni več naša bolečina, ampak njegova. Ideja je, da subjektivnih kvalitativnih izkustev, tj. takšnosti, nikoli ne moremo deliti z drugimi, jaz npr. ne morem imeti vaših vizualnih podob, vi pa ne mojih, vaše vizualne podobe pripadajo vam, moje pa meni.

To pa je za fizikalista, ki trdi, da so takšnosti fizične lastnosti, problematično, saj mora razložiti, kako so lahko *mentalni predmeti* izkustva, npr. bolečina, žeja, lakota itd., zasebni, če pa po njegovi teoriji niso nič drugega kot fizične stvari, ki so javne: če so javne fizične stvari, zakaj jih ne more imeti nihče drug razen mene in zakaj ne more obstajati podoba, srbenje ali paslika, ki ne pripada

nikomur. Problem lastništva, ki ga mora fizikalist rešiti, je tako dvojen:

(i) fizičnih stvari ne posedujemo na enak način kot mentalne predmete: moje kolo imate lahko tudi vi, mojo kapo imate lahko tudi vi, moj stol imate lahko tudi vi itd., medtem ko moje bolečine, moje žeje, moje lakote pa ne;

(ii) mentalni predmeti imajo vedno lastnika (so 'od nekoga'), fizične stvari pa lahko obstajajo tudi brez njega: vsaka podoba, srbenje ali paslika vedno nujno nekomu pripada, medtem ko pes, drevo ali otrok ne.

Fizikalist ima kot odgovor na voljo vsaj dve možnosti, obe pa namigujeta na to, da med mentalnimi predmeti izkustva in fizičnimi stvari sveta ni razlik:

(a) trdi lahko, da so tudi nekatere fizične stvari nujno zasebne, npr. moj smeh, moj krik ali moj jok itd., zaradi česar zasebnost ne more biti dokaz, da nekaj ni fizično;

(b) zanika lahko, da so mentalni predmeti nujno zasebni, saj je logično možno, da obstajata dve bitji, ki imata isto bolečino ne samo kvalitativne ampak tudi številčno (Unger, 1990).

(b) Problem subjektivnega vidika

Vrnimo se na primer s čebelo in izkustvo, ki smo ga doživeli ob njenem piku. Zdi se, da je za popolno razumevanje tega specifičnega občutka, ki se pojavi, potrebno vedeti, kako je, če si takšnemu izkustvu podvržen. Vedeti, kako je, ko te piči čebela, pa zahteva, da smo v stanju na to gledati z določenega vidika: ali smo bili kot v našem primeru sami žrtev čebeljega pika in smo imeli podobno fenomenalno izkušnjo ali pa si jo lahko zamislimo na podlagi drugih bolečin, ki smo jih občutili v času našega življenja, ne da bi nas kdajkoli pičila čebela.

Skratka, za to, da bi lahko zavestna stanja v celoti razumeli, moramo nanje pogledati s subjektivnega vidika. Problem za fizikalista je, da fizična stanja, kakršna naj bi bila tudi zavestna stanja, niso takšna, tj. njihovo popolno razumevanje ne vključuje nobenega subjektivnega vidika. Da bi npr. v celoti razumeli, kaj je blisk ali zlato, nobenega od njiju ne rabimo sami izkusiti. Tudi slepec lahko v celoti doume blisk ali zlato: prvi je sprostitev električne napetosti med oblaki, drugo pa element z atomskim številom 79. Zaradi tega fizična stanja niso subjektivna, ampak objektivna (Tye, 1995: 13).[17]

Naloga fizikalista, ki pravi, da so izkustva in občutki fizični, je tako pojasniti, kako lahko na objektiven način, tj. s fizičnimi pojmi opišemo nekaj, kar je izkustveno in od nas zahteva subjektivni vidik. Zateče se lahko k odgovoru, ki pravi, da je za to, da bi vedeli vse o bolečini, dovolj, da vemo, da je to in to možgansko stanje, ne pa, kako jo je izkusiti, sicer pa bomo o tem in njegovih

[17] Zaradi tega npr. naš zombi, ki ni sposoben čutiti, nima nobene možnosti, da bi kadarkoli doumel pojem takšnosti, saj to od njega zahteva vednost o tem, 'kako je biti', ki pa je nekaj subjektivnega, nekaj, kar njemu znana objektivna dejstva ne vključujejo.

drugih možnostih več govorili v zadnjem delu članka.

(iii) Problem mehanizma

Videli smo, da naj bi bila fenomenalna stanja subjektivna, možganska stanja pa objektivna, zato je eno izmed vprašanj za fizikalista (McGinn, 1991), kako lahko sivo-bela snov naših možganov ustvari bogato zavestno izkustvo. Odgovor na to vprašanje zahteva obstoj nekega posebnega mehanizma, ki je odgovoren za produkcijo subjektivnih stanj s strani objektivnih in ki zapolni vrzel, ki jo čutimo med enimi in drugimi. Za njeno ponazoritev vzemimo naslednji miselni eksperiment Predpostavimo, da so znanstveniki razvili napravo, ki jo lahko priključimo na našo glavo ter nam omogoča, da vidimo fizične spremembe v naših možganih, npr. ko premaknete roko, vidite na ekranu proženje nevronov s področja *A*, ko vas nekdo požgečka, vidite na ekranu proženje nevronov s področja *B*, ko vas nekdo udari vidite na ekranu proženje nevronov s področja C itd. Ko se ne zgodi nič od tega, pa omenjeni nevroni mirujejo. Ali se vam ne bi zdelo neverjetno, da električna aktivnost vaših možganov proizvede subjektivni občutek žgečkanja? (Tye, 1995: 15). Problem mehanizma opisujemo tudi kot vprašanje zapolnitve tako imenovane *pojasnjevalne vrzeli*, kako določeno proženje nevronov sploh povzroči občutek in zakaj to občutimo tako in ne drugače? O njej je Huxley leta 1866 zapisal naslednje: »... [K]ako je lahko nekaj tako posebnega, kot so zavestna stanja, rezultat draženja živčnega tkiva, je tako nemogoče dokazati, kot pojav duha, ko mu Aladin ukrade svetilko« (Huxley, 1866: 193).

