Filozofija uma: suvremene rasprave o odnosu uma i tijela

Jurjako, Marko; Malatesti, Luca

FILOZOFIJA UMA: SUVREMENE RASPRAVE O ODNOSU UMA I TIJELA

MARKO JURJAKO I LUCA MALATESTI

FILOZOFIJA UMA: SUVREMENE RASPRAVE O ODNOSU UMA I TIJELA

Marko Jurjako i Luca Malatesti

Sveučilište u Rijeci, Filozofski fakultet
Rijeka, 2022.

FILOZOFIJA UMA: SUVREMENE RASPRAVE O ODNOSU UMA I TIJELA
Marko Jurjako i Luca Malatesti

Izdavač:
Sveučilište u Rijeci
Filozofski fakultet
Sveučilišna avenija 4, 51000 Rijeka

Za izdavača:
Izv. prof. dr. sc. Aleksandar Mijatović

Recenzenti:
Doc. dr. sc. Tomislav Janović, Sveučilište u Zagrebu, Fakultet hrvatskih studija
Doc. dr. sc. Ljudevit Hanžek, Sveučilište u Splitu, Filozofski fakultet

Lektura:
Dr. sc. Martina Blečić

Dizajn naslovnice:
Eva Šustar

Kompjuterska priprema sloga:
Luca Malatesti

ISBN 978-953-361-057-3 (e-izdanje)

Sadržaj

Zahvale

Htjeli bismo zahvaliti recenzentima Tomislavu Janoviću i Ljudevitu Hanžeku na konstruktivnim komentarima i sugestijama kako poboljšati tekst knjige koja se nalazi pred vama.

U sedmom poglavlju (odjeljak 7.3) koristili smo objavljeni materijal iz rada Malatesti, Luca i Nela Malatesti. 2013. „Supervenience, mind and chemistry“. U Filozofija u dijalogu sa znanostima, uredili Luka Boršić i Ivana Skuhala Karasman, 253–71, Institut za filozofiju. Ovim putem zahvaljujemo Ivani Skuhala Karasman, Luki Boršiću i Neli Malatesti na dopuštenju da koristimo dijelove tog poglavlja.

Naš rad podržava Hrvatska zaklada za znanost u okviru triju projekata: RAD, HRZZ-IP-2018-01-3518 (MJ, LM); HIRe, HRZZ-UIP-2017-05-4308 (MJ) i JOPS HRZZ-IP-2020-02-8073 (LM). Sveučilište u Rijeci podržava nas u okviru dvaju projekata: KUBIM, uniri-human-18-265 (MJ) i DPV uniri-human-18-151 (LM). Ovim putem im zahvaljujemo na financijskoj podršci. Također, zahvale idu Izdavačkom centru Filozofskog fakulteta u Rijeci na djelomičnom pokrivanju troškova pripreme i lekture ove knjige.

Marko Jurjako napravio je završne dorade rukopisa prema sugestijama recenzenata tijekom istraživačkog boravka na Sveučilištu u Grazu. Ovim putem zahvaljuje se Austrijskoj akademiji za znanost (Österreichische Akademie der Wissenschaften) na velikodušnoj stipendiji koja je omogućila istraživački posjet. Također, velike zahvale idu Norbertu Paulou i Thomasu Pözleru na pozivu i pomoći pri pripremi istraživačkog projekta te Lukasu Meyeru na gostoprimstvu u Grazu.

Na kraju, htjeli bismo istaknuti da najveća hvala ide našim obiteljima na svesrdnoj biopsihosocijalnoj podršci. Luca se zahvaljuje Neli, Antoniju, Ankici, Stefaniji, Marcu, Paoli, Giuliju, Leonardu, Micheleu i Maggie. Marko se zahvaljuje Zdenki, Davidu, Luciu, Ivanki, Ivanu te Sonji i Ivu, domaćinima neprikosnovenog BIAS instituta (Nerezine).

Predgovor

Ova knjiga nastala je na temelju priprema za predavanja u sklopu kolegija Uvod u filozofiju uma koji se izvodi na Filozofskom fakultetu u Rijeci. Predavanjem tema iz filozofije uma uvidjeli smo da u Hrvatskoj ima dosta objavljenih radova iz tog područja, no da nedostaje monografskih djela koja daju iscrpniji pregled klasičnih i suvremenih rasprava o problemu odnosa uma i tijela. Stoga je ova knjiga nastala kao plod našeg napora da donekle popunimo taj intelektualni prostor.

U knjizi se bavimo raspravom o odnosu uma i tijela koja svoj moderan oblik poprima tijekom 17. stoljeća u radovima Renéa Descartesa. Stoga pregled ovih tema započinjemo razmatranjem znanstvene slike svijeta koja poprima novi oblik u Descartesovo vrijeme i načina na koji Descartes počinje razmišljati o prirodi uma i njegovom mjestu u svijetu. Knjigu završavamo suvremenim debatama koje su i dalje uvelike pod utjecajem Descartesove argumentacije, naročito kada razmišljamo o prirodi svjesnih mentalnih stanja.

U suvremenoj filozofiji uma obično se smatra da postoje dva glavna problema: problem prirode svijesti i problem intencionalnosti. U ovoj knjizi, problemom određivanja prirode svijesti bavit ćemo se kroz razmatranje problema odnosa uma i tijela. Pojam intencionalnosti odnosi se na to da neka mentalna stanja posjeduju sadržaj, tj. mogu biti o nečemu. Problem intencionalnosti se, između ostalog, odnosi na pitanje kako naša mentalna stanja mogu predstavljati stvari koje su izvan nas. Htjeli bismo istaknuti da se u knjizi nećemo potanko baviti ovim važnim problemom. Stoga, čitatelje koji bi detaljnije htjeli istražiti problem intencionalnosti upućujemo na skriptu „Filozofija uma: intencionalnost u suvremenim filozofskim raspravama“ (Malatesti 2014) koju je napisao jedan od nas te je dostupna na ovoj poveznici.

Cilj ove knjige nije toliko dati konačno razrješenje problema odnosa uma i tijela koliko njegovo razumijevanje i mapiranje pojmovnog terena koji ga generira. Kada se pojmovna struktura problema dobro analizira daju se jasniji obrisi samog problema, mogućnost njegovog rješavanja i smjer u kojem bi trebalo investirati naše kognitivne resurse radi proširenja našeg

korpusa znanja. To je poruka koju bismo htjeli da čitatelji usvoje nakon čitanja ove knjige.

1 Uvod

1.1 Dva glavna pitanja u filozofiji uma

Filozofija uma fokusira se na pitanje koja je priroda uma:[1] što je um i koja su njegova svojstva? Kako bismo ilustrirali oblik koji je pitanje poprimilo u suvremenoj filozofiji uma možemo započeti razmatranjem onoga što inače smatramo da su mentalna stanja. Standardno prepoznajemo nekoliko tipova mentalnih stanja. Na primjer, uobičajeno je da govorimo o bolovima, mislima, vjerovanjima, željama, nadama, raspoloženjima i osjećajima kao istaknutim dijelovima našeg mentalnog života. Ona čine različita mentalna stanja ili procese.

Prije nego se usredotočimo na raspravu o prirodi uma, osvrnut ćemo se na neka obilježja za koja se tipično smatra da karakteriziraju mentalna stanja. Mnogi filozofi uma smatraju da postoje dvije glavne vrste obilježja koje karakteriziraju mnoga mentalna stanja. Prvo svojstvo koje ćemo spomenuti je *intencionalnost*. Slijedeći tradiciju koju je započeo Franz Brentano (1838. – 1917.), neki filozofi uma smatraju da je osnovno obilježje koje razlikuje mentalna stanja od drugih stvari njihova intencionalnost, tj. činjenica da su o nečemu ili imaju sadržaj (Hanžek 2017; za uvod u suvremenu raspravu o intencionalnosti, vidi Malatesti 2014). Intencionalnost je filozofski termin koji se odnosi na svojstvo „biti o nečemu“. Svojstvo biti o nečemu ili „ovost“ može se objasniti na sljedećem primjeru. Zamislimo da neka osoba, nazovimo je Franz, vjeruje da gospodin David nosi sandale. Franzovo vjerovanje je o stanju stvari koje uključuje gospodina Davida koji nosi sandale. Kod intencionalnih stanja možemo razlikovati mentalni stav od onoga u pogledu čega zauzimamo određeni stav. U tom smislu možemo razlikovati vjerovanje koje predstavlja određenu vrstu mentalnog stava od njegovog *sadržaja* ili onoga o čemu je to vjerovanje. Franz ima stav vjerovanja koji je definiran time da se neki sadržaj uzima kao istinit. Sadržaj

[1] Nekad se u filozofskoj terminologiji koristila riječ „duh“ pa se govorilo o filozofiji duha, problemu odnosa duha i tijela i tome slično. U ovoj knjizi pratimo noviju praksu koja se uvriježila u Hrvatskoj da se umjesto riječi „duh“ koristi „um“ (vidi Berčić 2012, fusnota 2). Dakle, treba imati na *umu* da kada govorimo o umu onda mislimo na cjelokupna mentalna svojstva osobe, poput njezinih intelektualnih, emotivnih i perceptivnih sposobnosti.

Franzovog vjerovanja je da gospodin David nosi sandale. To drugim riječima znači da Franz smatra da je istinita tvrdnja da gospodin David nosi sandale. Možemo navesti još nekoliko primjera rečenica kojima se opisuju intencionalna mentalna stanja:

1) Nadam se da će Ivan doći sutra na ručak.

2) Mislim da su dva i dva četiri.

3) Bojim se da Marica ima loše mišljenje o meni.

U rečenici 1) možemo razlikovati mentalni stav nadanja i sadržaj tog stava koji izražavamo surečenicom da će Ivan doći sutra na ručak. Rečenica 2) izražava općeniti stav mišljenja čiji sadržaj je matematički iskaz da kada zbrojimo dva i da dobijemo broj četiri. Rečenica 3) izražava mentalni stav straha čiji je sadržaj da Marica ima loše mišljenje o meni. Dakle, postoji cijeli niz mentalnih stanja, poput vjerovanja, nadanja, strahova, želja, namjera, itd. čija je osnovna karakteristika da imaju određeni sadržaj.

Filozofi koriste tehnički termin „propozicijski stav“ kako bi referirali na mentalno stanje koje ima intencionalnost te stoga i sadržaj. Ova terminologija se koristi kako bi se zahvatila činjenica da neka mentalna stanja uključuju relaciju koju mentalni stav ili mentalno stanje ima prema sadržaju, koji se još naziva propozicija. Razmotrimo, na primjer, vjerovanje da su dva i dva četiri. Taj propozicijski stav ima kao sadržaj propoziciju „dva i dva su četiri“, dok je stav u ovom slučaju vjerovanje. Također, možemo imati različite stavove prema istim sadržajima. Na primjer, ja mogu htjeti da sutra bude sunčano, mogu vjerovati da će sutra biti sunčano, mogu se nadati da će sutra biti sunčano, bojati se da će sutra biti sunčano i tako dalje.

Pitanje koje možemo postaviti u ovom kontekstu jest imaju li sva mentalna stanja intencionalnost? Brentano (1874) je smatrao da je upravo intencionalnost glavno obilježje koje razlikuje mentalne fenomene od fizičkih fenomena. Na primjer, svi ćemo se složiti da kamen kao vrsta fizičke stvari nema nikakav intencionalni sadržaj. Drugim riječima, kamen nije o nečemu. Naravno, za određenu stvar poput kamena možemo reći da ima značenje *za* nekoga. Na primjer, dijamant kao određena vrsta minerala može prenositi informaciju da je osoba koja ga posjeduje bogata. Međutim, dijamanti ne prenose taj sadržaj po sebi, već ga eventualno prenose jer im ljudi *pripisuju* određeno značenje. U tom smislu dijamant nema intrinzičnu intencionalnost, u smislu nečega što je svojstveno samom dijamantu. Prema Brentanu, jedino što ima izvornu ili intrinzičnu intencionalnost su mentalna stanja. Ako je Brentano u pravu onda imamo jedan općeniti kriterij koji razlikuje mentalna stanja od svih drugih stanja.

Međutim, neće se svi složiti da baš sva mentalna stanja posjeduju intencionalnost. Na primjer, možemo se pitati jesu li bolovi propozicijski stavovi? Kada nas nešto boli, ima li ta bol nekakav sadržaj, tj. je li ona o nečemu? Intuitivno se čini da bol ne mora imati sadržaj. Njezina priroda se

određuje time kako je osjećamo i doživljavamo. Mnogi će reći da slično vrijedi i za raspoloženja i neke emocije. Kada smo loše raspoloženi ili kada se nalazimo u stanju depresije čini se da ta raspoloženja ne moraju imati neki sadržaj, već se njihova priroda očituje u tome kako ih doživljavamo.[2] Ova razmatranja nas dovode do drugog važnog obilježja za kojeg će se mnogi autori složiti da karakterizira mnoga mentalna stanja.

Drugo važno obilježje mnogih mentalnih stanja je njihov *fenomenalni* ili *pojavni karakter* (engl. *phenomenal character*). Često se umjesto o fenomenalnom karakteru iskustva govori o *qualia* ili sirovim osjetima koje osobe imaju pri pojedinačnim svjesnim iskustvima (lat. mn. *qualia*, jd. *quale*). Fenomenalni karakter ili *quale* se odnosi na kvalitativna svojstva koja određuju neke iskustvene ili osjetilne doživljaje. Fenomenalni karakter mentalnih stanja trebao bi odrediti kako je to *doživjeti mentalna stanja* ili kako je to *biti u nekom stanju*. Uzmimo, na primjer, degustaciju vina. Kada probamo vino poput malvazije steći ćemo iskustvo koje će kvalitativno biti dosta različito od iskustva okusa vina poput plavca. Drugim riječima, ova dva iskustva imaju različite fenomenalne ili pojavne karaktere. Slično tome, možemo reći da doživljaj glavobolje ima svoj specifičan fenomenalni karakter koji je različit od iskustva zubobolje te je još različitiji od degustiranja vina.[3]

Suvremeni filozofi uma istražuju prirodu intencionalnosti i fenomenalnog karaktera iskustva te odnose među njima. Međutim, mnoga se od ovih istraživanja fokusiraju na pokušaj davanja odgovora na drugo temeljno pitanje u filozofiji uma. To se drugo pitanje tiče odnosa uma i prirodnog svijeta kako nam ga otkrivaju prirodne i druge znanosti.

Ovaj problem se općenito naziva problem odnosa uma i tijela. Znanosti poput biologije, kemije i genetike su otkrile mehanizme koji objašnjavaju anatomiju i fiziologiju tijela. Nadalje, multidisciplinarna istraživanja živčanog sustava otkrivaju mnoge aspekte funkcioniranja ljudskog mozga. Ova otkrića podržavaju optimizam u pogledu budućeg proširenja našeg znanja i razumijevanja procesa koji se odvijaju u tijelu i mozgu. Uzmimo samo u obzir impresivne korake koji su poduzeti u istraživanju molekularnih temelja života ili biokemijskih mehanizama koji se nalaze u podlozi funkcioniranja i interakcije tjelesnih pa i moždanih stanica. Ne čini se stoga problematičnim pretpostaviti da naša tijela, sa svojim organima i mozgom pripadaju prirodnom svijetu koji se može, barem u principu, objasniti u okviru znanosti poput biologije, kemije i fizike. Međutim, što je s umom? Kada uzmemo u

[2] Ovi primjeri intuitivno nas navode na prihvaćanje tvrdnje da neka mentalna stanja nisu intencionalna. Međutim, postoje li stvarno mentalna stanja koja nisu intencionalna je i dalje predmet diskusije u filozofiji uma (vidi, npr. Farkas 2009). Ovdje se ne želimo opredijeliti prema tom pitanju, već samo naglašavamo da neće svi autori prihvatiti tvrdnju da intencionalnost predstavlja esencijalno svojstvo koje karakterizira sva mentalna stanja.

[3] Idejom fenomenalnog karaktera iskustva detaljnije se bavimo u poglavlju 8.

obzir skup mentalnih stanja koje smo dosad razmatrali, jesmo li opravdani smatrati da ćemo i njih jednom moći opisati i objasniti koristeći metode i spoznaje iz prirodnih znanosti?

Filozofi uma se upravo bave tim pitanjem te raspravljaju mogu li se mentalna stanja poput boli i propozicijskih stavova u potpunosti opisati i objasniti oslanjajući se na suvremenu znanost. Naročito se fokusiraju na istraživanje mogu li se intencionalnost i fenomenalni karakteri iskustva smjestiti unutar prirodnog svijeta koji se opisuje i objašnjava putem prirodnih znanosti.

U ovoj knjizi pretežno ćemo se baviti ovim drugim pitanjem koje se odnosi na prirodu mentalnih stanja i njihovog odnosa s prirodnim svijetom. Osnovni cilj knjige je upoznati domaćeg čitatelja s glavnim argumentima, pozicijama i debatama o odnosu uma i tijela slijedeći način na koji su se one odvijale u suvremenoj analitičkoj filozofiji uma. Pod analitičkom filozofijom uma mislimo na dominantne rasprave iz filozofije uma kako se one vode na engleskom govornom području. U knjizi ćemo razmotriti različite argumente koje su filozofi iz analitičke tradicije iznijeli kako bi kritizirali pozicije i argumente drugih filozofa. Iako ćemo kroz pregled rasprava iz ovog područja nastojati ukazati na različite pozicije i gledišta te raspraviti njihovu uvjerljivost s obzirom na dostupne argumente, krajnji cilj nam je ohrabriti čitatelja da sam dođe do zaključaka i gledišta o teorijama, pozicijama i argumentima kojima se bavi filozofija uma. Kroz pregled rasprava raznih pozicija i argumenata, nastojat ćemo iznijeti na površinu metodološke pretpostavke koje razni filozofi podrazumijevaju kada brane svoje pozicije i kritiziraju tuđe. Osvrt na metodološke pretpostavke će nam omogućiti da shvatimo pozadinu raznih argumenata, motivacije autora koji ih iznose te da uvidimo kako odabir metodološkog okvira utječe na shvaćanje problema kojim se bavimo i njegovog značaja u široj raspravi o odnosu uma i tijela.

Prije nego damo pregled rasprava kojima ćemo se baviti u sljedećim poglavljima, u nastavku uvoda dajemo osnovni prikaz metodologije koja se koristi u analitičkoj filozofiji uma.

1.2 Metodologija filozofije uma

Poput drugih filozofskih područja, filozofi uma nastoje dati informirane i argumentirane odgovore na probleme koji su tipični za njihovu disciplinu. U nastavku ćemo iznijeti osnove argumentacijske metodologije koja se koristi u analitičkoj filozofiji.

Argumenti su sastavljeni od rečenica ili propozicija. Rečenice ili propozicije koje nazivamo premisama daju razloge za prihvatiti drugu rečenicu ili propoziciju, koju nazivamo konkluzija ili zaključak. Na primjer, razmotrimo sljedeći kratki argument:

U umu se nalaze procesi koji su brži od svjetlosti. Ništa što

pripada fizičkom svijetu nije brže od svjetlosti. Dakle, um ne može biti entitet u fizičkom svijetu (tj. um nije fizička stvar).

Ovaj argument uključuje dvije premise, koje se mogu formalno predstaviti na sljedeći način:

1) U umu se nalaze procesi koji su brži od brzine svjetlosti.
2) Ništa što pripada fizičkom svijetu nije brže od svjetlosti.

Ove premise se navode kao razlozi u prilog konkluzije:

3) Um ne može biti entitet u fizičkom svijetu (ili um nije fizička stvar).

Ne tvrdimo da je ovo dobar argument. Međutim, netko bi mogao koristiti ovakav argument kako bi opravdao tvrdnju ili tezu da um nije fizička stvar. U tom pogledu, cilj je ovog argumenta uvjeriti slušatelja u istinitost njegove konkluzije.

Generalno, argumenti se dijele na induktivne i deduktivne. Induktivni argumenti su oni čije premise opravdavaju konkluziju na način da je čine vjerojatnom, ali nikad ne dovode do potpune sigurnosti. Primjerice, sociolozi i psiholozi navode da je ponašanje u prošlosti najbolji prediktor toga kako će se ljudi ponašati u budućnosti. Stoga, osoba koja ima povijest antisocijalnog ponašanja najvjerojatnije će i u budućnosti kršiti društvene norme (Međedović 2015). Međutim, ovaj argument ne daje stopostotnu sigurnost da će se osoba nastaviti ponašati kako se ponašala dosad. Sasvim je moguće da neke antisocijalne osobe, zbog provođenja vremena u zatvoru, otkrivanja novih mogućnosti u životu, pronalaska smisla ili ljubavi i tome sličnog, promjene obrasce ponašanja u budućnosti. Unatoč njihovoj probabilističkoj prirodi, induktivni argumenti i rasuđivanje temelje se na empirijskim podacima te kao takvi predstavljaju temelj moderne znanosti.

S druge strane, tradicionalna forma argumenata u filozofiji je deduktivna. Tendencija korištenja deduktivnih argumenata je vjerojatno povezana s nastojanjima mnogih filozofa kroz povijest da pronađu sigurne i nepobitne temelje našeg znanja, koje nam induktivno rasuđivanje, zbog svoje probabilističke prirode, nikako ne može jamčiti.

Deduktivni argument je valjan ako i samo ako konkluzija nužno slijedi iz premisa. Da „nužno slijedi“, u ovom kontekstu, znači da ako su premise istinite, onda i konkluzija mora biti istinita. Razmotrimo kao primjer sljedeći argument:

4) Ako je Rim u Francuskoj, onda je Kolosej u Francuskoj.
5) Rim je u Francuskoj.

Dakle:

6) Kolosej je u Francuskoj.

Ovo je valjan deduktivni argument. Kad bi premise 4) i 5) bile istinite, onda bi i konkluzija 6) morala biti istinita. Naravno, valjanost argumenta ne jamči da su premise istinite. Štoviše, valjanost argumenta nije dovoljna da ga uopće učini uvjerljivim. Unatoč tome što je gornji argument valjan, mi znamo da je njegova konkluzija neistinita. S obzirom na to da je argument valjan, a znamo da je konkluzija neistinita, onda slijedi da barem jedna premisa mora također biti neistinita. U ovom slučaju, poznavanje istinosne tablice za logički operator kondicionala nam ukazuje na to da je prva premisa istinita (vidi Tablicu 1).

A	B	$A \rightarrow B$
$\top$	$\top$	$\top$
$\bot$	$\top$	$\top$
$\top$	$\bot$	$\bot$
$\bot$	$\bot$	$\top$

Tablica 1: Tablica istinosnih vrijednosti za kondicionalne rečenice koje izražavaju materijalnu implikaciju ($\top$ = istina, $\bot$ = neistina)

Naime, u logičkom smislu, kondicionalna rečenica je neistinita samo u slučaju kada je njezin antecedens istinit, a konsekvens neistinit. U ovom slučaju antecedens premise 4) kojim se tvrdi da je Rim u Francuskoj je neistinit, što znači da je cijela kondicionalna rečenica formalno istinita. Budući da je prethodni argument valjan i da je premisa 4) istinita, onda slijedi da premisa 5) mora biti neistinita.

Deduktivni argument je pouzdan (engl. *sound*) kad je valjan i ima samo istinite premise. Razmotrimo sljedeći argument:

7) Ako su inducirane promjene u mozgu neke osobe popraćene primjetnim promjenama u mentalnim stanjima te osobe, onda postoji korelacija između tih promjena u njezinom mozgu i njezinih mentalnih stanja.
8) Konzumiranje alkohola uzrokuje promjene u mozgu osobe koje su popraćene promjenama u njezinim mentalnim stanjima.

Dakle:

9) Postoji korelacija između promjena u mozgu osobe i promjena njezinih mentalnih stanja.

Ovaj argument je valjan. Ako pretpostavimo da su premise 7) i 8) istinite, onda i konkluzija 9) mora biti istinita. U ovom slučaju obje premise su istinite. Dakle, možemo zaključiti da je argument pouzdan.

Ovdje nećemo detaljno objašnjavati pojmove valjanosti i pouzdanosti deduktivnih argumenata. Postoji više od jednog načina na koji deduktivni argumenti mogu biti nevaljani. S njima ćemo se susreti kroz knjigu kako

budemo raspravljali različite argumente i teorije iz filozofije uma. Ako čitatelj želi dobiti sistematičniji pregled različitih argumentacijskih formi, može konzultirati neki dobar udžbenik iz neformalne ili formalne logike.[4]

Cilj knjige je omogućiti čitateljima da kritički analiziraju utjecajne argumente iz filozofije uma. Opisat ćemo mnogo deduktivnih argumenata, i neke induktivne argumente, koje su filozofi ponudili kako bi odgovorili na osnovna pitanja o prirodi uma. Nadalje, ocjenjivat ćemo valjanost i pouzdanosti tih argumenata, no također i njihovu relevantnost i uvjerljivost u pojedinim kontekstima rasprave. U idealnom slučaju, ovaj postupak bi trebao omogućiti da čitatelj dođe do informiranih i obrazloženih gledišta o problemima iz filozofije uma.

U ovoj knjizi bavit ćemo se i nekim induktivnim argumentima. Ugrubo govoreći, ako se premise induktivnog argumenta temelje na uvjerljivoj dokaznoj građi, onda je konkluzija vjerojatno istina. Razmotrimo, na primjer, sljedeći argument:

> Većina odraslih ljudi zna koji broj cipela nosi. Marica je odrasla osoba. Dakle, vjerojatno zna koji broj cipela nosi.

Čak i ako su sve premise ovog argumenta istinite, on nije deduktivno valjan. U najboljem slučaju njegova konkluzija je vjerojatna, što znači da nije nužno istinita. Dakle, iako možemo razumno očekivati da će Marica znati veličinu svojih cipela, ne možemo u to biti sto posto sigurni. Na primjer, ako se ispostavi da Marica pati od amnezije onda se može dogoditi da neće znati koji broj cipela nosi.

Još jedna važna forma nededuktivnog zaključivanja koja se često koristi u filozofiji uma jest abdukcija. Ona se još naziva zaključak na najbolje objašnjenje (vidi Harman 1965). Ova vrsta zaključivanja je slična indukciji jer se oslanja na probabilističko rasuđivanje. No, razlikuje se od tradicionalnog shvaćanja indukcije gdje se kroz opažanja nastoji doći do nekakve generalizacije o karakteristikama koje obilježavaju određenu vrstu entiteta. Kod abduktivnog zaključivanja kreće se od razmatranja određene činjenice, događaja ili pojave koju želimo objasniti te tražimo hipoteze koje najbolje objašnjavaju tu činjenicu, događaj ili pojavu.

Kriteriji za najbolje objašnjenje mogu se razlikovati ovisno o kontekstu istraživanja. No, najčešće će najbolje objašnjenje biti ono koje pojavu od interesa čini vjerojatnom, tj. koje nam pokazuje zašto je bilo za očekivati tu pojavu s obzirom na naša trenutna znanstvena i teorijska znanja. Na primjer, zamislimo da čujemo nekakve čudne zvukove koji dolaze s tavana kuće u

[4] Za uvod u logiku prvog reda, čitatelja upućujemo na Kovač i Žarnić (2008) i Cauman (2004). Postoje i druge knjige čiji je cilj omogućiti čitateljima da razviju sposobnosti pisanja i čitanja filozofskih radova kako se to radi u analitičkoj tradiciji. U tom pogledu čitatelje upućujemo na knjigu Malatesti, Gavran Miloš i Čeč (2015).

kojoj živimo. Trebali bismo biti sami u kući jer smo sve ostale ukućane već ranije ispratili na posao i u školu. Stoga razmišljamo što može proizvoditi te čudne zvukove i dolazimo do, recimo, tri hipoteze koje bi ih mogle objasniti. Jedna hipoteza je da se pojavio duh prijašnjih vlasnika kuće kojima se ne sviđa da mi tu živimo pa nas nastoje uplašiti ne bismo li se iselili. Druga hipoteza je da je provalnik ušao u kuću te da lupa po prozoru. Treća hipoteza je da puše jak vjetar zbog kojeg prozor lupa. Prva hipoteza nam se čini vrlo nevjerojatna jer dosad nismo vidjeli duha te s obzirom na sve ostalo što znamo o tome kako funkcionira prirodni svijet nije vjerojatno da postoje duhovi. Druga hipoteza nam je malo uvjerljivija jer znamo da postoje lopovi. Međutim, u usporedbi s hipotezom da zbog vjetra lupaju prozori, druga hipoteza nam se čini manje vjerojatnom. Naime, znamo da živimo u mjestu gdje je stopa kriminala vrlo niska, a vjetar često puše. Također, tavan se nalazi na trećem katu te nije jednostavno do njega doći izvana. Uostalom, nije jasno zašto bi lopov uopće proizvodio takve zvukove kad mu je vjerojatno cilj ukrasti nešto i ostati neprimijećen. Stoga zaključujemo da je najbolje objašnjenje zašto čujemo čudne zvukove na tavanu činjenica da puhanje vjetra uzrokuje lupanje.

Ovakva vrsta zaključivanja se poziva na razne kriterije koji nam pomažu odvagati alternativne hipoteze koje objašnjavaju dostupnu dokaznu građu. Istaknut ćemo dva važna kriterija. Jedan je kriterij to koliko se hipoteza dobro uklapa u naše postojeće znanje ili teorije o svijetu. Na primjer, hipoteza da duh bivših vlasnika proizvodi čudne zvukove na tavanu nije uvjerljiva jer sve što iz znanosti znamo govori nam da ne postoje duhovi. Drugi važan kriterij je da preferiramo hipoteze ili objašnjenja koja su u nekom smislu jednostavnija od drugih objašnjenja koja su nam dostupna. Ovaj kriterij podrazumijeva jednu varijantu Ockhamove britve. To je princip prema kojemu kada objašnjavamo pojave u svijetu ne smijemo postulirati više entiteta nego što je potrebno da se pojava objasni. U našem slučaju, hipoteza koja pretpostavlja da postoje duhovi, u odnosu na hipotezu da je lopov ili vjetar uzrokovao lupanje prozora, narušava princip Ockhamove britve jer pretpostavlja postojanje entiteta, naime duhova, koje ne pronalazimo ni u jednoj domeni znanosti. Budući da čudne zvukove koje čujemo na tavanu možemo objasniti, a da ne postuliramo duhove, Ockhamova britva nam govori da trebamo preferirati druga, jednostavnija objašnjenja.

Vidjet ćemo da određeni autori u filozofiji uma smatraju da ova vrsta abduktivnog zaključivanja opravdava očekivanje da priroda uma mora biti u određenom smislu materijalna ili takva da se može objasniti i istraživati metodama prirodnih znanosti. U tom smislu neki filozofi smatraju da nam kriteriji koji određuju što je najbolje objašnjenje u kontekstu istraživanja uma pokazuje da um trebamo povezati i reducirati na određenu vrstu bioloških, tj. moždanih procesa. Ta gledišta i ova vrsta zaključivanja najviše će doći do

izražaja u petom poglavlju gdje ćemo se baviti zastupnicima teorije identiteta tipova koji smatraju da od dostupnih objašnjenja prirode uma, najuvjerljivija će biti ona koja pretpostavljaju da su mentalna svojstva jedna vrsta fizičkih svojstava mozga.

Ovim kratkim pregledom nekih od najistaknutijih načina zaključivanja nismo iscrpili sve metode koje filozofi koriste pri razvijanju teorija i obrane svojih gledišta. Štoviše, kao što smo ranije najavili, kroz razna poglavlja isticat ćemo metodološke pretpostavke kojima se vode pojedini filozofi kada formuliraju svoje argumente. No, unatoč raznim razlikama i metodološkim nijansama koja karakteriziraju pojedina filozofska gledišta, smatramo da ranije predstavljeni načini zaključivanja predstavljaju argumentativnu jezgru koju filozofi s različitim gledištima na prirodu odnosa uma i tijela podrazumijevaju u svojim raspravama. Stoga nam one daju dobar generalan okvir za kategorizaciju pojedinih argumenata i njihovo vrednovanje.

U ostatku uvoda, predstavit ćemo glavne teme kojima se bavimo u narednim poglavljima. Iz pregleda tema kojima se bavimo u knjizi, postat će jasno da u suvremenoj filozofiji uma dominiraju takozvane fizikalističke pozicije. To su ona gledišta koja podrazumijevaju da su um i njegova svojstva u nekom smislu fizički. To nije slučajno. Kao što ćemo vidjeti nefizikalistička gledišta susreću se s problemom objašnjenja zdravorazumske slike uma koji ima određene uzročne moći te može proizvoditi fizikalne učinke u prirodnom svijetu. Mnogi autori će tvrditi da je najbolje objašnjenje te činjenice to da um jest jedna vrsta fizičke stvari. U tom pogledu, teme i gledišta koje ćemo razmotriti u ostatku knjige u različitim će aspektima odražavati tu temeljnu postavku.

1.3 Struktura knjige

U drugom poglavlju, započet ćemo raspravu pregledom glavnih problema, teorija, pojmova i argumenata iz filozofije uma koje je formulirao jedan od najutjecajnijih novovjekovnih filozofa i znanstvenika, René Descartes. Descartes je ponudio vrlo utjecajne argumente u prilog teze da um mora biti nešto suštinski različito od tijela (ili općenito fizičkog) te s njim započinje suvremena rasprava o onome što se naziva problem uma i tijela. U ovom poglavlju vidjet ćemo da se najznačajniji problemi odnose na to kako shvatiti uzročne odnose između uma i tijela. Zdravorazumski se čini da se mentalna stanja nalaze u uzročnim odnosima s fizičkim stanjima. Na primjer, želja da popijemo čašu vode tipično uzrokuje fizičke radnje koje dovode do toga da popijemo čašu vode. Obrnuto, percepcija nekog fizičkog predmeta tipično uključuje uzročne odnose koje od percipiranog predmeta vode do toga da vidimo taj predmet. Ovdje se javlja sljedeći problem za kartezijanski dualizam: ako je um nematerijalna stvar, a tijelo u kojem taj um djeluje je sastavljen od fizičke stvari, kako te dvije suštinski različite vrste stvari uopće mogu interagirati. Ovaj problem postaje poznat kao problem uma i tijela koji

se u suvremenim debatama najčešće nastoji riješiti kroz različita filozofska gledišta.

U trećem poglavlju bavimo se biheviorizmom. Biheviorizam se javlja početkom 20. stoljeća kao metodološki pokret unutar psihologije. U tom pogledu, biheviористи stavljaju naglasak na pojave koje su javno opažljive i ponovljive te se mogu istraživati znanstvenim metodama. Stoga se biheviористи fokusiraju na ponašanje organizama. Biheviористи u filozofiji su otišli korak dalje te su tvrdili da opisi mentalnih stanja nisu ništa drugo nego opisi određenih vrsta ponašanja ili dispozicija za ponašanje. Ima više motivacija za takvo gledište. Jedna važna motivacija se temelji na filozofiji logičkog pozitivizma. Logički pozitivisti su prihvaćali hijerarhijsku sliku znanosti u čijem se temelju nalazi fizika. Kako bi se psihologija uklopila u takvu sliku znanosti ona prema njima mora biti disciplina koja se bavi opisima fizičkih fenomena koji su javno opažljivi. Smatrali su da se ta razina opisa odnosi na ponašanja ili neurobiološke procese koji se mogu istraživati znanstvenim metodama. Druga motivacija za prihvaćanje neke varijante filozofskog biheviorizma dolazi iz filozofije običnog jezika. Kao glavnog zastupnika te struje uzet ćemo Gilberta Rylea. Ryle je smatrao da zapravo ne postoji problem odnosa uma i tijela te da ćemo to uvidjeti kada se osvrnemo na to kako obično govorimo o umu i koju ulogu pojam uma ima u našim svakodnevnim međuljudskim interakcijama i razmišljanjima.

U četvrtom poglavlju prelazimo na temu teorije identiteta tipova. Ovo je fizikalistička teorija prema kojoj su mentalna stanja vrsta fizičkih stanja. Ova teorija se javlja kao reakcija na filozofski biheviorizam, naročito onu varijantu koju je zastupao Ryle, i probleme s kojima se susreće kartezijanski dualizam. S jedne strane, problem s biheviorizmom je što se čini da ne može zahvatiti iskustva koja povezujemo s perspektivom prvog lica. Ako se mentalna stanja na neki način svode na dispozicije za ponašanja onda se čini da iz ove slike uma ispadaju privatna iskustva koja se neće nužno manifestirati u ponašanju. S druge strane, dualisti se susreću s problemom uzročnosti između uma i tijela. Zastupnici teorije identiteta tipova rješavaju ove probleme tako što poistovjećuju iskustva i mentalna stanja s procesima i događajima u mozgu. Općenito, teoretičari identiteta tipova kreću od uvida da nam znanost daje najbolja objašnjenja toga kako funkcionira prirodni svijet. Budući da su ljudi dio tog prirodnog svijeta imamo dobrih razloga smatrati da ćemo ljudski um i njegova svojstva moći objasniti kao još jedan prirodni fenomen. Stoga smatraju da je najbolja metodološka pretpostavka za istraživanje uma tvrdnja da se tipovi ili vrste mentalnih stanja mogu reducirati na tipove moždanih stanja.

U petom poglavlju bavimo se funkcionalizmom. Kao što se teorija identiteta tipova javlja kao reakcija na biheviorizam tako se i funkcionalizam javlja kao reakcija na teoriju identiteta tipova. Mnogi smatraju da je teorija identiteta tipova neuvjerljiva pozicija jer se mentalna stanja mogu realizirati

kroz različite fizičke supstrate. Na primjer, hobotnica i čovjek mogu osjećati bol, ali budući da ljudi i hobotnice nisu u potpunosti fizički identični moguće je da ona fizička stanja i procesi koji realiziraju ili ostvaruju bol kod jedne vrste neće nužno biti ista fizička stanja koja realiziraju bol kod druge vrste. Funkcionalisti definiraju prirodu mentalnih stanja pomoću uzročnih uloga koje one igraju u povezivanju vanjskih podražaja, drugih mentalnih stanja i ponašanja. Kada tako definiramo mentalna stanja onda prema funkcionalistima postaje jasnije zašto ih ne možemo reducirati na svojstva mozga; naime, fizički različite vrste stvari u principu mogu igrati iste funkcionalne uloge. U tom pogledu, funkcionalizam dobro zahvaća intuiciju koja se nalazi u podlozi argumenta iz mogućnosti višestruke ostvarljivosti mentalnih stanja. Nadalje, značajna motivacija za funkcionalistička gledišta dolazi iz razvoja umjetne inteligencije i općenitije kognitivne znanosti, gdje se um počinje promatrati kao apstraktni stroj za procesiranje podataka. Stoga se funkcionalisti u svojim raspravama često pozivaju na usporedbu uma i računalnog softvera. Slično kao što softver treba razlikovati od hardvera na kojem se izvodi, tako treba razlikovati um od mozga. Problemi koji se javljaju za funkcionalizam često se odnose na pojavna svojstva iskustva, tj. njihove fenomenalne karaktere za koje se čini da se ne mogu funkcionalno objasniti. Stoga mnogi autori smatraju da čak i ako uspijemo ponuditi iscrpnu funkcionalnu analizu mentalnih stanja, svejedno nećemo uspjeti zahvatiti sva njihova obilježja; naročito ne ona obilježja koja karakteriziraju osjetilna iskustva poput percepcije boja, osjećaja boli i slične vrste osjetilnih iskustava.

U šestom poglavlju vraćamo se nekim temama koje smo otvorili u četvrtom poglavlju gdje smo se bavili teorijom identiteta tipova. Problemi koji se javljaju za funkcionalističke teorije uma odnose se na svjesne aspekte našeg iskustva za koje nije jasno kako se mogu reducirati na uzročne uloge mentalnih stanja. Za dualiste to predstavlja argument u prilog tezi da postoje neka mentalna svojstva koja nisu fizičke prirode. Međutim, za fizikaliste to predstavlja razlog da se vratimo nekoj varijanti redukcionizma u filozofiji uma. Općenito, redukcionizam je gledište da se mentalna svojstva mogu reducirati na određena svojstva fizičkih entiteta. U tom pogledu, fizikalisti probleme s funkcionalizmom uzimaju kao poziv za ponovno razmatranje argumenata kojima se nastoji pokazati neuvjerljivost teorije identiteta tipova. Stoga se u ovom poglavlju detaljno bavimo argumentom višestruke ostvarljivosti mentalnih stanja. Argument razmatramo u kontekstu odnosa psihologije kao znanosti i mogućnosti redukcije njezinih objašnjenja na neuroznanstvene teorije. Tu ćemo vidjeti da unatoč inicijalnoj uvjerljivosti ideje da se mentalna stanja mogu višestruko realizirati, ona postaje manje uvjerljiva kada je razmotrimo u kontekstu odnosa između psiholoških i neuroznanstvenih teorija.

U sedmom poglavlju razmotrit ćemo probleme koji se javljaju za fizikalistička gledišta koja ne prihvaćaju tezu da se mentalna stanja mogu reducirati na neku vrstu moždanih stanja. Prema poziciji koju možemo nazvati antiredukcionistički fizikalizam, mentalna stanja su povezana relacijom supervenijencije s fizičkim stanjima. Prema ideji supervenijencije, mentalna stanja na neki način ovise o fizičkim stanjima, no ne mogu se na njih reducirati. U ovom poglavlju razmotrit ćemo može li se pomoću pojma supervenijencije formulirati uvjerljiva antiredukcionistička varijanta fizikalizma. U tom pogledu razmotrit ćemo argumente koji pokazuju da je vrlo teško formulirati antiredukcionističku poziciju koja će uspjeti sačuvati zdravorazumsku ideju da mentalna stanja imaju uzročne moći.

U osmom poglavlju bavimo se antifizikalističkim argumentima kojima se nastoji pokazati općenita nemogućnost fizikalizma da objasni svjesne aspekte mentalnih stanja. Kada se govori o svjesnim aspektima mentalnih stanja onda se misli na ona svojstva mentalnih stanja koje ujedinjuje činjenica da postoji nešto takvo kao što je način doživljavanja tih stanja. Na primjer, postoji nešto takvo kao što je okus sladoleda od čokolade, izgled crvene boje ili osjet boli. Kao što smo ranije naveli, za taj aspekt iskustva, koji odgovara pitanjima *kako je to biti nešto* ili *kako je to doživjeti nešto*, u filozofiji uma koristi se izraz „fenomenalni karakter“ iskustva. Postoje mnogi argumenti kojima se nastoji pokazati da fizikalizam ne može objasniti fenomenalni karakter iskustva. Mi ćemo raspraviti dva vrlo utjecajna argumenta kojima se to nastoji pokazati, a radi se o argumentu iz znanja i argumentu pojmljivosti.

U devetom poglavlju se nastavljamo baviti problemima koje svijest generira za fizikalizam. Mnogi suvremeni autori smatraju da fenomenalni karakter iskustva predstavlja principijelni kamen spoticanja za fizikalizam. U tom pogledu, jedan od najistaknutijih suvremenih filozofa, David Chalmers, smatra da se fizikalizam suočava s tzv. teškim problemom svijesti prema kojemu postoji načelna nemogućnost fizikalističkog objašnjenja toga kako iz nesvjesne materijalne stvari kao što je mozak mogu nastati svjesna iskustva. Taj teški problem svijesti daje novi poticaj razvoju nefizikalističkih gledišta. Mi ćemo se posebice usredotočiti na dva gledišta: naturalistički dualizam i panpsihizam. U zadnjem dijelu poglavlja ćemo razmotriti gledište prema kojemu je teški problem svijesti za fizikaliste samo posljedica određene kognitivne iluzije. Stoga ćemo razmotriti neke od argumenata kojima se nastoji opravdati radikalni zaključak da na kraju dana možda ni ne postoje svjesna iskustva, barem ne onako kako ih shvaćaju nefizikalisti u filozofiji uma.

2 Kartezijanski[5] dualizam

2.1 Uvod

René Descartes (1596. – 1650.) zasigurno je jedan od najvažnijih osoba na intelektualnoj sceni sedamnaestog stoljeća. Osim što je udario temelje suvremenoj filozofiji uma, bio je i znanstvenik koji je značajno doprinio razvoju znanosti svoga vremena (Shea 1991; Garber 2001). Štoviše, kritizirajući tradicionalna srednjovjekovna i renesansna gledišta o prirodi, pružio je filozofske temelje modernoj znanosti. U Descartesovo vrijeme znanstvene revolucije su značajno promijenile način na koji percipiramo i razumijemo prirodni svijet. S obzirom na izuzetan razvoj fizikalnih znanosti i njihovog potencijala da objasne sav prirodni svijet, Descartes počinje razmatrati problem određivanja prirode uma. Upravo se u tom kontekstu formulira pitanje koje bi bilo mjesto uma u prirodnom svijetu kakvog nam znanost opisuje i objašnjava.

Descartesova pretpostavka je da se ovaj problem može najbolje adresirati koristeći spoznajne uvide koji su neovisni, i na neki način fundamentalniji, od onih koji se temelje na znanstveno-empirijskim istraživanjima. Štoviše, ova pretpostavka da istraživanje uma zaslužuje poseban epistemološki pristup i dalje je prisutna u suvremenim raspravama o prirodi uma. Kao što ćemo vidjeti, srž ovog pristupa je primjena *a priori* refleksije na zamislive i navodno moguće situacije koje nadmašuju naše uobičajeno iskustvo i načine razmišljanja o njemu. Osim prihvaćanja ovog epistemološkog pristupa u istraživanju uma, neki suvremeni filozofi uma također prihvaćaju i brane varijante Descartesovog dualističkog gledišta da um ili neka njegova svojstva ne pripadaju prirodnom svijetu kakvog nam opisuju fizikalne znanosti.[6]

Cilj ovog poglavlja je predstaviti Descartesova gledišta i pojasniti opća obilježja njegove filozofije uma. Iako u predstavljanju Descartesovih gledišta težimo vjernom predstavljanju njegove misli, naš primarni cilj ipak nije egzegetika Descartesove filozofije. Osnovni cilj nam je izdvojiti neka opća

[5] Konvencija je da se od latinske verzije Descartesovog prezimena (Cartesius) pravi pridjev koji karakterizira različite aspekte Descartesove filozofije i tradicije koju je stvorio u filozofiji uma. Tako je običaj govoriti o *kartezijanskoj* filozofiji uma i *kartezijanskom* dualizmu.

[6] Za daljnju raspravu, vidi poglavlja 8 i 9.

tematska i metodološka pitanja kojima se Descartes bavio, a utjecajna su u suvremenim raspravama iz filozofije uma.

2.2 Dualizam supstancija i interakcionizam

Descartes je smatrao da se prirodni svijet i ljudsko tijelo u njemu mogu objasniti oslanjanjem na jednostavna mehanička načela. U okviru te koncepcije prirodnog svijeta, formulirao je problem odnosa uma i tijela koji je, u određenim aspektima, još uvijek relevantan u suvremenim raspravama.

U temelju Descartesove filozofije uma je *dualizam supstancija.* To je teza da su um i tijelo dvije različite vrste stvari. Prema Descartesu, osoba je sastavljena od dvije različite supstancije: materijalnog tijela i nematerijalnog uma. Supstancija je entitet čije postojanje ne ovisi o postojanju drugih stvari. Ovu ovisnost možemo objasniti sljedećim primjerom. Zdravlje neke osobe postoji samo ako postoji i ta osoba. Ako osoba ne postoji onda ne postoji ni njezino zdravlje. Dakle, zdravlje osobe ovisi o postojanju te osobe. Međutim, osoba ne prestaje postojati kada nije zdrava (iako ponekad nedostatak zdravlja može dovesti do smrti osobe). Budući da osoba može postojati neovisno o postojanju drugih stvari, a zdravlje ne može, osoba je jedna vrsta supstancije, dok zdravlje nije. Slično tome, Descartes je smatrao da su um i tijelo supstancije; vrste stvari koje mogu postojati neovisno jedna o drugoj i o drugim vrstama stvari.

Prema Descartesu, supstancija u strogom smislu je nešto što može postojati čak i kad *ništa* drugo ne bi postojalo. On je smatrao da je samo Bog supstancija u tom strogom smislu, zato što samo Bog može postojati neovisno o bilo čemu drugom (Descartes 2014, 81, odlomak 51). Stoga, striktno govoreći, um i tijelo nisu supstancije jer, prema Descartesu, ni um ni tijelo ne mogu postojati neovisno o Bogu. Međutim, Descartes je smatrao da su um i tijelo supstancije u *sekundarnom smislu.* One mogu postojati neovisno o drugim stvarima osim Boga. Stoga bi se teza dualizma trebala odnositi na um i tijelo kao supstancije u sekundarnom smislu. No, ova razlika nije toliko bitna za razumijevanje Descartesovih argumenata. Stoga ćemo u nastavku nastaviti govoriti o dualizmu uma i tijela bez da tu vrstu dualizma supstancija kvalificiramo kao „sekundarnu“.

Pod “tijelom” Descartes ne misli samo na ljudsko tijelo nego i na supstanciju koja je zajednička svim materijalnim predmetima. Prema njegovom shvaćanju materijalnog svijeta, *ekstenzija* ili protežnost je *esencijalno* ili *nužno* svojstvo tijela. Esencijalno svojstvo je ono koje karakterizira bit nekog predmeta te ga u tom smislu taj predmet nužno posjeduje. Na primjer, esencija kruga je biti skup točaka u ravnini koje su jednako udaljene od središnje točke. Esencijalna svojstva suprotna su *akcidentalnim* ili *kontingentnim* svojstvima. Akcidentalna svojstva su ona koja predmeti mogu imati, ali ne definiraju njihovu bit. Na primjer, krug

može imati svojstvo biti oblik koji je najviše volio Giotto. No, nijedan krug ne mora imati to svojstvo kako bi bio krug.

Kao što smo vidjeli, prema Descartesu tijelo, kako bi bilo to što jest, mora biti protežno. Stoga on naziva tijelo *res extensa;* na latinskom to znači „protežna stvar". Protežne stvari su takve da se svi njihovi dijelovi mogu reducirati na opise u terminima veličine, oblika i kretanja. Tijelo je, dakle, stvar koja se može istraživati u fizici i drugim znanostima koje se bave mjerljivim prostornim i temporalnim dimenzijama. Descartesovo gledište na prirodu tijela i objašnjenje njegovog funkcioniranja nalazi se u temeljima moderne matematički formulirane znanosti.

S druge strane, Descartes je smatrao da je um esencijalno misleća stvar koja, za razliku od tijela, nije protežna. Latinski izraz koji je koristio za misleću stvar jest *res cogitans*. Budući da nije protežan, um za Descartesa nije materijalan; nema poziciju u prostoru, stoga se ne može kretati u prostoru i ne sastoji se od djeljivih materijalnih čestica koje se ponašaju u skladu sa zakonima prirode. Um je, prema Descasrtesu, misleća, samosvjesna stvar koja koristi jezik. Misleća stvar ima sposobnost da reflektira o samoj sebi. Razmotrimo, na primjer, osobu koja misli da je vrijeme za ići na odmor. Prema Descartesu, ova misao je nematerijalna modifikacija nematerijalne *res cogitans* koja konstituira tu osobu.

Descartes je smatrao da su samo ograničeni skup onog što mi zovemo mentalnim stanjima modifikacije *res cogitansa*. Na umu je imao dvije glavne modifikacije misleće supstancije. To su intelektualne aktivnosti poput mišljenja i volicijske ili voljne aktivnosti, koje se odnose na našu sposobnost donošenja odluka i djelovanja. Nasuprot tome, Descartes je smatrao da senzorna ili osjetilna iskustva, poput svrbeži, osjećaja boli i tome sličnog, ne mogu biti *samo* modifikacije uma ili tijela. Smatrao je da se one mogu pripisati samo kontingentnom *jedinstvu* ovih dviju supstancija (Descartes 2015, 151, odlomak 6.13). U tom smislu, one nisu čisto fizičke ni čisto psihičke. Primjerice, u knjizi *Meditacije o prvoj filozofiji*, Descartes o jedinstvu uma i tijela navodi sljedeće:

> Priroda me također uči preko tih zamjedbi boli, gladi, žeđi itd. da u svojem tijelu nisam samo prisutan onako kako je brodar prisutan na brodu, nego da sam s tijelom najtješnje povezan i gotovo pomiješan, tako da s njim tvorim nešto jedinstveno. Inače kada se tijelo ozlijedi, ja – koji nisam ništa drugo nego stvar koja misli – ne bi osjećao bol zbog ozljede, nego bi tu ozljedu percipirao čistim intelektom, kao što brodar vidom percipira ako se nešto slomi na brodu. (…) Jer ti osjeti žeđi, gladi, boli, itd. sigurno nisu ništa drugo nego zbrkani modusi mišljenja, nastali od spajanja i kao neke mješavine uma i tijela. (Descartes 2015, 151)

Ovo Descartesovo gledište predstavlja važnu razliku u odnosu na suvremena dualistička gledišta pri kojima se obično smatra da *fenomenalni* ili *pojavni* karakter osjetilnog iskustva, poput doživljaja boli, predstavlja nematerijalni ili nefizički aspekt uma.[7] Descartes, nasuprot tome, smatra da su samo misli i volicijski procesi paradigmatske nematerijalne modifikacije nematerijalnog uma.

Druga centralna pretpostavka Descartesove filozofije uma je *uzročni interakcionizam*. Prema ovome gledištu, bestjelesna mentalna supstancija i tjelesna supstancija mogu uzročno djelovati jedna na drugu. Na primjer, vatra može uzrokovati opekline na tijelu koje uzrokuju modifikacije *res cogitansa*; na primjer, uzrokuje bol i želju da odmaknemo tijelo od vatre. S druge strane, modifikacije nematerijalnog uma mogu uzrokovati promjene u tijelu. Na primjer, misao kako bi bilo zanimljivo dotaknuti vatru te odluka da to učinimo može uzrokovati da pomaknemo ruku *prema* izvoru vatre.

U *Meditacijama o prvoj filozofiji*, Descartes navodi kako „ne utječu svi dijelovi tijela neposredno na um, nego samo mozak, ili možda samo jedan njegov neznatan dio“ (Descartes 2015, 159). Konkretnije, Descartes je smatrao da se interakcija između uma i tijela odvija u dijelu mozga koji se naziva pinealna žlijezda ili epifiza. Descartes je imao ideju da se kroz epifizu tijelo može pokretati „putem finoga plina koji struji kroz živce“ (Descartes 2014, 38). Iako je Descartes imao dobru hipotezu da se dijelovi tijela pomiču zato što dobivaju impulse iz mozga, pogriješio je u vezi značajnosti epifize za odvijanje viših kognitivnih sposobnosti koje karakteriziraju njegov *res cogitans*. Danas se smatra da su psihološke funkcije implementirane kroz različite dijelove korteksa mozga. No, detalji neurofiziologije mozga nam nisu toliko bitni za trenutačnu raspravu.

Ovdje možemo sumirati dvije osnovne teze kartezijanske filozofije uma:

> (1) um i tijelo su dvije različite supstancije (dualizam supstancija);
> (2) postoji uzročna interakcija između uma i tijela (uzročni interakcionizam).

U sljedećem odjeljku razmotrit ćemo neke od značajnijih argumenata kojima je Descartes nastojao opravdati tezu dualizma supstancija.

2.3 Argumenti za dualizam supstancija

Descartes je ponudio više vrsta argumenata u prilog dualizmu supstancija. Suvremeni komentatori Descartesovih djela naglašavaju važnost razmatranja različitih konteksta u kojima Descartes iznosi svoja mišljenja i argumente (Cottingham 2005; Clarke 2005). Neke argumente u prilog dualizmu Descartes je stvarao u kontekstu razvoja nove prirodne znanosti

[7] Neke od ovih argumenata detaljnije razmatramo u poglavlju 7.

koja se suprotstavljala skolastičkoj tradiciji. Druge argumente formulirao je kako bi odgovorio na određene metafizičke i teološke probleme koji se raspravljaju u njegovo doba. Kao što ćemo vidjeti, ovi se argumenti oslanjaju na tradicionalne principe i pojmove. Mi se nećemo potanko baviti problemom povijesne interpretacije različitih argumenata koje Descartes razmatra i nudi te koje bi bilo njegovo konkretno gledište na prirodu uma (za pregled rasprave, vidi Clarke 2005). Raspravu ćemo poglavito usmjeriti na opis nekih od važnijih argumenata koji se pripisuju Descartesu te posebice onih koje razvio u sklopu razmišljanja o teorijskim pretpostavkama znanstvenih projekata kojima se bavio.