Potreba po mehanizmu izhaja iz tega, da je to običajen način, kako pojasnjujemo dogodke v svetu, zaradi česar se od nas pričakuje, da tako razložimo tudi nastanek zavesti. Tye (1995), recimo, navede tri dejanske primere iz znanosti, ki to potrjujejo: lomljivost, tekoče stanje in prebavo. Kako se npr. steklo zlomi? Pritisk na krhke vezi med kristali, iz katerih je sestavljeno, povzroči njihovo tresenje in zlom. Zakaj je npr. voda tekoča? Ker molekule v njej nimajo fiksne lokacije, ampak s premikanjem menjavajo svoja mesta. Kako prebavljamo? Encimi poskrbijo za to, da se hrana razgradi na preproste kemične komponente, ki telesu zagotovijo energijo za njegovo delovanje. Vsi primeri kažejo na to, da obstajajo v naravi na nižjih ravneh (nevrofiziološki, biološki, kemični) mehanizmi, s katerimi pojasnjujemo pojav stanj, dogodkov ali procesov na višjih ravneh (psihološki, sociološki ali ekonomski). In če je zavest možganski pojav, mora tudi zanjo na fizični ravni obstajati takšen mehanizem, ki predstavlja povezavo med subjektivnim in objektivnim.

Vprašanje je, ali obstaja? Intuicija nam pravi *ne*, znanost pa ga ne more dokazati. Ne glede na to, kako globoko v kemične procese, ki se odvijajo, ko se prožijo nevroni, lahko prodremo, še vedno se nam zdi, da ostane nekaj, česar s tem ne moremo pojasniti, tj. kako ta niz kemičnih sprememb sploh proizvaja kakršenkoli subjektivni občutek in zakaj ravno ta subjektivni občutek in ne katerega drugega.

»Lahko raziskujemo živčne mehanizme in korelate zavesti in to bo nedvomno obogatilo naše znanje o tem, kako so mentalna stanja odvisna od kemičnih in drugih značilnosti možganov. Vendar tega ne smemo zamenjevati z reševanjem globokega pojmovnega problema, in sicer pojasnitvijo, kako možgani izpeljejo proizvodnjo zavesti. Vemo, *da* to storijo, pa tudi precej podrobnosti o vključenih vzročnih korelacijah in odvisnostih poznamo, ne razumemo pa, *kako* to storijo« (McGinn, 1999: 69).

Možnosti fizikalista sta dve:

(a) lahko zanika, da obstaja kakršenkoli mehanizem, ki zagotovi zvezo med subjektivnim in objektivnim, in trdi, da je preprosto *golo dejstvo*, da nekatera možganska stanja realizirajo nekatera zavestna stanja, ter da tukaj ni nadaljnje razlage tega, *kako* se to zgodi.[18] »Med zavestjo in možgani ne obstaja 'zveza' nič bolj kot med tekočim stanjem vode in molekulo H2O. Če je zavest lastnost višjega reda možganov, potem vprašanja o obstoju zveze med lastnostjo sistema in tem sistemom ne more biti« (Searle, 1992: 105).

(b) lahko sprejme obstoj takšnega mehanizma, hkrati pa trdi, da je to, da bi ga razumeli, onkraj naših kognitivnih zmožnosti. Zakaj pa tudi mi podobno, kot se to dogaja ostalim bitjem v naravi, nečesa ne bi bili sposobni razumeti? »Vemo, da imamo duha in vemo, da so možgani zanj nekako odgovorni, ne vemo pa, kaj je tisto glede možganov, kar povzroči nastanek zavesti. Mehanizem zavesti je misterij« (McGinn, 1999:197). Ker je problem takšnosti, tukaj definiran kot vsota problema lastništva, problema subjektivnega vidika in problema mehanizma, najtežji problem zavesti, ki ga mora fizikalist rešiti, če želi ohraniti svojo ontološko doktrino in ker se omenja kot tisto, kar ljudje največkrat navajajo kot nekaj, kar kakršnemukoli stroju, računalniku, sistemu, robotu ali androidu manjka, se bomo v nadaljevanju osredotočili zgolj nanj, ostale pa pustili za kakšno drugo priložnost.

3 Napad na fizikalizem: netopir in Marija

Kot smo v napovedali v uvodu, je naša strategija za potrditev tega, da lahko fenomenalno zavest pripišemo tudi strojem, najti prepričljiv argument za to, da so takšnosti fizične, saj bi v tem primeru, da bi se izognili šovinizmu vrst, fenomenalno zavest morali pripisati tudi njim. Vendar nam na poti do uspeha stojita dva znana argumenta, ki sta bila prvič predstavljena v 70. in 80. letih 20. stol, in ki dokazujeta ravno nasprotno, tj. da so takšnosti nefizične, s čimer podvomita o fizikalizmu in s tem namigujeta na to, da obstaja del človekovega mentalnega življenja, ki nima materialnega sidra.

Kako je biti netopir?

[18] Podobno je vznik lastnosti višjih ravni iz lastnosti nižjih ravni svojčas zagovarjal *emergentizem*.

Nagel (1974/1990) na primeru netopirja dokazuje, da je problem subjektivnega vidika za fizikalizem nerešljiv, saj so vse funkcionalne analize mentalnega nezdružljive s prisotnostjo takšnosti, tj. s tem, 'kako je biti'. »Če naj zagovarjamo fizikalizem, morajo tudi fenomenološke strukture pridobiti fizikalne sodbe. A ko preiskujemo njihov subjektivni značaj, se zdi, da je takšen rezultat nemogoč. In to zato, ker je vsak subjektivni pojav bistveno povezan z enim gledišče, in se zdi neizogibno, da bo objektivna, fizikalna teorija to gledišče opustila« (Nagel, 1974/1990: 401). Pravi, da nikoli ne moremo izkusiti, 'kako je biti' netopir, tj. kako netopir »vidi« svet, saj nimamo za to na voljo nobenih drugih izkušenj razen lastnih, te pa nam pri tem ne koristijo, ker so posledica našega čutnega aparata, ki se povsem razlikuje od netopirjevega, saj imam mi za obdelavo vizualnih podatkov na volj oko, on pa sonar.

Še več, kljub temu, da lahko netopirjeve možgane, njegov živčni sistem in njegov zaznavni aparat do potankosti razume tudi človek, tj. glede funkcioniranja njegovih možganov vemo vse, da lahko razumemo, kako »vidi« svet, pa tega, 'kako je biti' netopir, ne vemo, četudi bi se nekaj časa obnašali tako kot on in bi dneve preživljati viseči z glavo navzdol, ponoči pa bi na pol slepi lovili mrčes. V tem primeru bi vedeli zgolj, kako je nam biti netopir, nas pa zanima, kako je njemu biti netopir. Tega pa ne moremo vedeti, če ne zasedemo posebnega gledišča, tj. gledišča prve osebe. Res je, da lahko naša fenomenalna stanja poznajo tudi drugi subjekti pripisovanja, vendar le pod pogojem, da so nam dovolj podobni, kar pa za netopirje seveda ne velja.

Nagel zaključi, da »Če oznanimo, da mora fizična teorija duha upoštevati subjektivni značaj izkustva, moramo priznati, da nam nobena od koncepcij, ki so nam na voljo sedaj, ne razlaga, kako je to mogoče« (Nagel, 1974/1990: 407). Razen tega bistveno vprašanje ni, kako zapolniti vrzel med subjektivnim in objektivnim, ampak, ali je razglabljanje o tem, da ima izkustvo objektivni značaj, sploh na mestu. »... [A]li ima smisel spraševati, kakšna so izkustva v *resnici*, v nasprotju s tem, kakšna se mi kažejo?« (Nagel, 1974/1990: 409).

Odgovor v prid fizikalizmu pravi, da se argument sicer zdi prepričljiv, vendar fizikalizma ne spodbije, saj ta ne trdi, da si moramo biti sposobni zamisliti takšnosti, ampak zgolj, da so te fizične.[19] Trditev, da so vsi organizmi, tudi netopirji, fizična bitja, nas namreč ne zavezuje k trditvi, da moramo v primeru, ko vemo vse o njihovi nevrofiziologiji, vedeti tudi, kako je v resnici biti netopir ali neko drugo bitje. Normalno, da imamo ljudje težavo zamisliti si občutke bitja, ki svet v osnovi zaznava povsem drugače od nas, tj. s pomočjo lociranja odmevov, vendar to še ne dokazuje, da je fizikalizem neresničen.

[19] Resnici na ljubo je treba priznati, da Nagel ne misli, da je katerikoli fizikalizem neresničen, ampak samo fizikalizem, kot ga razumemo sedaj oziroma danes: »Napačno bi bilo sklepati, da mora biti fizikalizem v zmoti. Neustreznost fizikalističnih hipotez, ki stavijo na lažno objektivne analize duha, ne dokazuje ničesar. Bliže resnici bi bilo reči, da je fizikalizem stališče, ki ga ne moremo razumeti, ker za zdaj nimamo še nobene koncepcije, kako naj bo resničen« (Nagel, 1974/1990: 407–408).

Zagovornik tega, da so takšnosti nekaj nefizičnega, pa lahko v luči omenjenega svoj argument zaostri tako, da vključi bitja, občutke katerih si lahko zamislimo, in Še dalje vztraja pri tem, da fizikalizem ne drži, kar prikazuje naslednji miselni eksperiment.

Marijina soba

Jackson (1982/1999, 1986) v okviru problema subjektivnega vidika dokazuje,

obstajajo telesne značilnosti, takšnosti, ki jih ne moremo spoznati oziroma opisati zgolj s fizikalnimi količinami, zaradi česar je fizikalizem teorija, ki ji ne uspe pojasniti vseh fenomenov našega sveta. »Povejte mi vse fizično, kar je treba povedati o tem, kaj se dogaja v živih možganih, o vrstah možganskih stanj, o njihovih funkcionalnih vlogah ... toda kljub temu mi ne boste mogli ničesar povedati o neugodju bolečine, o srbenju srbeža, o muki ljubosumja ali pa o značilnih izkušnjah okušanja limone, duhanja vrtnic, poslušanja glasnega ropota ... (Jackson, 1982/1999: 68). Idejo, da fizikalno znanje (znanje ene vrste) samo po sebi še ne zadostuje za fenomenalno znanje (znanje druge vrste), imenujemo *intuicija o znanju* in je tisto, na čemer temelji njegov argument zoper fizikalizem. Ta je danes najbolj znan pod imenom *argument iz znanja*, njegovo ozadje pa tvori

miselni eksperiment z nevrofiziologinjo Marijo, ki je iz nekega razloga zaprta v črno-beli sobi in raziskuje svet iz nje preko črno-belega televizijskega ekrana. Specializira se za vid in sčasoma pridobi o mehanizmu, ki ga uporabljamo za obdelavo vizualnih informacij in o tem, kako tvorimo vidno izkustvo, vse znanje, ki ga je možno pridobiti o teh stvareh s fizikalnega vidika. Ugotovi npr. kateri razpon valovnih dolžin vzdraži našo mrežnico tako, da vidimo »rdeče« in kako osrednji živčni sistem vpliva na krčenje glasilk in izdih zraka iz pljuč, ko rečemo »Paradižnik je rdeč«. Ključno vprašanje, ki ga sedaj Jackson zastavi, je, ali se bo Marija, ko bo zapustila črno-belo sobo in prvič v življenju zagledala rdeč paradižnik, *naučila* kaj novega?

Njegov odgovor je prej ko slej *da*, kar pa pomeni, da obstaja še nekaj, česar fizikalizem ne vključuje, kar ga dela neresničnega. »Zdi se očitno, da se bo naučila nekaj o svetu in naših vizualnih doživetjih sveta. Toda potem moramo reči, da je bilo njeno prejšnje vedenje nepopolno. Vendar pa je v sobi imela *vse* vizualne informacije. *Torej*, obstaja še nekaj več od tega, kar je treba imeti, in zato je fizikalizem zmoten« (Jackson, 1982/1999). *Argument iz znanja (v* njegovi izvirni obliki) v standardnem zapisu različni avtorji povzemajo različno, vsi pa ohranjajo njegovo bistvo, tj. iz intuicije o znanju sklepajo na zmotnost fizikalizma oziroma iz resničnosti epistemološke predpostavke na neresničnost ontološke:

1. Pred svojo izpustitvijo, ima Marija *vso fizikalno znanje*, ki ga je mogoče imeti o zaznavi rdeče barve.

2. Pred svojo izpustitvijo, Marija nima *vsega fizikalnega znanja*, ki ga je mogoče imeti o zaznavi rdeče barve (ker se po izpustitvi o njej nekaj nauči). () 3. Obstajajo nekatere resnice o zaznavi rdeče barve, ki jih fizikalizem ne zajema.

4. Fizikalizem je neresničen in fenomenalne lastnosti ne morejo biti pojasnjene kot fizikalne lastnosti ali z njimi identificirane (Van Gulick, 2004: 369–370).

Ali:

1. Marija ima, pred svojo izpustitvijo, popolno fizikalno znanje o človeškem videnju barv.

2. Obstajajo fizikalne informacije o človeškem videnju barv, ki jih pred svojo izpustitvijo nima.

∴ 3. Vse informacije niso fizikalne informacije.