2.4 Argumenti iz „fleksibilnosti“

Descartes je proširio mehanističku sliku svijeta, koja od Galileja počinje dominirati u znanstvenim objašnjenjima prirode, na ljudsku fiziologiju i psihološke sposobnosti. Osnovna ideja mehanističkog projekta je da se fenomeni u prirodnom svijetu mogu objasniti putem pojmova veličine, oblika i kretanja čestica koje tvore materijalni svijet. Descartes opisuje ovaj mehanistički pristup ljudskim bićima u svojem neobjavljenom djelu *Rasprava o čovjeku* (fr. *Traité de l'homme*) koje je dovršio 1630. godine. U tom je djelu tvrdio da se neke ljudske sposobnosti mogu svesti ili reducirati i time objasniti u terminima mehaničkih procesa koji se odvijaju u ljudskom tijelu i mozgu, a da ne se pretpostavi postojanje duše. Na primjer, među funkcije koje se mogu na taj način objasniti uključuje probavu hrane i rast. Međutim, zanimljivije je iz naše perspektive da je također tvrdio da se takav pristup može uspješno proširiti na istraživanja određenih „psiholoških” funkcija, poput osjetilne recepcije svjetla, zvukova, mirisa, okusa i topline te njihovo pohranjivanje u pamćenju. Prema Descartesu, čak i određene radnje, poput hodanja i pjevanja, kada se odigravaju, a da um ne obraća pozornost na njih, u potpunosti se mogu objasniti u terminima dispozicija unutarnjih organa (Descartes 1985, 1:108). Slično tome, u svojoj knjizi *Rasprava o metodi* koja je objavljena 1637. godine, Descartes (1951, 45–46) argumentira da se pozivanjem na mehanička svojstva mozga može objasniti cijeli niz ljudskih ponašanja.

Ovakav mehaničko-redukcionistički projekt predstavljao je zamjetan otklon od tradicionalne aristotelijanske slike uma. Prema toj slici uma, čak i kada govorimo o neljudskim životinjama, objašnjenje ranije spomenutih psiholoških funkcija zahtijevalo je pozivanje na aktivnosti „senzitivne“ ili „vegetativne“ duše (Descartes 1985, 1:108).

Descartes je također bio svjestan otpora s kojim bi se njegovo mehaničko gledište na ljudske sposobnosti moglo susresti u tadašnjim intelektualnim krugovima. Stoga u *Raspravi o metodi* daje argumente kojima nastoji neutralizirati intuicije da se takve kompleksne sposobnosti i povezana ponašanja ne mogu objasniti u mehaničkim terminima. U tu svrhu povlači

analogiju s kompleksnim ponašanjem automata koji je izradio čovjek. Kod nekih ljudi, budući da im nedostaju relevantna znanja o funkcioniranju stvari, činjenica da posjedujemo kompleksne psihološke sposobnosti čini intuitivnom misao, koja je prema Descartesu netočna, da se takve sposobnosti ne mogu objasniti pomoću mehaničkih dispozicija materijalnih dijelova automata (Descartes 1951, 46). Prema toj ideji, otpor njegovom redukcionizmu proizlazi iz nesposobnosti ljudi da shvate kompleksnosti mehaničkih procesa koji se odvijaju u živčanom sustavu.

Unatoč tome, i sam je Descartes nudio argumente u prilog dualizma supstancija koji se temelje na onome što je on smatrao nepremostivim praktičnim ograničenjima njegovog redukcionističkog projekta. Štoviše, Descartes nudi te argumente svjestan da se temelje na našem trenutnom nepoznavanju kompleksnosti materijalnog svijeta. Zbog toga je na Descartesu teret dokaza da pokaže da, čak i uzimajući u obzir naše ograničeno znanje prirodnog svijeta, imamo razloga smatrati da se određene ljudske sposobnosti ne mogu objasniti u sklopu redukcionističkog pristupa kojeg je inače prihvaćao i nastojao primijeniti u svojim ostalim istraživanjima.

U tom smislu, Descartes nudi argumente kojima nastoji pokazati da um ima svojstva koja ne možemo objasniti oslanjajući se na materijalna svojstva tijela. Kada bi um imao svojstva koja nadilaze materijalna svojstva tijela, onda bismo morali zaključiti da um i tijelo nisu ista supstancija. Takvo se zaključivanje temelji na principu da identični entiteti moraju imati sva ista svojstva. Taj princip se često naziva Leibnizov zakon, u čast filozofa Gottfrieda Wilhelma Leibniza (1646. – 1716.). Ovaj zakon je bikondicional koji se sastoji od dvije komponente. Prva komponenta se naziva nerazlučivost identičnih predmeta; njome se tvrdi da ako su predmeti *x* i *y* identični, onda su im sva svojstva ista. U formalnom zapisu ta komponenta se izražava na sljedeći način:

(NI) $\forall x \forall y\ (x = y \rightarrow (Fx \leftrightarrow Fy))$

Druga komponenta se odnosi na identičnost nerazlučivih stvari; njome se tvrdi da ako predmeti *x* i *y* dijele ista svojstva onda su identični, tj. predstavljaju isti predmet.

(IN) $\forall x \forall y\ ((Fx \leftrightarrow Fy) \rightarrow (x = y)$

Generalno, Leibnizov princip se izražava bikondicionalom koji se sastoji od iskaza (NI) i (IN). U formalnom zapisu to se izražava na sljedeći način:

(LZ) $\forall x \forall y\ (x = y \leftrightarrow (Fx \leftrightarrow Fy))$

Kako bismo ilustrirali Descartesove argumente bitna nam je prva komponenta kojom se izražava nerazlučivost identiteta (NI). Ako su dvije

stvari identične onda ih se ne može razlučiti. Stoga, ako se pokaže da x ima neko svojstvo koje y nema, onda slijedi da x i y nisu ista stvar. Na primjer, ako su jabuka *a* i jabuka *b* ista jabuka onda moraju imati sva ista svojstva. Što znači da ako je jabuka *a* crvena onda je i jabuka *b* crvena. Ako se *a* nalazi na stolu ispred nas onda se i *b* nalazi na stolu ispred nas. Kada bi bio slučaj da je jabuka *a* crvena, a jabuka *b* zelena onda bismo imali jasan razlog smatrati da to nije ista jabuka. Slično tome, ako možemo pokazati da um ima neka svojstva koja tijelo nema i obrnuto onda možemo zaključiti da um i tijelo ne mogu biti ista stvar.

Pomoću dvaju argumenata ovog tipa, koja Descartes iznosi u 5. dijelu knjige *Rasprava o metodi* (Descartes 1951), zaključuje da se ljudske sposobnosti za fleksibilnu upotrebu jezika i formiranje pojmova ne mogu objasniti u mehaničkim terminima, tj. terminima koji se odnose na kretanje materijalnih dijelova mozga i živčanog sustava. Prvi argument u tom kontekstu, Descartes iznosi na sljedeći način:

> Kad bi (…) postojala stvorenja, koja bi imala sličnosti s našim tijelima i toliko oponašala naše radnje, koliko bi moralno (praktično) bilo moguće, mi bismo ipak imali uvijek dva sasvim sigurna sredstva, da raspoznamo, da ipak ta stvorenja još nisu pravi ljudi. Prvo je, da se ona nikad ne bi mogla služiti riječima ili drugim znakovima, da ih sastavljaju kao što to činimo mi, da drugima saopćujemo svoje misli. Možemo naime lako zamisliti, da je neki stroj udešen tako, da izgovara riječi, pa čak da izgovara neke povodom tjelesnih radnji, koje izazivaju u njegovim organima izvjesne promjene (ako ga udarimo poviče da boli, itd.), ali nipošto da bi riječi raspoređivao na razne načine, kako bi znao odgovoriti na smisao svega, što se u njegovoj prisutnosti govori, kao što to mogu činiti i najgluplji ljudi. (Descartes 1951, 46–47)

Argument iz ovog odlomka može se rekonstruirati na sljedeći način:

1) Ljudi imaju sposobnost da korištenjem jezičnih izraza prikladno odgovaraju na jezične podražaje iz okoline.
2) Moralno (praktički) je sigurno da strojevi, tj. materijalni entiteti, ne mogu imati takvu jezičnu sposobnost.

Dakle:

3) Moralno (praktički) je sigurno da ljudi ne mogu biti isto što i strojevi ili drugi materijalni mehanizmi.

Razmotrimo ovaj argument malo detaljnije. Korištenje izraza „moralna sigurnost“ je opravdano s obzirom na znanstveni kontekst unutar kojeg Descartes daje argument. Sam Descartes pod „moralnom sigurnošću“ misli na sigurnost:

> koja je dovoljna da regulira naše ponašanje ili koja doseže sigurnost koju povezujemo sa stvarima koje se odnose na ponašanje u životu u koje normalno ne sumnjamo, iako znamo da je moguće da su, doslovno govoreći, neistinite. (Descartes 1985, 1:289, fusnota 2, prijevod autora)

Iz ovoga možemo vidjeti da ovaj argument nije deduktivan jer čak i kad bi izneseno rasuđivanje bilo pouzdano (tj. čak i ako su premise istinite), ono što bismo mogli zaključiti jest da je vrlo *vjerojatno* da su materijalne stvari i um različite supstancije. U nastavku ćemo razmotriti je li konkluzija ovog argumenta zaista u visokoj mjeri vjerojatna s obzirom na premise.

Premisa 1) se, s određenim preinakama, čini istinitom. Naime, jasno je da postoje neki ljudi koji nemaju sposobnost jezične komunikacije s drugim ljudima. No, postojanje takvih osoba nije nužno problematično za Descartesov argument. Argument se može preformulirati tako da se njime tvrdi da prosječni jezično kompetentni ljudi mogu fleksibilno i bez previše napora odgovarati na jezične podražaje drugih ljudi. Sjetite se razgovora s prijateljicom ili prijateljem o prednostima i manama odlaska u kino tijekom pandemije. Vođenje takvog razgovora zahtijeva sposobnost fleksibilnog korištenja i razumijevanja jezika.

Premisa 2) izgleda problematično. Kakvom vrstom dokaza bismo je mogli opravdati? U gornjem citatu, čini se da Descartes nudi sljedeći argument za ovu premisu. Kreće se od uvida da je doseg mogućih rečenica na koje ljudsko biće može prikladno odgovoriti stvarno zapanjujuć. Ako vas netko pita „Je li istina da tigrovi lete oko Zemlje u zelenim čajnicima?“, vi biste vjerojatno odmah odgovorili „ne“. Slično, kad bismo zamijenili u prethodnoj rečenici riječ „tigar“ s riječi „lav“, jednako lako biste mogli reagirati koristeći se nekim jezičnim izrazom. Postavlja se pitanje na koji način bismo takvu vrstu jezične sposobnosti mogli ugraditi u stroj. Descartes, oslanjajući se na modele strojeva koji su bili dostupni u njegovo doba, smatra da se takva sposobnost može implementirati u stroj na samo jedan način. Prema Descartesu, stroj koji može prikladno odgovoriti na svaku rečenicu koja bi mu se mogla uputiti mora imati za svaku rečenicu poseban mehanizam koji u odnosu na jezični podražaj daje ispravan odgovor. Prema ovoj ideji, kada bi netko rekao stroju „Dobro jutro“, stroj bi trebao imati mehanizam M_1 koji bi na ovaj lingvistički podražaj odgovarao s „Dobro jutro i vama” ili nekim sličnim prikladnim odgovorom. Slično, ako mu se uputi rečenica „Kako se zoveš?”, drugi mehanizam M_2 bi odgovarao „Moje ime je Stroj koji govori“. I tako u nedogled za svaku moguću rečenicu prirodnog jezika. Budući da bi takav stroj trebao imati potencijalno beskonačan broj mehanizama koji korespondiraju svakoj mogućoj rečenici prirodnog jezika, jasno je da ga ne bi bilo moguće napraviti.

Možemo primijetiti da je Descartesovo mišljenje o praktičnoj nemogućnosti konstruiranja inteligentnog stroja utemeljeno na ideji stroja

iz 17. stoljeća. Tada dostupni strojevi i njihovi osnovni mehanički principi funkcioniranja, predstavljali su temelje za razumijevanje načina na koji se općenito ponašaju materijalna tijela. Stoga je Descartes smatrao da materijalna tijela općenito ne mogu fleksibilno koristiti jezik poput nas.

Na temelju ovog argumenta Descartes također zaključuje da su životinje koje nemaju sposobnost fleksibilnog korištenja jezika vrlo vjerojatno samo obični materijalni entiteti (ili strojevi) bez *res cogitansa*. Stoga njima nedostaju volicija i intelekt koji predstavljaju esencijalne modifikacije misleće supstancije. Štoviše, ako su osjetilna iskustva proizvod unije ili povezanosti uma i tijela, a životinje nemaju umove, onda slijedi da životinje ne mogu imati osjetilna iskustva. Stoga, iz Descartesovih gledišta slijedi da životinje zapravo ne mogu imati osjetilna iskustva poput osjećaja boli.

Kao što su već neki Descartesovi suvremenici primijetili, problem s ovim argumentom leži u tvrdnji da nijedan stroj ne može voditi razgovor.[8] Štoviše, ova tvrdnja je problematična iz istog razloga zbog kojeg su problematični prigovori redukcionističkom projektu koji je Descartes inače prihvaćao. Iako se Descartes mogao jedino voditi primjerima strojeva nastalih u njegovo vrijeme, kada nisu postojali strojevi koji mogu voditi razgovore, svejedno to nam ne daje induktivno uvjerljive razloge da smatramo da u *principu* nikakva organizacija materije ne može proizvesti sustav koji bi bio sposoban obavljati takvu vrstu zadatka. Nije teško zamisliti da materija djeluje u skladu s mehanizmima koji su suptilniji nego strojevi koje su ljudi stvorili u Descartesovo vrijeme. U novije doba možemo čak i smatrati da nam računala nude razloge da pretpostavimo kako će u doglednoj budućnosti postojati strojevi koji će moći voditi inteligentne razgovore s ljudima.[9]

Nadalje, ono što Descartes nije uzeo u obzir jest činjenica da je prirodni jezik rekurzivan. To znači da se temelji na pravilima gramatike koja omogućuju da iz konačnog broja temeljenih pravila i osnovnih jezičnih elemenata proizvedemo beskonačan broj smislenih rečenica. Ovo je svojstvo jezika vrlo moćno te nam omogućuje da razumijemo i proizvedemo potencijalno beskonačan broj rečenica. Kako bismo ilustrirali ovo svojstvo prirodnih jezika uzmimo u obzir sljedeću rečenicu: „Sjedim na kamenu koji se nalazi na Marsu i razmišljam da ovdje nema dovoljno kisika“. Iako je sasvim moguće da čitatelj ovog teksta nikada prije nije čuo ili pročitao ovu

[8] Keith T. Maslin (2001, 45) spominje Barucha de Spinozu (1632. – 1677.) koji je argumentirao da ne znamo što su materija (*res extensa*) i prirodni zakoni sposobni proizvesti (Spinoza 2000, Dio III, propozicija 2, scholium), pa stoga nemamo dovoljno razloga tvrditi da nije moguće da materijalni predmeti ne mogu imati intelektualne sposobnosti koje karakteriziraju misleću stvar.

[9] Alan Turing (1912. – 1954.), začetnik suvremene informatike i istraživačkog polja umjetne inteligencije, smatrao je da se ljudsko mišljenje može simulirati kompjuterskim programom te je i osmislio ono što se po njemu nazvalo Turingovim testom koji bi nam trebao omogućiti da odredimo je li stroj inteligentan poput ljudi te smijemo li mu pripisati mentalna stanja. Tom raspravom ćemo se baviti u poglavlju 4.

rečenicu, svejedno bez problema može razumjeti njezino značenje. Ono što ljudima omogućuje ovakvu vrstu razumijevanja je upravo (implicitno) poznavanje osnovnih elemenata ove rečenice i pravila kako se ona kombiniraju (vidi Fodor i Pylyshyn 1988). S obzirom na to svojstvo jezika nema potrebe pretpostaviti da prikladno odgovaranje na beskonačan broj rečenica zahtijeva postuliranje beskonačnog broja mehanizama koji korespondiraju tim rečenicama. To može biti jedan mehanizam koji se kroz razvoj i učenje prilagođava prikladnoj upotrebi lingvističkih izraza koji su mu potrebni za adaptaciju u toj okolini. U tom smislu čini se da je otvoreno empirijsko pitanje mogu li se strojevi programirati da nauče fleksibilno koristiti jezik.

Descartes je ponudio još jedan induktivan argument za razlikovanje uma od tijela koji se temelji na ideji *kognitivne fleksibilnosti*:

> A drugo sredstvo je, da, iako bi oni činili mnoge stvari isto tako dobro ili možda bolje nego itko od nas, oni bi zacijelo otkazali u nekim drugima, uslijed čega bi se otkrilo, da ne postupaju svjesno, već samo zbog takvog rasporeda svojih organa. Dok je naime um opće oruđe, koje može služiti u svim mogućim prilikama, ovi organi moraju biti za svaku pojedinačnu radnju i na poseban način udešeni. Stoga je moralno (praktično) nemoguće, da bi bilo dovoljno različitih organa u jednom stroju, da postupa u svim slučajevima u životu na isti način kao što postupamo mi, jer imamo um. (Descartes 1951, 47)

Ovaj se argument oslanja na pretpostavku prema kojoj je intuitivno da stroj ne može iznutra biti kompleksan koliko je, prema Descartesu, potrebno da manifestira općenitu kognitivnu fleksibilnost koju posjeduju ljudi.

Ovdje se opet čini da premise ne opravdavaju induktivan zaključak. Štoviše, moglo bi se prigovoriti da nam se ovaj zaključak čini donekle uvjerljivim samo zato što nemamo dovoljno znanja i nismo sposobni pojmiti ili zamisliti kako bi mogao postojati takav stroj. No, opet nije jasno da naša nemogućnost zamišljanja takvog stroja daje induktivne dokaze da jedan dan nećemo biti u stanju napraviti ga.

Argumenti koje smo dosad razmatrali bili su induktivni. Njima je Descartes nastojao pokazati da je vrlo vjerojatno da su um i tijelo različite supstancije. No, Descartes je još poznatiji po svojim deduktivnim argumentima u prilog dualizma supstancija. Tim je argumentima nastojao pokazati ne samo da je vrlo *vjerojatno* da je um različit od tijela već da su oni *nužno* različiti. U sljedećem odjeljku, razmotrit ćemo neke od tih argumenata koje Descartes raspravlja u kontekstu svojih metafizičkih istraživanja.

2.5 Argument iz sumnje

Descartesu se često pripisuje deduktivni argument za dualizam supstancija koji se može nazvati *argument iz sumnje*. U nastavku slijedi formulacija tog argumenta iz Descartesove knjige *Rasprave o metodi*:

> Zatim sam pažljivo proučavao, šta sam, i vidio sam, da mogu pretpostaviti, da nemam tijela i da ne postoje ni svijet ni mjesto, gdje se nalazim, ali da ne mogu zato pretpostaviti, da ja ne postojim i da, naprotiv, baš iz toga, što namjeravam sumnjati o istini drugih stvari, slijedi veoma očito i veoma sigurno, da ja postojim. Da sam pak samo prestao misliti premda bi sve ostalo, što sam ikada predočivao, bilo istinito ne bih imao nikakva razloga misliti, da sam postojao. Iz toga sam spoznao, da sam supstancija, koje je čitava bit ili priroda u tome da samo misli i kojoj za bivanje nije potrebno nikakvo mjesto niti zavisi od bilo koje materijalne stvari. Prema tome je ovo ja, t.j. duša, koje me čini onim, što jesam, potpuno različna od tijela i može se čak lakše spoznati nego tijelo, a da njega i nema, ona bi ipak ostala upravo to, što jest. (Descartes 1951, 52)

Formalnije, argument iz sumnje može se formulirati na sljedeći način:

1) Mogu sumnjati da protežne stvari postoje, stoga i da moje tijelo postoji.
2) Ne mogu sumnjati da ja postojim kao misleća supstancija.

Dakle:

3) Ja sam misleća supstancija koja može postojati neovisno o protežnoj supstanciji.

Premise ovog argumenta temelje se na primijeni metode koju Descartes naziva „metoda sumnje“. Svrha ove procedure je da odredi nedvojbena vjerovanja koja mogu pružiti temelje za sve znanje. Descartes opisuje ovu metodu na sljedeći način:

> […] pošto sam tada želio da se posvetim samo traženju istine, smatrao sam, da moram postupiti upravo obrnuto i odbaciti kao sasvim krivo sve, o čemu bih mogao i najmanje sumnjati, da vidim, ne će li nakon toga ostati nešto u mom uvjerenju, što bi bilo sasvim izvan svake sumnje. (Descartes 1951, 31)

Jedan od načina da ocijenimo argument iz sumnje jest da prvo provjerimo jesu li mu premise istinite. Međutim, kao što ćemo vidjeti u nastavku, taj korak neće biti potreban zato što argument nije valjan. Ako argument nije valjan, onda istinitost premisa ne garantira istinitost konkluzije.

Već su Descartesovi suvremenici ukazali na to da argument iz sumnje nije valjan. Kako ističe John Cottingham, ovaj se argument može smatrati „jednim od najnotornijih *non sequitura* u povijesti filozofije“ (Cottingham 2005, 242). Descartes, u predgovoru čitatelju u *Meditacijama o prvoj filozofiji,* knjizi koja je objavljena četiri godine nakon *Rasprave o metodi*, spominje kako je jedan kritičar (čije ime je s vremenom zaboravljeno te je ostao anoniman), primijetio sličnu slabost u argumentu iz sumnje:

> […] iz toga što ljudski um kada je okrenut samomu sebi ne spoznaje da je nešto drugo osim misleće stvari, ne slijedi da se njegova narav ili *bit* sastoji samo u tome da je misleća stvar, naime u smislu da bi riječ „samo“ [lat. *tantum*] isključivala sve ostalo za što bi se možda također moglo reći da pripada naravi duše. (Descartes 2015, 17)

Ukratko, prigovor se odnosi na tvrdnju da spoznaju o nekom stanju stvari ne smijemo temeljiti na sumnji koja se temelji na našem neznanju. U nastavku ćemo malo detaljnije razmotriti u čemu se sastoji problem s takvim načinom razmišljanja.

Jedan od načina da pokažemo da neki argument nije valjan jest da pronađemo argument koji ima istu formu, a znamo da mu je konkluzija neistinita. Razmotrimo sljedeći primjer:

1) Mogu sumnjati da sam ja četvrta osoba koja je ušla u dizalo danas.
2) Ne mogu sumnjati da sam ja osoba koja se nalazi u dizalu.

Dakle:

3) Nisam četvrta osoba koja je ušla u dizalo.

Jasno je da ovaj argument nije uvjerljiv. Konkluzija logički ne slijedi iz premisa. Razmotrimo još malo što čini ovakav tip argumenta nevaljanim.

Argument iz sumnje uključuje nedozvoljeni prijelaz od premisa o tome u što možemo i ne možemo *sumnjati* na konkluziju o tome kakvo je *stanje stvari*. Možemo reći da argument uključuje prijelaz iz *epistemičke* domene, koja se odnosi na ono što smatramo dvojbenim (npr. mogu sumnjati da imam tijelo), u *ontološku* domenu, koja se odnosi na to kakvo je zapravo stanje stvari (ja sam različit od svog tijela). Međutim, argumenti ove vrste nisu uvijek uspješni, jer naše sumnje o postojanju i prirodi nečega mogu ovisiti o našem ograničenom znanju ili netočnim vjerovanjima. Na primjer, razmotrimo prvu premisu argumenta. Očito mogu sumnjati u to da sam četvrta osoba koja je ušla u dizalo danas, jer ne znam tu činjenicu. Međutim, iz toga ne slijedi da ja nisam četvrta osoba koja je danas ušla u dizalo.

U ovom kontekstu vrijedi istaknuti da nije jasno da je Descartes stvarno prihvaćao ovu vrstu zaključivanja. Naime, nerijetko se u filozofskim raspravama suparnicima pripisuju argumenti koje zapravo ne prihvaćaju.

Kada namjerno ili nenamjerno dajemo nepreciznu rekonstrukciju suparničkih argumenata ili gledišta, onda činimo pogrešku znanu kao „slamnati čovjek“.[10] To je vrlo nekorektan potez, jer se nudi pojednostavljena i oslabljena verzija suparničkog gledišta ili argumenta koju se može lako kritizirati. Rizik da činimo pogrešku takve vrste je još veći kada se radi o filozofskim djelima iz daleke prošlosti gdje vremenska udaljenost i kulturne razlike mogu dovesti do ozbiljnih nesporazuma.

Sam je Descartes primijetio da se argument iz sumnje, kako ga je predstavio u *Raspravi o metodi*, čini problematičnim jer u toj knjizi nije eksplicirao sve njegove premise. Tvrdio je da se njegova puna snaga može shvatiti tek u kontekstu rasprave argumenata koje skeptici u epistemologiji općenito nude protiv mogućnosti spoznaje. No, unatoč tome, Descartes navodi da se u knjizi *Rasprava o metodi* ne bavi tim skeptičkim argumentima jer je ona napisana na francuskom, jeziku koji je dostupan manje obrazovanima, te se bojao da bi ova vrsta publike bila više zavedena ili impresionirana skeptičkim argumentima nego odgovorima koje on daje protiv njih (vidi Clarke 2006, 187). Zbog ovih razmatranja ostaje otvoreno pitanje je li Descartes stvarno prihvaćao argument iz sumnje kako smo ga ranije formulirali.

Neki suvremeni stručnjaci smatraju da Descartes nije prihvaćao argument iz sumnje. Na primjer, Margaret Wilson (1978, 167), istaknuta stručnjakinja za Descartesovu filozofiju, smatra da se Descartesu ne bi trebalo pripisivati argument iz sumnje. Tu tvrdnju temelji na tome što je Descartes ponovio sličan argument četiri godine kasnije u svojoj najpoznatijoj knjizi *Meditacije o prvoj filozofiji.* U odgovoru kritičaru, Descartes navodi da cilj ovog argumenta nije bio pokazati da su um i tijelo različite supstancije, već je cilj bio ustanoviti slabiju konkluziju prema kojoj možemo *znati sa sigurnošću* da smo misleća supstancija, a da istodobno ne posjedujemo znanje da imamo tijelo.[11] Nasuprot tome, John Cottingham (2005, 242), još jedan suvremeni stručnjak za Descartesa, tvrdi da je jasno da je Descartes podržavao argument iz sumnje kako je naveden u odlomku iz *Rasprave o metodi* koji smo prethodno spomenuli. Također, Cottingham problematizira tvrdnju da se Descartes uspješno distancirao od tog argumenta u *Meditacijama o prvoj filozofiji* (vidi Cottingham 2005, 244–45).

Iako nije sigurno je li Descartes prihvaćao argument iz sumnje, svejedno ga je bilo korisno razmotriti iz dvaju razloga. Kao prvo, na temelju njega vidjeli smo kako se argumenti tipično vrednuju u raspravama iz suvremene filozofije uma. Drugo, argument iz sumnje, poput ranije spomenutih induktivnih argumenata, oslanja se na intuiciju da se priroda uma ne može adekvatno objasniti u okviru prirodnih znanosti. Međutim, prigovor koji se stalno pojavljuje jest da ta vrsta intuicije više odražava naše neznanje o tome

[10] Na engleskom ova se pogreška naziva „Straw man“ ili „Straw person“.

[11] Vidi (AT VII, 225; HR II, 101), vidi i Predgovor čitatelju u Descartes (2015, 17 i 19).

kako funkcioniraju stvari i zakoni u prirodnom svijetu nego što bi bile izvor pouzdanog znanja o razlici između uma i tijela. Kao što ćemo vidjeti u sljedećem odjeljku, Descartes je ponudio puno razrađeniji argument u prilog dualizma kojim zapravo brani pouzdanost te intuicije.[12]

2.6 Argument iz jasnih i odjelitih ideja (ili percepcija)

Descartes je u *Meditacijama o prvoj filozofiji* (2015) u šestom poglavlju ponudio još jedan utjecajni argument u prilog dualizmu supstancija koji se naziva argument iz jasnih i odjelitih ideja ili percepcija.[13] Kako bismo shvatili njegov način zaključivanja nije naodmet ukratko opisati ciljeve i sadržaj *Meditacija o prvoj filozofiji*, knjige koja predstavlja jedno od temeljnih djela zapadne filozofije. Kako ističe u podnaslovu, Descartes u *Meditacijama* nastoji demonstrirati dvije konkluzije: postojanje Boga i odvojenost ljudskog uma od tijela. Budući da je Descartes prihvaćao dualizam moglo bi se pomisliti da je ujedno smatrao da je um besmrtan. Zapravo, Descartes je prihvaćao slabiju tezu. Smatrao je da je moguće da um i tijelo postoje neovisno jedno o drugome. Descartes je smatrao da pitanje nastavljaju li um ili duh postojati nakon smrti tijela pripada teološkim raspravama koje nadilaze njegovu ekspertizu (vidi Descartes 2014, 81-82 I, odlomak 51).[14]

Nadalje, Descartes pokazuje kako ove dvije teze, koje prema njemu čvrsto pripadaju području metafizike, nisu samo konzistentne s njegovom mehanističkom slikom prirodnog svijeta, nego su joj i nužno potrebne. U knjizi sve to nastoji postići pomoću rasprava koje su međusobno povezane kroz šest meditacija. Neke od njih ćemo ukratko sažeti fokusirajući se na

---

[12] Ono što ćemo još vidjeti jest da su upravo priroda i pouzdanost naših intuicija o naravi uma i njegove relacije s prirodnim svijetom središnje i naširoko raspravljane teme u suvremenoj filozofiji uma. Za više vidi poglavlja 8 i 9.

[13] Descartes je na latinskom obično koristio izraz *clara et distincta perceptio*. Kasnije u tekstu gdje iznosimo sam argument objašnjavamo što je Descartes mislio pod „perceptio“. No, vrijedi imati na umu da pod „perceptio“ Descartes ne misli na ono što bismo danas nazvali osjetilnom percepcijom, već misli na subjektivni čin shvaćanja ili poimanja stvari. Stoga je za njega jasna i odjelita percepcija određena vrsta intelektualnog čina. Vjerojatno se zbog toga kod nas često samo govori o jasnim i odjelitim idejama, a termin percepcija se obično ne spominje. Međutim, treba imati na umu da je Descartes pravio razliku između ideja i percepcija te da za njega govor o jasnim i odjelitim *idejama* ne znači isto što i govor o jasnim i odjelitim *percepcijama*. Nama ta razlika nije toliko bitna te za potrebe ovog rada koristimo ove termine kao sinonime.

[14] Vrijedi spomenuti da je u prvom izdanju *Meditacija*, zbog intervencija Marina Mersennea (1588. – 1648.) koji je pripremio rukopis za objavljivanje, u podnaslovu pisalo da knjiga sadrži dokaz besmrtnosti duše. U naknadnim izdanjima knjige Descartes je zahtijevao da se u podnaslovu spomene slabiji zaključak koji se referira na odvojenost ljudske duše od tijela. To je točniji opis Descartesovih zaključaka; naročito zato što Descartes nije htio ulaziti u teološke debate o tome što se događa s ljudima nakon smrti (Clarke 2006, 202–3).

aspekte koji su relevantni za njegov argument u prilog dualizma supstancija.[15]

U *prvoj meditaciji*, Descartes odlučuje posumnjati u svako uvjerenje koje je prethodno imao, ako postoji i najmanji razlog za smatrati da bi ono moglo biti neistinito. Cilj je ovog skeptičkog stava, koji je postao poznat kao *metoda sumnje*, utvrditi postoji li znanje u koje ne možemo posumnjati te kao takvo predstavlja siguran temelj za izgradnju svega ostalog znanja. Na temelju tog stava Descartes počinje sumnjati da postoji vanjski svijet kako mu ga predstavljaju osjetila, uključujući i ideju da posjeduje fizičko tijelo. Dakle, u ovoj *Meditaciji* Descartes odbacuje kao nesigurna temeljna načela poimanja fizičkog svijeta koje je razvio u sklopu svojih znanstvenih istraživanja. Ova koncepcija uključuje ideju da fizička stvarnost ima svojstva koja se mogu proučavati koristeći matematiku. U tom smislu, ono što Descartes nastoji istražiti jest može li se njegova slika fizičkog svijeta ponovno izgraditi na sigurnim temeljima.

U *drugoj meditaciji* Descartes zaključuje da ne može primijeniti metodu sumnje na vlastiti proces razmišljanja. Primjećuje da sve dok sumnja on sam mora postojati jer nema sumnje bez nekoga tko sumnja. Na temelju ovog uvida Descartes dolazi do poznatog zaključka: *Cogito ergo sum* (u prijevodu: Mislim, dakle jesam). Nadalje, iz ovog uvida Descartes zaključuje da je esencijalno svojstvo nas kao umnih bića to da imamo mišljenje. Ne možemo sumnjati da mislimo s obzirom na to da je sumnja jedna vrsta mišljenja. Štoviše, kako bi Descartes mogao sumnjati, on sam mora postojati. Stoga, jedino u što ne može sumnjati jest da je on *misleća* supstancija. Međutim, u ovom stadiju meditacija, Descartes ne može ustanoviti daljnji zaključak da je ta misleća supstancija odvojena od materijalne supstancije. Kako bi to utvrdio potrebne su mu dodatne premise koje brani u meditacijama (tj. poglavljima) koje slijede u knjizi.

U *trećoj meditaciji*, Descartes daje argument u prilog Božjeg postojanja. Ugrubo, možemo reći da u tom dijelu daje niz razloga, koji se prema njemu temelje na nepobitnim osnovama, te pokazuju da Descartes posjeduje pojam Boga koji je takav da ga ne bi mogao posjedovati kada Bog ne bi zaista postojao. Nakon što zaključi da Bog mora postojati, dalje argumentira da nas Bog, budući da je savršen u svakom pogledu, nije mogao stvoriti kao kognitivno defektne (u smislu da su nam spoznajne sposobnosti defektne). Dakle, kada koristimo svoje kognitivne sposobnosti kako treba onda prema Descartesu imamo jasne i odjelite ideje ili percepcije stvari. A kada donosimo sud da je nešto slučaj na temelju jasnih i odjelitih ideja onda nam taj kriterij garantira da je taj naš sud istinit. U nastavku *Meditacija*, Descartes koristi taj princip kako bi pokazao da njegova matematičko-znanstvena slika prirodnog svijeta zadovoljava ovaj kriterij jasnoće i odjelitosti te time utemeljuje novu

[15] Pri tome se najviše oslanjamo na tekst od Sergia Landuccija (1997).

znanost na pretpostavci postojanja Boga. U nastavku ćemo se usredotočiti na argument za dualizam supstancija koji Descartes daje u *šestoj meditaciji*.

U *šestoj meditaciji*, Descartes nudi sljedeći argument u prilog tezi da su um i tijelo različite supstancije:

> Prvo, budući znam kako sve one što jasno i odjelito razumijevam takvim od Boga može biti kakvim ga razumijevam; dovoljno je što mogu jasno i odjelito razumijevati jednu stvar bez druge, da bih bio siguran da je jedna različita od druge, jer barem Bog ih može postaviti ponaosob; i nije važno kojom moći biva tako da se smatraju različitim; te zato, samo zbog toga što znam da egzistiram i što opažam da ništa drugo ne pripada mojoj naravi ili esenciji nego jedino to što sam *stvar koja misli* ispravno zaključujem da se moja esencija sastoji jedino u tome što sam stvar koja misli. Pa iako možda (ili pak sigurno, kao što ću poslije kazati) imam tijelo, koje je sa mnom veoma prisno povezano, ipak – kako s jedne strane – imam jasnu i odjelitu ideju o sebi samome, ukoliko sam tek stvar koja misli i nije protežna, a – s druge strane – odjelitu ideju o tijelu ukoliko je tek protežna stvar, a ne ona koja misli, sigurno je da sam odista različit od svojeg tijela i mogu egzistirati bez njega. (Descartes 1993, 154)[16]

Kako bismo bolje razumjeli argument, predstavit ćemo njegovu formalnu strukturu.

Argument se može rekonstruirati tako da se podijeli u tri dijela.[17] Prvi dio argumenta glasi:

1) Razumijem *jasno* i *odjelito* da samo mišljenje pripada mojoj esenciji.
2) Razumijem *jasno* i *odjelito* da samo protežnost pripada esenciji tijela.

Dakle:

3) Razumijem *jasno* i *odjelito* da um postoji bez tijela.

---

[16] Citat je preuzet iz izdanja *Meditacija* koje je preveo Tomislav Ladan (1993). Sintagma „clare et distincto intelligo“ se različito prevodi u hrvatskim edicijama Descartesovih *Meditacija*. Na primjer, Josip Talanga (2015) tu sintagmu prevodi s „jasna i razlučena spoznaja“, dok Ladan prevodi s „jasno i odjelito razumijevanje“. Smatramo da izraz „razumijevanje“ u ovom kontekstu bolje odgovara latinskom izrazu „intelligo“ kako ga Descartes upotrebljava. Posebice zato što kod Descartesa jasno i odjelito razumijevanje predstavlja razlog na temelju kojeg spoznajemo nešto ili opravdavamo znanje nečega.

[17] Naša se rekonstrukcija argumenta oslanja, uz pojednostavljenja, na one predložene u Landucci (1997, LII–LV), Maslin (2001, 57) i Wilson (1999). Međutim, važno je spomenuti da su neki stručnjaci za Descartesa pružili sasvim različita čitanja ovog argumenta. Vidi, na primjer, Rozemond (1998, 383–87) i Hatfield (2014, 255–63).

Drugi se dio argumenta može izložiti oslanjajući se na premise koje Descartes brani na drugim mjestima u *Meditacijama*:

4) Ako Bog postoji i nije maliciozan, te ako razumijem jasno i odjelito jednu stvar bez druge, onda su te stvari različite jedna od druge.
5) Bog postoji i nije maliciozan.

Dakle:

6) Ako razumijem *jasno* i *odjelito* jednu stvar bez druge, onda su te dvije stvari različite.

Konačno, u trećem dijelu argumenta koristimo konkluzije 3) i 6) iz prethodnih argumenata:

7) Razumijem *jasno* i *odjelito* da um postoji bez tijela.
8) Ako *jasno* i *odjelito* razumijem um bez tijela, onda je um različit od tijela.

Dakle:

9) Um je različit od tijela.

Prva premisa argumenta uključuje tehnički pojam *jasnog i odjelitog razumijevanja*. Ovaj tip razumijevanja temelji se na jasnim i odjelitim percepcijama (ili idejama). U ovom kontekstu, pojam percepcije ne označava sposobnost da postanemo svjesni nečega kroz osjetila, kao kad vidimo ekran računala ispred nas ili čujemo kako se automobil približava. Ovdje se percepcija treba razumjeti kao unutrašnji čin intelektualnog *shvaćanja* pomoću kojeg smo u stanju zahvatiti određene istine.

Prema Descartesu, percepcija je *jasna* kada predstavlja određenu istinu na način koji nam ne dopušta da sumnjamo u nju. Kao primjer spominje jasno razumijevanje koje nam omogućuje da zahvatimo činjenicu da sve dok mislimo mi postojimo ili da se povijest ne može promijeniti. Jasne percepcije su suprotne nejasnim, zbrkanim ili opskurnim percepcijama.

Među jasne percepcije spadaju i one koje su razlučene ili odjelite. To su one jasne percepcije koje su striktno odijeljene od drugih percepcija:

> Razlučenom pak nazivam onu [percepciju][18] koja je pored jasnoće tako od svega drugoga odijeljena i odsječena da u sebi sadržava samo ono što je jasno. (Descartes 2014, 77)

Dakle, imati jasnu i odjelitu percepciju da je nešto slučaj je poseban način razumijevanja da je to slučaj. Ako percipiramo jasno i odjelito da je nešto

[18] Na ovom i drugim mjestima gdje Descartes koristi izraze *Claram et distinctam perceptio*, Talanga (2014; 2015) u prijevodu koristi izraze „jasna i odjelita spoznaja“. Kako bismo ostali dosljedni ranijem korištenju termina, u Talanginom prijevodu smo riječ „spoznaja“ zamijenili s „percepcija“.

činjenica, ne možemo imati razloga sumnjati da stvari nisu takve kakve su nam predstavljene u percepciji.

U prvoj premisi Descartes tvrdi da imamo poseban pristup činjenici da je mišljenje esencijalno svojstvo uma ili onoga što nas čini onime što jesmo. U drugoj premisi tvrdi da je protežnost esencijalno svojstvo materijalnih stvari. Ove dvije premise, barem iz naše subjektivne perspektive, povlače tvrdnju da mi jasno i odjelito shvaćamo ili razumijemo um i tijelo kao različite supstancije.

Međutim, Descartes je svjestan da jasno i odjelito razumijevanje nekog stanja stvari, iako predstavlja *subjektivni* temelj za nedvojbenost, ne mora nužno biti razlog za zaključiti sa sigurnošću da je to stanje stvari zaista takvo kakvo nam se čini. Zato on dodaje daljnju premisu koja garantira valjanost zaključivanja iz subjektivno shvaćenog jasnog i odjelitog razumijevanja nečega na objektivno postojanje toga što se razumije na taj način.

Četvrta premisa Descartesovog argumenta slijedi iz pretpostavke da Bog postoji i toga da on nije zloban. Bog je onaj tko garantira da sve što se može jasno i odjelito razumjeti ujedno i jest tako. Sve što se može razumjeti kao jasno i odjelito je ono za što si um ne može pomoći nego da vjeruje u to. Drugim riječima, po prirodi stvari je primoran suditi kao istinito ono što percipira, poima ili razumije kao jasno i odjelito. Dakle, ono što se percipira kao jasno i odjelito mora biti istinito. U suprotnom, Bog bi bio varalica. No, kada bi bio varalica, ne bi bio neograničeno dobar. Dakle, zahvaljujući Božjoj dobroti, iz jasnih i odjelitih percepcija da je um različit od tijela, slijedi da su oni različite supstancije. U nastavku ćemo raspraviti uvjerljivost ovog argumenta.

Premisa koja se oslanja na benevolentnost Boga može privući određene prigovore. Pozivanje na postojanje i dobrotu Boga kao jamstvo da sve što percipiramo jasno i odjelito zaista jest tako, predstavlja kontroverznu tvrdnju koju neće prihvatiti svi sudionici u raspravi. Kako bi bili pošteni prema Descartesu, važno je naglasiti da u *Meditacijama*, prije nego što je izložio argument iz jasne i odjelite percepcije, daje argumente u prilog tvrdnje da Bog postoji i da je dobar. Isto tako je važno naglasiti da su ti argumenti bili često osporavani kroz povijest filozofije. Međutim, za našu raspravu nije se potrebno osvrtati na probleme koji se odnose na filozofiju religije.[19] Čak i kada pitanje postojanja Boga ostavimo po strani, postoje problemi s drugim premisama Descartesovog argumenta iz jasne i odjelite percepcije koje vrijedi istaknuti.

Prva premisa oslanja se na sporne pretpostavke koje se odnose na spoznaju uma. Nazovimo ih *epistemičkim* pretpostavkama. Descartes smatra da imamo sposobnost sigurnog shvaćanja da je samo mišljenje esencijalno svojstvo uma. Pretpostavlja se da um može dati epistemički sigurno razumijevanje samog sebe.

[19] Za opis i raspravu argumenata za Božje postojanje, vidi npr. Davies (2004).

Međutim, mnogi autori su skeptični u pogledu ove tvrdnje. Na primjer, Wilson argumentira da je:

> […] problem »jasnog i odjelitog razumijevanja« još uvijek pred nama. Kako mogu znati da moja sposobnost da razumijem sebe kao misleću stvar, neovisno o tjelesnim atributima, nije posljedica neznanja toga »što je misao« – koliko god se čini da je različita i koliko god se čini da imam »intimno« shvaćanje misli? […] Mislim da odgovor mora biti da, strogo govoreći, ja ne mogu to znati – posebice jer se čini savršeno manifestnim da nemam svu dokaznu građu. Mi jednostavno nemamo temeljito razumijevanje ljudske kognicije […], te nema načina da znamo kakvi nas empirijski i pojmovni šokovi čekaju u budućnosti. (Wilson 1978, 17, prijevod autora)

Dakle, čini se da se u argumentu jasne i odjelite percepcije, poput argumenta iz sumnje, nekritički pretpostavlja pouzdanost našeg intuitivnog shvaćanja uma i njegove prirode. Nekritički se oslanjamo na reflektivno razumijevanje uma i svijeta kako bismo donijeli zaključke o njihovoj prirodi.

Descartesovi pokušaji argumentiranja koji se temelje na njegovoj metodi sumnje izgledaju dakle *prima facie* nezadovoljavajuće. Naravno, čitatelj mora sam za sebe odlučiti je li to zaista tako. U svakom slučaju, vidjet ćemo da u suvremenoj filozofiji uma postoje drugi važni argumenti u korist dualizma. Mnogi autori nastoje pokazati da refleksija o prirodi mentalnih stanja, kako su nam ona dana u trenutku kada ih imamo, nudi razumijevanje koje opravdava zaključak da ona ne mogu biti dio svijeta koji se istražuje u prirodnim znanostima. Neke od tih argumenata ćemo razmotriti u narednim poglavljima.

2.7 Argument iz nedjeljivosti uma

U šestoj meditaciji Descartes (2015) nudi još jedan argument u prilog dualizma koji se može formulirati na sljedeći način:

1) Sva protežna tijela su djeljiva (tj., moje tijelo je djeljivo).
2) Umovi nisu djeljivi (tj., moj um nije djeljiv).

Dakle:

3) Um je različit od tijela.

Ovaj argument je zanimljiv jer na intuitivno privlačan način pokazuje da je um nešto različito od tijela, a da se ne oslanja na naše neznanje o pravoj prirodi uma i materije.

Međutim, čak i ako se složimo da um nije isto što i tijelo, svejedno nije jasno da Descartes ovim argumentom pokazuje da postoje dvije različite *supstancije*. Vidjeli smo ranije da je supstancija ono što može postojati

neovisno o drugim stvarima. Netko bi mogao tvrditi da se um ne može dijeliti kao što možemo dijeliti materijalne predmete jer um uopće nije *vrsta stvari* koja se može dijeliti. Razmotrimo sljedeći argument koji ima sličnu formu kao argument iz djeljivosti:

1) Sva protežna tijela su djeljiva.
2) Sposobnost za igranje tenisa nije djeljiva.

Dakle:

3) Sposobnost za igranje tenisa je različita od tijela.

Ovdje se čini da je zaključak u redu. Složit ćemo se da sposobnost koja nam omogućava da igramo tenis nije isto što i tijelo kao fizička stvar. Međutim, iz toga ne slijedi da je sposobnost za igranje tenisa, pa time i bilo koja druga ljudska sposobnost, nekakva vrsta nefizičke supstancije koja ima svoja nefizička svojstva. Prema analogiji može se tvrditi da ni um nije vrsta nefizičke supstancije, već da je to skup aktivnosti, procesa i sposobnosti koji omogućavaju fizičkim entitetima da obavljaju različite stvari i zadatke. Na ovakvoj vrsti odgovora i gledištu na prirodu uma će naročito inzistirati filozofski biheviorist kojima ćemo se baviti u poglavlju 3.

2.8 Argumenti protiv dualizma

U prethodnim odjeljcima razmotrili smo neke od Descartesovih argumenata za dualizam supstancija. Nastojali smo ukazati na problematične aspekte tih argumenata. Međutim, to ne dokazuje da teze kartezijanskog dualizma nisu istinite. To samo dokazuje da ti argumenti nisu dobri (tj. njihove premise nisu istinite ili njihov zaključak ne slijedi iz premisa). Razmotrimo, na primjer, sljedeći loš argument za istinitu tezu:

1) Upoznao sam Napoleona u svojim snovima.
2) Možemo sanjati samo osobe koje su umrle.

Dakle:

3) Napoleon je mrtav.

Kako bismo osporili neku tezu nije dovoljno pokazati da su neki argumenti za tu tezu loši. Kako bismo to učinili moramo pronaći argument koji direktno pokazuje da je teza neistinita. Na primjer, možemo argumentirati da je određena teza neistinita zato što su ona, ili teze koje logički slijede iz nje, nekompatibilne s tezama za koje znamo da su istinite.

U nastavku ćemo razmotriti niz argumenata koji osporavaju središnje teze kartezijanskog dualizma.

2.9 Interaktivni dualizam i problem uzročnosti

Poznata kritika upućena Descartesovoj teoriji jest da nije jasno kako se njegovo shvaćanje interaktivnog dualizma može povezati s intuitivnom

idejom da postoji uzročna veza između uma i tijela. Ovu kritiku neshvatljivosti među prvima je formulirala češka princeza Elizabeta.[20]

Elizabeta je svoju kritiku iznijela u poznatom pismu kojeg je uputila Descartesu (Descartes i Princeza Elizabeta od Boemije 2007). U pismu Elizabeta tvrdi da nije jasno kako bismo trebali shvatiti uzročni odnos između uma i tijela ako se prihvate dvije Descartesove tvrdnje. Prva je da je um nematerijalna supstancija. Druga tvrdnja se odnosi na Descartesov postulat koji glasi: ako A uzrokuje gibanje materijalnog predmeta B, onda A i B moraju biti u kontaktu u prostoru.[21]

Kako bismo potvrdili da je Descartes zaista prihvaćao ovakvo gledište na uzročni odnos među materijalnim predmetima možemo razmotriti sljedeći citat:

> […] pod tijelom razumijevam sve ono što se može ograničiti nekim oblikom, opisati mjestom i tako ispuniti prostor da se iz njega isključuje svako drugo tijelo; sve ono što se percipira dodirom, vidom, sluhom, okusom ili mirisom; isto tako i sve ono što se giba na različite načine, no ne mislim gibanje od sebe samog, nego gibanje koje nastaje dodirom od nečega drugoga […]. (Descartes 2015, 47)

S obzirom na prethodne dvije tvrdnje koje karakteriziraju Descartesovo gledišta, Elizabeta je formulirala svoj prigovor pitajući se kako Descartes može objasniti sljedeće:

> […] kako ljudski um može odrediti tjelesne duhove (tj. tekućine u živcima, mišićima, itd.) da proizvedu voljne radnje, s obzirom na to da je samo misleća supstancija. Čini se da su svi uzroci kretanja proizvedeni guranjem (dodirom) stvari koja se miče, zbog načina na koji je gurnuta, ili zbog njezinih svojstava te oblika površine stvari koja se kreće. Kontakt je potreban za prva dva uvjeta, te ekstenzija (protežnost) za treći. [Međutim] ti u potpunosti isključuješ posljednje iz svog pojma duše (duha ili uma), a ovo prijašnje se čini nekompatibilnim s postojanjem nematerijalne stvari. (Elizabeta od Boemije 2007, 62)

[20] Princeza Elizabeta (26.12.1618. – 11.02.1680.) bila je najstarija kćer njemačkog palatina Fridrika V, koji je nakratko bio kralj Boemije (područje današnje Republike Češke), i princeze Elizabete iz obitelji Stuart. Elizabeta je bila predstojnica protestantskog samostana Herford u zapadnoj Njemačkoj. Tijekom svog života putem pisama vodila je intelektualne rasprave te je imala značajan utjecaj na mišljenja mnogih intelektualaca koji su djelovali u 17. stoljeću.

[21] Za detaljniji opis Elizabetinog argumenta, vidi Garber (2001, pogl. 8).

Vidjeli smo da prema Descartesu, nematerijalna supstancija nema lokaciju u prostoru. No, prema Descartesu uzročni odnos između predmeta pretpostavlja mogućnost dodira u prostoru. Elizabeta ukazuje da nije jasno kako te dvije pretpostavke mogu zajedno biti istinite. Ako nematerijalna supstancija nema lokaciju u prostoru, onda nije jasno kako može uzročno utjecati na fizičke stvari koje se nalaze u prostoru. Na primjer, zdravorazumski nam se čini jasnim da kada imamo želju uzeti mlijeko iz hladnjaka, onda barem ponekad to mentalno stanje uzrokuje određene radnje i promjene u fizičkom svijetu. Međutim, Elizabetin prigovor ukazuje na to da nije jasno kako prema Descartesu takva želja može uzrokovati ikakvu fizičku radnju ako kao nematerijalna supstancija mora biti u kontaktu s materijalnom supstancijom. Ovim prigovorom pojmovne neshvatljivosti Elizabeta ukazuje na tenziju između dvije osnovne komponente kartezijanskog dualizma; a to su teza dualizma prema kojoj um i tijelo predstavljaju dvije radikalno drugačije supstancije i teza interakcionizma, tj. ideja da mentalna stanja stoje u uzročnim odnosima s fizičkim stanjima.

2.10 Obrana kartezijanskog dualizma

Descartes na prigovor neshvatljivosti odgovara da nije točna pretpostavka da se o uzročnoj relaciji između uma i tijela treba razmišljati kao o uzročnom odnosu između dvaju materijalnih tijela. U tom pogledu navodi sljedeće:

> Jer, kada želimo objasniti neku poteškoću pomoću pojma koji se na nju ne odnosi, ne možemo, a da ne pogriješimo; isto kao što radimo pogrešku kada želimo jedan od ovih pojmova objasniti drugim; budući da su primitivni, svaki od njih može se shvatiti samo kroz sebe samoga. Iako nam je uporaba osjetila dala pojmove ekstenzije, oblika i gibanja koji su nam mnogo bliži od ostalih pojmova, glavni uzrok naših pogrešaka leži u tome što obično želimo koristiti te pojmove za objašnjenje onih stvari na koje se ne odnose. Na primjer, kada želimo upotrijebiti maštu da pojmimo prirodu duše, ili još bolje, kada želimo pojmiti način na koji duša pomiče tijelo, pozivajući se na način na koji jedno tijelo pokreće drugo tijelo. (Descartes 2007, 65)

Vidimo da u korespondenciji s Elizabetom, Descartes odgovara da se pojam uzročne relacije koji uključuje kontakt u prostoru odnosi samo na materijalne predmete te da je pogrešno primijeniti taj pojam u kontekstima kada ne govorimo o materijalnim predmetima koji se nalaze u prostoru. Naše razumijevanje uzročnih relacija između uma i tijela prema Descartesu ovisi o, kako navodi, primitivnom pojmu „jedinstva uma i tijela“. Prema Descartesu taj pojam stječemo kroz normalan razvoj i učenje. U tom pogledu Descartes navodi sljedeće:

> Priroda me također uči preko tih zamjedbi boli, gladi, žeđi itd. da u svojem tijelu nisam samo prisutan onako kako je brodar prisutan na brodu, nego da sam s tijelom najtješnje povezan i gotovo pomiješan, tako da s njim tvorim nešto jedinstveno. Inače kada se tijelo ozlijedi, ja – koji nisam ništa drugo nego stvar koja misli – ne bi osjećao bol zbog ozljede, nego bi tu ozljedu percipirao čistim intelektom, kao što brodar vidom percipira ako se nešto slomi na brodu. (…) Jer ti osjeti žeđi, gladi, boli, itd. sigurno nisu ništa drugo nego zbrkani modusi mišljenja, nastali od spajanja i kao neke mješavine uma i tijela. (Descartes 2015, 151)

Ovaj odgovor na Elizabetin prigovor možemo smatrati zadovoljavajućim samo ako uspijeva objasniti zašto nismo u stanju shvatiti ili razumjeti na koji način um uzročno interagira s tijelom. S obzirom na to možemo se pitati je li uopće shvatljiv taj primitivni pojam jedinstva uma i tijela koji Descartesov pojam uzročnosti podrazumijeva? Ako nije, onda prigovor koji je Elizabeta uputila i dalje stoji. Naime, i dalje ostaje nejasno kako bismo točno trebali shvatiti uzročnu vezu između uma i tijela ako oni predstavljaju dvije kategorički različite supstancije. Iako je ovo pitanje izrazito zanimljivo iz povijesnih i teorijskih razloga, kao što smo ranije rekli ovdje nam nije cilj baviti se egzegetikom Descartesovih djela (za raspravu, vidi Garber 2001, 178–86). U nastavku ćemo razmotriti još jedan argument kojim se stavlja dodatni pritisak na tezu interakcionističkog dualizma.