∴ 4. Fizikalizem je zmoten (Nida-Rümelin, 2010).

Ali:

1. Vsaka fizikalna resnica je takšna, da jo Marija (pred izpustitvijo) pozna.

2. Ne drži, da je vsaka fizikalna resnica takšna, da jo Marija (pred izpustitvijo) pozna.

∴ 3. Obstaja vsaj ena resnica, ki je nefizikalna.

∴ 4. Fizikalizem je neresničen (Stoljar, Nagasawa, 2004: 14).

Če takšen argument, kjer premise izražajo *intuicijo o znanju*, sklep pa *neresničnost fizikalizma*, upravičeno imenujemo argument iz znanja, ga najdemo že npr. leta 1925 pri Broadu,[20] ki dokazuje, da bi tudi v primeru, ko bi bile mehanicistične teorije kemije resnične, še vedno obstajala lastnost amonijaka, tj. njegov vonj, ki ga niti bitje z neomejenim znanjem o njegovi atomski mikrostrukturi ne bi moglo napovedati. »Kljub temu, da bi natanko poznal mikroskopsko zgradbo amonijaka, še vedno niti slučajno ne bi mogel napovedati, da ima substanca s takšno zgradbo, ko jo zavohamo, takšen vonj, kot ga ima amonijak« (Broad, 1925: 71). Ni pa jasno, kako bi bil argument iz znanja tudi Nagelov »kako je biti netopir«, ki se od Jacksonove »Marijine sobe« bistveno razlikuje: Nagel pravi, da nobeno fizikalno znanje o netopirju ni dovolj, da bi vedeli, ‘kako je biti’ netopir, Jackson pa, da kljub vsemu fizikalnemu znanju, ki ga imamo o barvnem izkustvu, še vedno obstaja neka njegova lastnost, ki smo jo prezrli. In ko jo bomo spoznali, še vedno ne bomo vedeli, kako je npr. biti Marija (ni jasno, zakaj Nagel od nas, če smo fizikalisti, sploh zahteva, da bi to morali vedeti), res pa je, da bomo o nje vedeli več, kot smo do sedaj.

[20] *Glej* tudi Feigl (1958).

Kakorkoli, v naši razpravi smo sedaj prišli do točke, ko se moramo, če želimo biti uspešni pri zagovarjanju (možnosti) obstoja zavestnih strojev, spopasti z argumentom iz znanja in (vsaj tako se zdi) obraniti fizikalizem, bolj ali manj uspešne strategije pri tem, pa si bomo ogledali v naslednjem delu.

4 Obramba fizikalizma

Videli smo, da je znotraj problema takšnosti za fizikalizem najtrši oreh problem subjektivnega vidika, ki ga tako rekoč pooseblja argument iz znanja, poenostavljeno povedano pa gre za to, da iz intuicije o znanju sledi neresničnost fizikalizma. Strategije njegove obrambe so različne, vse pa se nanašajo na dve trditvi, ki jih vsebuje naša intuicija o znanju in ki ju miselni eksperiment z Marijo nazorno definira. To sta:

(i) trditev o popolnosti znanja: preden zapusti sobo, pozna Marija vse, kar je fizikalnega;

(ii) trditev o učenju: potem, ko Marija sobo zapusti, se nauči nekaj novega.

Če bi fizikalist uspel pokazati, da je katera izmed njiju neresnična, bi s tem podvomil o naši intuiciji o znanju, s tem pa bi preprečil tudi sklep, da je fizikalizem neresničen, ki v argumentu iz znanja iz nje sledi. Ker je verjetno najlažji način za prikaz takšne fizikalistične strategije tisti, ki ga je predstavil Van Gulick (2004), organiziran pa je okrog vprašanj, ki jih zastavlja argument iz znanja ter odgovorov nanje, si v nadaljevanju oglejmo nekatere izmed njih.

(i) Ali se Marija nauči kaj novega?

Kot pravita Stoljar in Nagasawa (2004), je trenutek, ko se je Jackson iz vnetega zagovornika takšnosti spreobrnil v gorečega fizikalista, pretresel filozofsko skupnost do te mere, da je o tem pod naslovom »Frank Jackson si je premislil« pisal celo »Philosophy News Service«. Njegova dokončna preobrazba v zvezi z argumentom iz znanja se je zgodila leta 1998, ko je v svoji knjigi *Mind, Method, and Conditionals*, v kratkem besedilu »PostScript on Qualia«, na zgornje vprašanje odgovoril z ne in obelodanil, da je zaznavna stran psihologije, kot je določena v našem svetu, načeloma izpeljiva iz njegove fizikalne narave (Jackson, 1998/2004: 418).

Jacksonova zavrnitev argumenta iz znanja sedaj temelji na prepričanju, da je to, da se Marija nauči o naravi duha nekaj, česar prej ni poznala, zaradi česar mora to biti nekaj več od fizičnega, zgolj videz. »Zaznavno izkustvo se nam predstavlja, kot da bi bilo pridobivanje podatkov o intrinzični naravi ... To je, mislim sedaj, izvor močne, a napačne intuicije, da se Marija po izpustitvi o tem, kakšne so stvari, nauči nekaj novega« (Jackson, 1998/2004: 419). V resnici gre zgolj za to, da Marija na takšen način, tj. preko doživljanja takšnosti, zbira relacijska in funkcionalna dejstva o tem, kaj se v njej dogaja. To pomeni, da je narava fenomenalne zavesti reprezentacijska, zaznavno izkustvo pa zgolj hiter in

enostaven način pridobivanja relevantnih relacijskih in funkcionalnih podatkov o svetu in *nič več*. Vidimo, da Jackson pri ugovoru argumentu iz znanja napade drugo trditev intuicije o znanju in preprosto pravi, da se Marija, ko zapusti sobo, ne nauči nobenega novega intrinzičnega dejstva o svojem duhu: gre zgolj za učinkovit mehanizem fenomenalne zavesti pri predstavljanju vzročnih relacij v svetu. V enem stavku: »Jackson namiguje na to, da je ideja o tem, da se Marija, ko pride ven, nauči nekaj novega, iluzija, ki je posledica dejstva, da je nova informacija predstavljena kot informacija, ki se nanaša na intrinzično lastnost, namesto kot informacija, ki opisuje relacijsko lastnost« (Stoljar, Nagasawa, 2004: 18).