2.11 Prigovor iz uzročne zatvorenosti fizičkog svijeta

Glavna premisa sljedećeg argumenta protiv dualizma temelji se na principu uzročne potpunosti fizike (skraćeno ga možemo zvati PUPF) (Papineau 2002; Smith i Jones 1988).[22] Tim Crane formulira ovaj princip na sljedeći način:

> Svaki fizički događaj ima fizički uzrok koji je dovoljan da ga proizvede, uzimajući u obzir zakone fizike. (Crane 2009, 45)

Ovim principom se pretpostavlja da su fizički uzroci *dovoljni* da proizvedu fizičke učinke. To znači da kada god imamo neki fizički događaj kao učinak ili posljedicu nekog drugog događaja, prema PUPF-u slijedi da postoji barem jedan fizički uzrok tog događaja. Na primjer, ako se opazi određena kemijska reakcija (kao jedna vrsta fizičkog događaja), prema PUPF-u slijedi da postoji neki kemijski, tj. fizički uzrok koji je dovoljan da proizvede tu kemijsku reakciju.

[22] Ovaj princip još se naziva princip uzročne zatvorenosti fizičkog svijeta.

Treba primijetiti da PUPF ne isključuje da fizički događaji mogu imati i neku drugu vrstu nefizičkih uzroka. Ono što se tvrdi PUPF-om jest da koji god bili uzroci fizičkog događaja, mora postojati barem jedan fizički uzrok koji je dovoljan da ga proizvede. PUPF također ne pretpostavlja da je determinizam istinit. Ovim se principom tvrdi da svaki fizički događaj B, ako ima uzrok, ima barem jedan dovoljan fizički uzrok A, koji uzrokuje B u skladu sa zakonima fizike. Prijelazi između fizičkih događaja mogu biti deterministički. To znači da, za događaj B, postoji samo jedan uzročni slijed koji uključuje prethodne fizičke događaje koji uzrokuju B. Međutim, prijelazi između fizičkih događaja mogu biti i nedeterministički. To bi značilo da, za događaj B, postoji više fizičkih uzroka, tj. prethodnih fizičkih događaja koji su mogli, s različitim vjerojatnostima, uzrokovati B. Za argument protiv dualizma koji slijedi nije važno odlučiti se između ovih dvaju tumačenja PUPF-a.

Postoje različita gledišta na ontološki i epistemološki status PUPF-a. Međutim, obično se PUPF brani kao opća metodološka pretpostavka koja je više od jednog stoljeća omogućavala uspješna empirijska i općenito znanstvena istraživanja (vidi Papineau 2002). U tom smislu PUPF treba prihvatiti zato što imamo induktivne dokaze da je on omogućio uspješna objašnjenja i istraživačke programe u različitim znanostima. Na primjer, fiziologija, neurobiologija, neuroznanosti i molekularna genetika primjeri su uspješnih znanosti koje su se razvile na temelju pretpostavke da se fizički fenomeni koje izučavaju mogu objasniti kao posljedice drugih fizičkih fenomena. U tom pogledu, PUPF se može shvatiti kao metodološka uputa: „Za objašnjenje fizičke pojave, traži fizički uzrok te pojave“.

Samo prihvaćanje PUPF-a nije dovoljno za osporavanje interakcionističkog dualizma. Kao što smo ranije istaknuli, PUPF sam po sebi ne negira da fizički događaji mogu uzrokovati mentalne događaje niti da mentalni događaji mogu uzrokovati fizičke događaje. Kako bismo to uvidjeli, razmotrimo sljedeći argument:

1) Mentalni događaji mogu uzrokovati fizičke događaje. (teza interakcionizma)
2) Mentalni događaji nisu fizički događaji. (teza dualizma)
3) Svi fizički učinci imaju dovoljne fizičke uzroke. (PUPF)

Dakle:

4) Neki fizički događaji imaju istovremeno dva različita dovoljna uzroka: jedan fizički i drugi mentalni.

Kako bismo objasnili ove premise oslonit ćemo se na sljedeći primjer. Moja želja za pivom, koja je mentalni događaj u trenutku t_0, uzrokuje moje otvaranje hladnjaka, što je fizički učinak u trenutku t_1. Ovo je primjer interakcije između mentalnih i fizičkih stanja. Međutim, prema PUPF-u mora postojati i neki fizički događaj F u trenutku t_0, koji je dovoljan da uzrokuje moje otvaranje hladnjaka u trenutku t_1. Dakle, ako prihvatimo premise 1) – 3), čini se da u trenutku t_0 postoji jedan mentalni događaj, moja želja za

pivom, te ujedno postoji jedan fizički događaj koji su različiti, ali ujedno dovoljni da uzrokuju moje otvaranje hladnjaka u trenutku t_1. U tom pogledu prihvaćanje PUPF-a kompatibilno je s prihvaćanjem interakcionističkog dualizma. Međutim, ako prethodnom argumentu dodamo još jednu uvjerljivu premisu da događaji ne mogu sistematski imati dva različita uzroka, onda ćemo vidjeti da prihvaćanje PUPF-a nije kompatibilno s kartezijanskim dualizmom. Drugim riječima, pod pretpostavkom da je PUPF istinit morat ćemo ili odustati od 1) teze interakcionizma ili 2) teze dualizma. U nastavku ćemo razmotriti tako nadopunjeni argument.

Ideja da jedan tip događaja ne može imati sistematski dva ili više različitih dovoljnih uzroka naziva se princip nepreodređenosti ili nepredeterminiranosti (engl. *non-overdetermination*). Prema ovome principu:

> 5) Nije moguće da jedan fizički događaj uvijek ima istodobno dva različita dovoljna uzroka. (princip nepreodređenosti)

Razmotrimo malo pobliže što je preodređenost općenito i zašto je neuvjerljivo tvrditi da ona postoji u slučaju odnosa mentalnih i fizičkih stanja.

Ideja uzročne preodređenosti nije po sebi sporna. Neki događaji mogu biti uzročno preodređeni. Radi ilustracije takvog događaja obično se koristi primjer sa streljačkim vodom. Zamislimo da imamo dva pucača koji ciljaju na zatvorenika osuđenog na smrt. Pretpostavimo da su oba pucača izvrsni strijelci te da je samo jedan hitac bilo kojeg pucača dovoljan da usmrti zatvorenika. Pretpostavimo da obojica pucaju na zatvorenika i ubiju ga. U tom slučaju ubojstvo zatvorenika ima dva dovoljna uzroka te možemo reći da je taj događaj uzročno preodređen. Preodređenost ovdje uključuje pretpostavku da čak i da prvi pucač nije pucao, hitac drugog pucača bi ubio zatvorenika i obrnuto; da drugi pucač nije pucao, hitac prvog pucača bi bio dovoljan da ubije zatvorenika. Takva vrsta preodređenosti nije čudna jer je na neki način kontingentna. Ono što bi bilo čudno jest da je jedan događaj *sistematski* ili *nužno* predodređen dvama ili više drugih događaja.

Međutim, ako se istodobno prihvati teza dualizma, teza uzročnog interakcionizma i PUPF, onda bi slijedilo da su neki fizički događaji sistematski preodređeni. Pretpostavka postojanja takve vrste sistematske preodređenosti kada govorimo o odnosu mentalnih i fizičkih događaja dovodi do problematičnih posljedica. Crane (2001, 50) neuvjerljivost te pretpostavke objašnjava na sljedeći način. Kad bi preodređenost bila istinita onda bi naš mentalni život bio samo kontingentno (tj. ne nužno) odgovoran za ono što radimo. Kada djelujemo zbog vjerovanja, želja, osjećaja, i tome sličnog, imamo doživljaj da su naše tjelesne aktivnosti dobro integrirane i usklađene s tim mentalnim događajima. Primjerice, nama se čini da je želja za sladoledom uzrokovala radnju da ga kupimo u trgovini. Da je bol nakon

uboda igle uzrokovala da jaučemo i tome slično. Kad ne bismo imali tu želju onda u tom trenutku ne bismo kupili sladoled u trgovini. Kad nas ne bi boljelo ne bismo jaukali itd. Međutim, kad bi preodređenost fizičkog i mentalnog uzrokovanja bila istinita onda čak i kad ne bismo imali želju za sladoledom mi bismo ga kupili; čak i kad ne bismo osjećali bol, jaukali bismo zbog uboda igle itd. Naime, budući da pretpostavljamo PUPF, slijedi da čak i kada ne bi postojali mentalni događaji kako ih pretpostavlja kartezijanski dualizam, svejedno bi postojali fizički događaji koji bi bili dovoljni da proizvedu tipične uzroke koje standardno pripisujemo tim mentalnim događajima. Dakle, barem u slučaju mentalnog uzrokovanja čini se da nije plauzibilno pretpostaviti postojanje sustavne preodređenosti. U suprotnom bi naš mentalni život bio suvišan kao objašnjenje toga što radimo. Na neki način, mentalni život bio bi eksplanatorno nerelevantan, zato što bi za objašnjenje fizičkih događaja i radnji bilo dovoljno pozvati se na fizičke događaje koji ih uzrokuju.

Da rekapituliramo. Ono što smo dosad pokazali jest da se sljedeće teze:

1) Mentalni događaji mogu uzrokovati fizičke događaje. (teza interakcionizma)
2) Mentalni događaji nisu fizički događaji. (teza dualizma)
3) Svi fizički učinci imaju dovoljne fizičke uzroke. (PUPF)

Mogu povezati u valjani argument prema kojemu slijedi da:

4) Neki fizički događaji imaju sistematski istovremeno dva različita dovoljna uzroka: jedan fizički i drugi mentalni.

Međutim, vidjeli smo da princip nepreodređenosti povlači da je zaključak 4) neistinit, s obzirom na to da:

5) Nije moguće da jedan fizički događaj uvijek ima istodobno dva različita dovoljna uzroka. (princip nepreodređenosti).

Ovaj argument nas navodi na zaključak da nije moguće istodobno prihvatiti 1), 2) i 3). Prvo, ako smatramo da je teza interakcionizma uvjerljiva i da PUPF predstavlja dobro utemeljen empirijski princip, onda se čini da moramo odustati od teze dualizma. To gledište ide prema nekoj varijanti fizikalizma u filozofiji uma.[23] Drugo, ako smatramo uvjerljivim teze kartezijanskog dualizma, onda moramo odustati od PUPF-a i znanstvene slike svijeta koji on podrazumijeva. Treće, ako smatramo da je PUPF uvjerljiv i da je dualizam istinit, trebali bismo odustati od teze interakcionizma. Ovo gledište više nije popularno jer se njime negira intuitivna ideja da mentalna stanja imaju uzročne moći te da ih se kao takve može znanstveno istraživati. No, u povijesti filozofije neki od poznatijih filozofa prihvaćali su ovakvo gledište. U

[23] Vidi poglavlja 4, 6 i 7.

nastavku ćemo spomenuti neka od gledišta koja su negirala tezu interakcije mentalnog i fizičkog.

Dualisti koji su odbacili interakcionizam (i time Descartesov dualizam) podijelili su se u dvije grupe. Jedni su zastupali *epifenomenalizam.* To je teorija prema kojoj mentalni događaji ne uzrokuju fizičke događaje, dok fizički događaji mogu uzrokovati mentalne događaje. U novije vrijeme varijantu epifenomenalizma branio je poznati filozof Frank Jackson (1982). No, kasnije je odustao od te pozicije te je prihvatio varijantu fizikalizma ili materijalizma prema kojemu um ovisi o tijelu (za pregled rasprave, vidi W. S. Robinson 2010).

Drugi su zastupali neku varijantu *paralelizma.* To je gledište prema kojemu se u domeni mentalnih stanja uzročni odnosi odvijaju paralelno s uzročnim odnosima u fizičkoj domeni. No, između mentalne i fizičke domene zapravo nema uzročnih odnosa. Jedan od najpoznatijih zastupnika ovog gledišta je Gotfried Leibniz (1646. – 1716.). Prema Leibnizu (1980, pogl. XIII), pretpostavka da mentalna stanja stoje u uzročnim odnosima s fizičkim stanjima posljedica je pogrešnog zaključivanja gdje iz korelacije između mentalnih i fizičkih događaja zaključujemo da oni stoje u uzročnim odnosima. Na primjer, kada nas igla ubode u prst, nama se čini da ubod kao fizički događaj uzrokuje bol kao mentalni događaj. No, zapravo to nije točno. Prema Leibnizu, mentalni i fizički događaji savršeno su korelirani (tj. odvijaju se paralelno) zbog prestabilirane harmonije između mentalnih i fizičkih događaja koju je uspostavio Bog u trenutku stvaranja svijeta.

Leibniz svoju varijantu paralelizma ilustrira s primjerom sata. Zamislimo da imamo dva starinska zidna sata s kukavicom koja uvijek pokazuju isto vrijeme. Kada se na jednom pomakne kazaljka istodobno se i na drugom pomakne kazaljka. Kada na jednom izađe ptica kukavica i na drugome izađe ptica kukavica itd. Budući da iz promatranja jednog sata možemo predvidjeti ponašanje drugog sata može se činiti da je rad tih satova na neki način povezan. U tom pogledu, možemo razmotriti tri opcije. Jedna je da ti satovi nekako uzročno utječu jedan na drugog te time održavaju sklad među svojim mehanizmima. Druga je da netko stoji kod tih satova te konstantno osigurava da se njihov rad odvija paralelno. Treća opcija je da je netko te satove izradio tako da funkcioniraju u savršenoj usklađenosti tako da ih ne treba konstantno podešavati. Leibniz smatra da ova treća mogućnost dobro opisuje odnos između uzročnih odnosa između mentalne i fizičke domene. Prema njemu, Bog pri stvaranju svijeta oblikuje mentalnu i fizičku supstanciju tako da one prema svojim zakonitostima postaju u potpunosti usklađene. Tako da unatoč prividu da postoje uzročni odnosi između mentalne i fizičke domene, njih zapravo nema. Slično kao što nam se može činiti da između savršeno podešenih satova kukavica postoji uzročni odnos iako ga zapravo nema. Ono što predstavlja pravo objašnjenja tog prividnog

uzročnog odnosa jest neovisan uzrok (Bog ili čovjek) koji ih pri stvaranju usklađuje.

Drukčiju varijantu paralelizma branio je Nicolas de Malebranche (1638. – 1715.). Njegovo gledište naziva se okazionalizam. Prema njemu, svaki uzročni odnos između mentalnog i fizičkog posljedica je božanske intervencije. Na primjer, ako se ubodemo iglom i osjetimo bol, onda nije tako da ubod igle direktno uzrokuje bol, već Bog intervenira i osigurava da se mentalno stanje pojavi kao uzročna posljedica. Koristeći Leibnizovu metaforu sata, Malebranchovo gledište moglo bi se opisati drugom opcijom prema kojoj ono što objašnjava savršenu usklađenost rada dvaju satova jest činjenica da neka osoba svakog časa intervenira kako bi se održao sklad među njima. Štoviše, Malebranche je općenito smatrao da kada god dva događaja stoje u uzročno-posljedičnom odnosu, Bog intervenira kako bi se on ostvario. Zato se njegova pozicija naziva okazionalizam; svaki uzročni događaj predstavlja priliku (lat. *occasio* = prigoda ili zgoda) za Boga da izvrši intervenciju u prirodnom svijetu. Dakle, za razliku od Leibniza, Malebranche nije vjerovao u prestabiliranu harmoniju, već je smatrao da svaki put kada nam se čini da dva događaja stoje u uzročnim odnosima zapravo Bog aktivno intervenira kako bi proizveo taj događaj.

Ova gledišta kao odgovori na problem odnosa uma i tijela više nisu popularna u suvremenim raspravama. Kako je sam Descartes tvrdio i što mnogi suvremeni filozofi uma prihvaćaju, čini se intuitivno jasnim da postoji uzročna interakcija između uma i tijela. Stoga svaka pozicija koja odbacuje tezu interakcionizma može djelovati neuvjerljivo do te mjere da i materijalizam u filozofiji uma postaje privlačno gledište. S obzirom na problem uzročnosti koji se javlja za kartezijanski dualizam, i sama Elizabeta izrazila je sklonost prihvaćanju neke vrste materijalizma. U tom pogledu navodi:

> […] priznajem da bi mi bilo lakše pridati umu materijalnost i ekstenziju, nego što bi mi bilo pridati nematerijalnoj stvari sposobnost pomicanja tijela […]. (Descartes 2007, 68).

I doista, mnogi suvremeni autori u filozofiji uma zbog problema uzročnosti koji se pojavljuje za dualističke pozicije prihvaćaju neku varijantu monizma prema kojem postoji samo jedna vrsta supstancije, pod čime se obično misli na onu vrstu supstancije koju bi Descartes okarakterizirao kao fizičku.[24] U

[24] Osim materijalističkog ili fizikalističkog monizma, također treba istaknuti da postoji i idealistička varijanta monizma. To bi bilo gledište prema kojemu postoji samo jedna vrsta supstancije i to ona duhovna. Najpoznatiji autor koji je zastupao tu vrstu idealizma je poznati filozof i katolički biskup George Berkeley (1685. – 1753.). Međutim, u suvremenim raspravama o odnosu uma i tijela nema istaknutih zastupnika ovog gledišta pa se njime nećemo baviti u ovoj knjizi.

sljedećim poglavljima bavit ćemo se nizom pozicija koje pripadaju ovoj skupini.

2.12 Zaključak

Filozofska djela Renéa Descartesa izrazito su važna za filozofiju uma. Prvo, proširujući mehanistička objašnjenja kako bi se opisalo i objasnilo funkcioniranje ljudskog tijela, uključujući i mnoge sposobnosti koje pripadaju domeni mentalnog, Descartes je ponudio gledište koje je bilo u suprotnosti s tradicionalnim načinima objašnjavanja ljudskog života i uma. U tom smislu Descartes s novom koncepcijom prirodnog svijeta povezuje tendenciju da se u znanstvenim terminima objasni veliki dio ljudskog ponašanja. Kao što ćemo vidjeti u narednim poglavljima, ova tendencija da se sve više aspekata ljudskog ponašanja i mentalnog života uključi u znanstvenu sliku svijeta predstavlja konstantnu težnju u filozofskim spekulacijama koja će konačno dovesti do materijalističkih teorija uma.

Unatoč tome, Descartes je smatrao da se ovaj mehanistički projekt suočava s nepremostivim ograničenjima. Ta ograničenja ispitivao je u sklopu svojih istraživanja temelja znanosti i ljudskog znanja. Kao što smo vidjeli, u srži je njegovih istraživanja pretpostavka da se spoznaja našeg uma i njegovih struktura može ostvariti pomoću privilegirane forme spoznaje ili kako Descartes navodi „jasnih i odjelitih percepcija“. Prema Descartesu, ove percepcije nude jasan i odjelit uvid u odvojenost naših materijalnih tijela i našeg duha. Ovaj uvid je utemeljen, poput matematike i matematičke koncepcije svijeta koju nudi znanost njegova vremena, na najboljim izvorima znanja koji su tada bili dostupni. No, kao što smo vidjeli, daljnji korak u Descartesovoj ratifikaciji ovog izvora znanja je postojanje benevolentnog Boga.

Međutim, čak i ako ovu pretpostavku ostavimo po strani, nastojali smo pokazati da se Descartesovi argumenti temelje na spornim *intuicijama* o tome koja je prava priroda naših umova i koje su granice do kojih ga možemo znanstveno istraživati i uklopiti u sliku prirodnog svijeta. U tom pogledu, ukazali smo na to da je uvijek moguće da su naše intuicije zapravo samo posljedica našeg neznanja o tome kako um, pa i prirodni svijet, zaista funkcioniraju. Međutim, u narednim poglavljima ćemo vidjeti da je pitanje pouzdanosti izvora na kojima temeljimo razmišljanja o prirodi uma i odnosa s prirodnim svijetom i dalje otvoreno te da predstavlja važan problem u daljnjem razvoju rasprava u filozofiji uma.

U sljedećem poglavlju, bavit ćemo se biheviorističkim gledištima na prirodu uma i mentalnih stanja te ćemo razmotriti kako različite varijante biheviorizma nastoje riješiti problem odnosa uma i tijela s kojima se suočava kartezijanski dualizam.

3 Biheviorizam u filozofiji uma

3.1 Uvod

U ovom poglavlju odmičemo se od rasprave kartezijanske filozofije uma koja se razvija u 17. stoljeća i prelazimo na gledišta koja se razvijaju u prvoj polovici 20. stoljeća pod skupnim nazivom biheviorizam. To ne znači da se ništa relevantno nije događalo u međuvremenu. Međutim, kao što navodi Gilbert Ryle (1949, x), čija ćemo gledišta razmotriti kasnije u ovom poglavlju, u tom vremenu su se, na pozadini kartezijanskog problema uma i tijela, izmjenjivala i razvijala poglavito dualistička i materijalistička gledišta uz povremeno isticanje idealističkih doktrina. Nasuprot tim gledištima, razvoj biheviorizma, u sadržajnom i metodološkom smislu, predstavlja zanimljivu i inovativnu alternativu.

Važno je istaknuti da je biheviorizam, u svojim filozofskim i psihološkim inačicama, motiviran inovativnom metodološkom pretpostavkom. Iako ćemo se u ovom poglavlju dotaknuti metodoloških uvida koji motiviraju biheviorizam u psihologiji, naš fokus će biti na isticanju utjecaja lingvističke analize koju koriste različite filozofske škole mišljenja pri elaboraciji svojih inačica biheviorizma. Vidjet ćemo da unutar tog teorijskog okvira filozofsko istraživanje uma podrazumijeva pažljivu lingvističku analizu svakodnevnih izraza koje koristimo kada govorimo i razmišljamo o umu. Ovakav analitički pristup koristio se kako bi se sustavno ispitali izvori, priroda i značajnost problema koji karakteriziraju filozofiju uma, a proizlaze iz kartezijanske slike mentalnog i fizičkog. U tom pogledu, filozofski biheviorizmi predstavljaju važnu polazišnu točku za razmatranje kasnijeg razvoja filozofskih rasprava o odnosu uma i tijela kojima ćemo se baviti u poglavlju 4.

3.2 Različiti tipovi biheviorizama

Početkom 20. stoljeća dolazi do zaokreta u filozofiji uma i znanstvenoj psihologiji. Mnogi se autori udaljavaju od kartezijanske koncepcije uma te nastoje utemeljiti psihologiju kao empirijsku znanost koja bi barem metodološki bila bliska prirodnim znanostima. Za mnoge autore to je značilo prihvaćanje ideje da se psihologija mora baviti javno dostupnim podacima te

da ne smije postulirati unutrašnja stanja koja nisu dostupna za promatranje u sklopu znanstvenog istraživanja. Stoga su neki zastupali ideju da se psihologija mora baviti ponašanjem te da se činjenice o mentalnim stanjima mogu, ili bi se trebale moći reducirati, na činjenice o ponašanju. Stoga se ta vrsta gledišta naziva biheviorizmom (engl. *behavior* = *ponašanje*).

Postoje različiti tipovi biheviorizama. U sljedećim odjeljcima razmotrit ćemo dva osnovna tipa. Metodološki i filozofski biheviorizam.

Metodološki ili psihološki biheviorizam jest gledište o tome kako bismo trebali formirati psihološke teorije i na koji način bi se trebala provoditi istraživanja u psihologiji. Osnovna pretpostavka ovog gledišta je da psihološke teorije smiju kao dokaznu građu koristiti samo javno ili intersubjektivno opažljive entitete. Pod javno ili intersubjektivno opažljivim entitetima misli se na podražaje, bihevioralne ili ponašajne reakcije te uvjete u okolini koji mogu utjecati na ponašanje organizama. Ovu viziju psihologije najjasnije ističe John B. Watson, jedan od začetnika metodološkog biheviorizma. Prema njemu:

> Psihologija […] je čisto objektivna eksperimentalna grana prirodne znanosti. Njezin teorijski cilj je predviđanje i kontrola ponašanja (Watson 1913).

Ovaj tip biheviorizma suprotstavlja se upotrebi introspektivnih metoda, kao i psihološkom istraživanju koje bi se temeljilo na takvim metodama. Različiti teorijski pravci u znanstvenoj psihologiji koji se razvijaju u drugoj polovici 19. i početkom 20. stoljeća stavljaju naglasak na introspekciju kao metodu za istraživanje psiholoških procesa (vidi, npr. James 1995). Introspekcija se obično shvaća kao sposobnost uvida u vlastita mentalna stanja i procese. Štoviše, neki psiholozi i filozofi shvaćaju introspekciju kao unutarnju percepciju ili oko našeg uma kojim promatramo privatne mentalne procese kojima nitko nema pristup osim nas samih. U tom smislu, smatra se da nam introspekcija jedina može otkriti pravu prirodu svjesnih doživljaja i iskustva.

Bihevioristi u psihologiji bili su skeptični prema metodama istraživanja koje se temelje na introspekciji. Kao što su i sami pobornici ove metode primijetili, introspekcija je nepouzdana (Wundt 1900). Ova se metoda koristila tako da bi se subjektu istraživanja predstavio nekakav podražaj te se od njega očekivalo da na temelju samopromatranja ponudi izvještaj o sadržaju svojih misli ili iskustva koje ima kada je izložen tom podražaju. Međutim, istraživanja su pokazala da se često događa da različiti subjekti daju različite odgovore i opise iskustava iako su bili izloženi istom podražaju. Stoga se ova metoda pokazuje kao nepouzdana za izgradnju čvrstih generalizacija o odvijanju psiholoških procesa. Nadalje, metoda ima ograničenja jer se ne može koristiti kod male djece ili životinja koje nam ne mogu dati izvještaje o svojim misaonim procesima te je od ograničene koristi

za istraživanje učenja, mentalnih poremećaja i razvoja. Stoga su bihevioristi smatrali da je introspekcija previše ograničena i nepouzdana metoda koja nema znanstvenu vrijednost te ne bi trebalo mjeriti uspješnost psihološkog istraživanja mogućnošću njezine primjene (Graham 2019).

Ovdje je važno istaknuti da metodološki biheviorizam ne nastoji objasniti temeljnu prirodu mentalnih stanja. Niti nastoji riješiti problem odnosa uma i tijela, ako se to uopće može uzeti kao problem. Možemo reći da zastupnici metodološkog biheviorizma prije ignoriraju ili ostavljaju po strani taj filozofski problem te nastoje razviti cjelovitu teoriju ljudskog ponašanja koja neće biti opterećena mentalističkim vokabularom i problemima koje njegovo korištenje nosi sa sobom. Takva pozicija slijedi iz odustajanja od korištenja introspekcije kao metode. To zaključivanje bi se moglo rekonstruirati na sljedeći način: introspekcija kao metoda jest nepouzdana i subjektivna te je stoga ne treba koristiti u znanstvenim istraživanjima. Introspekcija predstavlja osnovnu metodu istraživanja svjesnih mentalnih stanja kojima samo mi imamo pristup. Dakle, treba napustiti istraživanje svjesnih mentalnih stanja.

Analitički ili logički biheviorizam je gledište prema kojemu opisi mentalnih stanja jednostavno znače, ili se mogu prevesti u, iskaze o tome kako se neka osoba ponaša. Na primjer, prema logičkim bihevioristima značenje rečenice „Ivica osjeća bol“ može se u potpunosti prevesti koristeći rečenice poput „Ivica radi grimase i stenje“ ili opisom prema kojemu ima dispoziciju da se ponaša na taj način. Kada kažemo da Ivica ima dispoziciju ponašati se na određeni način želimo reći da bi se ponašao na taj način kada bi bio izložen određenim podražajima. Ovdje je središnja ideja da filozofske zagonetke koje se odnose na probleme odnosa uma i tijela generira nedostatno pridavanje pažnje upotrebi i značenju jezičnih izraza. Prema logičkom biheviorizmu, metoda filozofije uma je apriorna analiza značenja rečenica kojima opisujemo i ljudima atribuiramo mentalna stanja.

U ovom poglavlju najviše ćemo se baviti analitičkim ili logičkim biheviorizmom. Međutim, prije nego krenemo s tom raspravom iznijet ćemo neke od općenitijih razloga zašto je važno uzeti u obzir bihevioristička stajališta u pogledu prirode uma i znanstvene metodologije.

3.3 Temelji biheviorizma

Jedan od razloga zašto su neki psiholozi smatrali da bi trebalo usvojiti metodološki biheviorizam temelji se na prirodi dokazne građe koju koristimo pri pripisivanju mentalnih stanja različitim organizmima. Pripisivanje mentalnih stanja ljudima i životinjama temelji se na promatranju njihova ponašanja (Graham 2019, odjeljak 5). Budući da mentalna stanja nisu direktno dostupna promatračima, tj. nisu vidljiva ili provjerljiva, ponašanje nam daje dokaznu građu za vjerovanje da je prisutno određeno mentalno stanje. Isto tako, čini se da su uvjeti u kojima pripisujemo mentalna stanja,

poput vjerovanja i želja, neraskidivo povezani s određenim ponašajnim sklonostima ili dispozicijama. Primjerice, ako netko tvrdi da voli sladoled, a nikada ga ne pojede kada se nađe u prilici, onda se čini da ta osoba zapravo ne voli sladoled. Ovakva čvrsta veza između pripisivanja mentalnih stanja i ponašanja sugerira nam da je upravo ponašanje esencijalno ili nužno za postojanje mentalnih stanja.

Drugi razlog koji je bio relevantan za bihevioriste odnosio se na njihovo nezadovoljstvo u korištenju mentalnog vokabulara te pretpostavke da postoje neka stanja koja se ne mogu direktno promatrati znanstvenim metodama. Nezadovoljstvo korištenjem govora o unutrašnjim mentalnim stanjima proizlazilo je iz toga što je činilo ljude sklonijima da postuliraju postojanje bestjelesnih supstancija, slobodne volje koja narušava uzročni red u svijetu, homunkuluse (čovječuljke) koji upravljaju tijelima i tome slično. Na primjer, Gilbert Ryle (1949) kartezijansku je sliku uma opisivao metaforom duha u stroju. Prema kartezijanskoj slici uma stječe se dojam da je um nekakav čovječuljak koji živi unutar glava ljudi i upravlja njihovim tijelima. Prema bihevioristima, postuliranje ovakvih entiteta ne doprinosi objašnjenju psiholoških sposobnosti. Stoga ih treba izbjegavati kada nastojimo izgraditi ili formulirati znanstveno respektabilne teorije i hipoteze.

Nadalje, osim generalnih primjedbi protiv korištenja pojmovnog aparata kojim se koristimo pri referiranju na mentalna stanja kada radimo istraživanja u psihologiji, neki su bihevioristi ponudili i konkretnije argumente o tome zašto nam govor o mentalnim stanjima nije potreban kada nastojimo objasniti psihološke fenomene. Jedan od zanimljivijih argumenata u tom smislu ponudio je B. F. Skinner (1953). Smatrao je da nije korisno oslanjati se na mentalna stanja pri objašnjenju ponašanja jer su ona eksplanatorno irelevantna ili praktično nedostupna. Skinner navodi da uvijek postoje tri veze u uzročnim lancima (Skinner 1953, 34). Prva se odnosi na vanjske uvjete. Na primjer, organizam koji promatramo može neko vrijeme biti bez vode ili mu možemo u sklopu eksperimenta aktivno uskratiti pristup vodi. Druga veza u uzročnom lancu odnosi se na unutrašnja stanja. U ovom slučaju to može biti osjećaj žeđi. Treća se veza u uzročnom lancu odnosi na ponašanje. Na primjer, kada bi se naš organizam našao u prilici utažio bi žeđ uzimanjem neke tekućine. Prema Skinneru, ako ne znamo kako pomoću manipuliranja druge karike u lancu možemo kontrolirati ponašanje, onda je ona praktično irelevantna za psihologiju. Štoviše, za kontrolu ponašanja i druge karike u lancu, često je dovoljno poznavanje prvog uzroka u lancu koji se nalazi u okolišu izvan organizma ili djelatnika kojeg promatramo.

To nas dovodi do drugog razloga koji uključuje eksplanatornu irelevantnost. Ono što nastojimo objasniti otkrivanjem uzročnog lanca jest ponašanje. To često činimo pozivajući se na drugu kariku u lancu, tj. na neke unutrašnje procese. No, oni nisu uvijek dostupni za promatranje, tako da često imaju status postuliranih entiteta. S druge strane, prema pretpostavci,

unutrašnja stanja posljedice su prve karike u uzročnom lancu. Skinner daje primjer:

> […] kada nam kažu da je neka osoba ukrala hleb kruha zato što „je bila gladna“, i dalje moramo saznati kakvi su bili vanjski uvjeti koji su odgovorni za „glad“. Ti uvjeti bi bili dovoljni da objasne pljačku. (Skinner 1953, 35)

Budući da je prva karika u lancu dostupna za promatranje, a ujedno je dovoljna da preko druge karike proizvede ponašanje, onda je ona sasvim dovoljna da objasni ponašanje.

Unatoč određenoj uvjerljivosti ovih argumenata, razvoj znanstvene psihologije pokazao je da neke fenomene ne možemo uspješno objasniti bez pozivanja na unutrašnja mentalna stanja (Bermúdez 2014, pogl. 1). Jasna instanca ovog problema može se prikazati na primjeru učenja. Bihevioristi su smatrali da je u suštini svo učenje posljedica uvjetovanja te da se uvjetovanje temelji na procesima asocijacije i potkrepljenja. Uvjetovanje putem potkrepljenja generalno se može podijeliti na klasično i operantno. Iz poznatih Pavlovljevih eksperimenata poznato nam je da klasično uvjetovanje uključuje uparivanje neuvjetovanog podražaja i uvjetovanog podražaja koji stvaraju asocijaciju između uvjetovanog podražaja i neuvjetovane reakcije. Na primjer, u slučaju pasa, uparivanje hrane s određenim zvukom zvona dovodi do stvaranja neuvjetovane reakcije poput slinjenja te nakon nekog vremena samo izlaganje uvjetovanom podražaju poput zvuka dovodi do slinjenja kao uvjetovane reakcije. Operantno uvjetovanje uključuje kompleksnije učenje gdje se na temelju nagrada i kazni stvaraju asocijacije između ponašanja i posljedica tog ponašanja. Klasični eksperimenti koji ilustriraju operantno uvjetovanje uključuju tzv. Skinnerovu kutiju, u kojoj se nalazi polugica čije povlačenje pod određenim uvjetima donosi nagradu. Na primjer, miš koji se nalazi u takvoj kutiji mora naučiti da povlačenjem polugice dobiva hranu te time formira asocijacija između ponašanja i posljedica ponašanja.

Već tijekom 30-tih godina 20. stoljeća neki su istraživači primijetili da operantno uvjetovanje bez pretpostavke unutrašnjih reprezentacija ili predodžbi ne može objasniti sve obrasce učenja koje pokazuju štakori i druge životinje. Na primjer, u jednom istraživanju psiholozi su koristili tri grupe štakora koji se kreću po labirintu (Tolman i Honzik 1930). Prva bi grupa primila nagradu svaki put kad bi uspješno prošla na drugi kraj labirinta. Druga grupa ne bi nikad dobila nagradu. Treća grupa u početku nije bila nagrađivana za kretanje po labirintu. Nakon nekoliko dana bez nagrađivanja počeli bi davati nagradu štakorima u trećoj grupi kada uspješno prođu labirint. Otkrili su da treća grupa nakon što počne dobivati nagradu brže uči kretati se labirintom nego što je to bilo s prvom grupom koja je otpočetka bila nagrađivana. To pokazuje da postoji tzv. fenomen latentnog učenja koje

se ne može objasniti putem biheviorističke metodologije jer se čini da se štakori uče kretati po prostoru bez direktnog potkrepljenja. Štoviše, mnogi smatraju da ova vrsta eksperimenta pokazuje da štakori imaju unutrašnju predodžbu prostora koju pohranjuju u pamćenju te je prizivaju kasnije kada im to postane korisno (npr. za dobivanje nagrade) te da se uspješno objašnjenje latentnog učenja mora pozvati na takvu vrstu unutrašnjih mentalnih stanja (Bermúdez 2014).

Nadalje, razvoj kognitivnih znanosti te funkcionalizma u filozofiji omogućio je da se na znanstveno respektabilan način formuliraju te empirijski istražuju teorije i objašnjenja koja se pozivaju na unutrašnja mentalna stanja. O tome ćemo više govoriti u poglavlju 5. Stoga možemo reći da smo metodološki biheviorizam na neki način prerasli zbog njegovih eksplanatornih ograničenja, no također i zbog zbog pojmovnog i tehnološkog razvoja koji nam je omogućio spoznaju da govor o unutrašnjim mentalnim stanjima ne povlači nužno govor o misterioznim entitetima koji se ne mogu izučavati na objektivan način.

Nešto je drugačiji slučaj s filozofskim ili logičkim biheviorizmom. Kao što ćemo vidjeti u nastavku, zastupnici biheviorizma u filozofiji otišli su korak dalje od metodoloških biheviorista u psihologiji tvrdeći da govor o mentalnim stanjima nije ništa drugo nego govor o objektivno opažljivim fizičkim procesima ili javno dostupnim ponašanjima. S obzirom na ovo gledište uvjerljivost njihove pozicije ne ovisi o metodama psihološkog istraživanja koje su bile dostupne u prvoj polovici 20. stoljeća, već o filozofskim argumentima koji bi mogli podržati ideju da kada govorimo o mentalnim stanjima ne govorimo o nekim nedostupnim mentalnim entitetima, nego o stanjima, procesima i događajima koje možemo promatrati sredstvima koje koristimo kada promatramo druga javno opažljiva stanja, procese i događaje. U nastavku ćemo se dakle baviti logičkim ili filozofskim biheviorizmom.

3.4 Logički biheviorizam

Logički su bihevioristi tvrdili da govor o mentalnim stanjima nije ništa drugo nego govor o ponašanjima koje su ljudi u stanju izvoditi. Nije tako da se mentalna stanja nalaze u uzročnom lancu koji kreće od vanjskih uvjeta i završava izvođenjem određenih radnji. Mentalna stanja nisu uzroci ponašanja, već se sama na neki način svode na ponašanje. U literaturi se razlikuju dvije vrste filozofskog biheviorizma koji se nazivaju „tvrdi“ i „meki“. U nastavku ćemo krenuti obrazlaganjem tvrde verzije, koju su zastupali logički pozitivisti, tj. pripadnici poznatog Bečkog kruga (Berčić 2002).

3.5 „Tvrdi“ logički biheviorizam

Osnovna teza logičkog biheviorizma je da se iskazi o mentalnim stanjima mogu na neki način reducirati ili prevesti u iskaze koji se odnose na aktualno ili moguće ponašanje. Tvrdi biheviористi smatrali su da pri prevođenju psiholoških iskaza smijemo koristiti samo znanstvene opise ponašanja (Smith i Jones 1988, 144). Na primjer, ako želimo odrediti značenje iskaza kojim referiramo na vjerovanje da vani pada kiša, onda pri određivanju smijemo koristiti samo opise koji specificiraju, primjerice, položaj ruke te kojom se brzinom ona približava kišobranu, kako ga grabi, kojom brzinom se kreću noge i tome slično. Dakle, u definiranju pojmova kojima referiramo na mentalna stanja izbjegavaju se opisi radnji koji koriste ili podrazumijevaju nekakve intencionalne (mentalističke) pojmove te se smiju koristiti samo strogo znanstveni ili fizički opisi koji se odnose na javno dostupna stanja, procese ili događaje.

Pridjev „tvrdi“ odnosi se na ovo gledište jer su njegovi pobornici smatrali da se sve znanosti mogu reducirati, uključujući i psihologiju, na „tvrde“, tj. prirodne znanosti (engl. *hard sciences*) od kojih je najvažnija fizika (Maslin 2001, 111). Budući da u fizikalnim znanostima nemamo intencionalne opise[25] kojima zdravorazumski karakteriziramo psihološka stanja i radnje, poput toga da Ivica vjeruje da vani pada kiša ili da Ivica namjerava uzeti kišobran kada izlazi iz kuće, onda se očekuje njihova eliminacija iz znanstvene psihologije ili redukcija na pojmove kojima se opisuju fizički procesi.

Meki biheviористi su, nasuprot tome, smatrali da se pri definiranju pojmova kojima referiramo na mentalna stanja smijemo oslanjati na zdravorazumske opise radnji te nisu smatrali da se sve znanosti mogu reducirati na fiziku. Međutim, slično tvrdim biheviористima, smatrali su da se sav govor o mentalnim stanjima može interpretirati kao govor o ponašanjima i dispozicijama za ponašanje. Na primjer, meki biheviористi smatraju da se vjerovanje da vani pada kiša može jasnije odrediti pomoću zdravorazumskih opisa koji bi navodili da se u tim i tim okolnostima ljudi ponašaju na taj i taj način. Pri tome dopuštaju da se u obrazloženu govora o mentalnim stanjima možemo pozivati na termine koji pretpostavljaju intencionalne opise poput toga da je osoba koja ima vjerovanje da vani pada kiša osoba koja ima dispoziciju uzeti kišobran, skupljati robu koja se suši, ne zalijevati vrt u tim okolnostima i tome slično. O mekom biheviorizmu ćemo više govoriti kasnije; zasad ćemo se zadržati na osnovnim postavkama tvrdog biheviorizma.

Tvrdi logički biheviorizam bio je poglavito inspiriran gledištima koja se povezuju s logičkim pozitivizmom. Logički pozitivizam ili logički empirizam

[25] Ovdje se pojam intencionalnost treba shvatiti u specifično filozofskom smislu, poput onog kada govorimo o intencionalnim mentalnim stanjima (vidi Malatesti 2014).

filozofski je pravac koji su razvili članovi tzv. Bečkog kruga u prvoj polovici 20. stoljeća (Berčić 2002). Teze logičkog pozitivizma u kontekstu psihologije najjasnije ističu Rudolf Carnap (1995) i Karl Hempel (1980). U nastavku ćemo se pretežito osloniti na analizu koju daje Hempel.

Hempel (1980) razmatra problem znanstvenog statusa psihologije te se pita spada li psihologija u prirodne znanosti ili u ono što bismo danas nazvali humanističke i društvene znanosti? Logički pozitivisti su smatrali da će se u dobro uređenom sustavu znanosti sve znanosti u konačnici povezati i reducirati na fiziku te da će se svi smisleni iskazi iz posebnih znanosti moći u principu prevesti u iskaze kojima opisujemo fizičke procese i zakone prirode (vidi, npr. Carnap 1931). Carnap (npr. 1995, 44) to gledište naziva teza fizikalizma. Primjena fizikalizma na psihologiju stoga uključuje ideju da se, ako je psihologija znanost, onda svi iskazi kojima opisujemo psihološke procese mogu u principu prevesti u iskaze kojima opisujmo fizičke procese. Međutim, nasuprot ovakvom gledištu, početkom 20. stoljeća bilo je dosta izraženo gledište da se psihologija i prirodne znanosti razlikuju prema svojem sadržaju i metodologiji.

Osnovna ideja pobornika razlikovanja fizikalnih i takozvanih duhovnih znanosti (njem. *Geisteswisschenschaften*) jest da se prirodne znanosti oslanjanju na metode koje koriste opise i uzročna objašnjenja, dok se psihologija oslanja na mentalističke pojmove koji su različiti od fizikalnih pojmova (von Wright 1975). Smatrali su da je osnovna metoda u psihologiji empatijski uvid ili introspekcija. Pobornici introspektivne psihologije smatraju da se psihologija bavi strukturama koje su smislene ili imaju „značenje". U tom kontekstu metoda psihologije nastoji prodrijeti u značenje tih smislenih struktura. Hempel (1980) kao primjer daje čovjeka koji govori. Fizičko objašnjenje bi se oslanjalo na uzroke takvog ponašanja, koji se prema pretpostavci nalaze u fiziološkim procesima te osobe, tj. procesima u njezinom živčanom sustavu. Međutim, prema pobornicima introspektivne psihologije navođenje fizičkih uzroka ne daje odgovor na pitanja kojima se bavi psihologija. Ispravno korištenje psihološke metode primjenjuje se na razumijevanje *smisla* onog izrečenog te se taj smisao nastoji integrirati u širi sustav značenja. Osnovna ideja je da su procesi kojima se bave prirodne znanosti bez značenja ili smisla, dok su psihološki procesi smisleni i posjeduju značenje.

Logički pozitivisti su se suprotstavili ovakvom viđenju psihologije i podjeli između društvenih ili humanističkih i prirodnih znanosti (Carnap 1995; Hempel 1980). Za razliku od metodoloških argumenata na koje su se oslanjali bihevioristi u psihologiji, logički pozitivisti su ponudili pojmovne razloge zašto se psihologija mora shvatiti kao znanost o ponašanju. Prema logičkim pozitivistima, zadaća filozofije jest analizirati strukturu jezika znanosti. Smatrali su da ćemo, jednom kada uzmemo u obzir jezik psihologije, shvatiti da se psihologija kao znanost u principu ne razlikuje od prirodnih znanosti.

Štoviše, jednom kad shvatimo logičku strukturu psiholoških iskaza, uvidjet ćemo da se oni mogu svesti ili reducirati na iskaze o fizičkim procesima.

Kako bi opravdali takvo fizikalističko gledište, Hempel i Carnap se oslanjaju na princip verifikacije. Logički pozitivisti su smatrali da je sadržaj iskaza određen uvjetima u kojima možemo provjeriti istinitost, tj. verificirati taj iskaz. U tom smislu, princip verifikacije je trebao dati općeniti kriterij za određivanje kada je rečenica smislena (Berčić 2002). Prema njemu, rečenica je smislena ako i samo ako je analitička (logički istinita) ili se može empirijski provjeriti. Prema ovoj podjeli, matematika i logika se bave analitičkim iskazima. Nasuprot tome, kada govorimo o znanstvenim iskazima onda mislimo na one koji se mogu empirijski provjeriti. Hempel daje primjer sljedećeg iskaza: „danas u 13 sati, temperatura tog i tog mjesta u laboratoriju bila je 23.4 Celzijevih stupnjeva“. Prema principu verifikacije, mi razumijemo ovaj iskaz, onda i samo onda, kada znamo uvjete u kojima možemo provjeriti istinitost ili neistinitost tog iskaza.[26] Na primjer, razumijemo značenje prethodnog iskaza jer znamo da je on istinit kada termometar koji ima živu pokazuje na brojku „23.4“ ili kada elektronski termometar na ekranu pokazuje istu brojku (naravno ako koristi istu mjernu skalu). Prema principu verifikacije, iskaz nije ništa više nego skraćena formulacija svih tih testova pomoću kojih možemo provjeriti istinitost iskaza. Stoga, prema logičkim pozitivistima, princip verifikacije ukazuje na to da se iskaz koji specificira temperaturu neke točke u prostor-vremenu može prevesti, bez gubitka značenja, u iskaz koji ne sadrži riječ „temperatura“.

Logički pozitivist koristili su princip verifikacije kao metodološko sredstvo protiv onoga što su smatrali antiznanstvenim i metafizičkim gledištima (Carus 2009). Naime, prema principu verifikacije, iskaz za koji u principu ne postoji mogućnost empirijske verifikacije u potpunosti je lišen značenja. U tom slučaju govorimo o pseudoiskazima. Primjer pseudoiskaza koji je Carnap koristio, a navodno potječe od Heidegera, je „Ništa ništi“. S obzirom da nije jasno kako bismo mogli u principu provjeriti istinitost ove tvrdnje, Carnap (1959) je argumentirao da ona mora biti besmislena, unatoč tome što je rečenica koja je izražava gramatički ispravna.

Kako bi provjerili znanstveni status psihologije, logički pozitivisti primjenjuju princip verifikacije na iskaze u psihologiji. Kao primjer, Hempel uzima zubobolju. Razmotrimo iskaz „Pavla boli zub“. Možemo se pitati, koje su okolnosti u kojima možemo verificirati istinitost ovog iskaza? Prema

[26] Treba naglasiti da prema logičkim pozitivistima nije potrebno *stvarno* znati je li iskaz istinit ili neistinit, već je dovoljno da znamo u *principu* pod kojim uvjetima bismo mogli provjeriti je li istinit ili nije. Na primjer, trenutno ne znamo je li iskaz „Izvan naše galaksije postoji život“ istinit ili neistinit. Međutim, znamo kako bismo ga u principu mogli provjeriti. Recimo tako da putujemo od planeta do planeta izvan naše galaksije i provjeravamo ima li na njima živih bića.

Hempelu, sljedeći iskazi se mogu uzeti kao primjeri okolnosti u kojima se može provjeriti istinitost ovog iskaza:

a. Pavle plače i radi geste takve i takve vrste.
b. Na pitanje „Što ti je?“, Pavle odgovara „Boli me zub“.
c. Pobliža provjera pokazuje karijes na zubu.
d. Pavlov tlak, probavni procesi i brzina reakcije pokazuju takve i takve promjene.
e. Takvi i takvi procesi se događaju u Pavlovom središnjem živčanom sustavu. (Hempel 1980, 17)

Carnap daje kao primjer iskaz „Gospodin A je sada uzbuđen“. Prema njemu, taj se iskaz može prevesti bez gubitka značenja iskazom kojim se tvrdi da postoji određena fizička struktura koja ima dispoziciju da se ponaša na određeni način s obzirom na određene uvjete. Konkretnije, kada se kaže „Gospodin A je uzbuđen“, onda se taj iskaz može prevesti na sljedeći način:

a. Živčani sustav Gospodina A nalazi se u određenom stanju.
b. Gospodin A ubrzano diše i srčani puls mu je ubrzan.
c. Gospodina A karakterizira izražajno, uobičajeno intenzivno i faktički nezadovoljavajuće odgovaranje na pitanja.
d. Gospodin A naglo reagira na određene podražaje i tome slično. (Carnap 1995, 51)

Svrha ovih primjera je ukazati na to da uvjeti verifikacije uključuju javne i opažljive fizičke okolnosti koje mogu potvrditi ili osporiti psihološke iskaze. U tom pogledu, pozitivistička analiza pokazuje da psihologija nije u principu različita od fizike i drugih prirodnih znanosti. Nadalje, pojmovi poput zubobolje, boli, uzbuđenja i njima sličnih prema principu verifikacije se, poput pojma temperature, mogu zamijeniti iskazima koji verificiraju je li bol ili bilo koje drugo psihološko stanje prisutno. Uz to, pojmovi kojima govorimo o psihološkim stanjima mogu se smatrati skraćenicama kojima referiramo na te uvjete verifikacije. U tom smislu, logički pozitivisti prihvaćali su tvrdnju da se psihološki pojmovi mogu jednostavno prevesti u tvrdnje o ponašanjima i fizičkim procesima.

Nije na odmet istaknuti koliko je ovo snažna tvrdnja te koliko je bio ambiciozan projekt logičkih pozitivista. Logički biheviorizam uključuje semantičku tvrdnju da se psihološki pojmovi mogu prevesti, bez izostanka značenja, u tvrdnje o ponašanju. Dakle, ideja je da se iskaz „Pavla boli zub“ može iscrpno prevesti te svesti na tvrdnje iz primjera a. – e. Da bi taj projekt bio uspješan, prijevod ili redukcija psiholoških iskaza ne smije uključivati oslanjanje na tvrdnje koje koriste pojmove iz psihologije. Inače ne bismo imali uspješan prijevod koji se oslanja samo na javno dostupne fizičke opise

koji se mogu verificirati koristeći samo fizičke opise događaja i pojava. Dakle, u prijevodu iskaza „Pavla boli zub“ ne smijemo koristiti iskaze poput „Pavle želi otići kod zubara“ jer nam u tom slučaju prijevod ne bi bio potpun. Morali bismo dalje prevesti opis koji se oslanja na unutrašnje stanje poput želje oslanjanjem na neki drugi opis koji će se moći verificirati pozivanjem samo na javna ponašanja ili neka druga opažljiva fizička stanja. Općenitije, možemo reći da uspješan prijevod ne smije sadržavati opise koji referiraju na intencionalna (poput vjerovanja i želja) ili fenomenološka stanja (poput emocija) jer u tom slučaju redukcija na fizička stanja i opise ne bi bila potpuna.

Da sumiramo dosadašnju raspravu, način na koji logički pozitivisti argumentiraju mogao bi se formulirati na sljedeći način:

1) Značenje smislene rečenice iscrpljeno je uvjetima putem kojih možemo potvrditi ili verificirati sadržaj rečenice. (*princip verifikacije*).
2) Prema principu verifikacije, rečenica je smislena ako i samo ako izražava propoziciju koja je analitična ili empirijski provjerljiva.
3) Ukoliko je psihologija empirijska znanost, utoliko nastoji ponuditi empirijski provjerljive iskaze.
4) Empirijski provjerljivi iskazi su oni čiji su uvjeti verifikacije javno opažljivi.

Dakle:

5) Psihološki iskazi su smisleni onda kada su njihovi uvjeti verifikacije javno opažljivi.
6) Samo ponašanja i fizičke pojave su javno opažljive.

Dakle:

7) Značenje bilo koje smislene psihološke rečenice mora se moći specificirati koristeći iskaze čiji su uvjeti verifikacije javni, tj. iskazima koji opisuju ponašanja i fizičke pojave.

Čini se da konkluzija 7) logički slijedi iz prethodnih premisa. Stoga možemo zaključiti da je argument valjan. Međutim, ostaje pitanje jesu li mu premise istinite te može li se ovakav projekt redukcije kakvog su zamišljali logički pozitivisti u teoriji i praksi provesti. U nastavku ćemo se usredotočiti na neke prigovore koji upućuju na to da se ovakav projekt ne može ni u principu provesti.

3.6 Problemi s tvrdim biheviorizmom

Prvo što možemo primijetiti jest da prethodni argument pretpostavlja kontroverznu teoriju značenja. Kao što smo vidjeli, princip verifikacije čini važnu pretpostavku na kojoj se temelji argument. Međutim, u suvremenoj filozofiji princip verifikacije napušten je kao kriterij prema kojemu bi

prosuđivali je li neka rečenica smislena (Berčić 2002). Tradicionalni prigovor jest da nije jasno bi li sam princip prošao test verifikacije. S jedne strane, izgleda da se prema principu verifikacije ne može empirijski utvrditi je li neka rečenica smislena. Nema tog empirijskog istraživanja koje bi nam moglo odgovoriti je li istina da su sve rečenice smislene jedino ako se mogu empirijski provjeriti ili su analitičke istine. S druge strane, nije jasno da princip verifikacije predstavlja analitičku istinu koja bi bila slična iskazima iz logike u čiju istinitost ne možemo razumno sumnjati. Naime, intuitivno se čini da rečenica može biti smislena čak i ako nije empirijski provjerljiva ili analitička istina. Mnogi filozofi smatraju da upravo iskazi kojima govorimo o privatnim mentalnim stanjima predstavljaju takve iskaze. Mnogi smatraju da je trenutni osjećaj boli svojstvo našeg iskustva koje niti je javno dostupno, niti predstavlja analitičku istinu o našem iskustvu. Stoga mnogi autori odbacuju princip verifikacije kao kriterij smislenosti i jednostavno smatraju da značenje rečenice ovisi o njezinim istinosnim uvjetima, bez obzira na to jesmo li ih u principu u mogućnosti utvrditi (za raspravu, vidi Hempel 1976).

Međutim, čak i ostavimo po strani ovaj prigovor, nije jasno da se redukcionistički program kako su ga zamišljali logički pozitivisti može provesti. Štoviše, važan prigovor tvrdom biheviorizmu je to što si je dao pretežak zadatak. Kako bi uspio morali bismo pokazati da se svi psihološki iskazi i pojmovi moraju moći bez ostatka prevesti u iskaze i pojmove koji ne koriste psihološke termine. Međutim, nije jasno da je to moguće. Izuzetno je teško pronaći opis ponašanja koji neće pretpostavljati intencionalne opise, tj. atribuciju i pretpostavku mentalnih stanja. Uzmimo primjer s Pavlom. Kada kažemo da Pavle plače, podrazumijeva se da on pati od neugodnog iskustva. Što znači da opet pretpostavljamo nekakav mentalni fenomen čiji opis moramo prevesti u iskaze o fizičkim entitetima. Nadalje, problem je dati fizičke opise koji će predstavljati relevantnu analizu mentalnih termina, tj. koja će dati sinonimne termine. Naime, u primjeru s Pavlom, činjenica da netko ima karijes ne izgleda esencijalno važna za određivanje značenja iskaza „Boli me zub“. Nekoga može boljeti zub čak i ako nema karijes. Štoviše, netko može imati tzv. fantomsku bol koja je nepovezana s trenutnim uništenjem tkiva. Stoga nije jasno da tu rečenicu možemo koristiti kao dio prijevoda značenja rečenice „Pavle ima bol u zubu“.

Nadalje, protiv mogućnosti biheviorističke reduktivne analize psiholoških iskaza može se uputiti generalni argument. Kako bismo shvatili taj argument razmotrimo jedan zdravorazumski princip za koji će se mnogi složiti da ga podrazumijevamo kada objašnjavamo ljudsko ponašanje. Kada koristimo zdravorazumska objašnjenja i predviđanja ponašanja onda se obično pozivamo na dvije vrste mentalnih stanja. Na primjer, općenito možemo reći da je, ako neka osoba želi ostvariti cilj C i zna da će ga ostvariti ako poduzme radnju R, razumno očekivati da će osoba poduzeti radnju R. Slično vrijedi za objašnjenje radnje koja je već poduzeta. Ako se pitamo zašto je osoba

poduzela radnju R, često će objašnjenje biti da je osoba imala neku želju za koju je smatrala da će biti ostvarena ako učini R. Ova vrsta zdravorazumskog objašnjenja djelovanja često se naziva teorija uma ili psihologija vjerovanja i želja.[27] Njezino načelo se može izraziti na sljedeći način:

> Ako osoba A želi da p bude slučaj i vjeruje da je poduzimanje radnje R najbolji način da ostvari p onda će A, *ceteris paribus*, formirati namjeru da poduzme radnju R. (vidi, npr. P. M. Churchland 1993, 43)

Klauzula *ceteris paribus* govori nam da ovo načelo nema status nužnog prirodnog zakona. Njime se naglašava da će osoba namjeravati ili poduzeti određenu radnju samo u slučaju da su sve ostale okolnosti iste. Drugim riječima, poduzet će radnju samo u slučaju da osoba nema neku jaču želju, nije iracionalna, ima ostala prikladna vjerovanja, nema fizičkih prepreka i tome slično (vidi, npr. Jurjako 2020). Ovakva su ograničenja na zdravorazumska psihološka objašnjenja važna jer predstavljaju problem za mogućnost bihevioristične redukcije mentalnih stanja na ponašanja i neintencionalne opise.[28] Razmotrimo sljedeći primjer:

> Ako Ivan želi otići na konferenciju iz filozofije uma u Pariz i vjeruje da je najbrži i najisplativiji način da ode u Pariz let avionom s Krka, onda će Ivan rezervirati avionsku kartu za let s Krka.

Ako pretpostavimo da Ivan ima navedenu želju i vjerovanje, slijedi li iz toga da će Ivan stvarno rezervirati avionsku kartu za Pariz? Odgovor je: ne nužno. Kada bi Ivan to učinio onda bismo imali zdravorazumsko objašnjenje za njegovu radnju. Pozvali bismo se na njegovu želju i vjerovanje. Međutim, sasvim je moguće da, unatoč želji da ode na konferenciju u Pariz i vjerovanju da je najisplativiji način da leti avionom, Ivan ne kupi ili rezervira avionsku kartu za Pariz. Na primjer, moguće je da se Ivan brine za okoliš te da smatra da je avionski promet jedan od najznačajnijih uzročnika zagađenja (Teymoori i ostali 2020). S obzirom na to da mu je jako stalo do zaštite okoliša, spoznaja da avionski promet onečišćuje naš planet motivira Ivana da traži neko drugo prijevozno sredstvo.

Dakle, da bismo mogli zaključiti da će Ivan rezervirati avionske karte moramo pretpostaviti da su ostale okolnosti iste, u smislu da nema jaču želju

---

[27] U području etike, ovakvo zdravorazumsko objašnjenje se još naziva hjumovska teorija motivacije, u čast Davidu Humeu za koga se smatra da je dao jedno od prvih jasnih formulacija ovog gledišta (vidi Sušnik 2012).

[28] Ovu vrstu argumenta protiv mogućnosti bihevioristične redukcije mentalnih stanja daju Donald Davidson (1970) i Roderick Chisholm (1957). Naše izlaganje argumenta slijedi formulaciju koju daje Kim (1996, 66–67).

ne zagađivati okoliš ili da nema vjerovanje da je zračni promet jedan od najvećih uzročnika zagađenja. No, čak i ako pretpostavimo da Ivan ima ova vjerovanja i želju ne zagađivati Zemlju koja nadjačava želju da stigne u Pariz na najisplativiji način, svejedno ne slijedi da on neće rezervirati avionsku kartu i putovati avionom. Naime, moguće je da Ivan zna da će, ako ne putuje avionom u Pariz, propustiti vjenčanje svoje kćeri koja se udaje u Parizu. U tom bi slučaju bio sklon putovati avionom. Tako možemo u nedogled smišljati okolnosti koje bi mogle osporiti ili podržati vezu između mentalnih stanja i ponašanja.

Činjenica da odnosi između mentalnih stanja i ponašanja uključuju *ceteris paribus* ograničenja može se zahvatiti sljedećim načelom:

> Ako postoji odnos implikacije između mentalnih stanja M_1, … M_n, i ponašanja P, onda uvijek postoji neko daljnje mentalno stanje M_{n+1} koje je takvo da M_1, … M_n, M_{n+1} zajedno impliciraju ne-P. (Kim 1996, 67)

Postojanje ovakvog odnosa između mentalnih stanja i ponašanja ukazuje nam na sljedeće. Veza između ponašanja i mentalnih stanja je kompleksna te nije razumno očekivati da će se govor o mentalnim stanjima moći u potpunosti reducirati na govor o ponašanjima koja na ovaj ili onaj način neće podrazumijevati psihološke opise. Naime, za svaki skup mentalnih stanja koji bi u principu mogao implicirati neko ponašanje uvijek je moguće da postoji neko drugo mentalno stanje (kada *cetera* nije *paria*, tj. kada ostale okolnosti nisu iste) zbog kojeg se ponašanje ne bi manifestiralo. Kako bi redukcija bila uspješna, onda bi se i to dodatno mentalno stanje moralo moći reducirati na fizički opis ponašanja. Budući da je u principu moguće da postoji beskonačno mnogo mentalnih stanja koja bi mogla poništiti relaciju implikacije između izvornih mentalnih stanja i ponašanja, nije vjerojatno da ćemo ikada uspjeti u potpunosti reducirati ili prevesti opise mentalnih stanja na fizičke opise ponašanja.

Kao što ćemo vidjeti u nastavku, „meki“ logički biheviorist zaobilaze prethodno navedeni problem jer ne prihvaćaju redukcionističku tezu prema kojoj se valjana objašnjenja radnji i mentalnih stanja moraju prevesti u iskaze koji ne podrazumijevaju psihološke ili intencionalne opise radnji.

3.7 „Meki“ biheviorizam

Obično se smatra da je Gilbert Ryle (1900. – 1976.) jedan od najznačajnijih zastupnika mekog biheviorizma (Maslin 2001; Kim 2006). Ova vrsta biheviorizma nije utemeljena na verifikacijskoj teoriji značenja. Također, Ryle (1949) nije smatrao da se um može reducirati na fizičke stvari i svojstva niti da se psihološki vokabular može prevesti u vokabular kojim govorimo o fizičkim procesima i javno opažljivim ponašanjima. U tom smislu, Ryleovo

stajalište u filozofiji uma se ne mora nužno smatrati materijalističkim. Štoviše njegovo gledište toliko se razlikuje od toga kako su logički pozitivisti shvaćali prirodu uma da neki autori smatraju da Rylea ne bi trebalo nazivati logičkim biheviorističkim (Tanney 2009, odjeljak III).[29] Međutim, kako što ćemo vidjeti u nastavku, Ryle u svojoj kritici kartezijanskog dualizma i poimanja uma kako ga se shvaća u toj tradiciji ponekad ide toliko daleko da na neki način negira postojanje unutarnjih mentalnih stanja kojima bi pristup imao samo subjekt koji ih posjeduje. Nasuprot tome, Ryle je smatrao da, naročito kada govorimo o ljudskim bićima i sposobnostima za inteligentno djelovanje, ne govorimo o nekim skrivenim unutarnjim mentalnim stanjima, već govorimo o javno dostupnim ponašanjima i dispozicijama da se ponašamo na određene načine. Kako bismo bolje shvatili način na koji je Ryle došao do svojih gledišta i kako je argumentirao protiv kartezijanskog dualizma prvo ćemo se osvrnuti na to kako je on shvaćao filozofsku metodologiju.

3.8 Filozofija kao logička „geografija“

Ryle pripada tradiciji analitičke filozofije gdje se uloga filozofije očituje u rješavanju problema na temelju logičke analize pojmova. Važno je spomenuti da su na razvoj analitičke filozofije veliki utjecaj imali i logički pozitivisti. Međutim, za razliku od logičkih pozitivista koji su svoju metodologiju gradili na principu verifikacije, Ryle je smatrao da se filozofska analiza ili, bolje reći, filozofsko istraživanje pojmova temelji na ispravnom razumijevanju logičkih veza između pojmova kako ih razumijemo u njihovim svakodnevnim kontekstima upotrebe.[30]

Ryle razlikuje dvije razine na kojima možemo razumjeti pojmove i riječi kojima ih izražavamo. Prva razina uključuje znanje pojmova kroz korištenje riječi kojima ih izražavamo. Na primjer, svi kompetentni govornici hrvatskog jezika znaju što znači da nekoga *boli* nešto ili da *vjeruje* da je nešto slučaj. Ako kažemo da nekoga boli ruka to između ostalog znači da se ta osoba hvata za mjesto gdje osjeća bol, možda će reći „tu me boli“, imat će određeni izraz lica, možda neće moći podizati stvari tom rukom i tome slično. Ako kažemo

[29] Na primjer, M. R. Bennett i P. M. S. Hacker (2003, 413, fusnota 1) navode da Ryle nije bio logički biheviorist. Pod logičkim biheviorizmom čini se da podrazumijevaju gledište poput onog kojeg su zastupali logički pozitivisti (vidi Bennett i Hacker 2003, 416–17). Međutim, ako se dopusti da biheviorizam karakterizira i blaža gledišta prema kojima govor o mentalnim stanjima u normalnim slučajevima konstitutivno pretpostavlja njihovu manifestaciju u ponašanju, a ne nužno i mogućnost iscrpnog prijevoda psihološkog vokabulara u govor o fizičkim događajima i procesima, onda, kako ćemo vidjeti u glavnom dijelu teksta, ima smisla Rylea smjestiti u kategoriju logičkih bihevioristа.

[30] Stoga se Rylea obično svrstava u zastupnike filozofije običnog jezika (engl. *ordinary language philosophy*) u koju se obično ubrajaju autori poput Ludwiga Wittgensteina (kasniji radovi), Petera F. Strawsona i J. A. Austina (npr. vidi Milkov 2003). Ryle daje jasan prikaz svoje filozofske metodologije u radu „Abstractions“ iz (2009).

da netko vjeruje da vani pada kiša, onda to znači da će ta osoba vjerojatno to potvrditi kada se postavi pitanje pada li kiša vani, da će uzeti kišobran kako se ne bi smočila, da će biti u stanju zaključiti da su vjerojatno ulice mokre i tome slično. Dakle, prva razina uključuje razumijevanje pojmova koje se izražava s obzirom na to kako ih koristimo u govoru ili kako razmišljamo *pomoću* njih (Ryle 2009).

Druga razina razumijevanja pojmova odnosi se na reflektiranje o značenju pojma ili riječi gdje sam pojam postaje predmet istraživanja. U tom smislu više ne razmišljamo pomoću određenih pojmova nego počinjemo razmišljati *o* njima (npr. vidi Ryle 2009, 457). Prema Ryleu, ispravno razmišljanje o pojmovima i njihovom značenju ovisi o poznavanju značenja riječi kojima ih izražavamo u jezičnim praksama. Međutim, ispravno vladanje riječima i pojmovima na prvoj razini ne uključuje nužno razumijevanje pojmova na drugoj razini. Na primjer, mi možemo znati što znači psihološki termin *vjerovati* da je nešto slučaj ili *željeti* da nešto bude slučaj. No, to ne povlači da ćemo ujedno imati spreman odgovor na pitanje „što je vjerovanje?“ ili „što je želja?“ kada ih se shvati kao pitanja o značenju pojmova apstrahiranih od njihove konkretne upotrebe. Kako bismo odgovorili na ova apstraktnija pitanja, moramo imati određene vještine i sposobnosti pojmovne analize ili istraživanja koje tradicionalno povezujemo uz ono čime se bave filozofi.

Kako bi pojasnio što pod time misli, Ryle daje analogiju s poznavanjem geografije određenog područja. Zamislimo osobu, recimo farmera, koji je cijeli život proveo u određenom selu gdje živi i radi na zemlji te poznaje sve kuće, puteve i polja koji pripadaju tom selu. Ako netko treba pomoć u pronalasku nekog mjesta ili kuće u selu ta osoba mu može odmah pomoći. Zamislimo sada da se od našeg farmera traži da napravi kartu svog sela koja bi se mogla dobro povezati s kartama obližnjih sela te u konačnici povezati s kartama drugih mjesta i cijele zemlje. Samo poznavanje konfiguracije terena i seoske infrastrukture neće biti dovoljno za taj zadatak. Kako bi uspio obaviti kartografski zadatak, farmer mora steći novi skup vještina te usvojiti novi način razmišljanja. Između ostalog, mora usvojiti vokabular kartografa te opisivati stvari, mjesta i ceste u terminima koji će biti kompatibilni s načinima opisivanja koji se koriste u kartama drugih područja i regija. Drugim riječima, farmer će morati „*prevesti* i stoga preispitati svoje lokalno topografsko znanje na univerzalne kartografske termine“ (Ryle 2009, 454, kurziv u originalu).

Analogija s kartografijom ukazuje nam na nekoliko važnih stvari. Razlika između razumijevanja pojmova na prvoj i drugoj razini je poput razlike osobnog ili praktičnog poznavanja nekog geografskog područja i njegovog poznavanja kroz izradu ili čitanje karte za koje je potrebno poznavanje kartografskih termina. Nadalje, osobno znanje koje posjeduje farmer odnosi se na isto geografsko područje o kojem karta prenosi informacije. Međutim, poznavanje kartografije zahtijeva posjedovanje tehničkog znanja i

terminologije koja nam omogućuje obavljanje sofisticiranijih zadataka. Na primjer, osim što ćemo moći voditi ljude po selu, poznavanje kartografije nam omogućuje da na vrlo precizan način određujemo udaljenosti i smjerove kretanja. Slično tome, na prvoj razini razumijevanja pojmova koristimo iskaze koji se većinom odnose na svakodnevne stvari koje su nam potrebne za razmjenu informacija i davanje praktičnih savjeta. Razumijevanje u tom običnom smislu uključuje barem implicitno poznavanje pojmova, njihovih uloga u rečenicama i iskazima te logičkim vezama u kojima stoje s drugim pojmovima. Ryle to naziva implikacijske niti (engl. *implication threads*). U novije doba rekli bismo da se razumijevanje pojmova sastoji u osjetljivosti na inferencijalne veze koje određuju značenje pojmova. Znanje tih implikacijskih niti ili inferencijalnih veza omogućuje nam da razumijemo koji su iskazi kompatibilni a koji nekompatibilni, određuju dokaznu građu koja bi mogla biti relevantna za njihovu provjeru, kojem logičkom tipu iskazi pripadaju, koja je njihova domena primjene i tome slično. Međutim, slično kao i u primjeru s intuitivnim znanjem koje posjeduje farmer, implicitno razumijevanje pojmova nam ne daje samo po sebi i sposobnost njihovog ekspliciranja kada apstrahiramo od toga na što se odnose iskazi koje svakodnevno koristimo te počnemo razmišljati o značenjima pojmova koje koristimo u tim iskazima. Za taj posao nam je potrebna filozofsko/logička analiza pojmova. U tom smislu, Ryle ulogu filozofa u odnosu na običnog kompetentnog govornika nekog jezika vidi prema analogiji s farmerom i kartografom.

Pojmovna analiza koju koristi filozof uključuje razmatranje inferencijalnih veza između pojmova i sinoptičko nadgledanje načina na koji se implikacijske niti različitih pojmova preklapaju ili razlikuju. Prema Ryleu, filozofski problemi su pojmovni problemi koji nastaju zbog toga što u normalnim jezičnim praksama ponekad koristimo izraze i fraze čije implikacijske niti povlače u različitim smjerovima ili na neki način negativno interferiraju jedne s drugima. Kako bismo razriješili te probleme moramo eksplicitno reflektirati o pojmovima i njihovim inferencijalnim vezama koje inače prešutno pretpostavljamo te na sistematičan način odrediti njihove granice i međupovezanost koja će razriješiti pojmovni problem s kojim se suočavamo (Ryle 2009, 457).

Ovakvo shvaćanje filozofske analize važno je za naš kontekst upravo zato što Ryle smatra da problem uma i tijela nastaje zbog isprepletenosti različitih vokabulara kojima govorimo o mentalnim i fizičkim stvarima. Stoga njegovo razrješenje zahtijeva razmatranje implikacijskih niti ili inferencijalnih veza u kojima stoje termini i pojmovi kojima referiramo na mentalne i fizičke fenomene. U tom kontekstu, Ryle smatra da ćemo, kada se osvrnemo na logička svojstva psihološkog vokabulara, uvidjeti da se kartezijanski dualizam i problemi koje se vežu uz njega temelje na pogrešnom shvaćanju uloge koju psihološki vokabular ima u objašnjenju radnji i mentalnih stanja.

3.9 Duh u stroju: protiv kartezijanskog dualizma

Ryle je kartezijanski dualizam opisao kao teoriju koja se zasniva na „dogmi o duhu u stroju“ (Ryle 1949, 5). Kao što smo vidjeli u poglavlju 2, kartezijanski dualizam uključuje tvrdnju da postoje umna i tjelesna supstancija. Tjelesna supstancija nalazi se u vremenu i prostoru te je podložna fizičkim zakonima. Ona se može promatrati iz perspektive trećeg lica. Mentalna supstancija ne nalazi se u prostoru te nije podložna fizičkim zakonima. Njezina stanja su privatna te se ne mogu promatrati izvana. Ona su direktno dostupna samo iz perspektive subjekta koji posjeduje ta stanja. Nadalje, između uma i tijela postoji interakcija. Um svojom voljom može pokretati tijelo i uzročno djelovati na svijet te svijet može povratno djelovati na um.

Vidjeli smo da se kartezijansko gledište suočava s problemom uzročnog odnosa uma i tijela. Osim tog problema, prihvaćanje da postoje dvije različite supstancije dovodi do problema drugih umova. Ovaj problem pojavljuje se kada se pitamo kako mi uopće znamo da druge osobe imaju umove, osjećaje, želje, vjerovanja i druga mentalna stanja poput nas. Zdravorazumski se čini da ih oni imaju. Međutim, ako je kartezijanska slika uma ispravna, onda samo subjekti mentalnih stanja imaju direktan uvid u svoja mentalna stanja. Mi ostali možemo samo nagađati imaju li drugi ljudi stvarno svjesna mentalna stanja poput nas. Stoga je prema takvoj slici uma moguće da su druge osobe samo automati koji prema van izuzetno dobro simuliraju ljudsko ponašanje, misli i osjećaje, a da zapravo unutar sebe nemaju svjesna mentalna stanja. U filozofskoj literaturi takva bića se nazivaju zombijima. Oni na van djeluju kao normalni ljudi, no unutar njih nema svijesti. Kolokvijalno možemo reći da su u njima svjetla ugašena. Dakle, filozofski zombiji su fizički identični ljudima, ponašaju se na isti način kao i normalni ljudi, no nedostaju im unutrašnja svjesna mentalna stanja.[31] Ovaj problem javlja se za kartezijanski dualizam zato što su prema njemu mentalna stanja privatna. Njima imaju pristup samo osobe koje ih doživljavaju iz perspektive prvog lica. Dakle, mi možemo znati za sebe da smo svjesni jer imamo direktan uvid u svoja mentalna stanja. No, za druge osobe to ne možemo znati budući da nemamo direktan uvid u tuđe umove.

Ryle je smatrao da se svi problemi do kojih dovodi kartezijanski dualizam mogu jednostavno razriješiti kada uvidimo da se temelje na *kategorijalnoj pogrešci*. Prema njemu, dualisti rade pogrešku jer stavljaju termine koji se odnose na mentalna stanja u istu logičku kategoriju kojoj pripadaju fizička stanja.[32] Do tog zaključka bismo trebali doći razmatranjem implikacijskih niti

[31] Za više o zombijima i njihovoj ulozi u filozofskoj argumentaciji, vidi poglavlje 8.

[32] Ryle je smatrao da i materijalisti koji smatraju da je um identičan stanjima mozga rade istu kategorijalnu pogrešku. Materijalističkim teorijama identiteta bavit ćemo se u poglavlju 4.

ili inferencijalnih veza koje određuju načine na koji je legitimno koristiti pojmove (Ryle 1949, lxi).

Kako bismo jasnije shvatili ideju kategorijalne pogreške, Ryle (1949, 6–7) je ilustrira koristeći nekoliko primjera iz svakodnevnog života. Zamislite da osoba nakon što razgleda na kampusu različite fakultete, knjižnice, znanstvene odsjeke i administrativne urede, želi vidjeti gdje se nalazi sveučilište i gdje rade njegovi zaposlenici. Takvo pitanje se temelji na pogrešci i nerazumijevanju pojma „sveučilište“. Sveučilište se sastoji od svih navedenih institucija te njihove međusobne povezanosti. Sveučilište ne postoji kao još jedna institucija uz knjižnice, fakultete itd. Osoba koja stavlja sveučilište u taj kontekst čini kategorijalnu pogrešku. Prema Ryleu, sličnu pogrešku također radi osoba koja se pita tko igra ulogu timskog duha u nekom timskom sportu. Recimo da neka osoba često čuje da nogometna ekipa koja posjeduje timski duh obično dobro prolazi na natjecanjima. Pretpostavimo da nakon što sazna koja je uloga i pozicija napadača, braniča, golmana te kako su ostali igrači raspoređeni, postavi pitanje tko igra ulogu timskog duha. Za takvu osobu bismo rekli da ne razumije da timski duh nije dodatna uloga koju netko mora posebno igrati, već požrtvovnost s kojom netko obavlja svoje postojeće zadatke. Kategorijalnu pogrešku bi napravila i osoba koja se pita gdje živi prosječna osoba ili koliko ima djece. Doslovno govoreći, prosječna osoba nije vrsta stvari koja može imati kuću ili djecu, ona samo predstavlja način na koji govorimo o statističkim generalizacijama koje samo u nekoj mjeri karakteriziraju konkretne osobe.

Prema Ryleu, slično vrijedi za termine „um“ i „tijelo“. Ryle smatra da je problem odnosa uma i tijela posljedica kategorijalne pogreške koju mnogi filozofi čine. Dok se god razmišlja o umu kao nekakvom nefizičkom duplikatu tijela, dolazit ćemo u napast da razmišljamo o tome koja su svojstva i moći uma te u kakvoj relaciji oni stoje prema tijelu i njegovim svojstvima. Ryle navodi da osoba koja razmišlja na taj način radi sličnu pogrešku kao i osoba koja razmišlja o tome kako bi to bilo sresti prosječnu osobu na ulici. Razmišljanje o prosječnoj osobi kao stvarno postojećoj dovodi nas u probleme jer reificiramo ili postvarujemo nešto što ne postoji. Slično vrijedi i za mentalna stanja. Prihvaćanje kartezijanske slike uma navodi nas da o umu razmišljamo kao o nekakvom repozitoriju sposobnosti i mentalnih uzroka koji mogu i ne moraju biti vidljivi u javno dostupnom ponašanju. Štoviše, navodi nas da o umovima razmišljamo kao o *stvarima* koje, iako imaju uzročne moći posredno djelovati na vanjski svijet, zbog svoje nefizičke pirode ipak od njega zauvijek ostaju skrivene. Međutim, prema Ryleu govor o mentalnim stanjima i procesima ne pripada istom logičkom tipu kao govor o fizičkim stanjima. Stoga, nema smisla razmišljati o njima kao da posjeduju slična svojstva i uzročne moći poput fizičkih predmeta.

Nasuprot tome, Ryle smatra da logika pojmova kojima se služimo kada govorimo o mentalnim stanjima i procesima ukazuje na to da oni referiraju

na neskrivene, tj. javne radnje i dispozicije za manifestaciju tih radnji u određenim okolnostima. Na primjer, Ryle navodi da „[j]avne inteligentne izvedbe nisu indicije toga kako umovi funkcioniraju; oni jesu ta funkcioniranja“ (Ryle 1949, 57). Stoga, na um ne smijemo gledati kao na nefizički repozitorij skrivenih mentalnih stanja koja *uzrokuju* javno dostupna ponašanja nego kao skup sposobnosti koje, kada se manifestiraju u ponašanju, upravo manifestiraju sam rad uma. U tom smislu, um nije nešto skriveno i iza ponašanja; on *jest* ono što se manifestira u ponašanju. Ukoliko se rad uma poistovjećuje s njegovom manifestacijom u ponašanju, utoliko ima smisla govoriti o Ryleu kao zastupniku jedne varijante biheviorizma.[33]

Ryleov biheviorizam pruža zanimljivo rješenje problema odnosa uma i tijela. Ako prihvatimo da um ne pripada istoj logičkoj kategoriji kojoj pripadaju fizički predmeti, onda shvaćamo da ni ne postoji problem odnosa uma i tijela, tj. ne postoji pravi problem na koji način mentalna supstancija može biti uzročno povezana s fizičkom supstancijom. Prema Ryleu, govor o mentalnoj supstanciji predstavlja samo reifikaciju koja se temelji na pogrešnoj upotrebi pojmova. Umjesto toga, um se treba shvatiti kao skup sposobnosti da se djeluje na određeni način u određenim okolnostima. Štoviše, prema Ryleu, jednom kada shvatimo da smo napravili kategorijalnu pogrešku koristeći pojam uma kao da označava duplikat tijela u nefizičkoj formi, onda ćemo uvidjeti da je problem odnosa uma i tijela zapravo pseudoproblem. Drugim riječima, ispravna upotreba pojmova uma i tijela ukazuje nam na to da zapravo niti nema problema odnosa uma i tijela. Budući da um jednostavno nije supstancija, nego je skup sposobnosti, onda niti ne može biti u uzročnim relacijama s fizičkim i nefizičkim predmetima.

Dakle, Ryle smatra da govor o mentalnim stanjima često zapravo uključuje govor o sposobnostima koje se manifestiraju u ponašanju. U tom smislu, prema Ryleu, kada govorimo o umu i psihološkim stanjima, onda zapravo možemo govoriti o dispozicijama za ponašanje. Ako nekome pripisujemo neko mentalno stanje, onda time tvrdimo da je ta osoba *sklona* pod određenim uvjetima poduzeti određene radnje. Prema ovom gledištu, reći da nekoga nešto boli ili da vjeruje da je nešto slučaj znači da ta osoba ima dispoziciju ponašati se na taj i taj način. Stoga, prema Ryleu, kao i psiholozi, romanopisci, dramaturzi, biografi, povjesničari i antropolozi imaju jednako pravo tvrditi da se bave ljudskim psihološkim stanjima jer svako od njih na svoj način „prikazuje ljudske motive, misli, perturbacije i navike

[33] Sam Ryle prihvaća da se njegovo gledište na prirodu uma može okarakterizirati kao bihevioristično (vidi Ryle 1949, 308). Međutim, istodobno se želi distancirati od bihevioriste u psihologiji jer smatra da su oni gradili svoje teorije na pogrešnim temeljima. Na primjer, Ryle smatra da, ukoliko metodološki bihevioristi odbijaju koristiti introspektivne izvještaje zbog njihove subjektivnosti ili nedostupnosti objektivnom istraživanju, utoliko im se metodologija temelji na kartezijanskoj slici uma za koju smo vidjeli da je treba odbaciti (Ryle 1949, 309).

opisujući njihove radnje, govore i zamišljanja, njihove grimase, geste i tonove glasa“ (Ryle 1949, 309) te nam time različite discipline humanističkih i društvenih znanosti u svojoj oblasti istraživanja na svoj način otkrivaju nešto o ljudskoj psihi.

U Ryleovoj verziji logičkog biheviorizma pojam dispozicije igra važnu ulogu. Pojam dispozicije može se objasniti na primjeru lomljivosti. Reći da je čaša lomljiva ne znači da je čaša stvarno razbijena. To znači da kad bismo primijenili veću silu na čašu onda bi se razbila. Slično vrijedi za pojam topljivosti. Kad kažemo da je šećer topljiv u vodi, onda želimo reći da kad bismo ga stavili u vodu on bi se otopio. Međutim, šećer bi imao tu dispoziciju čak i da ga nikad ne stavimo u vodu. Slično vrijedi i za psihološke i karakterne osobine. Ryle daje primjer:

> To što sam redovni pušač ne povlači da u ovom ili onom trenutku pušim; to je moja trajna sklonost pušenju kada ne jedem, ne spavam, ne držim predavanja ili ne prisustvujem pogrebima, a nedavno nisam pušio. (Ryle 1949, 31)

Prema Ryleu, činjenica da neka stvar ili osoba ima neku dispoziciju znači prihvatiti kao istinitu kondicionalnu rečenicu koja ima sljedeću formu:

> Predmet P ima dispoziciju za M u uvjetima C ako i samo ako bi manifestirao M kada bi bio slučaj da C.

Na primjer, kocka šećera ima dispoziciju da se rastopi u vodi ako i samo ako, bi se rastopio kada bismo ga stavili u vodu. Ili osoba O želi pojesti sladoled, ako i samo ako, bi ga O pojeo kada bi se našao u prilici. Drugi dio rečenice koji slijedi nakon „ako i samo ako“ se naziva kontrafaktičkim ili protučinjeničnim kondicionalom. Njime se tvrdi da kada bi bili zadovoljeni uvjeti navedeni u antecedensu, koji trenutno nisu ostvareni, onda bi slijedilo ono što se tvrdi u konsekvensu takvog kondicionala. U našem primjeru, kada bi se osoba našla u slastičarnici (antecedens kondicionala), onda bi slijedilo da jede sladoled (konsekvens kondicionala).

Ono što je važno istaknuti u kontekstu Ryleova biheviorizma jest njegova tvrdnja da posjedovanje dispozicijskih svojstava nije utemeljeno na unutrašnjim stanjima osobe ili organizma. Stoga se smatra da Ryle prihvaća *fenomenalističko* ili *operacionalističko* shvaćanje dispozicija. Primjerice on navodi da:

> Posjedovati dispozicijsko svojstvo ne znači biti u određenom stanju, ili doživjeti određenu promjenu; znači biti vezan ili podložan biti u određenom stanju, ili doživjeti određenu promjenu, kada se određeni uvjeti realiziraju. (Ryle 1949, 31)

Dakle, prema Ryleu, pripisati nekome dispozicijsko svojstvo ne uključuje pretpostavku postojanja nekog nedispozicijskog stanja koje utemeljuje tu dispoziciju i koje je odgovorno za manifestaciju dispozicije u prikladnim okolnostima. Sve što je potrebno da bismo ispravno pripisali nekome dispozicijsko svojstvo jest da bude istinit protučinjenični kondicional koji specificira uvjete u kojima će se dispozicija manifestirati.

3.10 Prednosti Ryleova biheviorizma

Ryleov meki biheviorizam ne suočava se s prigovorima koji se upućuju tvrdom logičkom biheviorizmu. Vidjeli smo da Ryle ne prihvaća jaku tvrdnju da se psihološki iskazi mogu prevesti u iskaze kojima se referira na čisto fizičke procese. Stoga mu se ne može prigovoriti da u opisu radnji kroz koje se tipično manifestiraju mentalna stanja koristi bogate intencionalne opise koji podrazumijevaju pojmove vjerovanja, želje ili namjere. Štoviše, sama ideja da se govor o psihološkim stanjima i procesima može svesti na govor o fizičkim stanjima i procesima je u suprotnosti s Ryelovom analizom mentalnih pojmova. Ryle je smatrao da zastupnici materijalizma—teze da je um mozak ili da su psihološka svojstva fizička svojstva mozga—čine sličnu pogrešku poput filozofa koji prihvaćaju kartezijanski dualizam (vidi Ryle 1949, 22-23). Naime, kada bi um bio identičan mozgu to bi značilo da je um neka stvar koja se može identificirati s drugom stvari. Međutim, vidjeli smo da prema Ryleu um ne spada u istu logičku kategoriju u koju spada govor o fizičkim predmetima. Um nije stvar već skup sposobnosti koje se manifestiraju kada su zadovoljeni određeni uvjeti.

Osim što zaobilazi probleme s kojima se suočavaju logički pozitivisti, Ryleova analiza psiholoških pojmova elegantno rješava probleme s kojima se suočava kartezijanski dualizam. Budući da um nije supstancija koja postoji paralelno uz fizičku supstanciju, onda nema smisla ni pitati u kakvom su odnosu um i tijelo. Također, Ryle je svojom analizom mentalnih pojmova nastojao pokazati da su kriteriji za pripisivanje mentalnih stanja određeni njihovom tendencijom da se manifestiraju u javno opažljivim ponašanjima, a ne pretpostavkom postojanja opskurnih unutrašnjih stanja koja se paralelno odvijaju s fizičkim procesima u mozgu. Ako je tome tako, onda možemo vidjeti zašto problem tuđih umova, koji se javlja za dualistička gledišta, zapravo nije problem. Znamo što druge osobe misle, žele, osjećaju i tome slično jer ispravno pripisivanje tih mentalnih stanja ovisi o tome kako se manifestiraju u ljudskim radnjama kroz različite okolnosti, bez obzira na fiziološke procese koji se mogu, a i ne moraju, odvijati u ljudima.

Međutim, Ryleov se biheviorizam suočava s drugim problemima. Oni se mogu podijeliti u tri skupine, to su: problemi koji se odnose na njegovu fenomenalističku teoriju dispozicija, problemi koji se odnose na asimetriju spoznaje vlastitih i tuđih mentalnih stanja te problemi koji se odnose na prirodu kvalitativnih aspekata svjesnih mentalnih stanja, onoga što se u

filozofskoj literaturi naziva *qualia*. Krenut ćemo s problemima koji se odnose na njegovu fenomenalističku teoriju dispozicija.

3.11 Problem s fenomenalističkom teorijom dispozicija

Fenomenalističke koncepcije dispozicija podložne su prigovoru eksplanatorne ispraznosti koji je poznat pod nazivom *virtus dormitiva*. U Molièreovom dijelu „Umišljeni bolesnik“, jednog od likova se pita zašto opijum uspavljuje te on odgovara zato što opijum ima sposobnost da uzrokuje spavanje (lat. *virtus dormitiva*). Jasno je da ovakva vrsta objašnjenja zapravo ne objašnjava ono što smo htjeli znati. Ono što želimo saznati jest zašto opijum ima dispoziciju da uzrokuje pospanost kod ljudi. Da bismo dobili taj odgovor moramo saznati što se nalazi u kemijskom sastavu opijuma koji utemeljuje njegova dispozicijska svojstva.

Fenomenalističkom ili operacionalističkom gledištu suprotstavljeno je realističko gledište na dispozicijska svojstva. Realističko gledište podrazumijeva da postoje neka stanja ili kategorička (nedispozicijska) svojstva koja utemeljuju dispozicijska svojstva. Ovakva koncepcija dispozicija zaobilazi problem virtus dormitive jer pretpostavlja postojanje kategoričkog svojstva koje utemeljuje dispozicije i objašnjava zašto se manifestiraju u određenim okolnostima. Ovakvo gledište dobro zahvaća kako inače razmišljamo o dispozicijskim svojstvima. Na primjer, smatramo da ono što objašnjava zašto je staklo lomljivo jest mikrofizička struktura stakla, koja uključuje nedispozicijska ili kategorička svojstva. Kada bismo izmijenili tu strukturu, npr. kaljenjem stakla, onda bismo ujedno promijenili njegova dispozicijska svojstva.

Prema Ryleovom fenomenalističkom gledištu, nema pretpostavke da postoje takva temeljna svojstva koja bi bila odgovorna za dispozicije. Štoviše, prema njemu mentalna stanja nisu ništa drugo nego dispozicije za ponašanje. Međutim, ako je uvjerljivo da objašnjavalačka moć dispozicija općenito ovisi o njihovim kategoričkim svojstvima, kao što mikrofizika stakla objašnjava njegovu krhkost, onda postaje uvjerljivo tvrditi da i u slučaju ponašanja neka unutrašnja stanja i procesi koji se odvijaju, na primjer, u živčanom sustavu uzrokuju da se osoba u određenim uvjetima ponaša na taj i taj način. Stoga postaje uvjerljivo tvrditi da mentalna stanja nisu samo dispozicije za ponašanje nego unutrašnja stanja koja se nalaze u temeljima tih dispozicija.

Zastupnik mekog biheviorizma mogao bi odgovoriti da čak i ako se prihvati realistička slika dispozicija postojanje kategoričkih temelja za dispozicije svejedno nije bitno za pripisivanje mentalnih stanja. Logički kriteriji koji određuju kada je prikladno pripisati nekome neko mentalno stanje ovise samo o dispozicijama za ponašanje bez obzira na to koji su im potencijalni temelji u mozgu. Stoga se mentalna stanja ne mogu poistovjetiti s kategoričkim temeljima dispozicija.

No, prihvaćanje realističkog gledišta na dispozicije ide protiv ovog gledišta. Uobičajeno je da mentalna stanja i procese shvaćamo kao eksplanatorne. Pod time mislimo na činjenicu da pripisujemo mentalna stanja kako bismo objasnili druga psihološka stanja i radnje koje poduzimaju ljudi. Na primjer, Ivanov odlazak u kino možemo objasniti time što mu gledanje filmova u kinu pruža zadovoljstvo. Ili predviđamo Maričin odlazak u trgovinu zato što znamo da želi za doručak jesti palačinke i zna da kod kuće nema brašna. Kako bi mentalna stanja mogla igrati te uloge u objašnjenjima i predviđanjima radnji i drugih psiholoških stanja, ne smiju se poistovjetiti s dispozicijama za djelovanje jer bi u tom slučaju mentalna stanja bila identična onome što bi trebala objasniti. Ako je vjerovanje da vani pada kiša identično, na primjer, sklonosti da kada percipiramo da vani pada kiša uzimamo sa sobom kišobran, onda to vjerovanje ne može objasniti zašto uzimamo kišobran kada idemo van. Uzimanje kišobrana bi u tim okolnostima bila samo jedna moguća manifestacija vjerovanja da vani pada kiša, a postojanje sklonosti ne objašnjava zašto se ona manifestira u određenim okolnostima. Slično kao što krhkost stakla ne objašnjava zašto se ono razbilo u određenoj situaciji. Razbijanje stakla samo je manifestacija sklonosti da se ono razbije. Stoga izgleda da mogućnost objašnjenja eksplanatorne moći mentalnih stanja pretpostavlja da se ona odnose na kategorička svojstva koja utemeljuju i objašnjavaju zašto se dispozicija u određenim okolnostima manifestira. Čini se da meki biheviorizam, suočen s ovakvom situacijom, ili mora odustati od zdravorazumskog gledišta da mentalna stanja mogu igrati eksplanatorne uloge ili mora prihvatiti da su ona identična nekim unutrašnjim stanjima mozga u slučaju fizikalističke slike uma, koja se nalaze u temeljima naših dispozicija za ponašanje.

3.12 Problem s povlaštenim pristupom

Filozofi smatraju da postoji nešto neobično sa spoznajom koja se odnosi na (a) *pojavnost* i (b) *prirodu* mentalnih stanja. Štoviše, mnogima se čini da postoji asimetrija između spoznaje vlastitih mentalnih stanja i spoznaje mentalnih stanja drugih osoba (Boghossian 1989; Moran 2001; Pećnjak i Janović 2016). Drugim riječima, obično se ističe razlika između spoznaje mentalnih stanja iz *perspektive prvog lica* i spoznaje tuđih mentalnih stanja iz *perspektive trećeg lica*.

Pretpostavimo da *p* označava propoziciju „Subjekt *S* ima mentalno stanje *M*“ te da *q* označava „S-ovo mentalno stanje *M* ima esencijalno to i to svojstvo“. Prethodno navedena asimetrija može se objasniti pomoću sljedećih uvjeta:

(1) **Neposrednost znanja vlastitih mentalnih stanja**: dok se znanje iz perspektive trećeg lica o mentalnom stanju osobe *R* (koja je različita od *S*-a) koja vjeruje da je *p* slučaj temelji

> na promatranju i zaključivanju, S-ovo znanje iz perspektive prvog lica da je *p* slučaj je *direktno* ili *neposredno*.

Na primjer, Ivica prema ovom uvjetu zna da vjeruje da je Zagreb glavni grad Hrvatske jer ima neposredan uvid u svoja mentalna stanja. Međutim, Ivica može znati samo posredno da Marica ima isto vjerovanje. To može recimo saznati ako mu ona točno odgovori na pitanje koji je glavni grad Hrvatske.

> (2) ***Nepogrešivost* znanja iz perspektive prvog lica:** Ako *S* vjeruje da *p*, tada je *p* istinito (Ako *S* vjeruje da nije *p*, tada *p* nije istinito).

Ovim uvjetom se tvrdi da je spoznaja vlastitih mentalnih stanja nepogrešiva, za razliku od spoznaje tuđih mentalnih stanja za koje se pretpostavlja da uvijek ostaje prostora za grešku. Na primjer, ako Ivica vjeruje da osjeća bol, onda iz toga slijedi da on osjeća bol. Ili ako smatra da ga ništa ne boli, onda slijedi da ga ništa ne boli. Međutim, ako Ivica vjeruje da Maricu nešto boli, iz toga ne slijedi da je stvarno nešto boli.

> (3) ***Samoočiglednost* ili *samoobznanjivost mentalnih stanja***: Ako je *p* istinito, onda *S* vjeruje da *p*.

Ovaj uvjet predstavlja drugu stranu medalje u odnosu na prethodni uvjet. Na primjer, djeluje intuitivno reći da je Ivica, ako osjeća bol, toga i svjestan te u tom smislu vjeruje da ga nešto boli. Dakle, čini se da osjećati bol pretpostavlja imanje mentalnog stanja koje je samoočigledno.

> (4) ***Nepopravljivost***: *S* ima najveći epistemički autoritet donositi sudove o tome je li *p* istinito. Nitko drugi se ne nalazi u boljoj poziciji pokazati da *S* griješi u pogledu *p*-a.

Ovim se uvjetom ističe snaga autoriteta osobe o kojoj se radi. Ivica je najbolji autoritet za prosuditi cijeli niz pitanja koja se odnose na njegova mentalna stanja. Na primjer, ako želimo znati osjeća li bol, vrti li mu se u glavi, ima li određeno svjesno vjerovanje i tome slično, najbolje je pitati samog Ivicu.

Prigovor koji se može uputiti Ryleovom i sličnim vrstama biheviorizma jest da ne mogu objasniti privilegirani pristup koji ljudi imaju prema svojim mentalnim stanjima. Razmotrit ćemo ovu tvrdnju u odnosu na prethodno navedene uvjete.

Ryle bi se mogao složiti s tvrdnjom da barem za neka mentalna stanja vrijedi da, ako ih posjedujemo, onda direktno znamo da ih posjedujemo. Stoga nije nužno da uvjeti (2) i (3) predstavljaju nepremostivu prepreku za Ryleov biheviorizam. Međutim, čini se da Ryle ne prihvaća to da osoba ima privilegirani i inherentno *drugačiji* pristup svojim mentalnim stanjima u

odnosu na pristup koji ima prema tuđim mentalnim stanjima. Dapače, Ryle smatra da postoji simetrija između onoga što možemo saznati o sebi i o drugim ljudima. U tom pogledu navodi kako su „[v]rste stvari koje mogu saznati o sebi jednake […] vrstama stvari koje mogu saznati o drugim ljudima te su metode pronalaska približno iste“ (vidi Ryle 1949, 138). Štoviše, Ryle eksplicitno tvrdi da vlastita mentalna stanja spoznajemo jednako kao i mentalna stanja drugih osoba:

> Ako sada postavimo pitanje epistemologa: „Kako osoba spoznaje u kakvom je raspoloženju?”, Možemo odgovoriti da (...) on to otkriva na isti način poput nas. Kao što smo vidjeli, on ne [govori mrzovoljno] „Osjećam dosadu“ jer je otkrio da mu je dosadno, ništa više nego što pospani čovjek ne zijeva jer je otkrio da je pospan. Dapače, slično kao što pospani čovjek spoznaje da je pospan utvrdivši da, između ostalog, i dalje zijeva, tako i čovjek kojemu je dosadno spoznaje da mu je dosadno […], ustanovivši da između ostalog mrzovoljno govori drugima i sebi „Osjećam dosadu“ i „Kako mi je dosadno.“ (Ryle 1949, 87–88)

Ako se prihvati da govor o mentalnim stanjima zapravo u najvećem broju slučajeva implicira govor o dispozicijama za ponašanje, postaje jasno zašto se mnogima čini da Ryleovo gledište na neuvjerljiv način ukida asimetriju između vlastitih i tuđih mentalnih stanja. Naime, tuđa mentalna stanja spoznajemo kroz promatranje verbalnog i neverbalnog ponašanja kroz razne kontekste. Ako doslovno shvatimo Rylea kada kaže da su metode otkrića iste u slučaju spoznaje vlastitih i tuđih mentalnih stanja, onda ispada da i mnoga vlastita mentalna stanja spoznajemo na temelju promatranja toga kako se dispozicije za ponašanje manifestiraju u različitim kontekstima. Stoga ispada da ne možemo direktno znati što mislimo ili osjećamo. Ta posljedica je u suprotnosti s intuitivnom idejom da barem neka vlastita mentalna stanja spoznajemo na neposredan način.

Također, ako se doslovno uzme Ryleova tvrdnja da nema principijelne razlike između spoznaje vlastitih i tuđih mentalnih stanja, onda se čini da njegova varijanta biheviorizma negira uvjet nepopravljivosti. Naime, čini nam se da barem u nekim slučajevima imamo nepopravljivo znanje o tome što nam se događa. Kada osjećamo bol onda se čini da se ne možemo varati u pogledu toga. Međutim, druga osoba ne može imati slično nepopravljivo znanje u pogledu toga boli li nas nešto. Druga osoba može to činiti samo na temelju promatranja našeg ponašanja i zaključivanja o njemu iz svoje perspektive.

Nije sasvim jasno koliko su ovi prigovori uvjerljivi protiv Ryleova gledišta. Sam Ryle (1949, pogl. 7) tvrdi da se ideje nepopravljivosti, epistemičkog autoriteta i neposrednog uvida u vlastita mentalna stanja i procese temelje

na kartezijanskoj slici uma. Stoga ako je ta slika uma pogrešna onda možda ni prigovori nisu previše uvjerljivi.[34]

Međutim, čini se da čak i ako je kartezijanska slika uma pogrešna, to ne znači da neka mentalna stanja i procesi ne podrazumijevaju postojanje privatnog uvida koji je u suštini drukčiji od uvida u mentalna stanja i procese koji može imati neka druga osoba. Kao što smo vidjeli, barem se neka mentalna stanja čine privatnima. Ako nas nešto boli, onda barem ponekad to direktno i neupitno znamo. To nam ukazuje da barem u nekim situacijama postoji asimetrija između znanje o vlastitim i tuđim mentalnim stanjima. Ako Ryle i njegovi sljedbenici smatraju da odbacivanjem kartezijanskog dualizma odbacujemo postojanje epistemičke razlike u spoznaji vlastitih i tuđih mentalnih stanja, onda to može biti samo zato što pretpostavljaju, bez dokaza, da je govor o mentalnim stanjima analitički povezan s govorom o javno opažljivim ponašanjima i dispozicijama za ponašanje. Drugim riječima to može biti jedino zato što pretpostavlja neku vrstu biheviorizma. Ako tome nije tako, onda bi trebalo jasnije specificirati što nas točno sprečava smatrati da postoji, ne samo u stupnju, već i u principu razlika u načinu spoznaje vlastitih i tuđih mentalnih stanja. Stoga, ukoliko Ryleova pozicija podrazumijeva da *ne može* biti principijelne razlike između toga kako spoznajemo vlastita i tuđa mentalna stanja, utoliko njegovo gledište postaje neuvjerljivo.

3.13 Osjetilna mentalna stanja i Superspartanci

Meki biheviorizam može na dobar način objasniti kako ljudima pripisujemo intencionalna mentalna stanja. Naime, čini se uvjerljivim da će pripisivanje mentalnih stanja poput vjerovanja i želje igrati neku ulogu u objašnjenu toga kako i zašto ljudi poduzimaju određene radnje te će se njihova priroda barem djelomično iskazati u ponašanju. Međutim, nije jasno da se može dati slična analiza iskustvenih stanja čija je priroda u velikoj mjeri određena time kako ih doživljavamo. Na primjer, bol nije samo stanje koje se iskazuje u ponašanju. Bol je mentalno stanje koje doživljavamo na određeni način. Slično tome, raspoloženja se očituju u tome kako ih osoba doživljava te ih je često teško karakterizirati riječima i spoznati ako ih sami nismo doživjeli.

Nije jasno što bi pobornik mekog biheviorizma mogao reći o prirodi takvih iskustvenih i osjetilnih mentalnih stanja. Kartezijanski dualist bi rekao da su svjesna stanja zapravo svojstva ili modifikacije našeg privatnog uma. No, ako su kriteriji za pripisivanje iskustvenih doživljaja ovisni o javnim procedurama koje se odnose na javno opažljive karakteristike ljudi i njihove dispozicije za ponašanje, onda bi ispalo da poznavanjem tih procedura ujedno znamo sve što se ima znati o takvim mentalnim stanjima i procesima. No, čini se da je

[34] Takve argumente u kontekstu rasprave o prirodi svjesnih iskustava razvija Ryleov student Daniel Dennett (1988). Dennettovim argumentima ćemo se baviti u poglavlju 9.

upravo suprotno. To su stanja čiju prirodu možemo spoznati samo ako ih doživimo. Stoga se čini da se njihova priroda ne može iscrpiti načinom na koja se ona mogu ili ne moraju manifestirati u ponašanju.

Na prethodni prigovor bi se moglo odgovoriti da Ryle dopušta da postoje privatna mentalna stanja i procesi kojima pristup može u nekim situacijama imati samo subjekt koji ih doživljava (Tanney 2009, xxxii). Ono što je Ryleu važno istaknuti jest da ti privatni mentalni procesi i događaji ne mogu biti *paradigmatski* slučajevi znanja o mentalnim procesima. U normalnim slučajevima mi na temelju promatranja verbalnog i fizičkog ponašanja osobe možemo spoznati kada ona ima neku misao ili doživljava neko iskustvo. Upravo takvi javno opažljivi kriteriji za pripisivanje mentalnih atributa predstavljaju standardne epistemičke procedure putem kojih odlučujemo nalazi li se netko ili ne u nekom mentalnom stanju. Prema Tanney, ova ideja ima značajne posljedice jer „ide protiv bilo koje teorije [o mentalnim stanjima] koja bi je učinila nespoznatljivom u načelu ili u praksi […] (Tanney 2009, xxx).

Međutim, nije jasno zašto bi postojanje procedure koja je interpersonalno dostupna i u mnogo slučajeva nam daje točne rezultate o tome primjenjujemo li neki mentalni atribut ispravno govorilo protiv gledišta prema kojemu je moguće da postoji načelna asimetrija u spoznaji vlastitih i tuđih mentalnih stanja. Naime, iako možemo zamisliti da u većini okolnosti ispravno pripisujemo nekoj osobi da se nalazi u boli na temelju vidljivog ponašanja, iz toga ne slijedi da je vidljivo ponašanje nužan kriterij za postojanje boli ili bilo kojeg drugog mentalnog stanja. Hilary Putnam (1965) ilustrira tu mogućnost pomoću primjera hipotetičkih Superspartanaca. Superspartanci su osobe koje mogu normalno doživjeti osjećaj boli, no to nikada ne pokazuju kroz ponašanje na temelju kojeg bi im druga osoba mogla pripisati to iskustvo. U tom slučaju standardna procedura za pripisivanje boli Superspartancima navela bi nas na krivi put jer bismo mislili da Superspartanci ne osjećaju bol kada je zapravo osjećaju.

Ovdje bi se moglo odgovoriti da to nije standardni slučaj te da nas normalna procedura za pripisivanje boli obično dobro služi. Naime, kako trenutno koristimo pojmove kojima referiramo na svjesna i osjetilna stanja, njihovi kriteriji primjene se uvelike temelje na promatranjima ponašanja i konteksta u kojima se izvode.

No, čak i ako prihvatimo prethodnu tvrdnju, poanta ovog hipotetičkog primjera je malo suptilnija. On nam ukazuje na to da, ako dopustimo mogućnost postojanja Superspartanaca, smanjujemo plauzibilnost tvrdnje da čak i u normalnim aktualnim slučajevima kriteriji za pripisivanje, pa onda i spoznaju mentalnih i osjetilnih stanja, *konstitutivno* ovise o javno dostupnim procedurama za utvrđivanje postojanja takvih stanja. Nasuprot tome, skreće se pozornost na intuitivno gledište da, iako se mentalna stanja tipično manifestiraju kroz ponašanja te nam ona služe kao *bihevioralni*

dokazi za prisutnost tih stanja, ona nisu konstitutivno povezana s tom javno dostupnom dokaznom građom. Kako bismo pojasnili ove tvrdnje, malo ćemo detaljnije opisati Putnamov (1965) primjer sa Superspartancima.

U misaonom eksperimentu zamišljamo vrstu Superspartanaca ili Superstoika koji čine zajednicu u kojoj odrasle osobe imaju sposobnost sakriti svo ponašanje koje je povezano s doživljajem boli. Superspartanci mogu osjećati bol te to čak mogu i priznati. Međutim, čak i kada priznaju da osjećaju bol, oni to čine na vrlo miran način, koristeći ujednačen i dobro kontroliran glas, čak i kada pate od neizdržive boli. Dakle, niti jauču, niti skviče, niti rade grimase, niti se znoje, niti stišću zube. Priznaju da im treba veliki napor da izdrže, ali to čine zbog važnih ideoloških razloga te prolaze izraziti višegodišnji trening kako bi postali takvi. Nadalje, zamislimo da nakon više milijuna godina takvog načina odgoja i života Superspartanci počinju imati djecu koja se rađaju već prilagođena na njihov način života. Zamislimo još da nakon nekog vremena evoluiraju Supersuperspartanci koji jednostavno prestaju govoriti o boli i prave se da ne znaju što ta riječ znači. No, oni svejedno mogu osjećati bol.