Kakorkoli, večina kritikov argumenta iz znanja se z Jacksonovim odgovorom, da se Marija po izpustitvi ne nauči ničesar novega, ne strinja, prepričani so, da pridobi neko novo znanje, ugovarjajo pa temu, da bi to znanje predstavljajo neka nova dejstva o svetu in na takšen način blokirajo izpeljavo sklepa iz premis.

(ii) Ali pridobi zgolj »znanje, kako«?

Lewis je verjetno najbolj znan nasprotnik argumenta iz znanja, ki se sicer strinja s tem, da se Marija po izpustitvi iz sobe nauči nekaj novega, trdi pa, da tisto, kar se nauči, ni novo znanje o dejstvih,[21] ampak zgolj neka nova zmožnost.[22] Vse, kar pridobi, ko npr. prvič vidi rdeč paradižnik, je *znanje, kako* (ne pa *znanje, da)*, tj. spretnost in sposobnost prepoznati ter spomniti se enih mentalnih stanj kot npr. izkustva rdečega, drugih kot izkustva bolečine, tretjih kot izkustva srbenja itd. Marija je, ko vidi rdeč paradižnik, zanesljivo v drugačnem zavestnem stanju, kot prej, ko je o njegovi barvi zgolj brala, saj prej njeno možgansko stanje ni bilo identično z zaznavnim izkustvom rdečega. Zdaj pa ji znanje o tem, kako je videti rdeče, omogoča, da ga prepozna vedno, ko se pojavi in ga na tej osnovi tudi loči od kateregakoli drugega barvnega izkustva. »Tipično je, da ko vidimo ali slišimo nekaj novega, pridobimo neko sposobnost, da se kasneje tega spomnimo in da to prepoznamo kot isto, ko se ponovno pojavi« (Van Gulick, 2004: 375). Ker gre za neko spretnost, ki je Marija prej ni imela – noben, še tako podroben fizikalni opis, ji ni omogočal prepoznati pojava fenomenalnih stanj v njenih možganih »od znotraj« —, takšen odgovor imenujemo *odgovor iz sposobnosti*.

Odgovor iz sposobnosti temelji na tem, da se tega, kako je imeti neko izkustvo, ne naučimo izključno s pomočjo fenomenalnih podatkov – imenujmo to *fenomenalna hipoteza* ampak tudi tako, da smo si sposobni to izkustvo, kadar nas je volja, priklicati v spomin, kar imenujemo *hipoteza o sposobnosti [the ability hypothesis)*. Lewis jo opredeli takole: »Hipoteza o sposobnosti pravi, da vedeti. kakšno je neko izkustvo, ni nič drugega kot posedovanje sposobnosti, kot sc spominjanje, zamišljanje in prepoznavanje. Ni posedovanje nobenih podatkov, niti običajnih niti posebnih. Ni vedeti, da nekatere sposobnosti niso uporabljene.

[21] »Znanje o dejstvih je znanje o resnicah, ne o lastnostih« (Conee, 1994/2004: 202).
[22] *Glej* tudi Nemirow (2007).

Tc ni znanje, da. To je znanje, kako« (Lewis, 1988/2004: 100). Vedeti, kakšno je ne: ko izkustvo, je biti sposoben, da se ga spomniš, da si ga zamisliš in da ga prepoznaš. Marija po izpustitvi torej dejansko pridobi nekaj novega, tj. sposobnost predstavljanja in prepoznavanja čutnega izkustva, ne pridobi pa, ko npr. zagleda rdeč paradižnik, nobenega novega dejstva o svetu. Zaradi tega Lewis zavrača sklep argumenta iz znanja, ki pravi, da je fizikalizem zmoten, saj se nepopolnost Marijinega znanja nanaša izključno na *znanje, kako*, ki ga prej res ni imela, ne pa tudi na *znanje, da*, kar pa ne dokazuje, da je njenemu fizikalnemu znanju z vidika poznavanja dejstev o svetu prej kaj manjkalo. Kakorkoli, ne zdi se sporno, da Marija, ko prvič vidi nekaj rdečega, pridobi praktično spretnost vedeti, kako to je, zaradi česar je vsaj del tega, kar se nauči, *znanje. kako*, vprašanje je samo, ali je to *vse*, kar se nauči, pri čemer pa so epistemološke intuicije, ki jih filozofi imajo, tako različne, da soglasja o tem še zdaleč ni na vidiku[23] (Van Gulick, 2004: 375).

(iii) Ali se samo seznani z nečim?

Druga možnost nasprotnika argumenta iz znanja, ki misli, da se Marija sicer nauči nekaj novega, ni pa to, kar pridobi, znanje o dejstvih, je sklicevanje na to, da se po izpustitvi v prvi osebi zgolj seznani s tem, kako je videti nekaj rdečega.[24] Conee (1994/2004) zato poleg *znanja, da* in *znanja, kako*, loči še *znanje s seznanitvijo*, ki je tretja možnost, kako se naučimo tega, kako je imeti čutno izkustvo, predpostavlja pa, da imamo v nekem smislu *novo znanje o starih dejstvih*. Tako lahko z vidika znanja o tem, kako je biti', poleg hipoteze o sposobnosti in fenomenalne hipoteze govorimo še o *hipotezi o seznanitvi [the acquaintance hypothesis]*, ki pravi, da vedeti nekaj s seznanitvijo ni nič drugega, kot biti s tem v neposrednem čutnem stiku, ki nam omogoča, da se tega zavedamo (Conee, 1994/2004:198).

> In ker je zaznavno izkustvo najbolj neposreden način razumevanja lastnosti, za katerega je bila Marija prej, ko je bila zaprta v črno-beli sobi prikrajšana, to znanje pridobi šele potem, ko jo zapusti. To pa ne pomeni, da se je Marija naučila nekaj novega o dejstvih sveta, ampak zgolj, da se je s tem *znanjem, da*, ki ga je pridobila že prej s pomočjo teoretičnega izpeljevanja, sedaj seznanila na nov način. »Skladno s tem odgovorom je Marija poznala vsa fizikalna dejstva, ni pa vedela, kako je nekaj videti rdeče. To znanje je pridobila preprosto s tem, da je imela pravo vrsto izkustva, ne pa tako, da je pridobila neke nove [nefizikalne] informacije« (Conee, 1994/2004: 203). Njegov argument gre takole:
>
> 1. Takšnosti, npr. *P*, so fizikalne lastnosti izkustva (in izkustvo je fizikalni proces).

[23] Za ugovore *glej* npr. Conee (1994/2004), Gertler (1999), Tye (2000).
[24] *Glej* npr. Churchland (1989/2004), Bigelow, Pargetter (1990/2004), Tye (2009).