Ako je ovaj misaoni eksperiment koherentan, onda nam pokazuje da se iskustvo mentalnih stanja ni u principu ne mora manifestirati u javno opažljivim ponašanjima jer su Superspartanci osobe koje mogu osjećati bol, no nemaju dispozicije da manifestiraju bol u svom ponašanju. Zastupnik mekog biheviorizma mogao bi odgovoriti da čak i ako se složimo da je koherentan, misaoni eksperiment nam ne govori ništa o našoj trenutnoj praksi pripisivanja mentalnih atributa koja se nalazi u podlozi značenja termina kojima referiramo na mentalna stanja poput osjećaja boli. Međutim, čak i ako se složimo s tom tvrdnjom, ovaj primjer nam ukazuje na to da bez obzira na to kakve su naše trenutne prakse pripisivanja mentalnih stanja, svejedno ostaje otvorena mogućnost da nasuprot gledištu mekih biheviorista, javno dostupne procedure na temelju kojih u normalnim slučajevima pripisujemo mentalna stanja poput boli ne predstavljaju *konstitutivne* kriterije koji određuju kada se netko nalazi u stanju boli niti određuju značenje termina kojima pripisujemo iskustvo boli (Putnam 1965). Nasuprot tome, ovaj primjer nam ukazuje na to da se trenutni kriteriji za pripisivanje mentalnih stanja mogu plauzibilno smatrati dostupnom dokaznom građom na temelju koje možemo *zaključivati* nalazi li se netko u određenom mentalnom stanju. Trenutno nam je ta dokazna građa dostupna i pouzdana, no kada dođemo do slučaja Supersuperspartanaca ta praksa bi prestala biti pouzdana, budući da oni nemaju dispozicije za manifestiranje boli.

Nadalje, i možda još važnije za našu temu, mogućnost ovog slučaja ukazuje nam na to da činjenica da su uobičajeni kriteriji za pripisivanje mentalnih stanja i procesa javni i opažljivi ne osporava tvrdnju da postoje mentalna stanja koja su u načelu privatna i kojima direktan pristup ima samo

osoba koja ih doživljaja. Superspartanci su osobe koje mogu osjećati bol, a da je ne manifestiraju kroz ponašanje. Dakle, moguće je da druge osobe ne znaju kada oni osjećaju bol. To nam pokazuje da u tom slučaju postoji asimetrija u spoznaji vlastitih mentalnih stanja u odnosu na tuđa. Štoviše, ako postoji epistemička asimetrija u tom hipotetičkom slučaju, onda imamo razloga vjerovati da i u našem aktualnom slučaju postoji određena asimetrija u poznavanju naših mentalnih stanja u odnosu na tuđa. Budući da Ryle ne daje jasna razmatranja koja bi ukazivala na epistemičku razliku u samospoznaji kod Superspartanaca i nas, onda nije jasno zašto bismo odustali od teze da postoji asimetrija između spoznaje vlastitih i tuđih mentalnih stanja. Te ukoliko meki biheviorizam implicira da ne postoji privilegirani pristup, utoliko to gledište na prirodu uma postaje manje uvjerljivo.

3.14 Zaključna razmatranja

Zbog primjera i tvrdnji koje karakteriziraju meki biheviorizam, mnogi filozofi koji su bili nezadovoljni problemima u koje zapada kartezijanski dualizam odustaju od Ryleova gledišta. Bilo da to gledište karakteriziramo kao bihevioristično ili ne, čini se da se ono susreće s ograničenjima kada se radi o objašnjenju prirode mentalnih procesa koje vežemo uz svjesna i osjetilna iskustva.

Stoga se mnogi autori okreću pokušaju utemeljenja materijalističkih teorija uma koje bi s jedne strane izbjegle dualističke pretpostavke, a s druge uspjele ponuditi uvjerljivo objašnjenje načina na koji um i njegova svojstva možemo uklopiti u znanstvene spoznaje o funkcioniranju fizičkog svijeta. Takvim teorijama ćemo se baviti u narednim poglavljima.

4 Teorija identiteta tipova

4.1 Uvod

Teorije identiteta u filozofiji uma su materijalističke ili fizikalističke teorije koje se razvijaju tijekom 50-tih godina 20. stoljeća. Prema njima, mentalna stanja, događaji i procesi *identični* su stanjima, događajima i procesima u mozgu.[35] Treba istaknuti da prema teoriji identiteta tipova sva mentalna stanja su stanja mozga, no moguće je da postoje neka stanja mozga koja nisu mentalna. Ian Ravenscroft (2005, 39) daje primjer glija-stanica. Osim neurona, mozak se sastoji od glija-stanica koje podržavaju rad živčanog sustava, no ne provode električne impulse poput neurona. Glija-stanice formiraju mijelin koji okružuje i pruža potporu neuronima. Stoga je moguće da same glija-stanice nisu identične nijednom mentalnom stanju, već da samo podržavaju sustave koji čine mentalna stanja. Još treba istaknuti da se teorija identiteta tipova prvotno formulira i najčešće razvija, iako ne isključivo, kao teorija o fizičkoj prirodi *osjetilnih* mentalnih stanja, poput boli, koji imaju određeni pojavni karakter, a ne toliko uz intencionalna stanja poput vjerovanja, želja i namjera. Štoviše, kao što ćemo vidjeti u nastavku, izvorni pobornici teorije identiteta tipova smatrali su da logički biheviorizam Ryleova tipa može objasniti prirodu intencionalnih stanja. Ono što nam još preostaje objasniti jest priroda osjetilnih iskustava, koja nisu nisu nužno povezana s dispozicijama za ponašanje (Place 1956; Smart 1959; vidi također Sesardić 1984).

Teorija identiteta monistička je teorija jer se njome tvrdi da postoji samo jedna vrsta stvari. Istodobno je materijalistička jer se tvrdi da postoje samo materijalne stvari. U novije doba, ova vrsta materijalizma naziva se fizikalizam. Razlika između materijalizma i fizikalizma bi bila u njihovim konotacijama. Materijalizam se obično povezuje s teorijama prema kojima

[35] U nastavku ćemo skraćeno govoriti da su mentalna stanja identična stanjima mozga. Međutim, treba imati na umu da se ovdje govor o mentalnim i moždanim stanjima shvaća općenito kao govor o onome čiju će nam pravu prirodu tek otkriti znanost. Dakle, znanost će nam otkriti na što se točno referira naš govor o umu i njegovim svojstvima, radi li se stanjima mozga, dijelovima mozga, procesima u mozgu ili nekoj kombinaciji toga (Polger 2009, 824).

je stvarnost sastavljena od materijalnih čestica, koje se nalaze u prostoru, kreću se, određene su veličine i tome slično. Međutim, razvoj fizike nam ukazuje na to da postoje fizički procesi koji možda nisu materijalni u tom smislu. Primjerice, sila poput gravitacije jest fizički proces koji nije materijalan u smislu da se kreće u prostoru. Stoga, mnogi autori smatraju da je prikladnije govoriti o fizikalizmu u filozofiji uma. Prema fizikalizmu, sve se stvari, uključujući ljude i njihova mentalna stanja, mogu objasniti pozivanjem na fizičke događaje, procese i stanja. Osnovna ideja je da ne postoji nešto u prirodi što bi nadilazilo ili stajalo izvan fizičke stvarnosti.

U ovom poglavlju najviše ćemo se baviti varijantom teorije identiteta koja se naziva teorija identiteta *tipova*. Prema ovom gledištu, tipovi ili vrste mentalnih stanja identični su tipovima ili vrstama moždanih stanja. Počeci ove varijante teorije identiteta obično se povezuju s radovima U. T. Placea (1956), Jacka Smarta (1959) i Herberta Feigla (1967). Međutim, već kod Carnapa se može detektirati prihvaćanje određene varijante teorije identiteta tipova. Na primjer, Carnap, kada izlaže osnovne postavke svoje verzije logičkog biheviorizma, navodi da „rečenica o stvarima koje se odnose na tuđu psihu govori da se fizički proces određenog *tipa* odvija u ili na tijelu osobe o kojoj se govori“ (Carnap 1995, 54, kurziv dodan). Unatoč tome, Carnapa se ne smješta u zastupnike suvremene varijante teorije identiteta tipova. Dapače, izvorni zastupnici suvremene teorije identiteta tipova obično izgrađuju i brane svoja gledišta u suprotnosti s osnovnim postavkama logičkog pozitivizma te smatraju da ona može ispraviti nedostatke koje smo primijetili kod različitih verzija filozofskih biheviorizama (vidi Place 1956; Smart 1959).

Kako bismo bolje shvatili osnovne teze teorije identiteta tipova i argumente kojima se brane, u nastavku ćemo prvo objasniti neke od važnijih razloga za prihvaćanje teorije identiteta tipova. U sklopu tih razmatranja, osvrnut ćemo se na metodološke pretpostavke na temelju kojih istaknuti teoretičari identiteta razvijaju svoje teze i argumente. Nakon toga ćemo detaljnije objasniti teze koje karakteriziraju teoriju identiteta tipova. Naposljetku ćemo razmotriti probleme s kojima se suočava teorija identiteta tipova i načine na koje se može braniti.

4.2 Filozofski i metodološki izvori teorije identiteta tipova

Nasuprot tvrdim biheviorističkim, teoretičari identiteta tipova ne smatraju da njihova tvrdnja o identitetu mentalnih i moždanih stanja implicira tvrdnju o mogućnosti *prijevoda* psihološkog vokabulara u vokabular kojim govorimo o fizičkim stvarima. Smatraju da, kada govorimo o osjetilnim stanjima, onda govorimo o određenim procesima u mozgu, slično kao što kada govorimo o vodi, zapravo govorimo o molekulama H_2O. No, to ne znači da govor o osjetilnim iskustvima možemo *prevesti* bez gubitka značenja u govor o procesima u mozgu, slično kao što se ni ne očekuje da govor o vodi možemo

bez gubitka značenja *prevesti* u govor o molekulama H_2O (Place 1956; Smart 1959).

Nadalje, metodološke pretpostavke o ulozi i ciljevima filozofije također se razlikuju od shvaćanja filozofije koju zastupaju meki biheviorist poput Rylea. Sjetimo se da je kod Rylea osnovna filozofska metoda određivanje geografije pojmova, pod čime se mislilo na određivanje logičkih veza među pojmovima. Osnovni cilj filozofije, prema toj metodologiji, je razrješavanje pojmovnih problema u koje zapadamo zbog pogrešne ili konfuzne upotrebe pojmova. To se postiže tako da se pojmovnom analizom odvoje propozicije koje su smislene od onih koje su besmislene. Drugim riječima, uloga filozofije se odnosi samo na određivanje granice smisla rečenica i iskaza kojima govorimo o svijetu (u novije doba, ovakvu metodologiju zastupaju Bennett i Hacker 2003; vidi također Hacker 2009). Prema tom gledištu postaje jasno da se filozofija ne bavi otkrivanjem istinitih ili neistinitih propozicija koje bi govorile nešto supstancijalno o svijetu u kojem živimo. To bi bila uloga znanosti koja se striktno treba razlikovati od filozofije kao pojmovnog istraživanja.

Nasuprot takvom gledištu, teoretičari identiteta tipova smatraju da se filozofija ne bavi samo analizom pojmova u tom užem smislu, nego da su njezini sadržaji istraživanja u kontinuitetu sa znanstvenim istraživanjima. Smart (1963, pogl. 1) smatra da, osim pojmovne analize, filozofija nastoji dati sveobuhvatnu i sinoptičku sliku svijeta. Štoviše, uloga pojmovne analize se može vidjeti kao jedno od sredstava kojima možemo doprinijeti u znanstvenom istraživanju svijeta (Armstrong 1995). Osnovna razlika između filozofije i znanosti je u tome što se filozofija bavi apstraktnijim, teorijskim i spekulativnijim dijelovima istraživanja svijeta što nije u suprotnosti s onime čime se bave znanstvenici. Štoviše, Place (1956) pretpostavlja da je teza o identitetu uma i mozga znanstvena hipoteza kojoj u prilog govore znanstvena istraživanja. S druge strane, Smart (1963, pogl. 1) prihvaća slično gledište, no ipak smatra da doprinos filozofske argumentacije i analize najviše dolazi do izražaja kada ne postoji empirijski eksperiment ili test koji bi nam mogao pomoći odrediti je li neka teorija točna. Kao što ćemo kasnije vidjeti, Smart je smatrao da je teza o identitetu mentalnih i moždanih stanja upravo takva teza čija se istinitost ne može utvrditi empirijskim istraživanjem i eksperimentiranjem, već se treba prihvatiti na temelju filozofskih argumenata koji pokazuju da je ta teza *uvjerljivija* od alternativnog, dualističkog gledišta.

Ovakvo viđenje filozofske metodologije i uloge filozofije u istraživanju svijeta direktno motivira fizikalizam koji se nalazi u pozadini teorije identiteta tipova (Smart 2017, odjeljak 1). Ono što motivira fizikalistička gledišta je razvoj prirodnih znanosti te očekivanje da će slični pomaci biti postignuti u istraživanju ljudskog uma ako ga shvatimo kao još jednu vrstu fizičkog fenomena. Na primjer, znanost nam je otkrila da je voda nešto što ima

kemijsku strukturu H_2O; da je temperatura srednja kinetička energija; da je munja električno pražnjenje oblaka; da su geni nizovi DNA molekula itd. S obzirom na ovaj uspjeh znanosti, razumno je očekivati da će nam onda znanost pokazati da su ljudi samo još jedna vrsta kompleksnih fizičkih sustava (Place 1956; Smart 1959; 1963).

Malo konkretnije, Smart ističe da je njegovo gledište da ne postoje nereducibilno psihička stanja motivirano kriterijem jednostavnosti ili Ockhamove britve. U tom pogledu navodi sljedeće:

> Čini mi se da znanost sve više daje perspektivu prema kojoj se organizmi mogu promatrati kao fizikalno-kemijski mehanizmi: čini se da će se čak i ponašanje samog čovjeka moći jednog dana objasniti u mehanističkim terminima. Čini se da nema, barem što se znanosti tiče, ničega u svijetu osim sve složenijih kombinacija fizičkih stvari. Osim na jednom mjestu: u svijesti. […] Da sve treba biti objašnjivo u terminima fizike (zajedno s opisima načina na koje su dijelovi sastavljeni […] osim pojave osjeta, čini mi se iskreno nevjerojatnim. (Smart 1959, 142)

Ovdje Smart ističe da bi, s obzirom na izuzetan razvoj znanosti, bilo čudno pretpostaviti da samo svjesna mentalna stanja ispadaju iz fizikalističke slike svijeta. Smart (1959, 142–43) u nastavku navodi da, kada se svjesna stanja ne bi mogla objasniti u fizikalističkim terminima, bili bismo primorani zaključiti da su ona, kako ih je Feigl (1967) nazivao, samo nomološki privjesci. Nekakvi događaji koji stoje izvan zakona prirode, a opet su na neki regularan način povezani s njima.

Kod Smarta se Ockhamova britva koristi na dva načina. Jedan se odnosi na broj entiteta koji teorija pretpostavlja, a drugi na broj i kompleksnost zakona koji opisuju pravilne odnose između mentalnih i fizičkih stanja. Ima smisla pretpostaviti da je teorija identiteta tipova jednostavnija od dualističkih teorija jer pretpostavlja postojanje manjeg broja entiteta; budući da su mentalna stanja identična fizičkim stanjima, onda postoji samo jedna vrsta stanja. Međutim, ona je jednostavnija i u smislu da pretpostavlja manji broj prirodnih zakona. Kako bismo pojasnili potonju primjenu Ockhamove britve, razmotrit ćemo konkretnije primjere koji motiviraju teoriju identiteta tipova.

Općenito se prihvaća da postoje mnoge korelacije između metalnih stanja i fizioloških procesa koji se odvijaju u čovjekovom mozgu i tijelu. Na primjer, kada nas boli glava, skloni smo popiti tabletu koja otklanja bol. Kada se jako udarimo u glavu možemo izgubiti svijest. Nadalje, neurološka istraživanja pokazuju da lezije na predfrontalnom korteksu mozga dovode do promjene u karakteru osobe. Korištenjem funkcionalne rezonance i drugih metoda za snimanje i skeniranje mozga vidimo da se aktiviraju specifična područja u mozgu ovisno o tome što radimo ili što nam se događa. Na primjer, kod

zdravih ljudi koji dožive averzivni događaj ili vide lice koje je uplašeno aktivira se limbički sustav u mozgu koji uključuje i amigdalu. Oštećenje amigdale dovodi do smanjenja reagiranja na zastrašujuće podražaje i onemogućuje učenje putem averzivnih podražaja (Blair, Mitchell, i Blair 2008). Slične primjere korelacija između fizičkog i mentalnog možemo nabrajati u nedogled.

Ako se zapitamo što zapravo objašnjava mnogobrojne korelacije između mentalnih i moždanih događaja, tada nam se nude barem dva odgovora. Teoretičari identiteta tipova odgovaraju da zapravo ovdje nema korelacije jer mentalni procesi *jesu* jedna vrsta fizičkih procesa u mozgu. S druge strane, dualisti bi mogli reći da korelacije postoje jer postoje psihofizički zakoni posebne vrste koji povezuju mentalne i fizičke događaje u mozgu. Međutim, pretpostavka da postoje takvi psihofizički zakoni bi prema Smartu učinila osjetilna mentalna stanja samo nomološkim privjescima. Ne samo to, nego bi pretpostavka da postoje takvi psihofizički zakoni bila vrlo nevjerojatna jer bi to bili zakoni posebne vrste koji se ne bi mogli usporediti s drugim zakonima prirode. Naime, to su, prema pretpostavci, zakoni koji se ne mogu izvesti iz drugih fundamentalnih zakona. Stoga ako postoje, morali bismo pretpostaviti da oni čine dodatnu vrstu novih fundamentalnih zakona kakve ne susrećemo u fizici. I to je upravo ono što bi ih činilo čudnim i nevjerojatnim. Naime, oni ne bi povezivali jednostavne čestice koje čine temeljnu stvarnost našeg svijeta jer bi onda mogli biti svrstani u vrstu temeljnih zakona koji su kompatibilni s trenutno postojećim fizikalnim zakonima prirode. Dakle, prema pretpostavci, morali bismo povezivati mentalna stanja sa sinkroniziranim radom milijardi neurona koji se nalaze u podlozi naših mentalnih života. Prema Smartu upravo se u toj pretpostavci sastoji nevjerojatnost tih zakona:

> Takvi temeljni zakoni ne bi bili poput ičeg dosad poznatog u znanosti. Imaju „čudan“ miris. […] Kada bi nas neki filozofski argument primorao da povjerujemo u takve stvari, posumnjao bih da se nalazi neka kvaka u argumentu. (Smart 1959, 143)

Dosad nismo otkrili takve zakone, niti nam razvoj prirodnih znanosti i tendencije da sve objašnjavamo u terminima fizičkih uzroka, daje razloga vjerovati da ćemo u budućnosti susreti takvu vrstu novih fundamentalnih psihofizičkih zakona. Stoga, prema Smartu, Ockhamova britva daje prednost prihvaćanju jednostavnije teorije prema kojoj su mentalna stanja identična stanjima mozga.[36]

[36] Nasuprot Smartu, autori koji smatraju da postoji teški problem svijesti skloni su misliti da rješenje tog problema zahtijeva upravo pretpostavku da postoje fundamentalni psihofizički zakoni (vidi Chalmers 1996; 2010a). Razmatranjem tih gledišta, u kontekstu teškog problema svijesti, bavit ćemo se u poglavlju 9.

Osim metodoloških razmatranja i fizikalizma, ono što je posebno motiviralo razvoj teorije identiteta tipova jest nezadovoljstvo bihevioristočkom analizom osjetilnih i svjesnih iskustava. Place i Smart u svojim radovima daju do znanja da je u vrijeme kada su razvijali svoje teorije u njihovim filozofskim krugovima dominiralo ono što smo nazvali meki biheviorizam te su i sami aktivno djelovali unutar te paradigme (vidi, npr. Smart 2017, odjeljak 1). S nekim aspektima logičkog biheviorizma su se slagali. Na primjer, dijelili su simpatije s nastojanjima biheviorista u razotkrivanju nedostataka kartezijanskog dualizma. Također su smatrali da bihevioristočke analize dobro zahvaćaju mentalni diskurs, poput onog koji se odnosi na intencionalna mentalna stanja, inteligenciju, motive i crte ličnosti. U tom pogledu Place jasno ističe da

> […] u slučaju kognitivnih pojmova poput »znati«, »vjerovati«, »razumjeti«, »zapamtiti« i volicijskih pojmova poput »htjeti« i »namjeravati«, […] nema sumnje da je analiza u terminima dispozicija za ponašanje fundamentalno ispravna. (Place 1956, 44)

Ono oko čega je postojalo nezadovoljstvo ili pak skepticizam u pogledu uspješnosti bihevioristočke analize su unutrašnja iskustva. U tom pogledu Place ističe:

> […] čini se da postoji neukrotivi ostatak pojmova koji se grupiraju oko pojmova svijesti, iskustva, osjeta i mentalnih slika, gdje je neka vrsta priče o unutarnjem procesu neizbježna. Moguće je, naravno, da će se u konačnici pronaći zadovoljavajuće bihevioristočko objašnjenje ovog pojmovnog ostatka. Međutim, za naše trenutačne svrhe pretpostavit ću da se to ne može učiniti i da su iskazi o bolovima i probadanjima, o tome kako stvari izgledaju, zvuče i kako ih osjećamo, o stvarima o kojima se sanja ili ih zamišljamo u umu, iskazi koji se odnose na događaje i procese koji su u nekom smislu privatni ili unutarnji za pojedinca kojem se prediciraju. (Place 1956, 44)

Vidimo da u citatu, Place ostavlja otvorenu mogućnost da će se možda jednog dana pronaći zadovoljavajuća bihevioristička analiza iskaza kojima govorimo o unutarnjim iskustvima, no u slučaju da se to ne dogodi, nudi alternativno gledište u vidu teorije identiteta. Smart je u tom pogledu malo „otvoreniji“ te ističe svoju intuiciju da se neka mentalna stanja ne mogu objasniti putem bihevioristočke analize:

> Iako sam […] vrlo prijemčiv na […] „ekspresivno“ [bihevioristočko] objašnjenje iskaza o osjetima, ne osjećam da se njime može

> svemu doskočiti. Možda je to zato što ga nisam dovoljno promislio, ali čini mi se da, kada osoba kaže „Imam naknadnu sliku“, ona *doista* iznosi pravi sud, te kada kaže „imam bol“, ona *zaista* čini nešto više od „nadomještanja bolnog-ponašanja“ […]. (Smart 1959, 144)

Jednom kada se dopusti postojanje unutrašnjih mentalnih stanja koja se ne mogu bihevioristički analizirati onda se postavlja pitanje:

> […] jesmo li prihvaćanjem ove pretpostavke neizbježno obvezani na dualističko gledište prema kojemu osjeti i mentalne slike čine zasebnu kategoriju procesa koji nadilaze fizičke i fiziološke procese s kojima je poznato da su u korelaciji. (Place 1956, 44)

Naravno, Place u nastavku dodaje da

> […] prihvaćanje unutarnjih procesa ne podrazumijeva dualizam i da se teza da je svijest proces u mozgu ne može odbaciti na logičkim temeljima. (Place 1956, 44)

Ovdje treba istaknuti da se pod „logičkim temeljima“ misli na pojmovna razmatranjima koje je Ryle (1949) isticao, a ukazuju na to da materijalisti u filozofiji uma, slično dualistima, čine kategorijalnu pogrešku kada tvrde da su mentalna stanja identična moždanim stanjima. S obzirom na te argumente, teoretičari identiteta se suočavaju s dvije prijetnje; jedna se odnosi na dualističke intuicije da unutarnja mentalna stanja ne mogu biti fizički procesi, a druga se odnosi na Ryleova razmatranje iz filozofije običnog jezika prema kojima materijalisti čine kategorijalnu pogrešku kada tvrde da su osjetilna iskustva identična fiziološkim procesima u mozgu.

U nastavku ćemo detaljnije odrediti teze koje su prihvaćali izvorni teoretičari identiteta i probleme s kojima se njihovo gledište suočava.

No, prije nego krenemo dalje, vrijedi spomenuti da se u literaturi razlikuje više varijanti teorija identiteta tipova koje se razlikuju prema dosegu teze da su mentalna stanja identična moždanim stanjima. Prema varijanti koju možemo nazvati užom, samo neka mentalna stanja su identična stanjima mozga. Kao što smo vidjeli, u izvornim formulacijama, Place i Smart su smatrali da je glavna uloga teorije identiteta tipova da objasni osjetilna iskustvena stanja, poput osjećaja boli i mentalnih slika, dok biheviorizam može ponuditi uvjerljiva objašnjenja drugih mentalnih stanja poput vjerovanja, želja i namjera koja se tipično manifestiraju u ponašanjima. Prema široj varijanti teorije identiteta sva mentalna stanja identična su moždanim stanjima. Ova varijanta teorije identiteta u literaturi se naziva materijalizam centralnog stanja, jer se iskustvo svjesnih mentalnih stanja povezivalo s radom centralnog živčanog sustava, te se najčešće veže uz

filozofa Davida Armstronga (1968, pogl. 6). U kasnijim radovima, Smart (2017, odjeljak 4) također je prihvatio šire shvaćanje teorije identiteta.

Ovu razliku između šire i uže varijante teorije identiteta dobro je imati na umu iz nekoliko razloga. Kao što smo vidjeli, teorija identiteta tipova se izvorno javlja kao rješenje za problem postojanja unutarnjih mentalnih stanja i njihove prirode. Što znači da se teorija identiteta tipova u pogledu osjetilnih stanja u principu može kombinirati s drugim gledištima u pogledu drugih vrsta mentalnih stanja. To je važno spomenuti jer je moguće da neki argumenti protiv teorije identiteta neće biti od jednake važnosti ovisno o varijanti te teorije koju prihvaćamo.

Drugi razlog je što neki argumenti u prilog šireg shvaćanja teorije identiteta također podržavaju gledišta koja se obično uzimaju kao alternative teoriji identiteta. Tu posebice treba spomenuti Armstronga (1968; 1995; vidi, također Lewis 1966), koji razvija materijalizam centralnog stanja na temelju uzročne analize pojmova kojima referiramo na mentalna stanja. Smatra da, jednom kada odredimo uzročne uloge koje mentalna stanja poput vjerovanja, želja, namjera i percepcija igraju, možemo pronaći fizičke supstrate koji utjelovljuju te uzročne uloge i identificirati ih s njima. Međutim, kao što ćemo vidjeti u poglavlju 5, takva analiza mentalnog vokabulara se u novije vrijeme veže uz funkcionalistička gledišta na prirodu uma koja se često suprotstavljaju teoriji identiteta tipova (vidi Smart 2017, odjeljak 5).

4.3 Teze teorije identiteta

Teoretičari identiteta obično rade analogije sa znanstvenim otkrićima kako bi ilustrirali svoju teoriju. Neke od važnijih primjera smo ranije spomenuli a uključuju otkriće da je voda = H_2O, toplina = prosječna kinetička energija ili da je munja = električno pražnjenje oblaka. Slično tome, oni tvrde da će nam znanost otkriti da je, na primjer, osjećaj straha zapravo samo neki obrazac aktivacije određenog dijela mozga.

Ove analogije su važne jer nam pomažu shvatiti osnovne teze teorije identiteta. Prva je da se identiteti koje očekujemo pronaći između mentalnih i fizičkih stanja otkrivaju *a posteriori*. Drugim, riječima oni su nešto što će nam znanost otkriti empirijskim istraživanjem. U tom smislu, možemo reći da ti identiteti nisu očiti bez detaljnog znanstvenog istraživanja. Naime, bez znanstvenog istraživanja ne bismo otkrili da je voda ono što ima molekularnu strukturu H_2O ili da su munje električno pražnjenje oblaka ili da je temperatura srednja kinetička energija. To nije nešto što se moglo otkriti samo apriornom analizom pojmova vode, munje i temperature. U skladu s tim, teoretičari identiteta ne smatraju da je očito da su mentalna stanja identična stanjima mozga te još manje da će nam analiza riječi i pojmova kojima referiramo na mentalna stanja ukazati na njihovu pravu prirodu kao što su smatrali bihevioristi. Nasuprot tome, oni smatraju da će nam tvrdnju

da su mentalna stanja identična stanjima mozga potvrditi razvoj znanosti o mozgu, dakle, neuroznanost, neurobiologija i tako dalje (Smart 1959).

Druga važna stvar koja dolazi do izražaja u analogijama sa znanstvenim otkrićima je da se radi o identitetu između *tipova* ili *vrste* stvari. Dosad smo govorili o teoriji identiteta tipova, pod čime se misli da su *tipovi* mentalnih stanja identični određenim *tipovima* neuralnih stanja. Kako bismo bolje shvatili što se misli pod „tipovima" u teoriji identiteta moramo razlikovati primjerke od tipova. Ovu razliku ćemo prikazati na nekoliko primjera.[37]

Možemo zamisliti da kod kuće imamo tri bernardinca. Bernardinac je vrsta psa koji je dobar za pomaganje unesrećenima, poput slučajeva kada treba pronaći osobu koja je zbog nesreće ostala prekrivena snijegom. U ovom slučaju imamo tri primjerka jednog tipa ili vrste bernardinac. Ovdje se radi o *primjercima* pojedinačne životinje, dok su *tipovi* vrste kojima pripadaju ti primjerci. Naravno primjerci mogu pripadati različitim vrstama ili tipovima stvari. Na primjer, bernardinci pripadaju vrsti pasa, sisavaca, životinja, tipovima stvari koje mogu pomoći ljudima u nevolji i tako dalje.

Drugi primjer razlike između primjerka i tipa može biti razlika između konkretnog slova i tipa kojemu pripada. Na primjer, slovo „A" može biti tip slova i konkretan primjerak. Naime, slovo „A" se može oprimjeriti kroz više primjeraka riječi. Na primjer, različiti primjerci istog tipa slova „A" pojavljuju se u riječi „auto" i „marama".

Na temelju ovog razlikovanja trebali bismo moći shvatiti razliku između identiteta primjerka i identiteta tipa. Opet ćemo se poslužiti s nekoliko primjera. Trenutni predsjednik Republike Hrvatske je Zoran Milanović. Ako vam predsjednik Republike Hrvatske kaže da morate ostati kod kuće ili primiti cjepivo kako bi se suzbila pandemija virusa korone, onda vam istovremeno Zoran Milanović govori da morate ostati kod kuće kako bi se suzbila pandemija virusa korone. Trenutni predsjednik Republike Hrvatske i Zoran Milanović su ista osoba. Stoga možemo reći da su oni identični primjerci.

S druge strane, identiteti između vode i H_2O te između munje i električnog pražnjenja oblaka primjeri su identiteta tipova. Iskazom da je voda = H_2O tvrdi se da je voda kao *tip* stvari identična tipu stvari koja ima molekularnu strukturu H_2O. Važno je primijetiti da, ako su tipovi stvari identični, onda su i njihovi primjerci identični. Na primjer, ako je voda isto što i H_2O, onda je svaki primjerak vode identičan primjerku stvari koji ima strukturu H_2O. Slično vrijedi za munje. Svaki primjerak munje, to jest, svako pojedino sijevanje jest primjerak tipa električnog pražnjenja oblaka. Ili da koristimo aktualni primjer, možemo reći da je bolest Covid-19 isto što i bolest koju uzrokuje virus SARS-CoV-2. Dakle, kao tip bolesti Covid-19 je isto što i tip bolesti koju uzrokuje virus SARS-CoV-2. Međutim, obrnuto ne slijedi. Moguće je da su primjerci stvari identični, a da tipovi kojima pripadaju nisu. Na primjer,

[37] U ovom dijelu se oslanjamo na informativan pregled koji daje Ravenscroft (2005, pogl. 3).

komad gline koji se nalazi na stolu može biti identičan Descartesovoj bisti. Međutim, tip stvari koji je glina nije nužno identičan tipu stvari kao što je biti Descartesova bista. Descartesovu bistu mogli smo napraviti od metala, drveta ili nekog drugog materijala, stoga se taj tip entiteta ne može reducirati na određenu vrstu materijala od kojeg je napravljen.

Na temelju prethodnog razlikovanja između primjeraka i tipova trebali bismo moći bolje shvatiti drugu tezu koja karakterizira teoriju identiteta. Prema teoretičarima identiteta, svaki tip mentalnog stanja identičan je nekom tipu stanja mozga. Kako bismo ilustrirali ovu tezu koristit ćemo standardni primjer iz filozofije uma. Neka istraživanja su pokazala da se u našem mozgu aktiviraju živčana vlakna koja se nazivaju C-vlakna kada nas nešto boli.[38] Radi primjera, pretpostavimo da su ta istraživanja u pravu te da je svaki primjerak boli identičan nekom primjerku aktivacije C-vlakna u našem živčanom sustavu.[39] Teoretičari identiteta bi u tom slučaju tvrdili da je bol kao tip mentalnog stanja identičan tipu fizičkog stanja koji nazivamo aktivacija C-vlakana.

Treća važna teza koju prihvaćaju teoretičari identiteta odnosi se na prirodu odnosa identiteta. Kao što smo ranije vidjeli, pobornici teorije identiteta tipova ne tvrde da se naš svakodnevni govor o mentalnim stanjima može analitički reducirati na znanstveni govor o stanjima mozga. Na primjer, oni ne tvrde da iskaz „osjećam bol“ ima isto značenje kao i „moja C-vlakna su aktivirana“. Nasuprot tome, rani zastupnici teorije identiteta smatraju da iskazi poput „Bol je isto što i aktivacija C-vlakana“ izražavaju identitet koji je, osim što se spoznaje *a posteriori*, *kontingentan.* Kontingentne istine su one koje nisu nužne. Obično se smatra da su matematičke i logičke istine nužne, dok su empirijske istine (one koje nam otkriva znanost kroz istraživanja) kontingentne. S obzirom na to gledište, teoretičari identiteta su smatrali da aposteriorna znanstvena otkrića, poput toga da je voda = H_2O, predstavljaju

[38] C-vlakna su živčana vlakna koja čine dio perifernog živčanog sustava (skup živaca koji se granaju po tijelu i spajaju se s mozgom preko leđne moždine). Ona predstavljaju jedan od dva tipa nociceptivnih živčanih vlakana (drugi tip se naziva A-delta vlakna) koji reagiraju na neugodne podražaje te šalju signale u leđnu moždinu. István Aranyosi (2013, 40–41) ističe da je ironično da se u tradiciji teorije identiteta, naročito zastupnika materijalizma centralnog sustava, počeo koristiti ovaj primjer kao ilustracija onoga što se misli pod identitetom mentalnih i moždanih stanja jer C-vlakna uopće nisu direktno povezana s centralnim živčanim sustavom.

[39] Znanstvena istraživanja su pokazala da je neuralna podloga boli kompliciranija od tvrdnje da je bol isto što i aktivacija C-vlakana. Za pristupačnu i filozofski informiranu raspravu, vidi Aranyosi (2013, 41–44). Stoga se ovaj primjer ne smije shvatiti doslovno. Njegova svrha je ilustrirati općenitu karakteristiku teorije identiteta koja je razlikuje od drugih teorija poput onih da su mentalna stanja identična stanjima bestjelesne duše ili da mentalna stanja uopće nisu stanja kako tvrde logički biheviorist. Tvrdnju da je bol kao tip mentalnog stanja identična tipu stanja mozga treba shvatiti kao ontološki nespecificiranu referenciju na one mehanizme ili procese u mozgu za koje će se ispostaviti da su identični tipu mentalnog stanja koje nazivamo „bol“ (vidi Polger 2009).

kontingentne istine (Place 1956; Smart 1959). Prema tom gledištu, moguće je da su nam znanstvena istraživanja mogla pokazati da prava priroda vode ima neku drugu molekularnu strukturu koja ne uključuje dva atoma vodika i jedan atom kisika. Analogno tome, oni bi tvrdili da, pod pretpostavkom da je točno da je bol isto što i aktivacija C-vlakana, nije nužno da su oni jedan te isti tip stvari. Koherentno je zamisliti da u nekom mogućem svijetu bol nije identična aktivaciji C-vlakana (kao što u stvarnom svijetu nije), niti da je um općenito materijalan. U tom smislu moglo se pokazati da je dualizam točno gledište na prirodu uma. Međutim, u našem svijetu empirijska znanstvena istraživanja nam daju razloga tvrditi da su mentalna stanja zapravo identična stanjima mozga.

4.4 Prednosti teorije identiteta u objašnjenju svjesnih iskustava

Teorija identiteta tipova iz znanstvene perspektive ima dosta prednosti u odnosu na dualizam kartezijanskog tipa. Kao što smo vidjeli, teorija identiteta tipova je ontološki jednostavnija teorija jer postulira manji broj entiteta i zakona prirode.

Također, važna prednost u odnosu na dualizam je da lako može objasniti *korelacije* između mentalnih događaja i fizičkih događaja u mozgu. Znanost nam pokazuje da postoje značajne korelacije između mentalnih i fizičkih stanja. To se često otkriva u slučajevima oštećenja mozga. Ako su oštećenja nekog dijela mozga povezana s nedostatkom neke mentalne funkcije, onda možemo zaključiti da je taj dio mozga važan za izvođenje mentalne funkcije. Vrlo izraziti primjer toga daje Ravenscroft (2005). Phineas Gage je živio sredinom 19. stoljeća u SAD-u gdje je radio kao vođa ekipe za postavljanje željezničkih pruga. Gageov zadatak je bio namjestiti barut u stijene kako bi ih se raznijelo. Gage je barut u stijenama namještao čeličnim štapom. Jedan dan dogodila se nesreća jer je udaranje čeličnim štapom u rubove stijene gdje se stavljao barut proizveo iskre što je rezultiralo eksplozijom. U eksploziji je štap odletio i probio Gageov lijevi obraz te je vrh izašao na prednju stranu glave iznad čela. Taj dio mozga se naziva predfrontalni korteks. Nekim čudom Gage je preživio, no njegov karakter se promijenio. Izgubio je sposobnost samokontrole kako je normalno shvaćamo. Počeo se antisocijalno ponašati i pretjerano piti, postao je nemaran i nesmotren te je sve to dovelo do gubitka posla. Ubrzo je i umro od lošeg životnog stila. Ovaj primjer nam pokazuje da sposobnost kontrole ponašanja te procjena rizika ima svoj korelat u prednjem dijelu mozga.

Teorija identiteta tipova dobro objašnjava ovu činjenicu time što se tvrdi da mentalne sposobnosti nisu ništa drugo nego neki procesi u mozgu. Naime, oni su prema ovoj teoriji identični. Tako da u strogom smislu ne postoje korelacije između mentalnog i fizičkog jer stvar ne može biti u korelaciji sa samom sobom.

Treće, ova teorija može objasniti *uzročnu interakciju* između mentalnih i fizičkih stanja na način koji je konzistentan s uzročnom zatvorenošću fizičkog. Na primjer, osjećaj boli u ruci može uzrokovati trljanje mjesta gdje boli. Jasno je kako u ovom slučaju bol koju osjećamo može uzrokovati radnje i promjene u fizičkom svijetu. Budući da su mentalna stanja identična stanjima mozga, neka stanja mozga uzrokuju radnje. Također, ova teorija može objasniti kako to da mentalna stanja uzrokuju druga mentalna stanja. Na primjer, dobar okus čokolade može uzrokovati želju da pojedem još jedan komad čokolade što u konačnici dovodi do radnje. Ako su mentalna stanja samo neka stanja mozga onda je jasno da jedno stanje mozga može uzrokovati drugo stanje mozga.

Četvrto, ova teorija također objašnjava zašto ne dolazi do problema preodređenosti (vidi poglavlje 2). Sjetimo se da se dualizam suočava s problemom preodređenosti ako prihvatimo uzročnu zatvorenost fizičkog kao metodološki princip. Taj princip navodi da svaki fizički događaj koji je uzrokovan od strane nekog drugog događaja, mora imati dovoljan fizički uzrok. Ako dualisti vjeruju da um ima uzročne moći, onda se suočavaju s problemom preodređenosti. Čak i kada ne bismo imali mentalna stanja sve u fizičkom svijetu bi ostalo isto. Teorija identiteta tipova se ne suočava s problemom preodređenosti budući da poistovjećuje mentalna stanja s fizičkim stanjima. Štoviše, možemo čak reći da rješava taj problem tako što negira da su mentalna stanja nešto iznad i povrh fizičkih stanja.

Teorija identiteta tipova također rješava neke poteškoće s kojima se suočava biheviorizam. Prisjetimo se da, prema standardnom gledištu, logički biheviorizam pretpostavlja da bi se dispozicije za ponašanje trebale analizirati u terminima uvjeta koji se moraju ostvariti kako bi se manifestirala dispozicija. Međutim, ova analiza ne nudi nikakav uvid u uzroke našeg ponašanja. Nasuprot tome, teorija identiteta tipova objašnjava bihevioralne sklonosti pomoću pretpostavke da određena moždana stanja uzrokuju određena ponašanja.

Kao što smo vidjeli u kritici Ryleove fenomenalističke teorije dispozicija, realistička teorija dispozicija pretpostavlja da su temelji dispozicija neka unutrašnja stanja predmeta. Sa sličnim problemom bi se mogle suočiti Placeova i Smartova izvorna formulacija teorije identiteta koja se odnosi samo na svjesni aspekt mentalnih stanja. Međutim, ako se u obzir uzme šira varijanta kakvu je kasnije i Smart prihvatio, onda teoretičari teorije identiteta tipova mogu lako zahvatiti i objasniti intuitivnu ideju da mentalna stanja igraju značajne uloge u *uzročnim* objašnjenjima. Ako se prihvati šira varijanta teorije identiteta, onda se jednostavno može reći da su temelji dispozicija koje povezujemo s mentalnim stanjima poput vjerovanja i želja upravo stanja mozga koja čine kategorički temelj tih dispozicija. Na primjer, kada kažemo da želja za čokoladom pretpostavlja određenu dispoziciju za ponašanje koja se manifestira u povoljnim uvjetima, onda prema

teoretičarima identiteta tipova to ujedno znači da se mozak osobe nalazi u nekom stanju koje ima tendenciju da, kada se osoba nađe u prilici pojesti čokoladu, uzrokuje njezino posezanje za čokoladom. U tom slučaju, želja za čokoladom je zapravo neko stanje koje je identično određenom stanju mozga koje igra tu uzročnu ulogu (Armstrong 1968).

Nadalje, teorija identiteta tipova može objasniti epistemičku asimetriju između znanja mentalnih stanja iz prvog i trećeg lica koja predstavlja problem za bihevioriste, barem u jednom smislu ideje epistemičke asimetrije. Vidjeli smo da bihevioristi ne mogu objasniti asimetriju između znanja naših vlastitih mentalnih stanja i znanja tuđih mentalnih stanja. Budući da su prema njima mentalna stanja samo dispozicije za ponašanje, onda bi tako shvaćena mentalna stanja trebala biti epistemički jednako dostupna nama i drugim osobama koje nas promatraju. Teorija identiteta tipova može doskočiti tome problemu i reći da u nekim slučajevima postoji epistemička asimetrija određene vrste. Na primjer, kada se radi o stanjima mozga koje osoba doživljava na određeni način, onda ih subjekt iskustva može spoznati kroz čin introspekcije, dok ih druge osobe mogu spoznati samo na temelju promatranja iz perspektive trećeg lica. Dakle, epistemička asimetrija se odnosi na različita sredstva kojima možemo spoznati mentalna, tj. moždana stanja.

Unatoč tome što teoretičari identiteta tipova mogu donekle objasniti epistemičku asimetriju u pogledu vlastitih mentalnih stanja, važno je naglasiti da teorija identiteta tipova nije kompatibilna s jakom tezom u pogledu privilegiranosti pristupa našim mentalnim stanjima (za raspravu, vidi Sesardić 1984, pogl. 2). Na primjer, nije kompatibilna s tezom da postoje neka stanja koja bi bila u *principu* privatna i nedostupna drugim ljudima. Ako su mentalna stanja samo stanja mozga, onda se čini da u principu, s razvojem tehnologija za istraživanje mozga, možemo jednoga dana dobiti objektivan pristup tuđim mentalnim stanjima, kao što sada imamo pristup svojima. Također, treba istaknuti da prema zastupnicima ove teorije introspektivni pristup mentalnim stanjima ne otkriva njihovu pravu prirodu. Kao što smo ranije vidjeli, pravu prirodu mentalnih stanja možemo otkriti tek empirijskim istraživanjima koja nam nisu nužno introspektivno dostupna. Dakle, pravu prirodu mentalnih stanja nam može otkriti samo zrela neuroznanstvena istraživanja.

4.5 Protiv teorije identiteta

Sada ćemo razmotriti neke od utjecajnijih prigovora koji se odnose na teoriju identiteta tipova. Važan skup prigovora protiv teorije identiteta oslanja se na Leibnizov zakon identiteta. Prisjetimo se da nam Leibnizov princip kaže da, ako su *x* i *y* identični, onda moraju imati sva ista svojstva. Obrnuto, ako *x* i *y* imaju različita svojstva, onda slijedi da oni nisu ista stvar. Na primjer, ako su

Clark Kent i Superman ista osoba, onda jedan i drugi mogu letjeti. Ako su Zoran Milanović i trenutni predsjednik Republike Hrvatske ista osoba, onda jedan i drugi moraju trenutno biti vrhovni vojni zapovjednici u Hrvatskoj. Nasuprot tome, ako trenutni predsjednik ima bradu, a Zoran Milanović je nema, onda oni ne mogu biti ista osoba.

U ovim primjerima Leibnizov princip primjenjuje se na primjerke. Međutim, on također vrijedi za tipove stvari. Na primjer, ako se voda koristi za piće, onda se i stvar koja ima molekularnu strukturu H_2O koristi za piće. Ili, ako sijevanje prethodi grmljavini, onda električno pražnjenje oblaka prethodi grmljavini.

Općenito, ista razmatranja vrijede kada ih primijenimo na identitet mentalnih i fizičkih stanja. Kada bi neko mentalno stanje *m* imalo neko svojstvo koje ne posjeduje nijedno stanje mozga *b*, onda bi slijedilo da *m* nije identično s *b*. Da uzmemo prijašnji primjer, ako je bol isto što i aktivacija C-vlakana u živčanom sustavu, onda svako svojstvo boli mora biti svojstvo aktivacije C-vlakana. Ako se pokaže da bol ima neko svojstvo koje fizički proces aktivacije C-vlakana nema, onda bol kao tip mentalnog stanja ne može biti identičan tipu fizičkog procesa kao što je aktivacija C-vlakana.

Krenut ćemo s epistemičkom vrstom prigovora koji se mogu uputiti teoriji identiteta tipova. Prvi prigovor koji ćemo razmotriti može se lako odbaciti. Međutim, njegova primarna uloga je ta da dodatno pojasni neke od osnovnih teza teorije identiteta tipova.

Teoriji identiteta tipova bi se moglo prigovoriti na sljedeći način.

1) Ljudi su u srednjem vijeku znali kako je to doživjeti naknadnu sliku ili kako je to osjećati bol.
2) Međutim, nisu ništa znali o ljudskoj neurofiziologiji.

Dakle,

3) Iskustvena mentalna stanja ne mogu biti identična neurofiziološkim stanjima mozga.[40]

Ovaj se argument se oslanja na Leibnizov princip na sljedeći način. Čini se da su srednjovjekovni ljudi mogli imati puno istinitih vjerovanja o mentalnim stanjima, a da nemaju istinita vjerovanja o moždanim stanjima ljudi. Stoga, mentalna stanja moraju imati neka svojstva koja ih razlikuju od moždanih stanja. U suprotnom bi istinita vjerovanja o mentalnim stanjima morala implicirati istinita vjerovanja o moždanim stanjima. Ovaj primjer nam pokazuje koja bi to mogla biti razlikovna svojstva; upravo to da ljudi mogu imati istinita vjerovanja o njima, a da nemaju istinita vjerovanja o neurofiziološkim stanjima mozga. Ili formalnije, možemo reći da je to svojstvo biti predmet znanja srednjovjekovnog čovjeka, dok stanja mozga

[40] Ovo je prvi prigovor koji razmatra Smart (1959, 46). Naša formulacija oslanja se na Kimov (2006, 106) tekst.

nemaju slično svojstvo. Dakle, mentalna stanja ne mogu biti identična moždanim stanjima.

Ova vrsta prigovora nije valjana. Jednostavno je vidjeti u čemu je problem ako pokušamo primijeniti sličan način zaključivanja na druge primjere znanstvene identifikacije. Rekli smo da teoretičari identiteta smatraju da iskaz „bol = aktivacija C-vlakana" ima isti logički status kao i iskaz „voda je H_2O". Svi ćemo se složiti da su srednjevjekovni ljudi znali dosta toga o vodi, no da sigurno nisu znali da je voda H_2O. Iz toga zasigurno nećemo zaključiti da voda nije nešto što ima molekularnu strukturu H_2O. Stoga nešto mora biti pogrešno s prethodnim argumentom koji se odnosi na identitet mentalnih i fizičkih stanja.

Takva vrsta pogreške obično se naziva intenzionalna pogreška. Ona se javlja kada intenzionalni kontekst tretiramo kao da je ekstenzionalan. Ugrubo, možemo reći da su ekstenzionalni konteksti oni u kojima govorimo o samim stvarima. Intenzionalni konteksti su oni u kojima govorimo o stvarima pod određenim aspektom ili načinom na koji su nam prezentirane.[41] Malo preciznije, možemo reći da je ekstenzionalan kontekst onaj u kojem je dopuštena zamjena istoznačnih termina, a da se ne promjeni istinosna vrijednost cjelokupne rečenice u kojoj se pojavljuju ti termini. Uzmimo kao primjer činjenicu da je Samuel Clemens ista osoba kao i Mark Twain. S obzirom na to možemo reći da je rečenica „Samuel Clemens je napisao Avanture Toma Sawyera" ekstenzionalna zato što ćemo, ako u njoj zamijenimo ime „Samuel Clemens" s imenom „Mark Twain", opet dobiti istinitu rečenicu „Mark Twain je napisao Avanture Toma Sawyera". Ako zamjena istoznačnih termina dovodi do promjene u istinosnoj vrijednosti rečenice, onda govorimo o intenzionalnom kontekstu. Paradigmatski primjer intenzionalnih rečenica predstavljaju one koje uključuju psihološke predikate poput: „Ivica vjeruje da je Mark Twain napisao Avanture Toma Sawyera". Ako u prethodnoj rečenici zamijenimo ime „Mark Twain" sa „Samuel Clemens" moguće je da ćemo dobiti neistinitu rečenicu jer je sasvim uvjerljivo da Ivica ne zna ili ne posjeduje vjerovanje da je Mark Twain ista osoba kao i Samuel Clemens.

Primijenimo li ova razmatranja na prethodni argument, možemo vidjeti da pozivanje na vjerovanja srednjovjekovnih osoba kako bismo zaključili nešto o činjeničnim stvarima, poput toga jesu li mentalna stanja identična stanjima mozga, dovodi do intenzionalne pogreške. Naime, kao što riječ „voda" ne možemo zamijeniti riječi „H_2O" kada srednjovjekovnim ljudima pripisujemo vjerovanja o vodi, tako ne možemo zamijeniti riječ „bol" s „aktivacija C-vlakana" kada srednjovjekovnim ljudima pripisujemo

[41] Pod načinom prezentacije misli se na načine pod kojima poimamo predmete. U suvremenu filozofiju um ova sintagma dolazi iz filozofije jezika koja se temelji na radovima Fregea (1995). Izvorna njemačka sintagma je „Art des Gegebenseins", koja se na engleskom standardno prevodi s „mode of presentation".

vjerovanja o boli. Međutim, iz toga ne slijedi da u ekstenzionalnom smislu bol nije isto što i aktivacija C-vlakana.

Nadalje, ova razmatranja nam govore da Leibnizov princip ne bismo smjeli koristiti unutar intenzionalnog konteksta. Naime, Leibnizov princip se odnosi na svojstva *samih* predmeta, te ako dvije stvari imaju barem jedno različito svojstvo, onda one ne mogu biti ista stvar. Međutim, vjerovanja koja osoba može imati o nekom predmetu ne smiju se tretirati kao dodatna svojstva koja predmet ima. Vjerovanja predstavljaju naš način spoznaje svijeta. Drugim riječima, ona predstavljaju aspekte pod kojima spoznajemo svijet. Kao što znamo, isti svijet možemo spoznati pod različitim aspektima. Zdravorazumski, vodu karakteriziramo kao transparentnu tekućinu bez posebnog okusa i mirisa. No, znanost nam daje drukčiju perspektivu na vodu te nam obznanjuje da voda nije ništa drugo nego H_2O.

Razmotrimo sada sljedeći argument koji se također temelji na Leibnizovom principu, no nije ovisan o spoznajnim stanjima ljudi. Ovaj argument možemo nazvati argument iz prostorne lokacije.

1) Osjećam bol u stopalu.
2) Nijedno stanje mozga nije u stopalu.

Dakle:

3) Moja bol nije neko stanje mozga.

Ovaj argument ne pokazuje da je dualizam točna teorija, već da teorija identiteta kako je se standardno shvaća (da su mentalna stanja identična stanjima mozga) ne može biti točna. To je vrsta prigovora koju bi mogli uputiti biheviorist Ryleova tipa.

Zastupnici teorije identiteta bi ovdje mogli odgovoriti na barem dva načina. Jedan je da se modificira njihova teza te da se kaže da neka mentalna stanja nisu ograničena na procese u mozgu. Drugi način je da se pokaže što ne valja s ovakvom vrstom argumenta.

Što se tiče prvog odgovora, netko bi mogao zastupati tezu da nije u suprotnosti s teorijom identiteta tvrditi da je bol u stopalu. Naime, budući da se tamo nalaze završeci živaca iz perifernog živčanog sustava koju šalju signale kroz leđnu moždinu u mozak, ti živci mogu igrati ulogu koju povezujemo s tom konkretnom vrstom boli. U tom smislu, moglo bi se reći da doista neka mentalna stanja nisu u mozgu. Međutim, nema razloga ograničiti teoriju identiteta tipova samo na usku tezu da su mentalna stanja identična stanjima centralnog živčanog sustava već ima smisla tvrditi da ona čine međusobno povezanu strukturu koja, između ostalog, uključuje i periferni živčani sustav (vidi Aranyosi 2013).

Međutim, ova vrsta argumenta se lako može modificirati kako bi se dodatno stavio pritisak na uvjerljivost teorije identiteta tipova. Uzmimo u obzir sljedeći argument:

4) Imam narančastu naknadnu sliku[42].
5) Stanja mozga (uključujući periferni živčani sustav) nisu narančasta.

Stoga:

6) Naknadna slika koju vidim nije neko stanje mozga (ili perifernog živčanog sustava).

Ili

7) Bol koju osjećam je oštra.
8) Stanja mozga (ili perifernog živčanog sustava) nisu oštra.

Dakle,

9) Bol nije isto što i stanje mozga (ili perifernog živčanog sustava).

Ovdje postaje važna druga mogućnost odgovora. Place (1956) ukazuje da mnogi autori koji se pozivaju na ovakvu vrstu argumenata zapravo čine pogrešku koju je nazvao *fenomenološka pogreška*. Pogreška se sastoji u tvrdnji da iz iskustva ili doživljaja nečega s određenim obilježjima slijedi da postoji neka stvar s tim obilježjima. Iz iskustva narančaste naknadne slike ne slijedi da postoji neka nesvodivo mentalna stvar koja je narančaste boje. Iz iskustva oštre boli ne slijedi da postoji neka nesvodiva mentalna stvar koja je bol i ima svojstvo oštrine.

Nasuprot tome, rečenice o svjesnim doživljajima trebale bi se analizirati na način da pripisuju osobi određeni tip *iskustva*. Stoga bi se rečenice „Imam narančastu naknadnu sliku“ i „Osjećam oštru bol“ trebale razumjeti kao „Imam *iskustvo* ili *doživljaj* narančaste naknadne slike“ i „Imam *iskustvo* ili *doživljaj* oštre boli“. Ovi iskazi ne impliciraju da postoje svojevrsni mentalni predmeti, poput naknadne slike ili oštre boli, već se sugerira da postoje iskustva tih stvari koja, naravno nisu oštra ili narančasta te kao takva mogu biti identična s nekim fizičkim procesima u tijelu ili mozgu. Ovaj odgovor postaje još uvjerljiviji kada razmišljamo o postojanju fantomske boli. Neke osobe koje su ostale bez ruke ili noge i dalje imaju osjećaj da ih boli mjesto gdje im je nekad bila ruka ili noga. Na primjer, osoba može osjećati da je boli ruka koja joj je amputirana. Taj osjećaj boli može biti vrlo ozbiljan i nije uvijek jasno kako ga liječiti. No, jasno je u takvim slučajevima da nema boli na mjestu koje boli. Ovdje je uvjerljivo reći da osoba ima iskustvo boli. A iskustvo boli može biti neko stanje mozga.

4.6 Nesvodivost mentalnih svojstava

Na temelju Leibnizova principa mogao bi se dati suptilniji prigovor teoriji identiteta tipova. Sljedeći prigovor koji ćemo razmotriti temelji se na načinu

[42] Jedna od najuobičajenijih naknadnih slika je svijetla slika koja se čini da pluta ispred očiju nakon što se osoba zagleda u upaljenu žarulju na nekoliko sekundi.

na koji identificiramo stvari za koje tvrdimo da su identične (vidi Smart 1959, 148, prigovor 3). Rekli smo da nam prema teoretičarima identiteta jedino aposteriorna istraživanja mogu otkriti s kojim su fizičkim stanjima identični pojedini tipovi mentalnih stanja. Identitet boli i C-vlakana ne može se otkriti pojmovnom analizom termina „bol“ i „aktivacija C-vlakana“, ništa više nego što se može otkriti da pojam voda referira na isto što i pojam H_2O. Iz toga slijedi da, ako ćemo identificirati tip mentalnog stanja s nekim tipom fizičkog stanja, moraju postojati neovisni kriteriji pomoću kojih ćemo znati kako primijeniti termine kojima referiramo na mentalna stanja i termine pomoću kojih referiramo na fizička stanja. Štoviše, kada ne bismo imali posebne kriterije za primjenu termina koji referiraju na iste stvari onda bi njihov identitet bio spoznatljiv *a priori*. Bio bi sličan primjerima poput „Momak je neoženjeni muškarac“, gdje kriteriji za primjenu termina „momak“ nisu neovisni o kriterijima za primjenu termina „neoženjeni muškarac“.

U slučaju znanstvenih identifikacija, čini se da imamo neovisne kriterije za primjenu termina koji referiraju na iste stvari. Uzmimo za primjer identitet vode i H_2O. Prije nego smo saznali da voda = H_2O, znali smo kriterije za primjenu termina „voda“ i „H_2O“. Voda je nešto što je transparentno te nema okusa ni mirisa. H_2O je molekula koja se sastoji od dva atoma vodika i jednog atoma kisika. Štoviše, upravo poznavanje neovisnih kriterija za primjenu termina „voda“ i „H_2O“ nam je omogućilo da istraživanjem otkrijemo da oni referiraju na istu stvar. Prema toj analogiji, da bismo otkrili da je bol isto što i aktivacija C-vlakana morali smo znati neovisne kriterije za primjenu termina „bol“ i „aktivacija C-vlakana“. To nas dovodi do sljedećeg problema.

Čini se da moramo pretpostaviti da postoje fenomenalna ili pojavna svojstva za koja nije jasno da su fizička. Kako bismo prepoznali, na primjer, bol i razlikovali je od nekog drugog mentalnog stanja moramo se pozvati na to kako je imati iskustvo boli, tj. način na koji osjećamo ili doživljavamo bol. Čini se da to mora biti neko fenomenološko svojstvo iskustva boli jer u suprotnom ne bismo zadovoljili kriterij neovisne identifikacije. Na primjer, kada bismo bol mogli identificirati oslanjanjem na neurofiziološke kriterije, poput ideje da je bol aktivacija C-vlakana, onda naše otkriće da bol zaista jest aktivacija C-vlakana ne bi bilo empirijsko, već bi slijedilo iz apriornog znanja o kriterijima za primjenu termina. Budući da kriteriji za identifikaciju mentalnih stanja referiraju na fenomenološka svojstva (poput bolnosti boli), dok neurofiziološka stanja ne referiraju na takva svojstva, može se tvrditi da mentalna stanja imaju neka svojstva koja se ne mogu zahvatiti unutar objektivističke slike koju podrazumijeva fizikalizam. Drugim riječima, identitet tipova nije istinit jer postoje mentalna svojstva koja se ne mogu identificirati s fizičkim svojstvima.

Ovaj argument bi se mogao formulirati na sljedeći način:

1) Neka mentalna stanja razlikujemo i prepoznajemo

direktno pomoću njihovih fenomenalnih ili pojavnih svojstava. Na primjer, bol prepoznajemo preko bolnosti koju osjećamo, iskustvo boje pomoću doživljaja kako je to vidjeti boju, a slično vrijedi i za ostala iskustva.

2) Fizikalna stanja mozga ne razlikujemo niti prepoznajemo preko fenomenalnih ili pojavnih svojstava. Fizikalna stanja mozga razlikujemo i prepoznajemo pomoću njihovih neurofizioloških svojstava.

Stoga:

3) Određeni tipovi mentalnih stanja ne mogu se poistovjetiti s tipovima stanja mozga.

Ovim se argumentom nastoji pokazati da se, čak i ako možemo tvrditi da su mentalna stanja kao entiteti identični fizičkim stanjima, svejedno čini da *svojstva* mentalnih stanja nisu identična *svojstvima* fizičkih stanja mozga. Ako je ovaj argument uvjerljiv, onda bi se mogla braniti slabija verzija dualizma koja se naziva dualizam svojstava. To je gledište prema kojemu postoji jedna vrsta supstancije (ona fizička), koja može posjedovati fundamentalno različita svojstva (mentalna i fizička). Čovjek i druge životinje koje posjeduju svjesna iskustva bi onda bile takva vrsta bića koja može, uz fizička, imati i mentalna svojstva.[43]

Smart nudi dosta zanimljivo rješenje ovog problema. Nastoji ga riješiti pružanjem analize rečenica kojima govorimo o svjesnim iskustvima koja će biti neutralna u pogledu ontološkog statusa entiteta na koje referira. Drugim riječima, nastoji ponuditi analizu rečenica o svjesnim iskustvima koja neće pretpostavljati da postoje nesvodivo fenomenalna svojstva iskustva. Kako bismo bolje shvatili Smartovo rješenje, razmotrimo sljedeći iskaz:

„Ivan vidi žućkasto-narančastu naknadnu sliku.“

Smart predlaže da se ova vrsta rečenice kojom referiramo na pojavna iskustva analizira na sljedeći način (vidi Smart 1959, 149–50):

(A1) Nešto se događa u Ivanu što je slično onome što se događa kada ima otvorene oči, budan je i ispred njega se stvarno nalazi naranča.

Smart (1959, 150; vidi također 2017), prateći Ryleovu terminologiju, navodi da je niz riječi „Nešto se događa u osobi što je slično onome što se događa kada“ kvazilogičan ili tematski neutralan (engl. *topic-neutral*). U ovom slučaju pod tematskom neutralnošću misli se da značenje tih riječi ne implicira ništa o ontološkom statusu entiteta na koje referiraju. Stoga se

[43] Od suvremenih autora, ovakvu vrstu dualizma zastupa David Chalmers (1996; 2010a). Za više o tome, vidi poglavlje 9.

njegova analiza iskaza o svjesnim iskustvima naziva tematski neutralna analiza. Važno je istaknuti da, ako na ovaj način analiziramo iskaze o svjesnim doživljajima, onda se ne isključuje mogućnost da su ona identična fizičkim procesima u mozgu. Naročito zato što se ne pretpostavlja da postoje mentalna stanja koja spoznajemo putem njihovih posebnih fenomenalnih ili kvalitativnih svojstava. Ovdje se samo tvrdi da osoba ima iskustvo koje se može opisati s obzirom na njegove tipične uzroke koji se sami mogu opisati u fizikalističkim terminima. Slično gledište je prije Smarta istaknuo i Place. On navodi da

> Kad opisujemo naknadnu sliku kao zelenu […] kažemo da imamo onu vrstu iskustva koju normalno imamo kada […] gledamo zeleno osvjetljenu točku.

Daljnji korak su poduzeli David Lewis (1966) i David Armstrong (1968). Oni su, osim uzročnih podražaja, u analizu iskaza kojima referiramo na svjesna iskustva uveli i tipične *posljedice* koje ta mentalna stanja igraju. Drugim riječima, individuiraju svjesna iskustva pomoću njihovih tipičnih uzroka i ponašajnih posljedica. To im onda omogućuje da ih identificiraju s procesima u mozgu koji igraju te uzročne uloge. Na primjer, rečenica poput „Ivan ima oštru bol u ruci“ mogla bi se analizirati na sljedeći način:

> (A2) Ivan je u unutrašnjem stanju koje se tipično pojavljuje kao uzročna posljedica oštećenja kože na ruci te ima tendenciju uzrokovati podizanje ruke i jaukanje.

(A1) i (A2) uspijevaju specificirati tip svjesnog iskustva, a da se ne spominje bilo kakvo *kvalitativno* svojstvo iskustva. Stoga, ako su ove analize ispravne onda, kada govorimo o fenomenalnim ili fenomenološkim svojstvima, ne impliciramo da ona postoje kao nesvodivo psihička svojstva, već se ostavlja mogućnost da će nam znanstvena istraživanja ukazati na njihovu pravu fizičku prirodu.

4.7 Može li identitet između mentalnih i fizičkih stanja biti kontingentan?

Vidjeli smo da su Place i Smart tvrdili da je identitet između mentalnih i moždanih stanja kontingentan. Tvrdili su da imamo znanstvene i filozofske razloge vjerovati da su u našem aktualnom svijetu mentalna i moždana stanja identična. Međutim, također su smatrali da nije nužno da su mentalna stanja identična s fizičkim stanjima. Naime, čini se da možemo lako zamisliti mogući svijet u kojem je kartezijanski dualizam istinit te da postoje umovi koji nisu nužno vezani za fizička tijela.

Nasuprot tome, Saul Kripke (1971) daje utjecajan argument da identitet između dvije stvari mora biti nužan. Ako stvar postoji, onda je nužno identična samoj sebi. Na temelju tog uvida, Kripke (1997) razvija novu verziju kartezijanskog argumenta prema kojoj mentalna stanja ne mogu biti identična s fizičkim stanjima (za raspravu, vidi Pećnjak i Špiljak 2014). Ova vrsta argumenta bi se mogla formulirati na sljedeći način:

1) Ako su mentalna stanja identična fizičkim stanjima, onda ona moraju biti identična u svim mogućim svjetovima.
2) Postoje mogući svjetovi u kojima mentalna stanja nisu identična fizičkim stanjima.

Dakle:

3) Mentalna stanja nisu identična fizičkim stanjima. (Usp. Berčić 2012, 2:180)

Ideja argumenta je da, ako se prihvati da propozicije koje izražavaju identitet moraju biti nužne, tj. ako su istinite u svakom mogućem svijetu, onda slijedi da mentalna stanja nisu identična fizičkim stanjima ni u našem aktualnom svijetu. Ovaj argument, ako je valjan, pokazuje, ne samo da teorija identiteta tipova nije točna, već i da fizikalizam kao općenita materijalistička pozicija ne može biti istinit.

Iz perspektive izvornih teoretičara tipova, premisa 2) nije sporna. Njome se izražava tvrdnja koju oni prihvaćaju, naime da je je identitet između mentalnih i fizičkih stanja kontingentan. Dakle, ono što predstavlja nevolju za njih je premisa 1). U nastavku ćemo se osvrnuti na argument za koji mnogi smatraju da pokazuje da istiniti iskazi o identitetu predmeta moraju biti nužni.

Jednostavna verzija argumenta za nužnost identiteta ima dvije premise (vidi, npr. Lowe 2002, 85–86).[44] Prva premisa odnosi se na intuitivnu ideju da je svaka stvar nužno identična samoj sebi. Njome se tvrdi da ako je $a = a$, onda a mora biti identičan s a, tj. sa samim sobom. Kada a ne bi bio nužno identičan s a onda bi bilo moguće da a nije a. Što se čini apsurdnim. Na primjer, Samuel Clemens je ista osoba kao i Samuel Clemens te nije mogao biti slučaj da Samuel Clemens nije Samuel Clemens jer bi to značilo da on nije osoba koja jest. Slično vrijedi za tipove stvari poput vode. Iako nam se čini da je znanstveno istraživanje moglo pokazati da molekularni sastav vode nije H_2O, tome zapravo nije tako. Budući da je voda identična s H_2O to je nužna istina. Kada voda ne bi imala tu molekularnu strukturu, onda to ne bi zapravo bila naša voda, nego bi bio neki drugi entitet. Isto vrijedi za bilo koju drugu rečenicu kojom se izražava slična vrsta identiteta. Na primjer, ako vrijedi $a = b$ onda je nužno da je a identično s b. Drugim riječima, a i b su nužno isti predmet. Ili da ponovo ilustriramo s konkretnim primjerom, ako Samuel Clemens = Mark Twain, onda nije moguće da Samuel Clemens nije Mark

[44] Za formalnu derivaciju teze o nužnosti identiteta, vidi Jurjako i Brzović (2021).

Twain. Kada Samuel Clemens ne bi bio Mark Twain, onda Samuel Clemens ne bi bio Samuel Clemens, što smo rekli da nije moguće.

Druga se premisa odnosi na Leibnizov zakon o nerazlučivosti identiteta. Ako su dvije stvari identične onda one imaju sva ista svojstva. Na temelju te dvije premise možemo izvesti zaključak da su istinite propozicije o identitetu nužno istinite. E. J. Lowe daje konciznu formulaciju tog argumenta:

> Pretpostavimo, onda, da je određena propozicija kojom se izražava identitet istinita: recimo, propozicija da je *a* identično s *b*, gdje su *a* i *b* bilo koji predmeti. Sada, prema principu nužnosti samoidentiteta, možemo reći da je *a* nužno identično s *a*. Iz toga slijedi da je istinito za *a* da je nužno identično s *a*. Ali iz toga i pretpostavke da je *a* identično s *b* slijedi, prema Leibnizovu zakonu, da je također istinito za *b* da je nužno identično s *a* — iz čega slijedi da je *a* nužno identično s *b*. Ono što se, onda, čini da smo dokazali jest da ako je *istinito* da je *a* identično s *b*, tada je *nužno istinito* da je *a* identično s *b* — i posljedično da nema propozicije kojom se izražava identitet, a da je samo *kontingentno* istinita (istinita u nekim mogućim svjetovima, ali ne i u drugima). (Lowe 2002, 85)

Ključno za ovaj argument je ideja da se samoidentitet shvaća kao još jedno svojstvo koje predmet može imati (za raspravu, vidi Jurjako i Brzović 2021). Štoviše, ovdje se pretpostavlja da predmeti mogu imati *modalna* svojstva, kao što je biti *nužno samoidentičan*. Na primjer, ako Samuel Clemens ima svojstvo biti *nužno samoidentičan Samuelu Clemensu*, onda je jasno kako možemo izvesti zaključak da Mark Twain ima svojstvo biti *nužno samoidentičan Samuelu Clemensu*. Naime, prema Leibnizovu zakonu slijedi da, ako Samuel Clemens = Mark Twain, onda oni moraju imati sva ista svojstva. Budući da se pretpostavlja da biti samoidentičan Samuelu Clemensu jest svojstvo koje predmet može imati, onda slijedi da štogod je identično tom predmetu ima to svojstvo.

Ako se prihvati da ne postoje kontingentni identiteti, onda postaje jasno kako se može argumentirati protiv teorije identiteta u filozofiji uma. Budući da se čini da možemo zamisliti da svjesna iskustva, poput osjećaja boli, nisu fizički procesi u mozgu, čini se da postoji mogući svijet u kojem je to istina. Ako postoji mogući svijet gdje svjesna iskustva nisu identična s procesima u mozgu, nije nužno da su svjesna iskustva identična s procesima u mozgu. Ako nisu nužno identična, onda nisu uopće identična. Dakle, teorija identiteta je neistinita.

Kako bi zastupnici teorije identiteta mogli odgovoriti na ovaj argument? Čini se da imaju dvije opcije. Mogu tvrditi da je identitet mentalnih i fizičkih stanja ipak nužan. Ili mogu negirati premisu 1) u argumentu te tvrditi da su neki iskazi identiteti ipak kontingenti te da identitet između mentalnih i

moždanih stanja spada u tu kategoriju. Među suvremenim zastupnicima teorije identiteta trend je da se prihvati tvrdnja o nužnosti identiteta (Polger 2009). Prvo ćemo krenuti s razmatranjem te opcije.

Ako se prihvati tvrdnja da je identitet nužan, čini se da je problem riješen. Jednostavno se može reći da, ako je identitet općenito nužan, onda je nužno i da su mentalna stanja identična stanjima mozga. Međutim, za ovakav odgovor javlja se problem jer i dalje ostaje intuicija, koju su dijelili Place i Smart, da je moguće da mentalna stanja nisu fizička. Stoga zastupnici teorije identiteta moraju na neki načni objasniti zašto imamo intuicije da mentalna stanja nisu nužno identična fizičkim stanjima mozga, iako zapravo jesu nužno identična.

Jedan utjecajni odgovor je da koristimo različite pojmove kada referiramo na svjesna mentalna iskustva i kada referiramo na fizička stanja mozga. Na primjer, neki autori smatraju da kada govorimo i razmišljamo o svjesnim mentalnim stanjima, poput bolova, naknadnih slika, iskustva gledanja boje, i tome slično, koristimo fenomenalne pojmove (Malatesti 2012). To bi bili pojmovi kojima referiramo na pojavne aspekte naših iskustava i načine na koje doživljavamo mentalna stanja. Nasuprot tome, kada razmišljamo o fizičkim stvarima, poput mozga i njegovih svojstava, onda koristimo drugu vrstu pojmova, koje možemo nazvati fizikalnim pojmovima. Ovdje je krucijalno to što osoba može imati jednu i drugu vrstu pojmova, a da ne zna da referiraju na istu vrstu stvari. Kao što možemo imati pojmove Zornjača i Večernjača, a da ne znamo da oba referiraju na planet Veneru. U literaturi se odgovori ovog tipa nazivaju strategija fenomenalnih pojmova.

Ova vrsta odgovora na probleme s kojima se fizikalistička gledišta susreću dosta je utjecajna u suvremenoj raspravi te ćemo se njome detaljnije baviti u poglavlju 8. Ovdje ćemo spomenuti da, ako krenemo tim smjerom onda se ponovo susrećemo s ranijim problem. U slučaju Zornjače i Večernjače čini se da uspijevamo referirati na planet Veneru na način da Veneri pripisujemo dva različita svojstva. Kada razmišljamo o Veneri kao Zornjači, onda bi neki rekli da mi zahvaćamo Veneru pod određenim aspektom, a to je prva najsvjetlija točka na jutarnjem nebu. Kada razmišljamo o Veneri kao o Večernjači, zahvaćamo je pod aspektom najsvjetlije točke na večernjem nebu. No, da bismo mogli zahvatiti Veneru pod ta dva aspekta moramo joj pripisati dva različita svojstva, naime biti najsvjetlija točka na jutarnjem nebu i biti najsvjetlija točka na večernjem nebu. Ako ova analogija vrijedi u slučaju mentalnih stanja, onda se opet suočavamo s problem dualizma svojstava. Naime, iskustvena mentalna stanja identificiramo introspektivno prema tome kako ih doživljavamo, dok fizička stanja mozga identificiramo koristeći empirijska istraživanja. Ovdje se opet postavlja pitanje jesu li ta introspektivna svojstva mentalnih stanja ujedno i fizička svojstva mozga? Uspješnost strategije fenomenalnih pojmova, kako smo je ranije definirali,

ovisi o sposobnosti da ponudi odgovor na ovo pitanje koji neće pretpostavljati da govorimo o dvama različitim vrstama svojstava.

Ovaj problem nas opet vraća na Smartovo rješenje u vidu tematsko neutralne analize iskaza o svjesnim iskustvima. Međutim, njegova analiza je osmišljena na pozadini pretpostavke da je identitet između mentalnog i fizičkog kontingentan. Stoga bi se na temelju njegove analize možda ipak mogla sačuvati ideja da je odnos metalnih stanja i njihovih svojstava kontingentno identičan tipovima fizičkih stanja. U nastavku ćemo se osvrnuti na ovu mogućnost.

Prema Smartu (1959) sama analiza rečenica kojima pripisujemo svjesna iskustva nas ne bi smjela obvezati na određenu ontologiju tih stanja. Stoga one ostavljaju otvorenim na što točno referiramo kada govorimo o svjesnim iskustvima. Slična neodređenost bi trebala vrijediti i kod Armstrongove (1968) uzročne analize. Naime, sama uzročna analiza nam neće reći koje fizičke strukture igraju ulogu mentalnih stanja. U tom se smislu ostavlja otvorenim da je bol identična aktivaciji C-vlakana, no ne nužno jer neka druga fizička struktura može igrati ulogu boli kod ljudi i različitih životinja. Ako je tome tako onda možemo reći da je bol identična nekim fizičkim stanjima, no ne opredjeljujemo se kojima te se ostavlja otvorenim da su kod različitih bića koja mogu doživljavati bol realizirana kroz različita fizička stanja. Štoviše, mogli bismo reći da je zbog te vrste ontološke „otvorenosti“ bol samo kontingentno identična aktivaciji C-vlakana.

Ovakva tvrdnja nije *ad hoc* iz perspektive autora koji prihvaćaju da je identitet nužan (vidi Jurjako i Brzović 2021). Naime, čak i ako se prihvati tvrdnja o nužnosti iskaza o identičnosti predmeta, svejedno postoje iskazi kojima možemo izraziti sudove o identitetu koji bi bili samo kontingentno istiniti. U tu vrstu sudova spadaju tvrdnje identiteta u kojima govorimo o stvarima putem opisa. Na primjer, iskaz „Trenutni predsjednik Republike Hrvatske = Zoran Milanović“ kontingentno je istinit. Lako možemo zamisliti svijet u kojem trenutni predsjednik Hrvatske nije Zoran Milanović, nego neka druga osoba. Kako bi razlikovao ovu vrstu sudova od onih kojima izražavamo nužne identitete, Kripke (1997) uvodi razlikovanje između krutih i nekrutih označitelja. Kruti označitelji su one riječi koje referiraju na isti predmet u svim mogućim svjetovima. Ključno za našu raspravu, u krute označitelje spadaju imena stvari i vrsta, poput „Samuel Clemens“, „mačka“, „pas“, „voda“, „munja“, „toplina“ i tome slično. U nekrute ili smekšane označitelje spadaju izrazi koji ne referiraju nužno na isti predmet u svim mogućim svjetovima. Na primjer, nekruti označitelj je izraz „trenutni predsjednik Republike Hrvatske“ ili „najveći čovjek u prostoriji“ jer se njihova referencija može mijenjati kroz različite okolnosti.

Ako teoretičari identiteta žele obraniti tvrdnju da je identitet između mentalnih i fizičkih stanja kontingentan, onda izgleda da moraju tvrditi da riječi kojima referiramo na mentalna stanja predstavljaju nekrute

označitelje. Čini se da je to u skladu s tematsko neutralnom analizom iskaza o mentalnim stanjima koju prihvaća Smart i uzročnom analizom kakvu zastupa Armstrong. Međutim, Kripke (1997, 120–21) daje protuargument takvom shvaćanju termina kojima referiramo na mentalna stanja. Zadržimo se na izrazu „bol". Kada bi bol bila kontingentno identična aktivaciji C-vlakana, to bi značilo da je značenje riječi „bol" određeno opisom čija se referencija može mijenjati kroz moguće svjetove. Na primjer, putem uzročne analize riječi „bol", mogli bismo zaključiti da se radi o stanju koje je uzrokovano tim i tim podražajima te ima kao posljedicu ta i ta ponašanja.

Međutim, ovakvoj analizi riječi „bol" moglo bi se prigovoriti da ispušta nešto što je esencijalno za bol, a to su njezina fenomenalna svojstva. Drugim riječima, čini se da tematsko neutralna i uzročna analiza iskustva ispuštaju esencijalna svojstva svjesnih mentalnih stanja, a to je način na koji ih doživljavamo. Bol nije samo nešto što igra određene uzročne uloge, već je to iskustveno stanje koje identificiramo i razlikujemo od drugih iskustvenih stanja prema tome kako ih doživljavamo. Kada bi postojalo fizičko stanje koje zadovoljava sve uzročne uloge koje povezujemo s boli, a da osoba koja se nalazi u tom stanju ne osjeća bol, onda to jednostavno ne bi bila bol. Stoga se čini da riječ „bol" mora biti kruti označitelj koji u svim mogućim svjetovima referira na iskustva s određenim fenomenalnim svojstvima te kao takva ne može biti kontingentno identična s nekim fizičkim stanjima.

Opcija koja nam ostaje za razmotriti jest da se tvrdi da zapravo Kripkeova doktrina o nužnosti identiteta nije točna. Ako bismo mogli pokazati da se čak i kada govorimo o krutim označiteljima ne implicira da njihov identitet mora biti nužan, onda bismo mogli tvrditi da nema prepreka što se tiče prihvaćanja teze da su mentalna i fizička stanja kontingentno identična. Osporavanje doktrine o nužnosti identiteta nije bez presedana. Utjecajno gledište da identitet može biti kontingentan dao je Allan Gibbard (1975; usp. Lewis 1971; Noonan 1993; Jurjako i Brzović 2021). U novije doba kontingentnost identiteta brani i Hanoch Ben-Yami (2018). Bez ulaženja u detalje različitih kritika, u nastavku ćemo se osvrnuti na neke sporne pretpostavke argumenta da identitet mora biti nužan.

Ključan korak u dokazu da identitet mora biti nužan temelji se na primjeni Leibnizova zakona. Poznato je da Leibnizov zakon ne vrijedi u svim kontekstima. Vidjeli smo da ne vrijedi u kontekstima u kojima govorimo o vjerovanjima osobe. Ivica može vjerovati da Superman leti, a da ne vjeruje da Clark Kent može letjeti. Stoga, ako se može pokazati da je korištenje Leibnizova zakona nedozvoljeno u modalnim kontekstima (gdje izražavamo nužne ili kontingente propozicije) imali bismo razloga sumnjati da su svi iskazi identiteta nužni (uključujući one koji koriste krute označitelje). U nastavku ćemo razmotriti jedan razlog zašto možda nije prikladno koristiti Leibnizov zakon u modalnom kontekstu.

Leibnizov zakon odnosi se na svojstva predmeta. Ako su dva predmeta identična, onda imaju sva ista svojstva. Međutim, neće se svi složiti da biti nužno samoidentičan predstavlja svojstvo predmeta koje se čuva u iskazima u kojima zamjenjujemo termine koji referiraju na istu stvar. Gibbard (1975) daje primjer komada gline koji naziva Hrpić i kipa napravljenog od toga komada gline koji naziva Golijat. Ovaj primjer pokazuje nam da trebamo biti skeptični prema ideji da se Leibnizov zakon odnosi na modalna svojstva predmeta zbog sljedećeg razloga: budući da su napravljeni od istog materijala, dijele istu prostornu lokaciju i u drugim pogledima su nerazlučivi, ima smisla reći da je Golijat identičan Hrpiću. Međutim, njihov identitet nije nužan jer ako uništimo Golijata tako da ga pretvorimo u kuglu, Hrpić svejedno može nastaviti postojati kao okrugli komad gline. Stoga se čini da imamo primjer kontingentnog identiteta. Štoviše, čini se da imamo primjer gdje dva kruta označitelja (imena „Hrpić“ i „Golijat“) ne referiraju na istu stvar u svim mogućim svjetovima. Ako je ovo dobar primjer kontingentnog identiteta, onda to znači da se svojstvo biti *nužno samoidentičan Hrpiću* ne može putem Leibnizova zakona prenositi na stvari kojima je on identičan. Kao što smo vidjeli, iako se možemo složiti da je Hrpić nužno identičan samom sebi, on zasigurno nije nužno identičan Golijatu. Stoga se Leibnizov zakon ne može bez ograničenja primjenjivati u kontekstima gdje se predmetima pripisuju modalna svojstva.[45]

Da zaključimo ovaj dio rasprave, ako su kontingentni identiteti mogući, onda ništa ne sprečava teoretičare identiteta u filozofiji uma da tvrde da iskazi poput „Bol = aktivacija C-vlakana“ spadaju u takve iskaze. Naravno, neće se svi složiti da primjer Hrpića i Golijata predstavlja pravi primjer kontingentnog identiteta te će nastojati braniti Kripkeovu tvrdnju o nužnosti identiteta. Međutim, ovdje nam nije cilj raspraviti te argumente, već ukazati na moguće opcije koje zastupnici teorije identiteta tipova imaju za obranu svojih tvrdnji.

4.8 Prigovor višestruke realizacije

Unatoč tome što u suvremenim debatama u filozofiji uma dominiraju fizikalistička gledišta, mnogi autori smatraju da fizikalizam u smislu teorije identiteta tipova nije točna teorija o prirodi mentalnih stanja. Povijesno gledano, najznačajniji argument koji je doveo do napuštanja teorije identiteta tipova jest argument iz višestruke realizacije ili ostvarljivosti. Tim argumentom nastoji se pokazati da određeni *tip* mentalnog stanja može biti realiziran ili ostvarljiv u više različitih *tipova* fizičkih stanja. Već smo ranije

[45] Štoviše, možemo dodati da se Leibnizov zakon ne može se primjenjivati ni u kontekstima gdje govorimo o dispozicijskim svojstvima jer smo vidjeli da Hrpić može preživjeti pretvaranje u kuglu dok Golijat to ne može. Dakle, Hrpić ima barem jedno dispozicijsko svojstvo koje Golijat nema.

vidjeli da, ako se prihvati uzročna analiza pojmova kojima referiramo na mentalna stanja, onda se otvara mogućnost da tvrdimo da je identitet između mentalnih i fizičkih stanja kontingentan jer u principu različita mentalna stanja mogu zadovoljiti opise zadane uzročnom analizom. Međutim, mnogi smatraju da, ako se priroda mentalnih stanja analizira u terminima njihovih tipičnih uzročnih uloga onda, nasuprot ranim materijalistima poput Armstronga (1968) i Lewisa (1966), teorija identiteta *tipova* ne može biti istinita. Upravo zato što se čini da uzročna analiza dopušta da se *tipovi* mentalnih stanja mogu realizirati kroz različite *tipove* fizičkih stanja.

Hilary Putnam (1995) među prvima je formulirao ovaj utjecajan argument koji se temelji na ideji višestruke realizacije. Ova vrsta argumenta najčešće se temelji na misaonim eksperimentima koji testiraju naše intuicije o tome kako različita bića mogu imati mentalna stanja, a da nemaju istu fizičku konstituciju poput ljudi. U nastavku ćemo razmotriti nekoliko takvih misaonih eksperimenata.

Pretpostavimo da je kod ljudi bol zaista identična aktivaciji C-vlakana. Unatoč tome, možemo zamisliti da postoje životinje koje imaju u potpunosti drugačiji živčani sustav od našeg, no koje svejedno mogu osjećati bol. Pretpostavimo da hobotnice nemaju C-vlakna, tj. da imaju različiti živčani sustav u odnosu na nas. To je uvjerljiva hipoteza jer su ljudi dosta udaljeni na evolucijskom stablu od hobotnica, u smislu da bismo trebali ići dosta daleko u prošlost da pronađemo zajedničkog pretka. Međutim, ima smisla smatrati da hobotnice, unatoč tome što nemaju C-vlakna, osjećaju bol. Ako je tome tako, teorija identiteta tipova nalazi se u problemima. Ako hobotnice osjećaju bol, a nemaju C-vlakna, onda ne može biti slučaj da je bol kao *tip* mentalnog stanja identična aktivaciji C-vlakana kao *tipu* fizičkog stanja. Ako bismo htjeli inzistirati na tome da bol kao tip mentalnog stanja mora biti identična tipu fizičkog stanja koji je aktivacija C-vlakana, onda bi slijedilo da hobotnice ne mogu osjećati bol. Mnogima je to neuvjerljiva tvrdnja.

Razmotrimo drugu hipotetičku situaciju. Zamislimo grupu izvanzemaljaca koji su napravljeni od silicijskih čipova, za razliku od ljudi koji su sastavljeni od tvari na bazi ugljika. Budući da su napravljeni od silicijskih čipova, izvanzemaljci neće imati C-vlakna poput ljudi. Svejedno, možemo zamisliti da oni osjećaju bol kada nabiju nožni prst ili da ih boli grlo od deranja da treba uništiti sve Zemljane. Ako možemo zamisliti ovaj primjer, onda se čini da naš pojam boli ne referira na *tip* stanja koje se može identificirati s točno određenim *tipom* fizičkog stanja. Osjećaj boli kod hipotetičkih izvanzemaljaca će biti realiziran u silicijskim čipovima od kojih su sastavljeni, za razliku od ljudi gdje će biti realizirana u C-vlaknima.

Ovi primjeri ukazuju na to da bol kod različitih organizama može biti identična drugim vrstama fizičkih stvari. Zbog toga mnogi filozofi smatraju da bol može biti *višestruko realizirana*. Budući da bol predstavlja

paradigmatsko mentalno stanje, ovaj se prigovor onda generalizira na ostala mentalna stanja te se tvrdi da većina mentalnih stanja može biti višestruko realizirana. Iz toga slijedi da tipove mentalnih stanja ne možemo identificirati s tipovima fizičkih stanja.

Zastupnici teorije identiteta tipova na ovakve primjere mogu odgovoriti da su zapravo identiteti povezani s *vrstama* organizama. U tom smislu može se tvrditi da je bol kod ljudi identična tipu stanja koje uključuje aktivaciju C-vlakana, dok je bol kod hobotnica identična nekom drugom tipu fizičkih stanja. Možemo ga nazvati aktivacija D-vlakana. No, da bismo razlikovali vrste boli kod ljudi i hobotnica mogli bismo bol kod ljudi označiti s Bol_c, a kod hobotnica s Bol_d. Slično tome, možemo reći za hipotetičke izvanzemaljce, da je kod njih bol identična trećem tipu fizičkih stanja koje možemo nazvati aktivacija u silicijskim E-čipovima te da stoga oni mogu osjećati treći tip stanja boli koji možemo nazvati Bol_e.[46]

Protivnici teorije identiteta tipova odgovaraju teorijskim i znanstvenim istraživanjima koja ukazuju da tipove mentalnih stanja i procesa ne možemo poistovjetiti s točno određenim tipovima neuroloških struktura. Ravenscroft (2005) daje primjer višestruke realizacije vjerovanja kod različitih osoba. Na primjer, Ivica vjeruje da su tijekom prve korona-krize u Hrvatskoj uveli zabranu odlaska na otoke osim u slučaju da osoba ima prebivalište na otoku. Marica, budući da prati vijesti, ima isto vjerovanje. Dakle, Ivica i Marica imaju u svojim u glavama primjerke tipa mentalnog stanja koje možemo opisati kao vjerovanje sa sadržajem da je u Hrvatskoj zabranjeno odlaziti na otoke osim ako imate prebivalište na njima. Međutim, sasvim je moguće da je točan način na koji je to vjerovanje pohranjeno u Ivičinoj glavi malo drugačiji od načina na koji je ono pohranjeno u Maričinoj glavi. Unatoč tome što imaju fizički slične mozgove, nije vjerojatno da im mozak pohranjuje informacije na istim mjestima. Kako se točno informacije pohranjuju u mozgu vjerojatno ovisi o drugim informacijama koje su Ivica i Marica već usvojili, te je vrlo vjerojatno da su njihovi mozgovi usvojili različite informacije s obzirom na različita životna iskustva. Kao analogiju, Ravenscroft navodi kompjuterski *hard drive*. Točan obrazac pohrane na hard driveu ovisi o informacijama koje su već ranije pohranjene. Nova informacija se pohranjuje na dijelovima diska koji su slobodni. Posljedično tome, čak i ako imamo kopije istog dokumenta pohranjene na našim kompjuterima, nije vjerojatno da će dokument biti pohranjen na točno isti način i na istoj lokaciji u različitim kompjuterima.

Drugi autori koriste znanstveno utemeljene primjere koji ukazuju na to da zbog plastičnosti mozga nije realno očekivati da će tipovi mentalnih stanja biti identični tipovima fizičkih stanja (vidi, na primjer, Figdor 2010; Barrett 2013). David Barrett (2013, odjeljak 2.1) daje primjer studije u kojoj su

[46] Ovakvkom vrstom odgovora ćemo se detaljnije baviti u poglavlju 6.

koristili štakore kojima su odstranjeni dijelovi obje hemisfere frontalnog režnja. Poznato je da je frontalni režanj, kako kod ljudi tako i kod štakora, zadužen, između ostalog, za planiranje radnji, kalkulaciju posljedica radnji, usporedbu posljedica s obzirom na ciljeve, odabir i inhibiciju radnji. U studiji je korišteno više grupa štakora kojima su u različitim fazama i vremenskim periodima odstranjivani dijelovi frontalnog korteksa. Na primjer, u jednoj grupi najprije su odstranili jednu hemisferu frontalnog korteksa te drugu nakon 10 dana, u drugoj grupi su drugu hemisferu odstranili nakon 20 dana, u trećoj grupi nakon 30 dana, dok su u četvrtoj grupi obje hemisfere istodobno odstranili. Nakon toga su testirali koliko će vremena pojedinim grupama trebati da uspješno nauče riješiti zadatak prostorne alternacije, gdje je cilj prvo naučiti proći kroz labirint te naći nagradu na jednom kraju te u sljedećem krugu naučiti da je nagrada premještena na drugi kraj labirinata. Barrett (2013, 329) navodi da grupe kojima su dijelovi frontalnog režnja odstranjeni u razmacima od 20 i 30 dana nisu pokazivale značajne razlike u učenju u odnosu na kontrolne štakore koji nisu imali lezije na frontalnom korteksu. Detalji studije trenutno nisu bitni. Ono što je važno za našu raspravu jest da nam studija daje razloga smatrati da su barem neke psihološke funkcije (vezane za planiranje i inhibiciju radnji) višestruko realizirane. Budući da su štakorima odstranjeni dijelovi frontalnog režnja, mora biti da su drugi dijelovi mozga preuzeli funkcije povezane s planiranjem i inhibicijom ponašanja.

Čak i ako ovi primjeri uspješno pokazuju stvarnu mogućnost višestruke realizacije mentalnih stanja, svejedno možemo primijetiti da se većina ovih novijih argumenata temelji na primjerima mentalnih stanja poput vjerovanja i kognitivnih sposobnosti za učenje i zaključivanje. Nije jasno da se isti argumenti mogu dati za višestruku realizaciju osjetilnih stanja poput boli. Moguće je, stoga, da će teoretičari identiteta koji fokusiraju svoje teze na osjetilna stanja imati lakši posao pri obrani svoje verzije teorije. Ovdje se nećemo zadržavati na pregledu novijih rasprava koje u mnogim aspektima pokazuju na suptilnosti koje ostaju previđene u standardnim tumačenjima argumenta višestruke realizacije.[47] Trenutno nam je važno istaknuti da je argument višestruke realizacije motivirao potragu za drugim materijalističkim gledištima koja mogu izbjeći probleme s kojima se susreće teorija identiteta tipova.

4.9 Zaključna razmatranja

Mogućnost realizacije tipova mentalnih stanja u različitim tipovima fizičkih stanja navela je mnoge filozofe na mišljenje da fizikalizam tipova koji se

[47] Za detaljniji i kritički pregled rasprave o argumentu višestruke realizacije, vidi poglavlje 6.

nalazi u pozadini klasične teorije identiteta nije uvjerljivo gledište te da, ako hoćemo zadržati fizikalistička stajališta u filozofiji uma, moramo prihvatiti ograničeniju verziju fizikalizma. Stoga se nakon izvornih argumenata iz všestruke realizacije počela razvijati slabija fizikalistička teorija kojom se tvrdi da postoji identitet primjeraka, tj. da je svaki primjerak mentalnog stanja identičan nekom primjerku fizičkog stanja, no ne prihvaća se jača tvrdnja o identitetu tipova mentalnih i fizičkih stanja. U sljedećem poglavlju bavit ćemo se jednim takvim gledištem. Da budemo precizniji, bavit ćemo se funkcionalizmom, gledištem koje predstavlja jedno od najutjecajnijih pozicija u filozofiji uma kojim se nastoji pomiriti fizikalizam s neredukcionističkim intuicijama koje se nalaze u pozadini argumenta iz višestruke realizacije.

5 Funkcionalizam

5.1 Uvod

U prethodnom poglavlju vidjeli smo da je, povijesno gledano, ono što je dovelo do pada utjecaja materijalizma u obliku teorije identiteta tipova mogućnost višestruke realizacije mentalnih stanja. U kontekstu metafizičke debate o odnosu uma i tijela, veliki utjecaj u tom pogledu imao je Hilary Putnam kojeg neki nazivaju ocem funkcionalizma (Hill 1991, 45, fusnota 1). Putnam je u nizu radova isticao da ono što je specifično za prirodu uma nije materijalna realizacija već njegova funkcionalna organizacija (vidi Putnam 1995; 1975a; 1975c; 1975d; 1975f). Prema toj ideji, mentalna stanja su funkcionalna stanja cijelog organizma te upravo ta činjenica objašnjava kako se mentalna stanja i njihova svojstva mogu realizirati kroz različite fizičke supstrate. Vidjet ćemo da je funkcionalizam općenito neutralan u pogledu prirode uma. U principu, funkcionalizam ostavlja otvorenim je li funkcionalna organizacija koja definira prirodu različitih mentalnih stanja realizirana u materijalnim predmetima ili u nefizičkim supstancijama kakve pretpostavlja kartezijanski dualizam. Međutim, unatoč tome, funkcionalizam se obično svrstava u fizikalističke teorije te se često prikazuje kao okvir koji daje filozofske temelje istraživačkim programima unutar psihologije i kognitivnih znanosti te ih legitimira kao autonomne discipline u odnosu na bazičnije znanosti koje se bave biokemijskim temeljima naših mentalnih sposobnosti (Fodor 1974).

U nastavku poglavlja, prvo ćemo objasniti općenitu ideju što bi značilo da su mentalna stanja funkcionalna stanja nekog organizma. Nakon toga, razmotrit ćemo nekoliko načina na koje se prema funkcionalistima mogu odrediti uzročne uloge mentalnih stanja. Također ćemo razmotriti vezu između teorije identiteta tipova i funkcionalizma. U nastavku ćemo se osvrnuti na filozofsku metodologiju koju koristi Putnam i njegovo inzistiranje da se priroda uma može odrediti u odnosu na analogiju s Turingovim strojevima. Vidjet ćemo koja su ograničenja usporedbe uma s Turingovim strojevima razmatrajući poznati argument kineske sobe protiv strojnog funkcionalizma. U zadnjem dijelu poglavlja, bavit ćemo se prigovorima iz odsutnih i obrnutih *qualia* kojima se nastoje pokazati opća ograničenja

funkcionalizma kada govorimo o objašnjenju subjektivnih aspekata iskustvenih doživljaja.

5.2 Mentalna stanja kao funkcionalna stanja

Središnja ideja funkcionalizma je da je svako mentalno stanje definirano pomoću funkcije koju izvršava u organizmu (ili nekom drugom izomorfnom sustavu). Tipično se uzima da je funkcija mentalnog stanja specificirana uvjetima koji određuju *uzročne relacije* koje mentalna stanja imaju prema (a) podražajima (b) drugim mentalnim stanjima i (c) ponašajnim reakcijama. Kao tipičan primjer funkcionalnog karakteriziranja nekog tipa mentalnog stanja možemo uzeti bol. Bol može biti funkcionalno definirana kao mentalno stanje koje je uzrokovano oštećenjem tkiva, koje uzrokuje želju da se izbjegava takav podražaj te uzrokuje odmicanje ili povlačenje od štetnog podražaja. Na primjer, ako dodirnemo zagrijanu ploču štednjaka doći će do oštećenja tkiva na ruci, koje uzrokuje želju da odmaknemo ruku od štednjaka te u konačnici ovaj uzročni niz tipično završava odmicanjem ruke od štednjaka.

Funkcionalisti smatraju da se na sličan način mogu odrediti i sva ostala mentalna stanja. Na primjer, intencionalna stanja poput vjerovanja da je nešto slučaj mogu se funkcionalistički definirati na sljedeći način. Pretpostavimo da Ivica vjeruje da je opasan pas u blizini. Ulazni signal za to vjerovanje može biti to da Ivica čuje lavež pasa, vidi psa kako reži ili da mu je pouzdani prijatelj rekao da se u blizini nalazi opasan pas. Ovakvo vjerovanje tipično će stajati u određenim vezama s drugim mentalnim stanjima i ponašanjem. Na primjer, ovo vjerovanje će uzrokovati strah kod Ivice. Nadalje, povezano s vjerovanjem da se u blizini nalazi opasan pas, Ivica će se htjeti udaljiti od takve situacije. S obzirom na tu želju i vjerovanje da to može postići trčanjem često će dovesti do toga da Ivica pokuša pobjeći od opasnog psa. Međutim, u slučaju da Ivica ima neko drugo vjerovanje, recimo da vjeruje da treba stajati mirno jer je to najbolji način da ga pas ne ugrize, onda vjerovanje da je u blizini opasan pas može dovesti do toga da Ivica ostane mirno stajati na mjestu. U svakom slučaju, uloga vjerovanja kao i drugih mentalnih stanja definira se prema tipičnim inputima koje proizvode vjerovanja, njihovim vezama s drugim mentalnim stanjima i tipičnim outputima koji mogu dovesti do određenih ponašanja.

Ono što je izrazito važno kod funkcionalizma jest inzistiranje na neovisnosti funkcionalnog određenja mentalnih stanja o njihovoj materijalnoj realizaciji. Osnovna ideja jest da bilo koji sustav koji može zadovoljiti funkcionalnu karakterizaciju boli (ili nekog drugog mentalnog stanja) osjeća bol. Kako bismo još približili ovu ideju poslužit ćemo se drugim primjerima gdje je jasno da materijalna realizacija ne određuje prirodu stvari o kojoj govorimo.

Funkcionalisti često povlače analogije s drugim stvarima čija je priroda određena funkcionalno, a ne u odnosu na materijal koji ih realizira. Tipičan primjer koji se koristi je rasplinjač ili karburator. Automobili koji koriste klasična goriva imaju rasplinjač. Njegova uloga je da kombinira benzin i zrak te ubrizgava njihovu mješavinu u motor. U starim automobilima rasplinjač je bio napravljen od mesinga. To je legura bakra i cinka. U novijim automobilima rasplinjač je napravljen od sofisticiranijih legura. U još novijim automobilima moguće je da će rasplinjač biti napravljen od plastičnih dijelova. Svrha primjera je ukazati na to da nije bitno od čega je napravljen rasplinjač sve dok može kombinirati benzin i zrak te davati tu smjesu motoru. Dakle, rasplinjač se može višestruko realizirati. U tom smislu, rasplinjač je definiran funkcionalno prema ulozi koju igra u motoru od auta, a ne prema materijalu od kojeg je napravljen.

Ravenscroft (2005, 51) daje još jedan zanimljiv primjer. Antibiotik je tvar koja je određena ulogom koju igra u liječenju. Njegova uloga je da ubija bakterije koje uzrokuju bolesti, a da se ne nanosi šteta pacijentu. Primjer antibiotika je penicilin koji ubija bakterije koje uzrokuju bolesti bez nanošenja štete pacijentu. Još jedan primjer antibiotika je eritromicin koji također ubija bakterije koje uzrokuju bolesti bez nanošenja štete pacijentu. Ono što je važno u ovom kontekstu jest činjenica da penicilin i eritromicin imaju različite kemijske strukture, međutim svejedno igraju iste funkcionalne uloge u liječenju bolesti. To nam pokazuje da se antibiotik, kao vrsta stvari, može višestruko realizirati.

Upravo činjenica da funkcionalizam omogućuje definiranje mentalnih stanja i procesa neovisno o njihovim materijalnim realizacijama za mnoge funkcionaliste pokazuje kako psihologija može biti autonomna znanost u odnosu na neuroznanost i ostale znanosti koje se bave fizičkom realizacijom mentalnih stanja i procesa (Fodor 2001). Prema tom gledištu psihologija i kognitivne znanosti bi se bavile objašnjenjem uzročnih zakona i regularnosti koje upravljaju mentalnim pojavama na funkcionalnoj razini, dok bi, recimo, neuroznanost i ostale znanosti koje se bave mozgom objašnjavale procese koji implementiraju te zakonitosti na fizičkoj razini (za raspravu, vidi Weiskopf i Adams 2015 pogl. 2).

5.3 Kako se određuju funkcionalne uloge mentalnih stanja?

Vidjeli smo da je prema funkcionalizmu priroda mentalnog stanja određena njegovom uzročnom ulogom. Međutim, možemo se pitati koji su to tipični uzroci i posljedice nekog mentalnog stanja koji određuju prirodu tog tipa mentalnog stanja. Na primjer, zamislimo da Ivica svaki put kada čuje Odu radosti počne osjećati mučninu u trbuhu. Hoćemo li reći da je osjećaj mučnine određeno mentalno stanje koje je, između ostalog, definirano time da se pojavljuje kada Ivica čuje Odu radosti? Intuitivno bismo rekli da slušanje Ode radosti ne definira prirodu osjećaja mučnine, unatoč tome što

u ovom konkretnom slučaju slušanje Ode radosti ima tendenciju *uzrokovati* osjećaj mučnine. Dakle, ono što nas zanima jest kako odrediti uzroke koji su bitni za razlučivanje funkcionalnih uloga od onih koji nisu.

Klasičan odgovor na ovo pitanje dao je David Lewis. Lewis (1972) prilazi ovom pitanju iz perspektive semantike izraza kojima referiramo na mentalna stanja. U tom kontekstu daje proceduru kojom možemo formulirati predikate kojima referiramo na mentalna stanja. Te predikate možemo nazvati funkcionalnim predikatima.

Definicija funkcionalnog predikata može se odrediti na sljedeći način. U prvom koraku formuliramo psihološku teoriju koja sadrži sve rečenice koje govore o uzročnim relacijama koje neko mentalno svojstvo ima s drugim mentalnim svojstvima, podražajima i ponašanjima. Funkcionalisti se međusobno razlikuju u pogledu shvaćanja ove teorije (vidi Block 1980b). Analitički funkcionalisti nastavljaju metodološku tradiciju analitičkog biheviorizma te pretpostavljaju da se uzročne uloge mentalnih stanja mogu odrediti apriornim istraživanjem naših svakodnevnih pojmova koji se odnose na mentalna stanja. Jednu od najjasnijih formulacija tog gledišta dao je sam Lewis (1972). U tom pogledu, navodi da se značenje termina kojima referiramo na mentalna stanja temelji na njihovom zdravorazumskom shvaćanju, tj. na ulozi koju igraju u zdravorazumskoj psihologiji koju koristimo kada objašnjavamo i predviđamo tuđa ponašanja. Pri tome Lewis shvaća zdravorazumsku ili pučku psihologiju kao prototeoriju koju obični ljudi koriste kada interpretiraju, objašnjavaju i predviđaju tuđa pa i svoja ponašanja (za više o zdravorazumskoj psihologiji, vidi Biondić 2017; Jurjako 2020). Zdravorazumska teorija sastoji se od općih mjesta poput onog da ljudi imaju intencionalna mentalna stanja poput vjerovanja, želja i namjera na temelju kojih možemo objašnjavati zašto ljudi rade određene stvari. Znanja koja se temelje na pretpostavci da ljudi posjeduju takva stanja uključuju toliko očite i trivijalne stvari da ih, kada razmišljamo o drugim ljudima, rijetko ekspliciramo. Tu se misli na tvrdnje poput one da će osoba koja želi kupiti kruh, ako nema druge jače želje, otići u obližnji dućan za koji vjeruje da prodaje kruh. Ili ako osoba namjerava upisati medicinski fakultet možemo očekivati da će nastojati postići što bolje ocjene iz biologije i kemije i tome slično.

Kada govorimo o zdravorazumskoj psihologiji kao teoriji koja nam omogućuje da odredimo referencu psiholoških termina, onda Lewis navodi da je možemo shvatiti na sljedeći način:

> Predočite si zdravorazumsku psihologiju kao znanstvenu teoriju koja uvodi nove termine, ali kao teoriju koja je izmišljena mnogo prije nego što je postojalo nešto poput profesionalne znanosti. Skupite sva opća mjesta (engl. *platitudes*) koja vam padnu na pamet koja se tiču relacija mentalnih stanja, senzornih podražaja i motornih reakcija. Možemo razmišljati o njima kao da imaju

sljedeći oblik:

> Kada je netko u takvoj i takvoj kombinaciji mentalnih stanja i primi senzorne podražaje takve i takve vrste, on ima tendenciju da s tom i tom vjerojatnošću time bude uzrokovan da pređe u takva i takva mentalna stanja i da proizvede takve i takve motorne reakcije.

> Dodajte također sva opća mjesta o tome koje mentalno stanje spada pod drugo – »zubobolja je vrsta boli« i slično. […] Uključite samo opća mjesta koja su zajednička svima nama – svi ih znaju, svi znaju da ih svi ostali znaju, i tako dalje. (Lewis 1972, 256; dio prijevoda je preuzet iz Sesardić 1984, 53)

Prema Lewisu, zdravorazumsku psihologiju možemo koristiti kao izvor apriornog znanja o značenju termina kojima referiramo na mentalna stanja. Zdravorazumska psihologija nam daje opća mjesta ili floskule koje vežemo uz pojedina mentalna stanja, što zapravo definira značenje termina kojima referiramo na njih. U citatu, Lewis spominje, na primjer, floskulu ili opće mjesto da je zubobolja vrsta boli. Međutim, možemo spomenuti još neka opća mjesta koja svi povezujemo s mentalnim stanjima.

- Osoba koja doživi teške tjelesne ozljede će osjećati bol.
- Osoba koja naglo osjeti oštru bol će napraviti nekakvu grimasu.
- Bol je vrsta osjetilnog stanja.
- Osoba koja je dosta vremena bez hrane će osjećati glad.
- Osoba koja okusi limun će osjetiti kiselkasti okus.
- Ljuta osoba imat će tendenciju biti nestrpljiva.
- Percepcija uzrokuje vjerovanja.
- Vjerovanja su tip spoznajnih stanja, itd. (vidi P. M. Churchland 1993)

Ovakva vrsta znanja floskula ili općih mjesta nam, prema analitičkim funkcionalistima, u konačnici omogućuje da odredimo funkcionalne uloge koje pojedina stanja igraju u našim mentalnim životima.

S druge strane, psiho-funkcionalizam se temelji na razmatranju metodologije „kognitivnih znanosti“ (vidi Block 1980b). Prema pobornicima ovog gledišta, mentalna stanja i procesi su oni entiteti koji su postulirani u našim najboljim znanstvenim objašnjenjima ljudskog ponašanja. Gledište psihofunkcionalista je da ova objašnjenja postuliraju mentalna stanja u terminima funkcija koje one obavljaju unutar nekog organizma. U tom smislu, za razliku od analitičkih funkcionalista, na njihovu metodologiju možemo gledati kao da uključuje aposteriorno određivanje funkcionalnih uloga koje se temelje na našim najboljim znanstvenim teorijama o prirodi mentalnih stanja i njihovih uloga u objašnjenju ponašanja.