2. Marija ve vse o lastnosti *P* (in že pred izpustitvijo ve, da ima neko izkustvo lastnost P).

3. Po izpustitvi se seznani z lastnostjo *P* (saj pride z njo v neposreden čutni stik).

4. Marija je že prej vedela, pod katerimi pogoji imajo normalni spoznavalci izkustvo z lastnostjo *P, "*.

5. Pri seznanitvi z lastnostjo *P* ne pridobi novega *dejstva/podatkov* o svetu, q. *znanja*, *da* (ampak, to lastnost, za katero tako že ve, da je resnična, zgolj izkusi na lastni koži). Copeland (1993) v prid temu, da se Marija po izpustiti vi iz črno-bele sobe zgolj z nečim seznani, ne nauči pa se novih dejstev o svetu, navaja primer z nizozemskim arboristom Van Plankom, ki pozna absolutno vsa dejstva o novozelandskem drevesu kauri, ni ga pa še nikoli v življenju videl. Vprašanje je, ali je bilo njegovo znanje o drevesu kauri pred tem nepopolno? Copeland odgovarja, da v nekem smislu da, v nekem smislu pa ne, in da rabimo za razjasnitev tega semantično distinkcijo, ki jo ponazarja trditev »ne poznam ga, vendar vem veliko o njem« (Copeland, 1993: 177), ki namiguje na to, da obstaja znanje v smislu osebnega poznavanja (izraža jo njen prvi del) in znanje v smislu posedovanja dejstev (izraža jo njen drugi del). Z vidika argumenta iz znanja ima tako Jackson prav, ko pravi, da je bilo Marijino znanje nepopolno, manjkalo ji je znanje v smislu osebnega poznavanja, vendar iz tega še ne sledi, da je fizikalizem neresničen, saj to ne dokazuje, da obstajajo dejstva o svetu, ki nimajo fizikalne narave (nepopolnost posedovanja dejstev).

Kakorkoli, še zdaleč ni jasno, da je to vse, česar se Marija nauči, pa tudi definiranje seznanitve je kočljiva zadeva, tako da ponovno ni pričakovati, da bi med filozofi vladalo soglasje o tem, ali se res zgolj seznani s starimi dejstvi na nov način ali pa po izpustitvi resnično pridobi nove informacije o svetu.[25]

(iv) Ali se nauči novo dejstvo ali gre za staro dejstvo v *novi preobleki?*

Številni fizikalisti menijo, da sklicevanje na trditev o sposobnosti in trditev o seznanitvi ni pravi odgovor na argument iz znanja in da Marija potem, ko zapusti črno-belo sobo, kljub vsemu pridobi novo *znanje, da* oziroma se nauči novo dejstvo o svetu, ki pa ni nič drugega kot drugačen opis tega, kar je poznala že prej, zaradi česar gre bolj za staro dejstvo v novi preobleki, kot pa za nekaj resnično novega. kar fizikalizem še vedno ohranja pri življenju.[26]

Obstajajo sicer različne različice takšnega odgovora, vsem pa je skupno to, da so fizikalni pojmi in fenomenalni pojmi *spoznavno neodvisni* (Nida-Rümelin, 2010), tj. če poznamo fizikalni opis fenomenalnega značaja (tega, 'kako je biti'), to še ne pomeni, da a priori poznamo tudi njegov fenomenalni opis. Zaradi tega je možno, da ve Marija vse, kar je fizikalnega o rdečosti, tj. pozna rdečost pod

[25] Za ugovor *glej* npr. Gertler (1999)
[26] *Glej* npr. Loar (1990/2004), Nida-Rümelin (1995/2004,) Papineau (2002), Byre (2002), Levine (2007).

fizikalnim pojmom, ne da bi vedela karkoli fenomenalnega o njej, tj. ne da bi rdečost poznala pod fenomenalnim pojmom.

Horgan (1984) je bil eden izmed prvih, ki je tako odgovoril na argument iz znanja: ko Marija zapusti sobo in izkusi rdečost, ki je fizikalna lastnost izkustva, se seznani z njeno takšnostjo v prvi osebi, s čimer pridobi *fenomenalni pojem* rdečosti, s pomočjo katerega lahko potem tvori nova prepričanja o njej. Kljub temu, da obstaja očitna podobnost med tem odgovorom in prejšnjim odgovorom s seznanitvijo, pa je med njima tudi razlika: seznanitev s takšnostjo nekega izkustva, npr. rdeče barve, nam tukaj omogoča, da oblikujemo nov pojem rdečosti in pridobimo novo znanje o njej s fenomenalnega vidika. Seveda pa takšen odgovor namiguje na to, da se Marija po izpustitvi resnično nauči nova dejstva: prej je po znala vsa dejstva o rdečosti s fizikalnega vidika oziroma pod fizikalnim opisom, tj. imela je fizikalni pojem rdečosti, zdaj pa spozna še nova dejstva o rdečosti, in sicer s fenomenalnega vidika oziroma pod fenomenalnim opisom, tj. pridobi nov fenomenalni pojem rdečosti, s katerim oblikuje prepričanja in tako pridobiva novo *znanje, da*.

Vendar takšen odgovor ni zagovor argumenta iz znanja, kot bi se lahko nekomu zdelo, saj je tisto, kar se Marija nauči, ko zapusti črno-belo sobo, tj. vsako novo fenomenalno prepričanje o danem izkustvu, ki ga tvori iz prve roke, identično enemu izmed njenih starih fizikalnih prepričanj o tem izkustvu, ki ga je imela iz tretje roke in ima *isto* vsebino (Papineau, 2002), zaradi česar ne drži, da se Marija nauči novo dejstvo, ampak gre zgolj za staro dejstvo v novi preobleki, tako da fizikalizem ne utrpi nobenih posledic. Argument v prid takšnemu stališču, ki ga lahko imenujemo *stališče o novem znanju, a starih dejstvih*, je naslednji:

1. Fenomenalni značaj, tj. fenomenalna rdečost (takšnost), je *fizikalna lastnost* izkustva.

2. Pridobitev znanja o tem, kako je imeti izkustvo določenega fenomenalnega značaja, zahteva pridobitev fenomenalnega pojma o njem.

3. Kaj je za organizem pridobitev in posedovanje fenomenalnega pojma, lahko opišemo izključno v fizikalnih pojmih.

4. Fenomenalne pojme lahko pridobimo in posedujemo samo, če imamo izkustvo danega fenomenalnega značaja.

5. Po izpustitvi pridobi Marija znanje o fenomenalnem značaju, opisano v fenomenalnih pojmih.

6. Dejstva, ki naredijo te nove elemente njenega znanja za resnične, so fizikalna dejstva, opisana v fizikalnih pojmih, ki jih je Marija poznala pred izpustitvijo,.