I jedno i drugo gledište imaju svojih prednosti i mana (Pećnjak i Janović 2011). Na primjer, analitički funkcionalisti dobro zahvaćaju ideju da barem inicijalno funkcionalne uloge kojima definiramo mentalna stanja dolaze iz svakodnevnog korištenja vokabulara kojim referiramo na psihološke fenomene. U tom smislu, čini se da i psihofunkcionalisti moraju dopustiti da su analitički funkcionalisti barem donekle u pravu kada govorimo o izvoru definicija funkcionalnih uloga mentalnih stanja. Međutim, problem za analitički funkcionalizam je da suvremena znanost može pokazati da neke od stvari koje podrazumijevamo u zdravorazumskoj psihologiji zapravo ne postoje. Na primjer, u kemiji se nekad smatralo da postoji flogiston, tvar koja se oslobađa tijekom izgaranja. Suvremena kemija je pokazala da tijekom izgaranja ne postoji tvar koja bi se oslobađala poput flogistona, već se nasuprot tome pokazalo da postoji kisik koji se zapravo veže za stvar koja izgara. Slično tome, neki autori tvrde da nam suvremena neuroznanost sugerira da mentalna stanja poput vjerovanja, želja i namjera, kako ih se shvaća u zdravorazumskoj psihologiji, ne postoje (vidi P. M. Churchland 1993). Kako bi se izbjegle ovakve neintuitivne i radikalne revizije naše zdravorazumske psihologije, čini se da bi analitički funkcionalisti trebali biti otvoreni za mogućnost da se značenje termina kojima referiramo na mentalna stanja, a time i njihove funkcionalne uloge, mogu revidirati i ažurirati s obzirom na razvoj empirijskih znanosti o umu te time približiti psihofunkcionalističkom gledištu (vidi Miščević 1990).

U nastavku se nećemo baviti ovim problemom. Koju god opciju odaberemo za izvor funkcionalnih definicija, funkcionalisti moraju odgovoriti na problem cirkularnosti (Pećnjak i Janović 2011). Vidjeli smo da se, prema funkcionalizmu, mentalna stanja definiraju prema odnosima koje imaju prema podražajima, ponašanjima do kojih dovode, ali i prema drugim mentalnim stanjima. Stoga se kao potencijalni problem javlja cirkularnost u definiranju. Taj problem je možda najočitiji kada nastojimo funkcionalno definirati mentalna stanja poput vjerovanja i želja. Vidjeli smo da se funkcionalna uloga vjerovanja može odrediti kao ono stanje koje, kada se poveže s određenom željom, tipično uzrokuje određeno ponašanje. Slično se može definirati i želja. Na primjer, želja da se pojede sladoled može se definirati kao ono stanje koje, kada se spoji s vjerovanjem da se može kupiti sladoled odlaskom do obližnje slastičarne, dovodi do određenog ponašanja. Za rješenje tog problema važan nam je drugi korak u konstrukciji funkcionalnih definicija.

Drugi se korak sastoji od razmatranja Ramseyeve rečenice (engl. *Ramsey sentence*). Kako bismo pojasnili o čemu se tu radi, prvo moramo uzeti u obzir da se teorije općenito mogu shvatiti kao skupovi rečenica. Rečenice se u logičkom smislu sastoje od singularnih termina koji referiraju na predmete i predikate koji referiraju na svojstva predmeta. Teorije koje opisuju kako funkcionara neki aspekt svijeta će nastojati opisati svojstva koja

karakteriziraju svijet. Dakle, u logičkom smislu, kada gledamo formalnu strukturu teorije koja opisuje neki aspekt svijeta, važno nam je zahvatiti predikate koji označuju svojstva svijeta koji opisujemo. Uzmimo u obzir neke predikate $P_1 \ldots P_i \ldots P_n$ koji čine dio teorije T. Izrazom $T(P_1 \ldots P_i \ldots P_n)$ dobivamo rečenicu koja se sastoji od konjunkcije svih rečenica teorije T koja sadrži predikate $P_1 \ldots P_i \ldots P_n$. Ono što nam je ovdje bitno jest da se Ramseyeva rečenica može konstruirati tako da se zamjeni svaki predikat u n-torki $P_1 \ldots P_n$ koji se pojavljuju u $T(P_1 \ldots P_i \ldots P_n)$ s varijablama koje su vezane egzistencijalnim kvantifikatorom. Takvom zamjenom dobivamo sljedeću egzistencijalno kvantificiranu rečenicu koja se naziva Ramseyeva rečenica:

$$\exists X_1 \ldots X_n \, T(X_1 \ldots X_n).$$

Ovdje su predikati $P_1 \ldots P_i \ldots P_n$ kojima referiramo na svojstva mentalnih stanja zamijenjeni egzistencijalno kvantificiranim varijablama $X_1 \ldots X_n$. Rečenica nam kaže da postoje neke stvari koje zadovoljavaju relacije koje su opisane teorijom T. Međutim, važno je uočiti da ne govorimo o kojim se stvarima točno radi, već se samo tvrdi da postoje *neke* stvari koje teorija T točno opisuje.

Ova procedura nam omogućuje da predstavimo generalni okvir načina na koji se daju funkcionalne definicije predikata kojima referiramo na mentalna stanja i njihova svojstva. Ako je X_1 varijabla koja zamjenjuje mentalni predikat M iz psihološke teorije koju razmatramo, onda dobivamo sljedeću funkcionalnu definiciju predikata:

$$M(x) \leftrightarrow \exists X_1 \ldots X_n \, T(X_1 \ldots X_n) \wedge X_i \, (x).$$

Sadržaj ove formule jest da je mentalno stanje M definirano u terminima uzročnih relacija za koje teorija T tvrdi da opisuju svojstva koja izražavamo predikatom X_i i onima koje izražavamo predikatima $X_1 \ldots X_n$.

Ramzificiranjem psihološke teorije izbjegava se problem cirkularnosti jer više ne definiramo pojedine termine kojima referiramo na mentalna stanja koristeći druge termine kojima referiramo na mentalna stanja. Umjesto toga, sva imena (kojima referiramo na mentalna stanja i njihova svojstva) zamjenjujemo egzistencijalno vezanim varijablama. To nam omogućuje da implicitno definiramo značenje mentalnih predikata putem teorije koja opisuje relacije u kojima ta mentalna stanja stoje.

Kako bismo malo konkretnije ilustrirali proces ramzificiranja, razmotrimo sljedeći pojednostavnjeni primjer. Pretpostavimo da M označuje svojstvo biti u stanju boli. Ranije smo vidjeli da stanje boli možemo definirati kroz tipične uzroke i posljedice boli. U tom smislu, možemo reći da naša „teorija“ boli uključuje sljedeće tvrdnje (vidi Kim 2006, 152–54):

Teorija T_b
1. Ako osoba *O* ošteti tkivo i u stanju je normalne budnosti, onda osjeća bol.
2. Ako *O* osjeća bol, onda se ne može koncentrirati na obavljanje redovitih zadataka.
3. Ako osoba *O* osjeća bol, onda jauče i radi grimase.
4. Ako je *O*-u pažnja usmjerena na neki drugi predmet možda neće osjetiti bol zbog oštećenja tkiva.

Našu teoriju T_b možemo ramzificirati tako da uvedemo egzistencijalno vezane varijable umjesto imena kojima referiramo na mentalna stanja. Razmotrimo revidiranu teoriju T_{br} koja nastaje ramzificiranjem teorije T_b:

Teorija T_{br}
Postoje stanja M_1, M_2 i M_3 koja su takva da za svaki *O*, ako *O* ošteti tkivo i nalazi se u M_1, onda je *O* u M_2; ako je *O* u M_2 onda se ne može koncentrirati na obavljanje redovitih zadataka; ako je *O* u M_2 onda jauče i radi grimase; ako je *O* u M_3 onda nije u M_2.

Ramzificiranje nam omogućuje da ne referiramo direktno na mentalna stanja. Umjesto toga govorimo o *nekim* stanjima, koja smo ovdje označili s M_1, M_2 i M_3, koja su povezana međusobno i s određenim ponašanjima na način kako to opisuje teorija T_b. Ovdje je važno to da ramzificirana teorija T_{br} ne sadrži psihološke izraze nego samo fizičke opise kojima opisujemo podražaje i posljedice mentalnih stanja. Drugim riječima, kada uvedemo egzistencijalne kvantifikatore i mentalne termine zamijenimo varijablama postižemo nešto slično sadržajno neutralnim opisima koje smo spominjali u kontekstu Smartova (1959) prijevoda mentalnih iskaza u sadržajno neutralne iskaze (vidi poglavlje 4). Važno za naš kontekst jest da, budući da teorija T_{br} ne sadrži psihološke termine, možemo bez cirkularnosti definirati druge psihološke izraze. Na primjer, predikat *biti u stanju boli* možemo definirati na sljedeći način (modificirano prema Kim 2006, 153–54):

Osoba *O* je u boli = $\exists M_1, \exists M_2, \exists M_3, T(M_1, M_2, M_3) \wedge$ *O* se nalazi u M_2

Na sličan način možemo definirati i druge predikate. Na primjer, biti u stanju normalne budnosti možemo definirati na sljedeći način:

Osoba *O* je u stanju normalne budnosti = $\exists M_1, \exists M_2, \exists M_3\, T(M_1, M_2, M_3) \wedge$ *O* se nalazi u M_1

Prva definicija nam daje pojam boli tako da ga specificira u odnosu na uzročne uloge koje su povezane s oštećenjem tkiva, jaukanjem i grimasama

te nemogućnošću koncentracije. Druga definicija nam daje pojam stanja budnosti koja je opet definirana njezinim uzročnim ulogama koje su određene u odnosu na osjećaj boli, jaukanje i grimase te nedostatak koncentracije. Ramizificiranje nam na ovaj način omogućuje da mentalne termine definiramo u odnosu na međusobne relacije u kojima stoje mentalna stanja, a da se ne pozivamo direktno na ta mentalna stanja.

5.4 Funkcionalizam i fizikalizam

Vidjeli smo da su prema funkcionalizmu mentalna stanja ono što ima ili igra karakteristične uzročne uloge. Ta ideja objašnjava zašto je moguće da su mentalna stanja višestruko realizirana. Sada ćemo razmotriti kakvu ontologiju podrazumijeva funkcionalistička slika uma.

Općenito možemo reći da prihvaćanje funkcionalizma ne obvezuje na prihvaćanje određene ontologije u pogledu uma i tijela. To nam predočava i ramzificirana verzija teorije kojom opisujemo mentalna stanja. Ona samo kaže da postoje neke stvari koje stoje u uzročnim odnosima, a da se ne navodi nužno koje su to stvari koje realiziraju te odnose. U tom smislu, ramzificirana rečenica ostavlja otvorenim da različiti fizički pa i nefizički sustavi mogu zadovoljiti opise određene psihološke teorije. Drugim riječima, funkcionalizam je kompatibilan i s kartezijanskim dualizmom. Netko može tvrditi da ono što igra određenu uzročnu ulogu koja definira neko mentalno stanje jest zapravo nefizička supstancija. Uzmimo bol kao primjer. Organizam se nalazi u boli ako postoji neko stanje koje igra ulogu boli. Moguće je da je ono što igra ulogu boli neko svojstvo nefizičke supstancije. U tom smislu, moguće je da je funkcionalistički dualizam supstancija istinita teorija.

Funkcionalisti su, međutim, većinom fizikalisti (Hill 1991). Smatraju da su neka fizička stanja mozga ono što igra uzročne uloge karakteristične za mentalna stanja. U tom smislu, smatraju da, ako je funkcionalizam istinit, onda je neka varijanta teorije identiteta istinita.

U prošlom poglavlju vidjeli smo da su Armstrong (1968) i Lewis (1972) došli do ideje da funkcionalistička karakterizacija mentalnih stanja vodi do teorije identiteta tipova. Njihovo zaključivanje kreće od ideje da su pojmovi kojima referiramo na mentalna stanja funkcionalno definirani. Dakle, kada odredimo na koje uzročne uloge referiramo kada govorimo o pojedinim tipovima mentalnih stanja, onda nam preostaje za vidjeti koje su točno fizičke strukture koje igraju te uzročne uloge. Jednom kada ih otkrijemo možemo zaključiti da su mentalna stanja identična upravo tim fizičkim strukturama koje ih implementiraju.

Međutim, drugi funkcionalisti su ukazivali na to da se takva vrsta identifikacije ne može ostvariti zbog toga što su mentalna stanja višestruko ostvarljiva. U tom smislu, smatraju da funkcionalizam više odgovara fizikalizmu primjerka. Prema fizikalizmu primjerka svako pojedino mentalno

stanje je identično nekom fizičkom stanju koje ga realizira, no iz toga ne slijedi da možemo reducirati *tipove* mentalnih stanja na *tipove* fizičkih stanja. Vidjeli smo da, barem pojmovno, moramo dopustiti da druga bića mogu doživjeti i imati ista mentalna stanja poput nas, unatoč tome što fizički mogu biti jako različita od nas. Na primjer, možemo zamisliti da postoje izvanzemaljci koji mogu imati vjerovanja i imati osjetilna iskustva unatoč tome što su građeni od silicija a ne od ugljika poput nas. Kako bismo shvatili zašto mnogi povezuju funkcionalizam upravo uz varijantu fizikalizma koja se ne može reducirati na teoriju identiteta tipova, u nastavku ćemo se osvrnuti na korijene funkcionalizma koji nastaju na temelju takozvane računalne metafore (vidi Miščević i Smokrović 2001). Računalna metafora je ideja da usporedbom funkcioniranja uma s računalnim programima, koji dobivaju strogu definiciju razvojem Turingovih strojeva, možemo doći do novih otkrića o prirodi i funkcioniranju uma te postaviti temelje za razvoj materijalističke znanosti o uma i njegovih sposobnosti. Stoga ćemo u nastavku razmotriti ideju strojnog funkcionalizma.

5.5 Strojni funkcionalizam i Turingovi strojevi

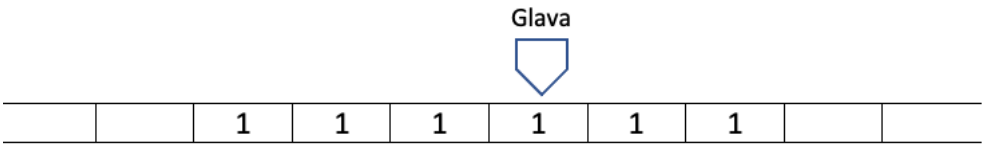
Putnam je među prvima uvidio mogućnost plodnog reformuliranja klasičnih rasprava iz filozofije uma u terminima Turingovih strojeva te je na pozadini te ideje razvio funkcionalističko gledište o prirodi uma. Putnamova filozofska karijera započinje u filozofiji znanosti te on općenito prihvaća naturalističku sliku svijeta. U tom smislu, njegova metodološka gledišta slična su onima koje zastupa Smart. Kada odabiremo znanstvenu teoriju ili gledamo možemo li napraviti neku znanstvenu identifikaciju, uključujući i one koje se odnose na odnos mentalnih i moždanih stanja, njihova uvjerljivost će ovisiti o znanstvenim i teorijskim razmatranjima, poput onih možemo li na taj način pojednostaviti naše teorije, omogućuju li nam nova znanstvena predviđanja ili na neki drugi način proširuju spoznaju (vidi, npr. Putnam 1995).

Međutim, kod Putnama je naglašen i drugi metodološki aspekt. U svojim razmišljanjima često se oslanja na pojmovnu analizu i analogije koje su sličnije Ryleovom (1949) načinu argumentiranja. No, u tom spoju pojmovne analize i naturalističkih sklonosti, Putnam je spreman spekulirati o prirodi mentalnih stanja na način koji nadilazi razmatranja logičkih biheviorista i teoretičara identiteta tipova. Te sklonosti se vide u području teoretiziranja o mentalnom životu strojeva i robota (vidi, npr. Putnam 1975c; 1975f). Štoviše, Putnam, slično Ryleu, smatra da je problem odnosa uma i tijela na neki način problem koji se rješava kada uvidimo logičku strukturu diskursa koji koristimo kada govorimo o mentalnim stanjima. Također, slično Ryleu, Putnam smatra da će nam ispravno razmatranje problema uma i tijela pokazati da ni dualizam ni materijalizam ne predstavljaju adekvatna rješenja. Međutim, za razliku od Rylea, Putnam se u svojoj argumentaciji oslanja na ideju Turingova stroja. Osnovna ideja je da ćemo, ako uvidimo da se klasični

problem odnosa duha i tijela može formulirati u terminima inteligentnih strojeva, uvidjeti i da rješenja koja se nude u obliku dualizma ili materijalizma mogu biti uvjerljiva samo ako su uvjerljiva kada govorimo o Turingovim strojevima. Kako bismo jasnije shvatili Putnamov način argumentacije, u nastavku ćemo prvo objasniti ideju Turingova stroja.

Turingov je stroj dobio ime prema Alanu Turingu (1912. – 1954.), poznatom britanskom matematičaru čiji su znanstveni i životni uspjesi te tragičan kraj života ovjekovječeni u filmu *The imitation game*. Turing je osmislio teorijske strojeve kako bi razvio matematički model uređaja koji može računati te je time udario temelje suvremenim informatičkim znanostima i razvoju umjetne inteligencije (vidi Miščević i Smokrović 2001).

Turingov je stroj zapravo misaoni ili teorijski stroj koji predstavlja model za razvoj fizičkih računala, međutim sam se nikada u stvarnosti ne može realizirati. Vidjet ćemo i zašto. Turingov stroj sastoji se od 1) skupa unutarnjih stanja, koje možemo označiti sa simbolima $q_1, \ldots, q_n$; 2) trake koja je podijeljena u polja koja mogu sadržavati određene simbole, primjerice a_1, ..., b_n, itd.; te 3) glave koja služi kao skener i pisač. Glava se nalazi iznad polja na traci; može se micati lijevo ili desno te može a) čitati simbole koji se nalaze na traci, b) brisati ih i c) pisati simbole po traci. Traka koju posjeduje Turingov stroj može se metaforički shvatiti kao da predstavlja njegovu memoriju te ona u principu može biti beskonačna u oba smjera (lijevo i desno od glave stroja, vidi sliku 1). Upravo zbog posjedovanja beskonačne trake, Turingov stroj se ne može ni u principu proizvesti u fizičkom obliku.



Slika 1 Turingov stroj koji ima traku koja ide beskonačno u lijevom i desnom smjeru. Na traci su zapisani simboli „1“ koje čita glava.

Ono što definira identitet Turingova stroja je tablica s instrukcijama. Tablica s instrukcijama definira program ili algoritam koji izvodi Turingov stroj. U odnosu na svaki par koji se sastoji od unutrašnjeg stanja stroja i simbola nad kojim je pozicionirana glava, pravila daju naredbu glavi (npr. „napiši određeni simbol“, „pomakni se udesno“, „pomakni se ulijevo“ i „stani“) i određuju u kojem bi se unutrašnjem stanju stroj trebao nalaziti.

Pravila u Turingovu stroju mogu se shvatiti kao „ako onda“ kondicionali. Kao primjer možemo uzeti sljedeće pravilo:

> (R1) Ako glava na traci čita simbol „1“ i stroj je u stanju q_0, tada ostavi simbol nepromijenjenim, pomakni se u desno i pređi u stanje q_2.

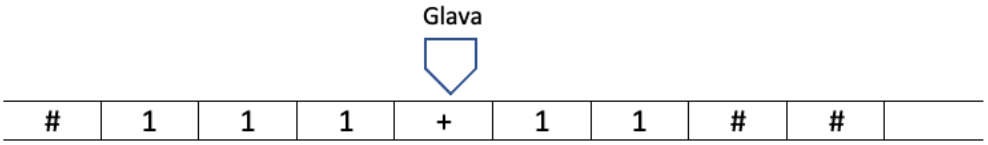
Tablica s instrukcijama određuje identitet Turingova stroja jer ona određuje način na koji će Turingov stroj računati neku funkciju. Međutim, ista funkcija može se izračunati na različite načine. Na primjer, postoje različiti načini na koje možemo množiti i dijeliti brojeve. Zato su načini računanja tj. pravila prema kojima se računa ono što određuje radi li se o istom ili različitom Turingovu stroju. Općenito, tablica s instrukcijama ima sljedeći oblik (Tablica 2):

	q_0	q_1
1	$1Dq_0$	#Halt
+	$1Dq_0$	
#	$\#Lq_1$	

Tablica 2 (modificirano prema Kim 2006, 126): Tablica se čita na sljedeći način. Gornji red nam govori u kojem se unutarnjem stanju nalazi stroj. U ovoj tablici imamo dva stanja q_0 i q_1. Lijevi stupac govori koji se simboli nalaze na traci. Matrica u unutrašnjem dijelu tablice daje instrukcije glavi što da radi s obzirom na simbol koji čita na traci i stanju u kojem se nalazi. U našem slučaju prva instrukcija „$1Dq_0$" daje sljedeću naredbu glavi: ako skeniraš simbol 1 i nalaziš se u unutrašnjem stanju q_0 zamjeni simbol 1 sa simbolom 1 i pomakni se u desno za jedno polje te pređi u stanje q_0 (tj. ostani u istom stanju). Instrukcija „$1Dq_0$" daje naredbu stroju: ako si u stanju q_0 i skeniraš simbol +, zamjeni simbol + simbolom 1, pomakni se za jedno polje u desno i pređi u stanje q_0. Instrukcija $\#Lq_1$ daje naredbu: ako čitaš simbol #, pomakni se u lijevo za jedno polje i pređi u stanje q_1. Instrukcija #Halt kaže da ako skeniraš simbol 1 u stanju q_1, zamjeni simbol 1 simbolom # i stani.

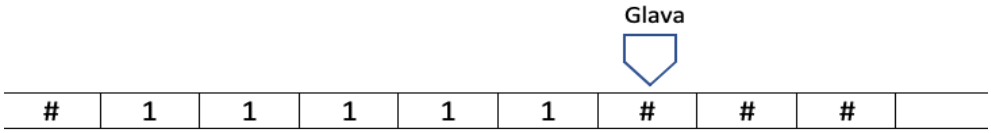
Unatoč svojoj jednostavnosti, Turingov stroj predstavlja vrlo moćno matematičko oruđe, te u teoriji može opisati i izvršavati svaki mogući računalni program. Drugim riječima, kompjutacijska moć Turingova stroja jednaka je ili veća od kompjutacijske moći bilo kojeg stolnog računala.

Kako bismo malo konkretnije vidjeli kako funkcionira Turingov stroj prema tablici za instrukcije, razmotrimo sljedeći primjer. Recimo da želimo napraviti Turingov stroj koji zbraja brojeve. Jedan od načina da ga napravimo je prema uputama iz Tablice 2. To možemo postići na sljedeći način. Recimo da nam se Turingov stroj nalazi u sljedećoj konfiguraciji (Slika 2, preuzeto iz Kim 2006, 125).



Slika 2

Cilj nam je konfigurirati Turingov stroj koji će simulirati operaciju zbrajanja, što znači da kao konačni ishod mora zapovjediti glavi da ostvari sljedeću konfiguraciju na traci (Slika 3).



Slika 3

Dakle, cilj je da brisanjem znaka „+“ na traci ostane samo pet jedinica. Upravo to se može postići instrukcijama koje su zadane u Tablici 2. Da je tome zaista tako ostavljamo čitateljima da kao vježbu provjere sljedeći instrukcije iz Tablice 2.

Dosad smo govorili o deterministčkim Turingovim strojevima. Njima, za svaki input, tablica s instrukcijama daje točno određeni output. Međutim, mogu se formulirati i kao probabilistički automati. To je tip Turingova stroja čije instrukcije ili naredbe i prijelazi između unutarnjih stanja imaju probabilistički karakter. Kao primjer možemo uzeti sljedeće pravilo koje opisuje primjer probabilističkog Turingova stroja:

> (R2) Ako glava na traci čita simbol „1“ i stroj se nalazi u stanju q_0, tada s vjerojatnošću r_1 ne mijenjaj simbol, pomakni glavu u desno i pređi u stanje q_2 ili s vjerojatnošću r_n, stani.

Putnam je smatrao da se upravo probabilistički automat može koristiti za opis rada ljudskih umova te da se modeliranjem uma pomoću Turingovih strojeva mogu izvući vrijedne pouke za filozofske probleme o odnosu uma i tijela.

5.6 Putnamova kompjutacijska teorija uma

Iako je Putnam kroz različite radove isticao da se ljudski um ne može poistovjetiti s Turingovim strojem (vidi, npr. Putnam 1975c; 1975f), svejedno je smatrao da se logički aspekti rasprave o odnosu uma i tijela mogu rasvijetliti ako o njima razmišljamo kao da um na apstraktnoj razini funkcionira analogno Turingovom stroju. Analogija se postiže na sljedeći način.

Zamislimo Turingov stroj, nazovimo ga T_1 čije tablice za instrukcije daju pravila koja opisuju karakteristične načine na koje ljudski um reagira na

podražaje i izvodi radnje. Na primjer, kada T_1 kao podražaj dobije rečenicu „Dobar dan, kako ste?“ u skladu s uputama koje ima u svojoj strojnoj tablici kao output daje rečenicu „Ja sam dobro, a kako ste vi?“. Ili ako mu se predstavi neki matematički zadatak on ga prema algoritmima koji su implementirani u njegovoj strojnoj tablici rješava i izbaci ispravan rezultat kao output. Kako bi ga se učinilo realističnijim i dalo mu mogućnost odlučivanja, možemo zamisliti da T_1 ima funkciju korisnosti (engl. *utility function*) koja predstavlja njegove preferencije i određuje načine na koje će djelovati. Recimo da je njegov sustav vrijednosti sličan vrijednostima koje prihvaća prosječna osoba. Na primjer, ako se nađe u situaciji gdje mora birati između vlastitog interesa i pomoći nekoj drugoj osobi, onda najčešće odlučuje pomoći drugoj osobi pod uvjetom da ga to ne košta previše u smislu resursa i vremena. Drugim riječima, njegova funkcija korisnosti rangira na više mjesto moralnu radnju u odnosu na sebičnu radnju. Međutim, ako pomoć drugoj osobi zahtijeva značajnija žrtvovanja vlastitog interesa, onda neće biti toliko sklon pomoći te će se to odražavati u njegovom rangiranju radnji koje određuje funkcija korisnosti.

Nadalje, kako bismo upotpunili analogiju, zamislimo da je T_1 povezan s mehaničkim tijelom kojim upravlja kako bi se snalazio u fizičkom okolišu. To mehaničko tijelo opremljeno je senzorima koji su funkcionalno identični ljudskim perceptivnim sposobnostima. Među njih spadaju i senzori za bol, tako da kada se, na primjer, T_1 udari ili opeče reagira poput ljudi te se udaljava od izvora boli, jauče, trlja dio tijela koji boli i tome slično. Ukratko, strojna tablica koja opisuje i regulira unutrašnja i vanjska ponašanja robota T_1 zapravo na vrlo apstraktnoj ili funkcionalnoj razini opisuje ponašanje i unutrašnja stanja običnog čovjeka.

Na temelju takve analogije, može se argumentirati da ni dualizam ni teorija identiteta tipova nisu uvjerljive pozicije. Argument se može formulirati na sljedeći način:

1) Postoji analogija između odnosa mentalnih i fizičkih stanja osobe i unutarnjih i fizičkih stanja Turingovih strojeva.
2) Ako neka tvrdnja o odnosu između unutarnjih i fizičkih stanja Turingovih strojeva nije uvjerljiva, onda neće biti uvjerljiva u slučaju odnosa mentalnih i fizičkih stanja kod ljudi.
3) Dakle, ako dualizam nije uvjerljiv u slučaju Turingovih strojeva, onda neće biti uvjerljiv ni u slučaju ljudi.
4) Dakle, ako teorija identiteta tipova nije uvjerljiva u slučaju Turingovih strojeva, onda neće biti uvjerljiva ni u slučaju ljudi.

Protiv dualizma bismo mogli argumentirati tako da zamislimo sljedeću situaciju. Zamislimo da postoji društvo robota koji poput T_1 imaju sve ljudske karakteristike, u smislu da njihove strojne tablice opisuju ponašanje i razmišljanje običnih ljudi. Zamislimo dodatno da se nalaze na stupnju znanja

na kojem tim robotima nije jasno da su oni roboti.[48] Što se njih tiče, oni su obične osobe poput nas ljudi koji imaju unutrašnje živote i moraju se snalaziti u društvu. Zamislimo sada da se među njima pojave roboti s filozofskim interesima te raspravljaju o prirodi mentalnih stanja i odnosu njihovih unutarnjih života s fizičkim tijelima. Pretpostavimo da i oni razviju dualističke i materijalističke pozicije. Dakle, prema jednima, um je nešto nematerijalno što se ne može svesti na fizičko funkcioniranje tijela, dok drugi smatraju da ne postoji mentalnost koja nadilazili fizičko, već da se mentalni procesi mogu reducirati na fizičke procese koje znanost još mora otkriti.

Ovakvim misaonim eksperimentom Putnam (1975c) ukazuje na to da mnogi argumenti u prilog dualizma nisu uvjerljivi. Sjetimo se fenomenološkog prigovora koji je bio upućen teoriji identiteta tipova. Moglo bi se tvrditi da osjećaj boli ne može biti isto što i aktivacija C-vlakana jer bol otkrivamo kroz introspekciju, dok aktivaciju C-vlakana možemo provjeriti samo kroz empirijsko promatranje mozga. Dakle, moglo bi se tvrditi da bol ima određeno introspektivno svojstvo koje aktivacija C-vlakana nema te da u tom smislu postoje neka psihička svojstva koja se ne mogu svesti na fizička svojstva mozga.

Putnam (1975c, 376) pokazuje da ovaj argument nije dobar jer se isti argument može formulirati oslanjajući se na analogiju s Turingovom strojevima. Zamislimo da je kod naših robota situacija sljedeća. Kada se T_1 ili bilo koji drugi od njegovih sugrađana nalazi u stanju q_0 u njihovom fizičkom tijelu aktivira se sklopka 47 koja regulira određeni strujni krug. Ono što još možemo pretpostaviti jest da kada je T_1 u stanju q_0 onda on to odmah zna, u smislu da ako dobije upit u kojem se stanju nalazi nije mu potrebno koristiti senzore kojima provjerava u kakvom stanju se nalazi vanjski okoliš ili njegovo fizičko tijelo, nego mu je odgovor dostupan samom činjenicom što se nalazi u stanju q_0. U tom pogledu, možemo zamisliti da, u kojem god stanju se T_1 nalazi, njegova tablica s instrukcijama uključuje naredbu da ako ga se pita u kojem se stanju nalazi on, uz ono što inače radi u tom stanju, odmah može odgovoriti i na to pitanje. Možemo pretpostaviti da materijalisti u toj zajednici robota predlažu da se svojstvo biti u stanju q_0 treba poistovjetiti sa svojstvom imati aktiviranu sklopku 47. Međutim, recimo da se dualisti usprotive takvom prijedlogu prigovarajući na sljedeći način. Budući da oni imaju direktan pristup unutarnjim stanjima te direktno mogu znati kada se nalaze u stanju q_0, dok samo indirektno putem senzornih organa mogu znati da je aktivirana sklopka 47, onda biti u stanju q_0 ne može biti isto svojstvo kao što je aktivacija sklopke 47. Naime, stanje q_0 ima introspektivno svojstvo koje je robotu direktno dostupno samom činjenicom što se nalazi u tom

[48] Zamislite nešto poput radnje filma *Zoe* iz 2018. u kojem glavne uloge igraju Ewan McGregor i Léa Seydoux. Radnja filma odvija se oko životnih i duhovnih problema s kojima se suočava jedan kiborg kada otkrije da nije biološki čovjek.

stanju. Sklopka 47 nema to svojstvo jer T_1 mora koristiti senzorne organe kada ispituje koje se sklopke aktiviraju u kojem trenutku.

Međutim, možemo primijetiti da u ovom slučaju argumentacija vodi do čudnih posljedica. Kada bi u slučaju ljudi dualisti bili u pravu da mentalna stanja imaju nereducibilno psihička svojstva koja moždana stanja nemaju, onda bi prema analogiji slijedilo da roboti poput T_1 imaju nereducibilno psihička svojstva koja njihova fizička tijela nemaju. Međutim, znamo da u slučaju robota taj argument nema previše smisla jer su oni prema pretpostavci fizička bića koja se ponašaju u skladu sa svojim strojnim tablicama. Budući da ovaj argument nije uvjerljiv u pogledu Turingovih strojeva, ne može biti uvjerljiv ni kada govorimo o ljudima. Ili riječima samog Putnama, „Svatko tko želi, (…), argumentirati [na temelju prethodnih razmatranja] da postoji duša, morat će biti spreman na svoje filozofske grudi prigrliti duše Turingovih strojeva!“ (Putnam 1975c, 376).

Naravno, to ne znači da treba prihvatiti materijalizam u obliku teorije identiteta tipova. Štoviše, Putnam na više mjesta ukazuje na to da analogija s Turingovim strojevima ide protiv prihvaćanja teorije identiteta tipova (Putnam 1995; 1975c). Jedna varijanta njegovog argumenta može se predstaviti na sljedeći način. Budući da su unutrašnja stanja koja putem strojne tablice određuju ponašanje Turingova stroja logički drugačija od fizičkih ili strukturalnih stanja u kojima stroj može biti utjelovljen, onda ne možemo poistovjetiti ili reducirati unutrašnja stanja na fizička ili strukturalna stanja stroja. Ponašanje robota T_1 određeno je strojnom tablicom koja definira odnose između podražaja i ponašanja. Međutim, identitet i priroda strojne tablice neovisni su o mogućoj realizaciji tog stroja u nekom fizičkom mediju. U tom smislu, nije relevantno je li tijelo robota T_1 napravljeno od metala, ugljika ili nekog trećeg materijala. Njegov „mentalni“ identitet određen je strojnom tablicom čija je priroda definirana skupom instrukcija.

Kako bismo došli do istog zaključka ne moramo se nužno oslanjati na hipotetičke primjere sofisticiranih robota poput T_1. Dovoljno je primijetiti sljedeće. Svako računalo se sastoji od fizičkih komponenti koje čine njegov hardver i algoritama ili procedura koje čine softver. Prema usporedbi s Turingovim strojevima, ideja je da je naš um analogan skupovima algoritama, dok je naše tijelo analogno hardveru na koji se softveri mogu instalirati. Dakle, ono što je relevantno za identitet nekog softvera neće ovisiti o hardveru na koji ga instaliramo, niti će identitet hardvera ovisiti o softverima koje može izvoditi. Štoviše, za prirodu softvera nije uopće bitna fizička struktura hardvera. Kao što se isti Turingov stroj može realizirati kroz različite fizičke sustave, tako se i konkretni softveri mogu izvoditi na različitim fizičkim kompjuterima. Na primjer, isti program poput onog za računanje može se izvoditi na strojnom računalu, mobitelu ili čak i ljudskom umu. Prema toj analogiji, slično vrijedi i za odnos uma i tijela. Mentalne sposobnosti analogne su procedurama koje definiraju softver. Mozak je

analogan hardveru na koji se mogu instalirati različiti softveri. Dakle, um je definiran prema svojim funkcionalnim svojstvima koja se mogu realizirati kroz različita fizička tijela te kao što softver ne možemo reducirati na hardver, tako ni ljudski um ne možemo reducirati na mozak.

Ovu vrstu argumenta može se povezati s idejom kategorijalne pogreške na koju upućuje Ryle (vidi poglavlje 3). Naime, argument se može shvatiti kao da se njime tvrdi da, kao što bi bila kategorijalna pogreška poistovjetiti softver s hardverom koji ga izvodi, tako bi izjednačavanje uma i tijela predstavljalo kategorijalnu pogrešku. Dakle, iz perspektive strojnog funkcionalizma ni klasični dualizam ni teorija identiteta tipova nisu previše uvjerljive. Ono što prema njemu slijedi jest da je um funkcionalno definirani entitet koji se realizira u fizičkom mediju, no ne može se na njega reducirati, kao što se kompjuterski algoritam ne može reducirati na fizički kompjuter.

S obzirom na ova razmatranja, možemo reći da suvremeni funkcionalisti najčešće prihvaćaju slabiju verziju fizikalizma prema kojoj su sva mentalna stanja na neki način povezana s fizičkim stanjima, no ne mogu se na njih reducirati kako su to smatrali pobornici teorije identiteta tipova. Ta verzija fizikalizma obično se naziva fizikalizam primjerka. Kao što smo ranije istaknuli, ideja je da će svaki primjerak mentalnog stanja biti identičan nekom primjerku fizičkog stanja. Na primjer, da se zadržimo na našem hipotetičkom slučaju, stanje našeg robota T_1 koje smo nazvali q_0 može biti identično aktivaciji sklopke 47. Međutim, identitet stanja q_0 kao *tipa* stanja može se jedino definirati u odnosu na strojnu tablicu koja određuje njegovu ulogu u kontekstu toga što stroj treba činiti u odnosu na određene inpute. Stoga q_0 kao *tip* stanja ne može se reducirati na *tip* fizičkog stanja koji ga trenutno realizira, tj. ne može se reducirati na aktivaciju sklopke 47. Slično bi trebalo vrijediti i za mentalna stanja. Bol možemo u svakom pojedinom slučaju poistovjetiti s primjerkom nekog fizičkog stanja. Na primjer, kod ljudi, ako osoba osjeti bol u trenutku t_1, onda ta bol u tom trenutku može biti identična aktivaciji C-vlakana. Međutim, ista ta bol se u drugom trenutku može identificirati s primjerkom nekog drugog fizičkog stanja, što će biti naročito izvjesno ako se radi o drugim bićima koja imaju različitu neuralnu strukturu mozga od ljudi.

5.7 Koliko je jaka analogija uma i Turingovih strojeva?

Putnam je često argumentirao na temelju ideje da je um na apstraktnoj razini analogan Turingovu stroju. No, možemo se pitati kako bi se ova analogija trebala shvatiti? Ovdje možemo razlikovati barem dvije interpretacije. Prema slaboj interpretaciji, probabilistički automat može *simulirati* um, u smislu da će, s obzirom na određeni skup inputa, um i automat proizvesti isti output. Prema jakoj interpretaciji, imati um jest *konstituirano* time da se zadovoljava opis nekog probabilističkog automata.

U ovom slučaju, unutrašnji procesi uma su oni koji karakteriziraju funkcioniranje automata.

Prema jako interpretaciji, um je vrsta računala. Kao oprimjerenje Turingovog stroja, um se sastoji od skupa simbola kojima barata prema određenim pravilima kao što Turingov stroj barata simbolima u skladu s tablicom instrukcija. Prema ovom gledištu, ljudsko mišljenje samo je jedna vrsta manipulacije simbolima i znakovima prema određenim pravilima. Zadatak psihologije kao znanstvene discipline sastojao bi se u određivanju teorije o tome koja je to vrsta simbola ili reprezentacija koje koristi um i prema kojim pravilima.

Čini se da je Putnam oscilirao između ova dva gledišta. U nekim radovima tvrdio je da Turingov stroj predstavlja samo analogiju za razmišljanje o umu, dok je u drugim radovima čini se zastupao tvrdnju da um jest jedna verzija Turingovog stroja (vidi, npr. Putnam 1975e). Ovdje nam nije toliko relevantno dati prikaz Putnamovih stajališta po ovom pitanju. U nastavku ćemo se osvrnuti na jaču interpretaciju te vidjeti koliko je uvjerljivo tvrditi da Turingovi strojevi konstituiraju um. Ovo je pitanje zanimljivo jer, ako je um stvarno jedna vrsta Turingovog stroja, možemo očekivati da će i dovoljno sofisticirana računala imati um te posjedovati svjesna mentalna stanja. Međutim, ako jaka interpretacija strojnog funkcionalizma nije uvjerljiva, imat ćemo razloga sumnjati u to da računala mogu posjedovati svjesna mentalna stanja.

5.8 Strojni funkcionalizam: jaka interpretacija i Turingov test

Možemo se pitati na koji način bismo mogli provjeriti mogu li računala misliti? Drugim riječima, pitanje je kada smijemo računalima i drugim sofisticiranim entitetima pripisati ljudska mentalna stanja? Sam Turing (1950) predstavio je test koji bi nam trebao omogućiti da odgovorimo na ova pitanja. Prema njemu ovaj se test naziva Turingov test. Osnovna ideja Turingova testa je da, ako ne možemo razlikovati ponašanje računalnog programa od onog normalne osobe, tada slijedi da taj računalni program i jest osoba koja poput nas posjeduje um i mentalna stanja. Preciznije, Turing (1950) predlaže da se pitanje „Mogu li strojevi razmišljati?“ operacionalizira kroz igru imitacije (engl. *the imitation game*).

Igra imitacije sastoji se od tri osobe, nazovimo ih A, B i C. U izvornoj inačici, osoba A je muškog roda, osoba B je ženskog roda, dok je osoba C ispitivač čiji rod nije određen. Ispitivač C se u odnosu na A i B nalazi u posebnoj sobi. Cilj igre je da C, koji se nalazi u odvojenoj prostoriji, odredi kojeg su roda osobe s kojima razgovara. Dakle, mora odrediti tko je od dvije osobe u susjednoj sobi A, a tko je B. C može postaviti bilo koje pitanje kako bi došao do točnog odgovora. Pri tome A ima zadatak da navodi C-a na krivi odgovor i time štiti svoj pravi identitet. Zadatak osobe B je da ispitivaču pomogne doći do pravog idsientiteta osoba A i B. Budući da C ne zna tko je A, a tko je B, ne

zna kome od dva glasa iz druge prostorije može vjerovati kada nastoji utvrditi njihov identitet. Turing se pita što bi se dogodilo da računalo preuzme ulogu osobe A te koliko često bi ispitivač pogriješio u pogledu identiteta računala koje igra tu ulogu. Ako bi bilo dovoljno često, kao i kada ispitivač igra igru imitacije s pravim ljudima, onda prema Turingu možemo zaključiti da to računalo posjeduje mentalna stanja. Drugim riječima, Turingov test nam kaže da, ako možemo napraviti računalo koje je toliko sofisticirano da u razgovoru s osobom ona ne može shvatiti da razgovara s računalom, onda smijemo pripisati mentalna stanja tom računalu.

Turing (1950) je dao predviđanje da će za 70 godina od trenutka pisanja njegovog rada postojati računala koja će u mnogo slučajeva moći prevariti ispitivača u pogledu svoje prirode. I doista, danas imamo dosta sofisticirane programe koji bi u mnogo slučajeva mogli prevariti ljude da misle da razgovaraju s drugom osobom. Jedan suvremeni primjer toga je GPT-3, računalni sustav za procesiranje prirodnog jezika koji može stvarati fikcijske tekstove, poeziju, glazbu, šale i tako dalje (Brown i ostali 2020). Temelji se na modelima neuralnih mreža dubokog učenja (engl. *deep learning*) koje koriste ogroman broj slojeva i parametara koji im omogućuju da na temelju dostupnih podataka na internetu uče i vode uvjerljive razgovore s ljudima na bilo koju temu. Mnogi se slažu da, iako GPT-3 možda ne bi sasvim prošao Turingov test, svejedno pokazuje sposobnosti korištenja opće inteligencije koja je u mnogim aspektima slična ljudskoj inteligenciji.[49]

Pretpostavimo u svrhu argumenta da stvarno napravimo stroj ili računalo koje može proći Turingov test, tj. zavarati osobu s kojom razgovara da se radi o drugoj ljudskoj osobi. Bi li to pokazalo da računala posjeduju inteligenciju i imaju mentalna stanja? Drugim riječima, bi li to značilo da je ljudski um samo jedna vrsta Turingova stroja koja je realizirana u ljudskom mozgu i tijelu? John Searle iznio je poznati argument da Turingovi strojevi, koliko god bili sofisticirani, ne mogu posjedovati sadržajna mentalna stanja.

Argument koji daje Searle usmjeren je protiv ideje da računalni programi bilo koje kompleksnosti mogu razumjeti jezik koji procesiraju te u tom smislu ne mogu razumjeti misli ili posjedovati sadržajna mentalna stanja. Argument se temelji na misaonom eksperimentu kineske sobe. Searle ga opisuje na sljedeći način:

> Pretpostavimo da sam zaključan u sobu s velikom hrpom kineskog pisma. Nadalje pretpostavimo (što je i slučaj) da ne znam niti pisati niti govoriti kineski, i da nije izvjesno da ću raspoznati kinesko pismo kao pismo različito, od recimo, japanskog ili besmislenih vijuga. […] Sada pretpostavimo da mi se nakon prve hrpe dostavi još jedna hrpa kineskog pisma sa skupom pravila za uspostavu korelacije između druge i prve

[49] Vidi, npr. https://dailynous.com/2020/07/30/philosophers-gpt-3/

> hrpe. Pravila su na engleskom i ja ih razumijem […]. Ona mi omogućavaju da uspostavim korelaciju između jednog skupa formalnih simbola i drugog skupa formalnih simbola, a sve što ovdje „formalno“ znači jest da simbole mogu u potpunosti identificirati s njihovim oblikom. Pretpostavimo nadalje, da mi je dostavljena treća hrpa kineskih simbola skupa s pravilima, opet na engleskom, koja omogućuje da uspostavim korelaciju između elemenata treće hrpe i prvih hrpa, i da ta pravila upućuju kako da kineskim simbolima određenih oblika odgovorim na određene oblike treće hrpe. Nepoznato meni, ljudi koji mi dostavljaju simbole prvu hrpu nazivaju „pismom“, drugu „pričom“, a treću „pitanjima“. Nadalje, simbole kojima odgovaram na treću hrpu nazivaju „odgovori na pitanja“, a skup pravila na engleskom „programom“. […] Nadalje pretpostavimo da se nakon određenog vremena toliko usavršim u praćenju uputa za manipuliranje kineskim simbolima, a programeri se toliko usavrše u pripremi programa, da se iz vanjske točke gledišta – to jest, iz točke gledišta nekog izvan sobe u kojoj sam zaključan – moji odgovori na pitanja ne mogu razlikovati od odgovora osoba kojima je kineski materinski jezik. Međutim, [...] kod kineskog […] odgovore dajem manipulirajući neinterpretiranim formalnim simbolima. Što se kineskog tiče, jednostavno se ponašam poput [računala]; izvodim [računalne] operacije nad formalno specificiranim elementima. Za svrhe kineskoga, ja jednostavno predstavljam oprimjerenje [računalnog] programa. (Searle 2001, 136–37)

Ovaj misaoni eksperiment dobro dočarava način rada računalnog programa ili općenitije Turingova stroja. Prisjetimo se da je rad Turingova stroja određen njegovom tablicom instrukcija. Ona je analogna priručniku koji Searlu govori što mora raditi u kineskoj sobi. Dakle, tablica s instrukcijama određuje glavi što da radi sa simbolima koje čita na traci. Ono što je važno jest da ti simboli koje čita glava Turingova stroja u principu ne znači ništa za sam stroj, slično kao što kineski simboli ništa ne znače Searleu jer ne razumije kineski. Stroj njima barata poput Searlea, tako da slijedi upute (tj. tablice instrukcija) koje nalažu kakav output treba izbaciti kad se susretne sa simbolom koji ima taj i taj oblik. Dakle, Searle u kineskoj sobi predstavlja analogiju za rad računala.

Intuitivan zaključak na koji nas navodi razmatranje ovog misaonog eksperimenta jest taj da oprimjerenje računalnog programa nije dovoljno za posjedovanje intencionalnih mentalnih stanja poput vjerovanja, a time onda niti za razumijevanje sadržaja misli. Ključna je karakteristika intencionalnih mentalnih stanja da imaju sadržaj. Na primjer, vjerovanje da kiša pada razlikuje se od vjerovanja da je vani sunčano upravo prema svojem sadržaju.

Prvo vjerovanje je o tome da kiša pada, dok je drugo o tome da je vani sunčano. Dakle, kako bismo razumjeli rečenice i misli koje razmatramo moramo biti sposobni imati mentalna stanja koja imaju sadržaj jer je upravo sadržaj ono što možemo razumjeti. Budući da Searle u kineskoj sobi oprimjeruje jednu vrstu računalnog programa, no ne razumije simbole kojima barata, onda niti računalo čije se funkcioniranje u potpunosti svodi na baratanje neinterpretiranim simbolima ne može posjedovati ljudsku formu razumijevanja. Štoviše, analogija sa Searleom u kineskoj sobi bi nam trebala pokazati da koliko god računala bila sofisticirana nikada neće moći posjedovati sadržajna mentalna stanja poput ljudi upravo zato što računala, poput Searlea u sobi, barataju simbolima koja za njih nemaju nikakvo značenje. U konačnici, to bi nam trebalo pokazati da Turingov test nije dobar test za određivanje mogu li računala misliti. Naime, čak i kada bi neki program mogao zavarati čovjeka da se radi o osobi, argument kineske sobe nam govori da bi to bila samo iluzija jer strojevi ne razumiju informacije koje procesiraju, a time ni ne posjeduju sadržajna mentalna stanja.

Naravno, mi često koristimo mentalistički jezik kada govorimo o računalima i drugim artefaktima. Na primjer, znamo reći da kalkulator *zna* zbrojiti ili oduzeti dva broja ili da termostat *opaža* kada je došlo do promjene temperature. Međutim, u svim tim slučajevima prema Searleu koristimo mentalistički jezik metaforički te zapravo proširujemo intencionalnost naših mentalnih stanja na vanjske predmete. U tom smislu mi možemo reći da računala posjeduju različite informacije koje govore o različitim stvarima. Međutim, pri tome moramo biti svjesni da je ono što legitimira takav govor činjenica da informacije koje računalo posjeduje imaju značenje za *nas*. Drugim riječima, ljudi su ti koji koriste računala i pomoću svoje sposobnosti interpretacije pridaju *značenje* formalnim strukturama kojima barataju računala.

Argument kineske sobe oslanja se na jaku intuiciju da računala ne mogu posjedovati sadržajna mentalna stanja poput ljudi. Time nas ovaj argument upućuje na zaključak da jaki kompjutacijski funkcionalizam ne može biti točna teorija jer ljudski um nije samo još jedno oprimjerenje apstraktnog Turingova stroja. Da budemo precizniji, ovaj argument osporava tvrdnju da je implementacija određene vrste Turingova stroja (tj. programa koji ga specificira) dovoljna da bismo nekom entitetu pripisali mentalna stanja ili procese. Ako je Searle u pravu, da bismo rekreirali um i njegova obilježja moramo moći rekreirati njegova fizička i biološka obilježja, a ne samo program koji implementira (Searle 2001, 147).

Međutim, unatoč intuitivnoj snazi ovog primjera, neki filozofi ne slažu se s njegovim zaključkom da se mentalni procesi ne mogu smatrati određenom vrstom kompjutacijskih procesa (za pregled rasprave, vidi Cole 2020). O argumentu kineske sobe i mogućim odgovorima na njega dosta se

raspravljalo od samog početka kada je Searleov članak objavljen.[50] Sam Searle u svom originalnom radu osvrnuo se na više mogućih odgovora na njegov argument. Ovdje ćemo razmotriti dva odgovora za koja nam se čini da najbolje pokazuju slabosti argumenta.

Prvi odgovor Searle (2001, 139–42) naziva sistemski odgovor. Prema njemu, Searle radi grešku jer na krivoj razini funkcioniranja sustava traži kome ili čemu pripisati razumijevanje. Pojedinac koji se nalazi u sobi samo je jedna komponenta sustava koji obrađuje informacije. Sustav koji obrađuje informacije (tj. kineske simbole) u skladu sa zadanim pravilima je taj koji posjeduje razumijevanje. Dakle, kineska sobi u cjelini bi trebala biti subjekt našeg pripisivanja razumijevanja, dok Searle čini samo jednu komponentu tog sustava.

Kako bi pokazao da ovaj odgovor nema previše smisla, Searle traži od nas da zamislimo pojedinca koji će na neki način internalizirati u svojoj glavi cijelu kinesku sobu. Na primjer, možemo zamisliti da se Searle nalazi izvan sobe te da je zapamtio sva pravila koja se odnose na baratanje kineskim simbolima, tako da kada mu netko izgovori ili napiše rečenicu na kineskom on zna koji simbol ili rečenicu mora dati kao odgovor. Unatoč tome, on neće razumjeti kineski jer i dalje samo barata simbolima koji nemaju za njega značenje. Samo baratanje simbolima prema pravilima ne omogućuje sustavu da razumije ono što radi.

Drugi odgovor Searle naziva robotski odgovor (Searle 2001, 143). Prema njemu trebamo zamisliti da se stroj za obradu kineskih simbola ugradi u robota koji ima ruke i noge, može se kretati u prostoru te ima osjetilne senzore koji mu omogućuju da vidi i snalazi se u okolini. Dakle, u ovom slučaju imali bismo robota kojim upravlja računalni program te u odnosu na određene inpute može izbacivati prikladne outpute te bi poput običnih ljudi mogao izvoditi fizičke radnje i tome slično.

Searle odgovara da čak i u ovom slučaju ne bismo mogli reći da robot posjeduje razumijevanje ili da ima sadržajna mentalna stanja. Kako bi to pokazao traži od nas da zamislimo kako se Searle nalazi u robotovoj glavi te da prema pravilima kineske sobe barata simbolima koje robot dobije kroz senzore kao input te nalaže robotu što da izbaci kao ponašajni output. Budući da Searle i dalje ne razumije kineske simbole, već njima barata kao formalnim znakovima, onda ni robot kojim Searle upravlja ne posjeduje razumijevanje onoga što govori ili radi.

---

[50] Tome naročito doprinosi format časopisa *Behavioral and Brain Sciences* u kojemu je Searlov članak izvorno objavljen (Searle 1980). Urednici BBS-a odabiru glavni, tzv. ciljani članak na koji drugi autori koji se bave tim područjem pišu komentare. U istom broju rada autor ciljanog teksta u zadnjem dijelu publikacije piše svoje odgovore na komentare.

Način na koji Searle reagira na ove odgovore ukazuje na to da su oni blisko povezani. Vidjeli smo da Searle kaže da možemo zamisliti osobu koja je internalizirala pravila koja definiraju kinesku sobu, a sama se nužno ne nalazi u sobi. Međutim, ona svejedno neće razumjeti te simbole. Dakle, ovdje zamišljamo osobu koja bi funkcionalno bila identična robotu iz drugog odgovora. Tako da se uvjerljivost ovih primjera temelji na istim intuicijama o tome može li sustav koji na temelju baratanja simbolima poduzima određene stvari posjedovati prava mentalna stanja. Nasuprot onome što nas Searle nastoji uvjeriti, vidjet ćemo da dodavanjem detalja ovim primjerima možemo utjecati na promjenu intuicija u pogledu toga mogu li računalni programi biti dovoljni da se omogući razumijevanje.

Daniel Dennett (1980) u svom je komentaru na Searleov argument ukazao na dvije važne stvari. Prva je ta da su njegovi revidirani misaoni eksperimenti nedorečeni u krucijalnim aspektima. Vidjeli smo da Searle na prvi prigovor odgovara tako da dosta modificira misaoni eksperiment te sada zamišljamo osobu koja internalizira sva pravila za odgovaranje na inpute u obliku kineskih simbola te više nije zatočena u kineskoj sobi. Dennett ukazuje da ovakva revizija misaonog eksperimenta nije bezazlena. Štoviše, ona ima potencijal da promjeni naše intuicije u pogledu toga posjeduje li sustav o kojem govorimo razumijevanje. Zamislimo da se Searle nalazi u Kini te da se nađe usred pljačke gdje mu pljačkaši na kineskom kažu „Ruke u zrak! Ovo je pljačka, predaj nam sav novac!“ Možemo očekivati da će Searle, budući da je internalizirao priručnik koji mu govori kako da se ponaša i što da kaže u odnosu na bilo koji jezični input, postupiti kako mu pljačkaši naređuju. Hoćemo li i u tom slučaju reći da Searle ne razumije što se događa, što mu govore pljačkaši i koje je značenje onog što on sam govori i radi? Intuicije ovdje više nisu jasne. Slično vrijedi i u slučaju robota. Ako se robot doista ponaša kao da razumije formalne simbole kojima barata poput obične osobe te ako ga u tom aspektu ne možemo razlikovati od obične osobe, onda nije jasno zašto ne bismo rekli da on razumije jezik koji koristi i radnje koje poduzima. U svakom slučaju, vidimo da se, kada modificiramo misaoni eksperiment kineske sobe i dodamo opisne detalje koje kinesku sobu na neki način čine dinamičnijom i sličnom običnom ljudskom djelatniku, izvorni zaključci na koje nas navodi Searle čine manje uvjerljivima.