∴ 7. Marija pridobi novo znanje, a o starih dejstvih.

∴ 8. Fizikalizem je resničen (Nida-Rümelin, 2010).

Z vidika veljavnosti ne gre temu argumentu ničesar očitati, iz navedenih premis sklep vsekakor sledi, vprašanje pa je, če je argument tudi dober. Spomnimo se, da je za fizikalista argument iz znanja sporen, ker v končni fazi zavrača fizikalizem, tj. da je vse, kar obstaja, fizično. To pomeni, da mora fizikalist, če želi ohraniti fizikalizem, na nek način pokazati, da sklep argumenta iz znanja iz premis ne sledi. Videli smo, da vsi poskusi, ki smo si jih do sedaj ogledali, napadajo tisti del trditev o intuiciji, ki pravi, da je Marija po izpustitvi pridobila novo znanje, tj. trditev o učenju. Isto strategijo je zapaziti tudi v zgornjem argumentu, ki naj bi dokazoval, da se Marija nauči zgolj staro dejstvo v novi preobleki, zaradi česar je fizikalizem še vedno resničen. Vendar, če sprejmemo prvo premiso, fizikalizem sploh ni več vprašljiv, saj, če so takšnosti fizikalne lastnosti, ne more biti nič, kar se Marija po izpustitvi nauči, nad fiziko ali več od nje. Ker pa je resničnost fizikalizma tisto, kar je sporno, se zdi, da argument že v prvi premisi predpostavlja tisto, kar bi moral šele dokazati, zaradi česar se »izogiba vprašanju« oziroma je krožen (lat. *petitio*) in tako slab.[27] Razen tega se zdi, da gre za rešitev, ki za zagovor fizikalizma uporablja dva različna načina predstavljanja prepričanj o fenomenalnih stanjih, fizikalni in fenomenalni opis, kar običajno vključuje tudi dve različni vrsti lastnosti, na katere se opisa nanašata, npr. fizikalno in nefizikalno, zato ni čudno, da med filozofi ni soglasja o tem, ali je ustrezna.[28]

4 Sklep

Vprašanje je, kaj nesoglasje glede tega, kako se mora fizikalizem spopasti z argumentom iz znanja, da bi takšnosti bile fizične, pomeni za našo razpravo o tem, ali so stroji lahko zavestni. Ali to, da ni prepričljivega argumenta za to, da je naše subjektivno kvalitativno izkustvo ‘kako je biti’ fizično, na kakršenkoli način predstavlja nevarnost za uspešnost našega projekta, ki bi pokazal, da lahko imajo zavest tudi npr. roboti? Odgovor je ne, ves problem pripisovanja zavesti umetnim sistemom in njegovo rešitev, ki je neodvisna od tega, ali so takšnosti fizične, pa lahko na kratko povzamemo s pomočjo naslednjega primera.

Predpostavimo, da so na Inštitutu Jožef Štefan naredili z interaktivno tehnologij opremljen računalnik, ki je z njeno pomočjo sposoben komunicirati z okoljem ga doživljati na enak način kot človek. Vprašanje je, ali Einstein, ko skozi svoje umetno oko na vrtu opazi rdeč paradižnik, doživi subjektivno izkušnjo rdečosti, nek poseben kvalitativni občutek, ali pa njegovo elektronsko vezje zgolj prepozna tipično binarno zaporedje, ki označuje rdečo barvo? Seveda je možno, bi zakričali nasprotniki zavestnih strojev in se to tudi trudili dokazati, da Einstein kljub temu, da paradižnik vidi, da se zaveda, da ga vidi, da ve, da je rdeč in ne moder, da prepozna njegovo tipično obliko itd., nima nobenega posebnega

[27] Enak očitek velja tudi za prejšnji argument, razvit v prid odgovoru s seznanitvijo.
[28] Za ugovore *glej* npr. McConnell (1994), White (2007), Chalmers (2007).

kvalitativnega izkustva, tj. nobenega ‘kako je občutiti rdeče’, ko je s paradižnikom v »čutni interakciji«.

Toda, ali obstaja kakršnokoli zagotovilo, da je moj občutek, ko zagledam nekaj, kar je rdeče, enak vašemu občutku, ko zagledate nekaj, kar je rdeče oziroma, kako vem, da vi, ko zagledate nekaj, kar je rdeče, sploh karkoli občutite? Res je, da vsi uporabljamo pojem ‘rdeče’, ko npr. opisujemo zrele paradižnike, zdi pa se, da je to tudi vse, kar imamo skupnega, saj so takšnosti subjektivno izkustvo, dostopno zgolj lastniku in neopisljivo v pojmih fizičnega mehanizma. Nikomur ne pade na kraj pameti, da bi drugemu človeku takšnosti odrekel, čeprav nima nobenega razloga za prepričanje, da jih ta ima, ko pa beseda nanese na stroj, smo vsi takoj prepričani, da jih takšen sistem niti slučajno ne more imeti.

Problem takšne drže je, da nimamo pravega razloga niti za odrekanje fenomenalne zavesti računalnikom, saj v primeru, če je ta fizična, ni nobenih ovir za to, da je umetni sistem ne bi mogel imeti, pa tudi naši ne bi rabila biti podobna, če pa je nefizična, ni nobenega razloga za to, da je ob naravnih možganih ne bi mogli proizvesti tudi umetni (Copeland, 1993: 179). Enostavno dilemo, pred katero se z vidika pripisovanja fenomenalne zavesti robotom znajdemo, lahko v standardni obliki zapišemo takole:

1. Takšnosti so fizične ali nefizične.
2. Če so takšnosti fizične, jih lahko ima tudi stroj.
3. Če so takšnosti nefizične, jih lahko ima tudi stroj.
4. Stroj ima lahko takšnosti.

Vidimo, da za govorjenje o zavestnih strojih sploh ne rabimo zavrniti argumenta iz znanja, saj je, če se želimo vzdržati šovinizma vrst, pripisovanje fenomenalne zavesti umetnim sistemom možno bodisi, da so takšnosti fizične bodisi, da so nefizične. Tako smo, mislim, odstranili še eno oviro na poti do tega, da bi na računalnike končno začeli gledati kot na misleča bitja, s čimer ne bi ničesar izgubili, kvečjemu pridobili; če ne drugega vsaj to, da bi ob izpolnjevanju enakih pogojev na različne organizme gledali enako.

MACHINES AND CONSCIOUSNESS: THE PROBLEM OF QUALIA

The article deals with the question whether the phenomenal consciousness, i.e. ‘what it’s like’, is something reserved for humans, or something that even a machine could have, at least in one of possible worlds. It introduces the problem of qualia as being derived from the three of consciousness’ many problems that

are usually associated with the so called *hard problem of consciousness*, i.e. the question, how qualia can be physical. *The knowledge argument* proving that qualia are non-physical is threatening to defeat physicalism, and philosophers disagree about the right way of how to reject it. By the help of the simple dilemma, the article suggests that for establishing whether or not there can also be conscious machines we do not have to agree about how to reject the knowledge argument from the physicalistic point of view for the nature of qualia does not play any role at ascribing phenomenal consciousness to, e.g. robots.

Keywords: phenomenal consciousness, qualia, 'what it's like', the knowledge argument, the inverted spectrum.

Literatura

Alter, T. (2005). »The Knowledge Argument Against Physicalism«. *The Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/know-arg/.

Berkeley, G. (1710/1976). *Razprava o načelih človeškega razuma*. Ljubljana: Slovenska matica.

Bigelow, J. in Pargetter, R. (1990/2004). »Acquaintance with Qualia«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 179–195.

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

Block, N. (1983). »Mental Pictures and Cognitive Science«. *Philosophical Review*, 92, str. 499–541.

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

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

Bregant, J. (2013). »Funkcionalizem, računalniki in 'kitajska soba'«. *Analiza*, 1 št. 3., str. 67–95.

Broad, C. D. (1925). *The Mind and its Plače in Nature*. London: Routledge & Kegan Paul.

Byrne, A. (2002). »Something about Mary«. *Grazer Philosophische Studien*, 62, str. 123–140.

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

Chalmers, D. (2007). »Phenomenal Concepts and the Explanatory Gap«. V Alter, T. in Walter, S. (ur.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, str. 167–194.

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

Churchland, P. M. (1989/2004). »Knowing Qualia: A Reply to Jackson«. V

Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 163–178.

Conee, E. (1994/2004). »Phenomenal Knowledge«. V Ludlow, P., Nagasawa. Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 197–215.

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

Davidson, D. (1970/1995). »Mental Events«. V Moser P. K. in Trout, J. D. (ur.), *Contemporary Materialism*. London: Routledge, str. 107 – 121.

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

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

Eccles, J. C. (1994). *How the Self Controls Its Brain*. Berlin: Springer.

Feigl, H. (1958). »The Mental and the Physical«. V Feigl, H., Scriven, M. in Maxwell, G., *Minnesota Studies in the Philosophy of Science II: Concepts, Theories, and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, str. 370–497.

Freud, S. (1932). *Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse*. Frankfurt: S. Fischer Verlag.

Gertler, B. (1999). »A Defense of the Knowledge Argument«. *Philosophical* Studies, 93, str. 317–336.

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

Heil, J. in Mele, A. (ur.) (1995). *Mental Causation*. Oxford: Clarendon Press.

Horgan, T. (1984). »Jackson on Physical Information and Qualia«. *Philosophical Quarterly*, 32, str. 127–136.

Huxley, T. H. (1866). *Lessons in Elementary Physiology*. London: Macmillan.

Jackson, F. (1982/1999). »Epifenomenalne takšnosti« *Analiza*, 2, str. 86–77.

Jackson, F. (1986). »What Mary Didn't Know«. *Journal of Philosophy*, 83, str. 291–295.

Jackson, F. (1998/2004). »PostScript on Qualia«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 417–420.

Kim, J. (1993). *»Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. (2001). *Mind in a Physical World*. Cambridge: MIT Press.

Kriegel, U. (ur.) (2013). *Phenomenal Intentionality*. Oxford: Oxford University Press.

Kripke, S. (1972/2000). *Imenovanje in nujnost*. Ljubljana: Krtina.

Levine, J. 1983. »Materialism and qualia: the explanatory gap«. *Pacific Philosophical Quarterly*, 64, str. 354–361.

Levin, J. (2007). »What is a phenomenal concept?«. V Alter, T. in Walter, S. (ur.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, str. 87–110.

Lewis, D. (1988/2004). »What Experience Teaches«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 77–103.

Loar, B. (1990/2004). »Phenomenal States« (Revised Version). V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 219–239.

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

McConnell, J. (1994). »In Defense of the Knowledge Argument«. *Philosophical Topics*, 22, str. 157–187.

McGinn, C. (1991). *The Problem of Consciousness*. Oxford: Blackwell.

McGinn, C. (1999). *The Mysterious Flame*. New York. Basic Books.

Nagel, T. (1974/1990). »Kako je biti netopir?«. V Hofstadter, D. R. in Dennett, D. C. (ur.), *Oko duha*. Ljubljana: Mladinska knjiga, str. 399–421.

Nemirow, L. (2007). »So this is what it' s like: a defense of the ability hypothesis«. V Alter, T. in Walter, S. (ur.), *Phenomenal Concepts and Phenomenal Knowledge. New Essay on Consciousness and Physicalism*. Oxford: Oxford University Press, str. 32–51.

Nida-Rümelin, M. (1995/2004). »What Mary couldn't know: Belief About Phenomenal States«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 241–267.

Nida-Rümelin, M. (2010). »Qualia: The Knowledge Argument«. *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/qualia-knowledge/.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. in Selina, H. (2000). *Introducing Consciousness*. Cambridge: Icon Books.

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

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

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

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.

Stoljar, D. in Nagasawa, Y. (2004). »Introduction«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 1–36.

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

Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge: The MIT Press.

Tye, M. (2006). »Absent Qualia and the Mind-Body Problem«. *Philosophical Review*, 115, str. 139–168.

Tye, M. (2009). *Consciousness Revisited*. Cambridge: The MIT Press.

Unger, P. (1990). *Identity, Consciousness and Value*. Oxford: Oxford University Press.

Van Gulick, R. (2004). »So Many Ways of Saying No to Mary«. V Ludlow, P., Nagasawa, Y. in Stoljar, D. (ur.), *There's Something About Mary*. Cambridge: The MIT Press, str. 365–405.

White, St. L. (2007). »Property Dualism, Phenomenal Concepts, and the Semantic Premise«. V Alter, T. in Walter, S. (ur.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, str. 210–248.

Wittgenstein, L. (1953/1977). *Philosophische Untersuchungen*. Frankfurt: Suhrkamp.