Druga važna stvar na koju Dennett ukazuje jest da Searle u misaonom eksperimentu miješa razine objašnjenja. Searle spominje taj prigovor u vidu sistemskog odgovora. Međutim, čini nam se da pri tome ne uzima dovoljno u obzir krucijalnu stvar. Prema sistemskom odgovoru, kada razmišljamo o tome kome ili čemu možemo pripisati razumijevanje, subjekt tog pripisivanja ne bi trebao biti Searle u sobi nego cijeli sustav koji ta soba predstavlja. Međutim, ono što je problematično jest to što u tim misaonim

eksperimentima, bilo da se radi o izvornoj verziji, sistemskom odgovoru ili robotskoj verziji, Searleovo ponašanje i što on zna uopće nisu relevantni. Kako bismo to uvidjeli možemo razmišljati na sljedeći način. Odnos Searlea i kineske sobe kao cjeline može se shvatiti kao odnos osobe i njezinog mozga. Searle je samo jedna komponenta kineske sobe koja kao cjelina izvodi neki program. Slično tome, mozak je samo dio osobe koja izvodi neki program i radi određene stvari. U slučaju odnosa čovjeka i mozga jasno je da nećemo pripisivati razumijevanje mozgu koji, iako je važan čimbenik, ipak ostaje samo jedan dio te osobe. Razumijevanje možemo jedino pripisati osobi koja ima taj mozak. Mozak je naravno stvar koja omogućuje različite kognitivne funkcije, uključujući i razumijevanje, ali nije sam *subjekt* psiholoških atributa (vidi Bennett i Hacker 2003, pogl. 3).

Stoga bi zastupnik strojnog funkcionalizma mogao tvrditi da se misaoni eksperiment kineske sobe temelji na kategorijalnoj pogrešci. Od nas se traži da zamislimo osobu koja se nalazi u sobi te barata simbolima za koje znamo da joj nisu razumljivi. Ponašanje te osobe bi trebalo ilustrirati procese koji su analogni funkcioniranju računalnog programa. I upravo ovdje se nalazi problem. Ako ta osoba predstavlja rad računalnog programa onda pitanje razumije li ona simbole kojima barata uopće nije prikladno. Ono se temelji na kategorijalnoj pogrešci jer razmišljamo o pripisivanju psiholoških atributa na krivim razinama objašnjenja funkcioniranja uma. To vidimo kada razmišljamo o ljudima kao o biološkim bićima. Ako bi cijela kineska soba trebala biti analogna ponašanju ljudske osobe, onda bi ono što Searle radi u kineskoj sobi trebalo biti analogno nekom procesu u mozgu. Međutim, jasno je da neki proces u mozgu koji bi implementirao računalni program ne posjeduje razumijevanje simbola kojima barata. To zapravo nije čudno jer, kako smo rekli, psihološki predikati se smisleno pripisuju samo osobama, a ne njihovom dijelovima.

Stoga ako se vratimo na izvorni misaoni eksperiment, činjenica da Searle ne razumije kineske simbole kojima barata nije relevantna za pitanje mogu li računala misliti i imati mentalna stanja. Činjenica da sustavi od kojih je računalo sastavljeno ne razumiju ništa ne implicira da sustav kao cjelina koja se sastoji od tih procesa neće posjedovati ljudsku razinu razumijevanja. Stoga, ako smo dovoljno liberalni i dopuštamo si da zamislimo postojanje izuzetno sofisticiranih robota koji se ponašaju poput ljudi te bi mogli proći Turingov test, onda nije jasno što bi nas spriječilo da im pripišemo intencionalnost i razumijevanje. Činjenica da njihove unutrašnje sklopke i računalni programi koji upravljaju njima ne posjeduju razumijevanje je nebitna, kao što je nebitno da procesi u ljudskim mozgovima ne posjeduju razumijevanje informacija koje procesiraju. Zastupnik strojnog

funkcionalizma bi mogao reći da je jedino bitno ono što se događa na razini osobe, bilo da je realizirana u biološkom ili nekom drugom supstratu.

To su neka razmatranja koja bi strojni funkcionalist mogao koristiti kako bi se obranio od Searleovog prigovora kineske sobe. Međutim, čak i ako se strojni funkcionalisti mogu obraniti od argumenata koji pokazuju da funkcionalno definirani entiteti mogu posjedovati intencionalna mentalna stanja, preostaje nam za razmotriti mogu li funkcionalisti objasniti druga važna svojstva mentalnih stanja, kao što su subjektivni aspekti ljudskih iskustava. U nastavku ćemo se baviti upravo tim pitanjem. Razmotrit ćemo neke argumente kojima se nastoji pokazati da funkcionalizam ne može dati dobro objašnjenje svjesnih mentalnih stanja i procesa te da kao takav ne može predstavljati potpunu teoriju uma.

5.9 Argument protiv funkcionalizma: obrnute *qualia*

Mnogi smatraju da čak i ako funkcionalizam može zahvatiti prirodu intencionalnih mentalnih stanja, zasigurno neće moći objasniti zašto određena svjesna mentalna stanja posjeduju određeni kvalitativni karakter (klasični radovi uključuju T. Nagel 1974; Jackson 1982; Block 1978; vidi također Pećnjak i Janović 2011). U tom smislu argumenti protiv funkcionalizma temelje se na osjetilnim *qualia*.[51] Funkcionalizam definira mentalna stanja pozivajući se na njihova relacijska svojstva. Upravo zato što mentalna stanja nastoji definirati relacijski, mnogi smatraju da funkcionalizam ne može zahvatiti kvalitativnu prirodu mentalnih stanja. Međutim, mnogi će reći da je priroda kvalitativnih mentalnih stanja zapravo određena njihovim intrinzičnim svojstvima. U tom smisli, mnogi će se složiti da je esencijalno svojstvo boli njezina bolnost, tj. način na koju je doživljavamo, a ne nužno uloga koju igra u mentalnoj ekonomiji. Slično će se reći za doživljaj boja, osjećaj temperature, visine tona i tako dalje. Mnogi će reći da je esencijalno svojstvo tih stanja ono kako ih prepoznajemo i individuiramo, tj. da je određeno njihovim kvalitativnim svojstvima ili načinom na koji ih doživljavamo, a ne njihovom ulogom i odnosom prema drugim mentalnim stanjima.

Jedan utjecajan argument ove vrste temelji se na misaonom eksperimentu obrnutog spektra (vidi, npr. Block i Fodor 1972). Prema ovom misaonom eksperimentu, čija se suvremena formulacija obično pripisuje Johnu Lockeu, možemo zamisliti da dvije osobe, koje su u objektivno-fizikalnim terminima identične gledaju istu površinu, no da su boje koje subjektivno doživljavaju različite. Štoviše, možemo zamisliti da su im cijeli spektri boja koje mogu doživjeti kroz vizualno iskustvo obrnuti. Na primjer,

[51] Za podsjetnik kako treba shvatiti pojam *qualia*, vidi poglavlje 1. Za detaljnije objašnjenje vidi poglavlja 8 i 9.

kada gledaju istu rajčicu osoba O_1 ima iskustvo gledanja crvene boje poput ostalih ljudi, dok osoba O_2 ima iskustvo gledanja zelene boje, tj. ono što bismo mi doživjeli kao zelenu boju. Slično tome, kada O_1 gleda bananu onda ima iskustvo žute boje, dok kada O_2 gleda tu istu bananu ima iskustvo plave boje. Nadalje, čini se da možemo zamisliti da su te dvije osobe u potpunosti funkcionalno identične; one mogu na isti način diskriminirati predmete koje promatraju i donositi iste sudove o njima. Štoviše, kada se osobu O_1 pita koje je boje rajčica ona će reći da je crvene boje. Slično, kada se O_2 pita koje je boje rajčica ona će također reći da je crvene boje, unatoč tome što ima različite unutrašnje doživljaje boje od O_1. Budući da ne postoji način na koji bi oni mogli usporediti svoje unutarnje kvalitativne doživljaje boja, onda nema načina da otkriju da je kod jedne osobe spektar boja koje vidi obrnut u odnosu na drugu osobu.

Problem za funkcionalizam je sljedeći. Kada bi mentalno stanje bilo definirano prema uzročnoj ulozi koju igra kod pojedinog djelatnika, onda bi se doživljaj boje trebao moći svesti na određenu funkcionalnu ulogu. Naime, prema funkcionalizmu mentalna stanja ne mogu biti ništa drugo nego definirana svojom funkcionalnom ulogom. Iz toga slijedi da kada bi funkcionalizam bio istinit, onda obrnuti spektar ne bi bio moguć. Međutim, čini se da je obrnuti spektar moguć u smislu da možemo zamisliti da unatoč tome što su spektri boja kod O_1 i O_2 obrnuti sva njihova mentalna stanja i ponašanja su funkcionalno identična. Na primjer, možemo zamisliti da kada god žele pojesti rajčicu oboje posegnu za njom, kada je vide oboje će reći da je rajčica crvena, kad vide bananu oboje će reći da je žuta i tako dalje. Stoga se, iz čisto funkcionalne perspektive, O_1 i O_2 nalaze u istom tipu mentalnog stanja kada vide i razmišljaju o bojama i obojanim stvarima. Iz toga slijedi da funkcionalizam ne može biti istinita teorija o prirodi svih tipova mentalnih stanja.

Mogućnost obrnutog spektra boja dovodi u pitanje čisto funkcionalističke karakterizacije uma. Dakle, one karakterizacije prema kojima je suština mentalnih stanja određena uzročnim ulogama koje igraju, a ne prirodom fizičkih struktura koje zapravo realiziraju te uzročne uloge. U tom pogledu, Paul Churchland (1988) navodi da funkcionalisti ovaj problem mogu riješiti prihvaćanjem ideje da neke fizičke realizacije funkcionalno definiranih mentalnih stanja imaju intrinzičnu prirodu te da naša introspektivna identifikacija tih stanja ovisi o toj intrinzičnoj prirodi. Na primjer, mogli bismo

> identificirati kvalitativnu prirodu […] osjeta crvene boje s onim fizičkim obilježjem (stanjem mozga koje ga instancira) na koje mehanizmi introspektivne diskriminacije zapravo reagiraju kada [donosimo] sud da [imamo] osjet crvenog. (P. M. Churchland 1988, 40)

Ovakvo gledište predstavljalo bi svojevrstan kompromis između funkcionalizma i teorije identiteta tipova. Neka mentalna stanja možemo definirati funkcionalno, dok njihova kvalitativna obilježja identificiramo s fizičkim strukturama koje ih instanciraju. Međutim, prihvaćanje ovog rješenja nam pokazuje da funkcionalizam nema dovoljno vlastitih resursa da pruži generalnu teoriju prirode mentalnih stanja (vidi Pećnjak i Janović 2011). Štoviše, prihvaćanje ovakvog rješenje nas zapravo ponovo približava izvorno funkcionalističkim teorijama koje se razvijaju u sklopu teorije identiteta tipova kakvu su zastupali Lewis (1972) i Armstrong (1968).

Sada se može činiti da se vrtimo u krug. Krenuli smo od ideje da se teorija identiteta tipova suočava s problemom višestruke realizacije mentalnih stanja za kojeg se činilo da ga funkcionalizam rješava. Sada vidimo da se funkcionalizam susreće s problemom za koji se čini da ga teorija identiteta tipova može riješiti. Ovaj problem bi nas mogao uputiti da preispitamo uvjerljivost argumenta višestruke realizacije. Pod pretpostavkom da je fizikalizam u filozofiji uma uvjerljiva pozicija ona nas upućuje na to da u nekim aspektima reduciranje određenih svojstava mentalnih stanja na svojstva fizičkih stanja nije toliko problematično kako nam se ranije činilo. Naravno, ako je dualizam točna pozicija onda bi ona mogla objasniti zašto se kvalitativni aspekti vizualnog iskustva boja ne mogu reducirati na funkcionalne uloge metalnih stanja. Međutim, kao što smo vidjeli, dualizam se susreće s dodatnim problemima koji se odnose na uzročnu moć mentalnih stanja (vidi poglavlje 2). To nam može sugerirati da stvarno ima smisla razmotriti dosege i ograničenja argumenata koji se temelje na višestrukoj realizaciji protiv fizikalizma u obliku teorije identiteta tipova. Tim problemom ćemo se više baviti u sljedećem poglavlju, gdje ćemo se osvrnuti na redukcionističke i neredukcionističke pristupe u filozofiji uma. U nastavku ćemo se usredotočiti na drugi problem za funkcionalizam kojim se nastoji pokazati da ne može zahvatiti kvalitativne aspekte svjesnih iskustava.

5.10 Argument odsutnih *qualia*

Još jedan prigovor funkcionalizmu temelji se na argumentu iz odsutnih *qualia*. Pretpostavka funkcionalizma je da se funkcionalna organizacija koja definira svjesno iskustvo može instancirati ili realizirati kroz različite fizičke sustave. Vidjeli smo da upravo ovo svojstvo funkcionalizma objašnjava mogućnost višestruke realizacije mentalnih stanja. To je naročito jasno kod strojne varijante funkcionalizma. Na primjer, dovoljno veliki i sofisticirani elektronički kompjuter trebao bi moći barem u principu realizirati funkcionalno definirani um. Štoviše, prema Putnamu (1975e) fizička realizacija ne igra nikakvu ulogu u određenju prirode mentalnih stanja jer je priroda mentalnih stanja u potpunosti određena njihovim funkcionalnim ulogama. Posljedica tog gledišta je da što god zadovoljava formalni opis relacija u kojima mentalna stanja stoje prema inputu, drugim mentalnim

stanjima i outputu (npr. određenom ponašanju) može konstituirati um koji ima ta mentalna stanja.

Kako bismo uvidjeli neintuitivne posljedice ovog gledišta, možemo se poslužiti jednim od misaonih eksperimenata koje je osmislio Ned Block (1978). Zamislimo da je milijardu Kineza na sat vremena organizirano i povezano preko radio prijamnika putem kojih mogu upravljati određenim robotskim tijelom. Ideja je da u tih sat vremena milijardu Kineza i njihove veze s robotskim tijelom instanciraju ili realiziraju funkcionalnu strukturu ljudskog mozga, gdje milijardu Kineza povezanih radioprijamnicima igra ulogu milijardu neurona i njihovih veza. Ako je to moguće, trebali bismo prihvatiti da robot kojim oni upravljaju ima mentalna stanja koja su instancirana u tim vezama između milijardu Kineza. Pretpostavimo da milijarda Kineza na sat vremena svojim signalima simulira funkcionalne uloge koje su povezane s funkcionalnom ulogom boli. Na primjer, kada se robot udari u nogu onda dobiva signale da se uhvati za udareni dio noge, radi grimase te mu se aktiviraju rutine koje ukazuju na to da se treba odmaknuti od izvora oštećenja i tome slično. Međutim, unatoč tome što bismo mogli zamisliti da reakcije milijarde Kineza zadovoljava funkcionalnu definiciju boli, svejedno se čini da robot ne bi imao subjektivni doživljaj boli. Drugim riječima, kod robota bi nedostajala *quale* boli. Ovim argumentom se opet nastoji pokazati da funkcionalizam daje nepotpunu teoriju mentalnih stanja jer je moguće instancirati funkcionalno definiranu bol u odsutnosti subjektivnog doživljaja boli koje u normalnim situacijama prati to stanje.

Ovaj argument možemo sumirati na sljedeći način:

1) Moguće je da se populacija Kine ustroji na način koji bi zadovoljio funkcionalnu definiciju uma.
2) Umu koji bi instancirala kineska nacija nedostaje esencijalno svojstvo *qualia*, tj. subjektivni doživljaj kako je to biti u tom stanju.

Dakle,

3) Funkcionalna definicija uma ne zahvaća kvalitativna svojstva nekih mentalnih stanja.

Argumentom odsutnih *qualia* pokazuje se da je funkcionalizam preliberalna teorija. Kada bi funkcionalizam bio istinit, dozvoljavao bi pripisivanje osjetilnih mentalnih stanja čak i onim stvarima za koje je intuitivno jasno da ih nemaju.

Jedan odgovor bi stoga mogao biti da ograničimo pripisivanje mentalnih stanja samo onim stvarima koje imaju određenu fizičko-biološku konstituciju. Slično kao i kod odgovora na mogućnost obrnutog spektra, mogli bismo tvrditi da su, osim funkcionalnom ulogom, mentalna stanja određena intrinzičnom prirodom svojih realizatora. Ljudi mogu osjećati bol

jer, osim što instanciraju funkcionalnu ulogu boli, napravljeni su od određenog biološkog materijala koji uključuje nociceptore, C-vlakna i tako dalje. Budući da su roboti napravljeni od metala kojima nedostaju važne biološke karakteristike ne mogu osjećati bol čak i kada instanciraju funkcionalne uloge povezane s boli.

Ranije smo vidjeli da se ova vrsta funkcionalizma, koju bismo mogli nazvati neurofiziološki funkcionalizam, suočava s prigovorom mogućnosti višestruke realizacije mentalnih stanja. Štoviše, Block (1978) ovakve odgovore naziva šovinističkima jer prema njemu ograničavaju pripisivanje osjetilnih stanja samo bićima koja imaju određenu biološku prirodu. Na primjer, ako sada kažemo da je za bol potrebno imati C-vlakna, onda se čini da mogućnost osjeta boli ograničavamo samo na bića koja posjeduju C-vlakna. Međutim, čini se intuitivno jasnim da će postojati bića koja mogu osjećati bol, a da nemaju identičnu biološku strukturu poput ljudi. Funkcionalizam je upravo trebao riješiti taj problem time što je prirodu mentalnih stanja odredio njihovom funkcionalnom organizacijom. Sada vidimo da takvo gledište previše liberalizira pripisivanje mentalnih stanja. Pitanjem koliki je to problem za fizikalistička gledišta u filozofiji uma, više ćemo se baviti u sljedećem poglavlju.

U završnom dijelu ovog poglavlja razmotrit ćemo jedan utjecajan odgovor koji je Sydney Shoemaker ponudio protiv argumenta odsutnih *qualia*.

5.11 Funkcionalistički model introspekcije

Shoemaker (1975; 1981) brani funkcionalizam razmatranjem odnosa između osjetilnih mentalnih stanja i mogućnosti njihove spoznaje putem introspekcije. Shoemaker primjećuje da u normalnim okolnostima osjeti uzrokuju u osobama koje ih doživljavaju vjerovanja da imaju osjet s određenim kvalitativnim obilježjem. Na primjer, kada se ubodemo iglom onda ta bol u nama uzrokuje vjerovanje da osjećamo oštru bol. Tu vrstu vjerovanja naziva kvalitativnim vjerovanjima (Shoemaker 1975, 295). Prema Shoemakeru, funkcionalne definicije kvalitativnih stanja će, između ostalog, uključivati i uzročne relacije koje ta stanja imaju prema kvalitativnim vjerovanjima. Na primjer, ako govorimo o funkcionalnoj definiciji boli, onda će to biti ono stanje koje će imati a) tendenciju utjecati na ponašanja na određeni način, b) tendenciju proizvesti vjerovanje da je nešto na razini organizma pošlo po zlu (na primjer, da se osoba porezala ili ubola i tome slično) i c) tendenciju proizvesti kvalitativno vjerovanje da osjeća bol određene vrste. Kada bi odsutne *qualia* bile moguće onda bi bilo moguće da postoji funkcionalno definirano stanje boli koje bi zadovoljavalo opise (a) – (c), no svejedno ne bi imalo kvalitativni karakter, te stoga ne bi bila prava

bol. U literaturi se govori o mentalnom stanju M i mentalnom stanju M-ersatz (njem. *ersatz* = *zamjena*) kada je funkcionalna definicija M i M-ersatz identična, međutim, M ima određeni kvalitativni aspekt dok ga M-ersatz nema (Block 1980a; Shoemaker 1981). Kada bi odsutne *qualia* bile moguće, onda bi prema Shoemakeru sljedeći kondicional bio istinit:

> Ako su odsutne *qualia* moguće, onda prisutnost ili odsutnost kvalitativnog karaktera boli ne bi radilo razliku u uzročnim posljedicama, što bi onemogućilo bilo kome da razlikuje slučajeve prave boli od slučajeva ersatz-boli. (Shoemaker 1981, 588)

Shoemaker argumentira da je konzekvens prethodnog kondicionala neistinit te stoga da odsutne *qualia* nisu moguće. Kako bismo to uvidjeli sjetimo se da kada bi odsutne *qualia* bile moguće onda bi bilo moguće da neki entitet zadovoljava uvjete (a) – (c), dakle funkcionalno je identičan osobi koja osjeća bol, no da svejedno njegovo stanje nema kvalitativni karakter. Međutim, to bi značilo da je kvalitativni karakter boli nužno nedostupan introspekciji. Drugim riječima, osoba koja bi zadovoljila uvjete (a) – (c) ne bi znala nalazi li se u stanju boli ili ersatz-boli. Ako je tako, onda zapravo kvalitativni karakter ne bi bio važan za znanje o stanjima uma drugih ljudi ili nas samih. Ako mi sami ne možemo znati nalazimo li se u pravom ili ersatz-stanju boli, čini se da kvalitativni karakter boli ne igra nikakvu ulogu u našim metalnim životima. Što se nas tiče možda ni ne postoji kvalitativni karakter boli. Međutim, kako navodi Shoemaker, čini se

> apsurdnim pretpostaviti da obični ljudi govore o nečemu što je u principu nespoznatljivo svima kada govorimo o tome kako se osjećaju ili o tome kako stvari izgledaju, miriše, zvuče itd. (Shoemaker 1975, 297)

Dakle, budući da ima intuitivnog smisla reći da ljudi znaju ima li njihovo mentalno stanje određeni kvalitativni karakter te je u tom smislu kvalitativni karakter spoznatljiv, onda nije moguće da postoje dvije funkcionalno identične osobe od kojih jedna ima prava iskustva dok druga ima ersatz-iskustva.

Shoemakerov argument pretpostavlja uzročnu teoriju znanja. Ono što je relevantno za uzročne teorije znanja jest da pretpostavljaju nužan uvjet prema kojemu osoba ima znanje samo ako je njezino vjerovanje istinito i uzročno povezano s predmetom znanja (vidi, npr. Čuljak 2015; Dancy 2001). Na primjer, ako Ivica zna da se auto nalazi ispred njega, onda slijedi da je njegovo vjerovanje da se auto nalazi ispred njega istinito i da je uzrokovano činjenicom da se auto nalazi ispred njega.

Uzročna teorija znanja ima svoje probleme kada govorimo o spoznajnim domenama za koje nije jasno da mogu stajati u uzročnim vezama sa spoznavateljima. Na primjer, nije jasno da istine o apstraktnim domenama poput matematike i logike spoznajemo tako što stojimo u uzročnim odnosima s matematičkim ili logičkim činjenicama.

Međutim, u slučaju kvalitativnih stanja ima smisla smatrati da njih spoznajemo tako da ona u nama uzrokuju određena vjerovanja do kojih dolazimo introspekcijom. Drugim riječima, ima smisla prihvatiti uzročnu teoriju introspekcije. Na primjer, ako Marica zna da je glava boli onda ima smisla smatrati da je Maričino vjerovanje da je glava boli uzrokovano činjenicom da je glava boli. Ovakvo uzročno objašnjenje naših introspektivnih sposobnosti je pretpostavljeno u ranije spomenutom uvjetu (c). Prema Shoemakeru, kada imamo neko kvalitativno mentalno stanje, u normalnim uvjetima to stanje u nama uzrokuje kvalitativno vjerovanje da se nalazimo u određenom stanju koje ima određeni kvalitativni karakter. Dakle, ako introspekcija podrazumijeva uzročnu vezu među mentalnim stanjima, onda će te uzročne uloge koje kvalitativna mentalna stanja mogu igrati u našoj spoznaji biti uključene u njihovu funkcionalnu definiciju. Iz toga slijedi da će, ako je moguća spoznaja kvalitativnih mentalnih stanja, funkcionalno identični entiteti biti u identičnim kvalitativnim mentalnim stanjima. Time se, dakle, isključuje mogućnost odsutnih *qualia*.

Shoemakerov argument pokazuje da nasuprot prvim dojmovima funkcionalisti imaju prostora za osmisliti uzročne uloge koje kvalitativna mentalna stanja igraju u našim mentalnim životima i time ih pridružiti putem funkcionalnih definicija ostalim mentalnim stanjima. Vidjeli smo da Shoemaker to postiže oslanjanjem na uzročnu teoriju introspekcije i idejom da posjedovanje kvalitativnih mentalnih stanja uključuje njihovu spoznaju putem kvalitativnih vjerovanja.

Uvjerljivost Shoemakerovog argumenta se može vrednovati iz više perspektiva (vidi, npr. Block 1980a; Shoemaker 1981; Hill 1991, pog. 6). Ovdje ćemo se osvrnuti na činjenicu da argument pretpostavlja kvalitativna vjerovanja koja nastaju kao uzročni učinci kvalitativnih mentalnih stanja. Vidjeli smo da prema Shoemakeru kvalitativna vjerovanja pripadaju funkcionalnim definicijama kvalitativnih mentalnih stanja. To znači da ih biće nužno mora imati kako bismo mu pripisali kvalitativna mentalna stanja. Međutim, ova pretpostavka ima neintuitivne posljedice. Ako je točno da kvalitativna mentalna stanja proizvode kvalitativna vjerovanja, onda se čini da životinje i mala djeca neće imati kvalitativna mentalna stanja. Naime, nije uvjerljivo tvrditi da životinje i mala djeca imaju *vjerovanja* o svojim kvalitativnim stanjima. Na primjer, ima smisla smatrati da će malo dijete, koje još ne posjeduje pojam boli, kada osjeti ubod igle u ruku doživjeti taj

ubod na određeni način. Međutim, nije vjerojatno da će imati nešto slično vjerovanju da ga je nešto ubolo. Vjerovanja su intencionalna stanja čiji su sadržaji sastavljeni od pojmova. Na primjer, vjerovanje da *osjećamo oštru bol* sastoji se od pojmova osjećati, oštro i bol. Ako dijete nema te pojmove, ne možemo mu pripisati kvalitativno vjerovanje koje uključuje te pojmove. To je još manje vjerojatno kod drugih životinja. Dakle, ako organizam nije dovoljno kognitivno kompleksan da možemo reći da ima pojam osjeta, onda mu nećemo moći pripisati vjerovanja o tom osjetu. Iz ovoga slijedi da, ako funkcionalna definicija kvalitativnih stanja uključuje posjedovanje kvalitativnih vjerovanja, ispada da sva živa bića osim odraslih ljudi i možda primata koji su dovoljno kognitivno sofisticirani da imaju pojmove o stvarima neće imati kvalitativna mentalna stanja. Ova posljedica Shoemakerovog argumenta vrlo je neintuitivna te baca sumnju na njegovu uvjerljivost.

Moglo bi se odgovoriti da kvalitativna vjerovanja treba shvatiti kao jednostavnija reprezentacijska stanja koja ne uključuju nužno pojmovni sadržaj. Međutim, ako to nisu klasični propozicijski stavovi, ostaje nejasno kakva je to vrsta svjesnih stanja. Na primjer, moglo bi ih se shvatiti kao neku vrstu osviještenosti da trenutno doživljavamo određeni osjet. Na primjer, to bi moglo biti ono znanje koje se odnosi na to kako je doživjeti to stanje koje ne moramo nužno moći konceptualizirati ili verbalizirati kroz nekakve opise. Međutim, ako odgovor ostane ovako neodređen, onda se može prigovoriti da se pretpostavlja ono što se tek treba dokazati. Naime, moglo bi se prigovoriti da je upravo taj pojam osviještenosti ili svjesnosti ono što se ne može zahvatiti funkcionalnim opisom mentalnih stanja. Drugim riječima, da je to upravo ona vrsta svijesti koju jedna osoba može posjedovati dok može biti odsutna kod njezinog funkcionalnog blizanca.

Kao što smo ranije rekli, problem odsutnih *qualia* mogao bi se riješiti ako kažemo da je intrinzična priroda fizičkih stanja koja realiziraju funkcionalna stanja također relevantna za pripisivanje mentalnih stanja. U tom smislu, možemo reći da kineska nacija koja upravlja robotom nema um jer ona nije sačinjena od takve vrste organizacije biološke tvari koja bi bila u stanju proizvesti kvalitativna mentalna stanja. Međutim ova je vrsta funkcionalističke teorije šovinistička jer se njome ograničava mogućnost višestruke realizacije mentalnih stanja. U sljedećem poglavlju razmotrit ćemo koliki je to problem kada razmišljamo o psihološkim objašnjenjima na različitim razinama opisa fizičke realizacije.

5.12 Zaključak

U ovom poglavlju razmotrili smo funkcionalizam kao opću teoriju mentalnih stanja. Vidjeli smo da postoje različite verzije funkcionalizma. Možemo ih razlikovati prema dvjema dimenzijama. Jedna se odnosi na pitanje kako

određujemo funkcionalne uloge. Neki smatraju da se one određuju prema našim zdravorazumsko-psihološkim gledištima o prirodi mentalnih stanja. Drugi smatraju da se funkcionalne uloge trebaju odrediti prema našim najboljim znanstvenim teorijama o prirodi uma. Druga dimenzija se odnosi na veze funkcionalizma i teorije identiteta tipova. Rane verzije funkcionalizma razvijene su u sklopu teorije identiteta tipova. Funkcionalizam kao posebna teorija uma razvija se na pozadini metodološke pretpostavke da su umovi jedna vrsta Turingovih strojeva. Izvorni zastupnici te varijante funkcionalizma poput Hilarya Putnama predstavljaju funkcionalizam kao alternativu teoriji identiteta tipova koja može bolje zahvatiti tezu da se mentalna stanja mogu višestruko realizirati. Na kraju smo vidjeli da se ovakva liberalna verzija funkcionalizma susreće s prigovorima koji ukazuju na to da funkcionalizam ne može zahvatiti kvalitativna obilježja svjesnih mentalnih stanja. Budući da se funkcionalizam primarno razvija unutar fizikalističke ontologije onda nas ti problemi ponovo upućuju da pri određivanju prirode mentalnih stanja u obzir uzmemo njihovu materijalnu realizaciju.

Ta razmatranja nas ponovo približavaju redukcionističkim aspiracijama koje se obično povezuju s teorijom identiteta tipova. Stoga ćemo se u sljedećem poglavlju baviti općenitijim problemom koji se javlja unutar fizikalističke slike svijeta, a odnosi se na problemom redukcionizma i antiredukcionizma u filozofiji uma.

6 Redukcionizam i antiredukcionizam u filozofiji uma

6.1 Uvod

Prije nego krenemo na temu ovog poglavlja, vrijedi se osvrnuti na našu dosadašnju raspravu u kojoj smo se bavili temeljenim ontološkim pitanjima iz filozofije uma. Da se prisjetimo, bavili smo se pitanjima koja se tiču prirode uma i tijela te njihovih međusobnih odnosa. Također smo se bavili epistemološkim pitanjima koja se odnose na našu spoznaju tih odnosa. Konačno, razmatrali smo različita metodološka gledišta koja pojedini autori prihvaćaju u pogledu načina na koji bismo trebali pristupiti istraživanju ovih filozofskih pitanja.

U ovom poglavlju nastavljamo s istraživanjem odnosa uma i tijela u kontekstu rasprave između reduktivnih i nereduktivnih varijanti fizikalizma. Vidjeli smo da funkcionalizam i teorija identiteta tipova predstavljaju gledišta koja bi trebala biti kompatibilna sa znanstvenim spoznajama o umu te svako na svoj način podražava ideju da prirodni ili fizički svijet ima, u ontološkom smislu, prednost i na neki način određuje prirodu mentalnih, tj. psiholoških pojava. Međutim, ova dva gledišta razlikuju se prema tome smatraju li da se mentalna stanja mogu reducirati na fizički bazu. U tom pogledu, teorija identiteta tipova obično se predstavlja kao redukcionističko gledište, dok se funkcionalizam obično predstavlja kao antiredukcionističko gledište.

Iz fizikalističke perspektive postoji nekoliko općenitih razloga za odbacivanje redukcionističkih teorija uma. Jedan se razlog odnosi na tezu o psihofizičkom anomalizmu. To je teza da ne postoje zakoni prirode koji bi povezivali mentalne i fizičke pojave. Ovu je doktrinu razvio Donald Davidson (2001a). Davidson je tvrdio da fizikalne teorije kojima opisujemo uzročne veze među događajima pretpostavljaju postojanje deterministčkih zakona prirode. Nasuprot tome, smatrao je da, kada uzročne događaje opisujemo jezikom psihologije, zapravo nema govora o deterministčkim zakonima. Štoviše, Davidson je smatrao da su psihološki fenomeni određeni intencionalnošću, u smislu da se psihološka objašnjenja temelje na pretpostavci da ljudi posjeduju sadržajna mentalna stanja. I ne samo to nego čin pripisivanja mentalnih stanja podrazumijeva da su ljudi racionalna bića

sposobna za zaključivanje i ciljno-usmjereno djelovanje. Upravo ta pretpostavka racionalnosti prema Davidsonu ukazuje da psihološki fenomeni ne mogu biti regulirani determinističkim zakonima (za raspravu, vidi Miščević 1988). Oni su stoga anomalni (grč. *anomos* = bezakonje).

Davidsonovo gledište ima važan i utjecajan položaj u povijesti suvremene rasprave o odnosu psiholoških i fizičkih objašnjenja (za uvod u te rasprave, vidi Glüer 2011; za raspravu o Davidsonovoj filozofiji uma, vidi Child 1996). Međutim, ova rasprava pretpostavlja postojanje radikalne razlike između psiholoških i fizikalnih objašnjenja koja su većinom napuštena u suvremenoj filozofiji psihologije (Weiskopf i Adams 2015; vidi, također Miščević 1990). Stoga se u nastavku poglavlja nećemo baviti Davidsonovim gledištima.

Drugi važan argument protiv reduckionizma je već spomenuta mogućnost višestruke realizacije mentalnih stanja. U ovom poglavlju, usredotočit ćemo se na novije rasprave koje razmatraju ontološke aspekte psiholoških objašnjenja i odnosa između psiholoških i fizikalnih razina opisa mentalnih pojava. U tom kontekstu ćemo razmotriti predstavlja li argument višestruke realizacije, i u kojoj mjeri, problem za redukcionistička objašnjenja prirode mentalnih stanja.

U raspravi o redukcionizmu i antiredukcionizmu glavno pitanje je može li psihologija opisati, objasniti i predvidjeti mentalne procese i s njima povezana ponašanja neovisno o neuroznanstvenim teorijama koje se bave procesima i svojstvima mozga. Stoga ćemo prvo objasniti različite pojmove redukcije koji se javljaju u filozofiji uma. Usredotočit ćemo se na pojam redukcije kao odnosa između znanstvenih teorija te njezinu primjenu u filozofiji uma. Ova vrsta filozofske rasprave je u kontinuitetu s općim raspravama o prirodi znanstvenih teorija, objašnjenja i zakona. U tom pogledu, vidjet ćemo kako se znanstvena istraživanja uma koriste kako bi se informirale filozofske rasprave o tome možemo li objasniti funkcioniranje uma u terminima funkcioniranja mozga. Nakon toga ćemo razmotriti poznati argument koji daje Jerry Fodor (1974) protiv mogućnosti redukcije psiholoških teorija na teorije iz prirodnih znanosti, koji se temelji na mogućnosti višestruke realizacije psiholoških svojstava. U ostatku poglavlja bavit ćemo se raznim prigovorima Fodorovu argumentu koji nastoje pokazati da redukcionizam u filozofiji uma nije toliko neprihvatljiva pozicija kako se na prvi pogled čini.

6.2 Redukcionizam u filozofiji uma

Kako bismo krenuli s našom raspravom trebamo razlikovati nekoliko pojmova redukcije. Jedan od pojmova redukcije koji se koristi u filozofiji uma jest pojmovna redukcija. Zastupnici te vrste redukcije smatraju da se jedan pojam može u potpunosti analizirati koristeći neki drugim pojam. Tu vrstu redukcionizma vidjeli smo u slučaju logičkih pozitivista, ali i u slučaju Smartove teorije identiteta tipova (vidi poglavlja 3 i 4). Dok su logički

pozitivisti (usp. Carnap 1995; Hempel 1980) smatrali da se psihološki pojmovi mogu analizirati u terminima pojmova koji se odnose na ponašanja ili dispozicije za ponašanja, Smart (1959) je tvrdio da se pojmovi kojima referiramo na osjetilna iskustva mogu analizirati koristeći sadržajno neutralne pojmove čijom se upotrebom ne obvezujemo na dualistička ili fizikalistička gledišta na prirodu mentalnih svojstava. Slično tome, zastupnici analitičke varijante funkcionalizma smatraju da se mentalistički pojmovi mogu analitički svesti na pojmove kojima referiramo na funkcionalna stanja (usp. Lewis 1966).

Međutim, u dosadašnjoj raspravi nismo se susreli s filozofima koji su podržavali pojmovno reduktivni fizikalizam. To bi bila doktrina prema kojoj se naši mentalistički pojmovi mogu *a priori* reducirati na pojmove koji se odnose na fizička svojstva mozga. Čak su i teoretičari identiteta tipova poput Smarta prepoznali da mentalistički pojmovi imaju kognitivnu dimenziju značenja koja ih razlikuje od fizikalnih pojmova. To se vidi iz činjenice da oni dopuštaju da osoba može, na primjer, na kompetentan način koristiti pojam „bol“ te znati da osjeća bol, a da pritom ne misli da se nalazi u određenom fizičkom stanju mozga. Kada bi se mentalni pojmovi mogli reducirati na fizikalne pojmove onda takva misao ne bi bila moguća. Upravo je jedna od temeljnih stavki njihovog gledišta da mentalni pojmovi referiraju na određena fizička svojstva, stanja ili procese koje treba otkriti putem aposteriornog istraživanja.

Vrijedi spomenuti da su neki filozofi u novijim raspravama argumentirali da se fizikalizam treba shvatiti kao teza da se mentalni opisi i svojstva mogu izvesti *a priori* iz potpunog fizikalnog opisa našeg svijeta (Jackson 2007; usp. McLaughlin 2007). Ta varijanta fizikalizma se može nazvati apriorni fizikalizam. Međutim, čak i prema zastupnicima tog gledišta ne slijedi da se mentalni *pojmovi* mogu *a priori* svesti ili reducirati na fizikalne.

Kako bismo to uvidjeli razmotrimo primjer koji je Frank Jackson (2007) koristio kako bi ilustrirao svoju apriornu varijantu fizikalizma. Na temelju posjedovanja znanja dovoljnog broja lokacija točaka na površini predmeta moguće je deducirati *a priori* geometrijski oblika predmeta. Međutim, sam pojam geometrijskog oblika i dalje nije identičan pojmu lokacija točaka na površini predmeta čiji raspored moramo znati kako bismo zaključili o kakvom se obliku radi. Dakle, iako nam poznavanje nekih činjenica omogućuje da *a priori* izvedemo znanje nekih drugih činjenica, iz toga ne slijedi da se pojmovi kojima referiramo na jedan skup činjenica mogu reducirati *a priori* na pojmove kojima referiramo na drugi skup činjenica. U tom smislu, iz posjedovanja pojma mentalnog stanja ne slijedi da istodobno imamo i pojam fizikalnog stanja, kao što iz posjedovanja pojma geometrijski oblik ne slijedi da imamo pojam lokacija skupa točaka na površini predmeta.

Pojam redukcije relevantan za fizikalizam u filozofiji uma jest *ontološka* redukcija. Ova vrsta redukcije uključuje metafizičku relaciju između dvije

vrste entiteta. U tom smislu, reducirati neki entitet na neki drugi znači pokazati da je prvi entitet sastavljen ili napravljen od potonjeg. Kao primjer možemo uzeti činjenicu da je motor automobila reducibilan, tj. u potpunosti se svodi na različite dijelove koji ga čine. Još jedan primjer ontološke redukcije je kada jedan predmet možemo reducirati na drugi zato što je prvi identičan drugom. Ovdje kao klasičan primjer možemo uzeti činjenicu da je svojstvo biti voda identično svojstvu biti molekula H_2O.

Vidjeli smo da teoretičari identiteta tipova prihvaćaju ontološku redukciju entiteta, poput mentalnih procesa, događaja ili svojstava na fizičke entitete u mozgu. U slučaju funkcionalizma, stvari su nešto složenije jer njegovi zastupnici tvrde da se mentalno svojstvo, na primjer biti u boli, ne može poistovjetiti s određenim svojstvom mozga. Štoviše, funkcionalizam se obično veže uz gledište da različita fizička svojstva mogu biti realizatori istog mentalnog svojstva. Međutim, funkcionalizam se i dalje može shvatiti kao da podrazumijeva skromniju vrstu ontološke redukcije jer mnogi njegovi zastupnici prihvaćaju identitet primjerka, tj. gledište da je bilo koja instancijacija mentalnog svojstva identična instancijaciji određenog fizičkog svojstva (vidi poglavlje 5).

Među različitim pojmovima redukcije, pojam interteorijske redukcije pokazao se vrlo važnim u raspravama o tome kako formulirati fizikalizam u filozofiji uma. Ova vrsta redukcije odnosi se na čitave znanstvene teorije umjesto na pojmove ili entitete. Klasičnu formulaciju pojma interteorijske redukcije u filozofiji znanosti dao je Ernest Nagel (1974). U nastavku ćemo prikazati ključne aspekte ove vrste redukcije.

U Nagelovu objašnjenju redukcije između teorija središnju ulogu ima pojam prirodnog zakona. Postoje određena neslaganja oko toga kako bi trebalo formulirati pojam prirodnog zakona (Carroll 2020). Međutim, za naše potrebe dovoljno je reći da se prirodni zakon može smatrati općom rečenicom koja ima sljedeći oblik: „Za svaki *x*, ako *x* je F onda *x* je G". Ovom općom rečenicom navodi se da za bilo koji predmet *x*, ako *x* ima neko svojstvo F, onda ujedno ima svojstvo G. Osim što ima ovaj opći oblik, svojstva F i G ne smiju biti povezana slučajno. Drugim riječima, korelacija između F i G ne smije biti kontingenta. Na primjer, činjenica da Ivica ima crvene kamenčiće u džepu te da svi imaju masu jedan gram može se izraziti općom rečenicom „Za svaki *x* u Ivičinom džepu, *x* je crveni kamenčić i *x* ima masu jedan gram". Međutim, ta rečenica ne izražava zakon prirode jer se njome izražava samo slučajna ili kontingentna činjenica da Ivica ima crvene kamenčiće u džepu čija je masa jedan gram. Mogli smo zamisliti da crveni kamenčići imaju veću ili manju masu. Dakle, da bi rečenica izrazila prirodni zakon, korelacija između F i G ne smije biti na taj način slučajna. Nadalje, neki autori smatraju da bi zakonolike korelacije trebale podržavati kontrafaktičke ili protučinjenične kondicionale. Kontrafaktički kondicional je rečenica oblika „Kada *x* ne bi bio slučaj, onda *y* ne bi bio slučaj" ili „Da *x* nije bio slučaj, onda

y ne bi bio slučaj“. Konkretniji primjer bi bila rečenica „Da Hrvatski sabor nije izglasao neovisnost 1991. godine, izglasao bi ga neke druge godine.“ Dakle, takvom vrstom rečenice izražava se odnos između situacije za koju pretpostavljamo da nije aktualna ili stvarna (u tom smislu je kondicional *protu*činjeničan) i posljedice te situacije. Klasičan primjer takve vrste prirodnog zakona je Newtonov zakon univerzalne gravitacije. Prema ovom zakonu, svako fizičko tijelo određene mase m_1 na nekontingentan je način podložno privlačnoj gravitacijskoj sili te utjecaj gravitacijske sile ovisi o masi m_2 drugog fizičkog tijela i njihovoj udaljenosti. Ovdje imamo nekontingentnu vezu između predmeta mase m_1 i mase m_2 koja podržava kontrafaktički kondicional, jer kada bi prvi predmet imao neku drugu masu, onda bi se u skladu s time promijenio i odnos gravitacijske privlačnosti prema drugom predmetu. Upravo taj kontrafaktički međuodnos između svojstava nedostaje u primjeru s kamenčićima jer čak i kada Ivica ne bi imao crvene kamenčiće u džepu ne slijedi neki poseban zaključak u pogledu njihove mase. To je zato što odnos između boje kamenčića i njihove mase nije zakonolik, već je samo kontingentan. Nasuprot tome, odnos između mase dva predmeta relevantan je za utvrđivanje načina na koji će gravitacijska sila djelovati na njih.

Nagel svoj prikaz interteorijske redukcije temelji na modelu objašnjenja pokrivajućih zakona (engl. *covering law model*) koji je razvio Karl Hempel (1965; za pregled, vidi Salmon 1989). Ovaj se model objašnjenja još naziva *nomološki* model objašnjenja (grč. *nomos* = zakon). Prema ovom modelu, tipično znanstveno objašnjenje ima formu valjanog deduktivnog argumenta čije premise uključuju zakone prirode i početne uvjete, dok se u zaključku nalazi opis pojave koju želimo objasniti.[52] Pojavu koju nastojimo objasniti Hempel naziva *eksplanandum*, dok faktore koji objašnjavaju tu pojavu naziva *eksplanans*. Na primjer, razmotrimo objašnjenje fenomena da se određeni komad metala proširio zagrijavanjem. Objašnjenje bi imalo sljedeći oblik, gdje su premise 1) i 2) dio eksplanansa, a zaključak 3) je eksplanandum:

1) Svi metali zagrijavanjem se šire. (primjer zakona prirode)
2) Ovaj komad metala bio je zagrijan. (početni uvjet)

Dakle,

3) Ovaj komad metala se proširio. (zaključak)

Sada imamo sve sastojke koji ulaze u Nagelovo objašnjenje interteorijske redukcije. Stoga ćemo u nastavku razmotriti kako se prema Nagelu jedna teorija može reducirati na drugu teoriju.

[52] Kada se radi o probabilističkim fenomenima kakvi se proučavaju u društvenim znanostima onda prema Hempelu (1965) takva objašnjenja trebaju imati formu induktivno-statističkih argumenata, čije premise sadržavaju probabilističke generalizacije i početne uvjete, a zaključak slijedi kao njihova probabilistička posljedica.

Jezgra Nagelova gledišta na interteorijsku redukciju jest da se jedna teorija svodi na drugu teoriju kada se zakoni prirode koji su formulirani unutar prve, a time i njezina objašnjenja, mogu logički izvesti iz zakona formuliranih unutar druge teorije. Radi jednostavnijeg izražavanja nazovimo prvu teoriju koju reduciramo *reducirana teorija*, a drugu teoriju na koju prvu reduciramo *reducirajuća teorija*. Iz ovoga slijedi da reducirajuća teorija mora biti općenitija te imati veću eksplanatornu moć od reducirane teorije. Osnovna ideja je da se zakoni reducirane teorije mogu objasniti pomoću zakona reducirajuće teorije.

Formalnim se jezikom odnos interteorijske redukcije može prikazati na sljedeći način. Prema Nagelu, znanstvene teorije mogu se shvatiti kao skupovi rečenica koji opisuju pojave koje istražujemo. Neke od tih rečenica odnosit će se na pojedinačne događaje koji opisuju početne uvjete, neke na pojave koje objašnjavamo i neke na zakone različite općenitosti. U tom smislu, teorija T_1 reducira se na teoriju T_2 kada se svi fenomeni i zakoni koji ih objašnjavaju iz T_1 mogu logički izvesti putem rečenica i objašnjenja iz teorije T_2. Međutim, moguće je da neki pojmovi ili termini iz T_1 nemaju direktnih pandana u T_2. Drugim riječima, Nagel ukazuje na to da svaka redukcija podrazumijeva da pojmovi i iskazi, tj. vokabulari u kojima izražavamo teorije T_1 i T_2 moraju biti sumjerljivi, u smislu da znamo da termini iz T_1 referiraju na iste stari kao i termini iz T_2. Kako bismo to postigli moramo moći formulirati princip premošćivanja (engl. *bridge principle*) koji će na smislen način povezivati predikate kojima opisujemo stanja stvari koristeći teoriju T_1 s predikatima iz T_2. Principi premošćivanja se u tom smislu mogu shvatiti kao načela prevođenja koja nam omogućuju da povežemo vokabulare različitih teorija i time omogućimo njihovu međusobnu redukciju.

Kako bismo pojasnili taj aspekt interteorijske redukcije razmotrimo Tablicu 3. Ovdje imamo slučaj da reducirana teorija T_1 može objasniti činjenicu opisanu rečenicom $M_2(x)$ na temelju zakona (i). Ovdje možemo vidjeti da se relevantni termini koji se spominju u Objašnjenju 1 mogu pomoću principa premošćivanja PP_1 i PP_2 povezati s terminima koji se pojavljuju u zakonima reducirajuće teorije T_2. Stoga se pojava $M_2(x)$ može objasniti na temelju zakona iz T_2. Ako se to može postići za sve zakone iz T_1, a time i za sva objašnjenja koja se pomoću nje mogu formulirati, onda se T_1 reducira na T_2.

Teorija T_1	Principi premošćivanja	Teorija T_2
Objašnjenje 1		Reduktivno objašnjenje 1*
(i) Ako $M_1(x)$ onda $M_2(x)$ (Zakon prirode). (ii) $M_1(x)$. Dakle: (iii) $M_2(x)$.	(PP_1) $M_1(x)$ ako i samo ako $F_1(x)$. (PP_2) $M_2(x)$ ako i samo ako $F_2(x)$.	(i*) Ako $F_1(x)$ onda $F_2(x)$ (Zakon prirode) (ii*) $F_1(x)$. Dakle: (iii*) $F_2(x)$. Dakle, prema PP_2: (iii) $M_2(x)$.

Tablica 3

Nagel je ilustrirao interteorijsku redukciju na primjeru termodinamike, znanstvene discipline posvećene proučavanju toplinskih pojava, za koju se dugo vremena smatralo da se ne može reducirati na mehaniku koja je određena Newtonovim zakonima gibanja i gravitacije. Termodinamika neke fizičke pojave opisuje u terminima koje dijeli s mehanikom, poput obujma, tlaka i zakona koji povezuju te fizičke veličine. No, u termodinamici se također pojave opisuju u terminima kao što su toplina, temperatura ili entropija koji nemaju svojeg direktnog pandana u klasičnoj Newtonovoj mehanici. Štoviše, ti pojmovi su vrlo bitni jer se na temelju njih mogu formulirati zakoni prirode koji objašnjavaju termodinamičke pojave. Kao primjer možemo uzeti Boyle-Charlesov zakon $pv = kt$, koji su otkrili Robert Boyle (1627. – 1691.) i Jacques Charles (1746. – 1823.). Njime se tvrdi da je za određenu količinu plina, umnožak obujma v i tlaka p proporcionalan apsolutnoj temperaturi t (k je konstanta). Budući da se u termodinamici govori o temperaturi i drugim stvarima koje se definiraju u terminima temperature za koje ne postoje prijevodi u Newtonovoj mehanici, pretpostavljalo se da je termodinamika disciplina neovisna od klasične mehanike.

Zahvaljujući radu znanstvenika poput Daniela Bernoullija (1700. – 1782.) i zatim Jamesa C. Maxwella (1831. – 1879.) te Ludwiga E. Boltzmanna (1844. – 1906.) shvatilo se kako je moguće izvesti zakone termodinamike na temelju zakona mehanike (vidi Hanlon 2020). Na primjer, Nagel ilustrira slučaj izvođenja Boyle-Charlesova zakona iz principa statističke mehanike. Središnji korak u ovom zaključivanju sastojao se u formulaciji principa premošćivanja koji povezuje apsolutnu temperaturu t idealnog plina (pojam koji je svojstven termodinamici) s fizikalnom veličinom koja je proporcionalna prosječnoj kinetičkoj energiji molekula za koje se pretpostavlja da sačinjavaju određenu količinu idealnog plina (pojam koji pripada mehanici). Stoga se ovaj zakon može shvatiti kao

> […] logička posljedica principa mehanike, kada su oni

> nadopunjeni hipotezom o molekularnoj konstituciji plina, statističkom pretpostavkom o kretanju molekula i postulatom koji povezuje (eksperimentalni) pojam temperature s prosječnom kinetičkom energijom molekula. (E. Nagel 1987, 345)

Jednom kada su se ovaj i drugi zakoni termodinamike, uz pomoć principa premošćivanja, logički derivirali iz principa statističke mehanike, sve pojave objašnjene u sklopu termodinamike mogle su se objasniti pomoću zakona mehanike. U nastavku ćemo se osvrnuti na to kako bi mogla izgledati interteorijska redukcija u kontekstu relacije između psiholoških teorija i onih koje se tiču znanosti o mozgu.

6.3 Interteorijska redukcija u znanostima o umu

Iako je postojanje zakona i njihova relevantnost u psihologiji i dalje stvar spora, može se pretpostaviti da je znanstvena psihologija uspjela otkriti barem neke zakonolike regularnosti u ponašanju i mentalnim procesima koji omogućuju objašnjenja mentalnih pojava (Weiskopf i Adams 2015, 23–30). Ovdje ćemo razmotriti kako bi mogla izgledati interteorijska redukcija u znanstvenoj psihologiji koja bi se temeljila na zakonolikim regularnostima koje slijede iz fenomena asocijativnog učenja. Kako bismo ilustrirali principe asocijativnog učenja, prvo ćemo se upoznati s nekim pojmovima iz psihologije učenja.

Istraživanje asocijativnog učenja spada u tradicionalna i dobro razvijena područja empirijske psihologije (Houwer i Hughes 2020). Jednu od osnovnih formi asocijativnog učenja spomenuli smo u poglavlju 3 kada smo se bavili metodološkim biheviorizmom. Ruski fiziolog Ivan Pavlov (1849. – 1936.), u svojim je poznatim eksperimentima ustanovio vrstu učenja koja je danas poznata kao klasično uvjetovanje. U eksperimentima sa psima, pokazao je kako pokazivanje hrane psima, koja kod njih prirodno proizvodi slinjenje, popraćeno zvukom zvona rezultira činjenicom da sama proizvodnja zvuka, bez popratnog pokazivanja hrane, uvjetuje slinjenje kod pasa. Formalnijim rječnikom, klasično uvjetovanje istražuje kako početno neutralan podražaj (zvuk zvona u Pavlovljevom eksperimentu), kada se poveže s neuvjetovanim podražajima (NP) (pokazivanje hrane psima) koji prirodno izazivaju neku neuvjetovanu reakciju (NR) (salivacija), postaje uvjetovani podražaj (UP) (zvuk zvona) koji izaziva uvjetovanu reakciju (UR) (sam zvuk zvona proizvodi slinjenje).

Rescorla-Wagnerov (1972) model predstavlja jedno od najutjecajnijih formulacija pravila kojim se nastoji zahvatiti dinamika povezanosti između UP-a i NP-a tijekom faza asocijativnog učenja. Neformalno, ovo pravilo kaže da će tijekom faze učenja kada se organizmu predstave UP i NP jačina asocijacije između njih biti proporcionalna tome koliko je prisutnost NP-a s

obzirom na UP *iznenađujuća* ili *neočekivana* za organizam. Dakle, kada se prvi put UP i NP u potpunosti neočekivano predstave zajedno, organizam će ih početi jače asocirati nego pri svakoj sljedećoj prezentaciji kada postaje sve očekivanije da će pojavljivanje UP-a, zbog prijašnjih asocijacija, biti popraćeno pojavljivanjem NP.

Rescorla-Wagner pravilo ima sljedeću formu:

(RW) $\Delta V = \alpha\beta\,(\lambda - \Sigma V)$

Običnim jezikom, ΔV predstavlja promjenu u očekivanoj ili prediktivnoj vrijednosti uvjetovanog podražaja (UP). Izraz u zagradi ($\lambda - \Sigma V$) predstavlja količinu iznenađenja povezanu s predstavljanjem podražaja *V*. λ označuje maksimalnu količinu asocijativne snage koju podržava NP. α označuje istaknutost UP-a. β predstavlja stopu parametra NP. α i β su konstante koje kontroliraju brzinu učenja, u smislu da moderiraju utjecaj količine iznenađenja koja je zahvaćena izrazom ($\lambda - \Sigma V$) na promjenu prediktivne vrijednosti podražaja *V.*

Ostavljajući po strani tehničku terminologiju, ovdje nam je važno istaknuti da se jednadžba (RW) može shvatiti kao da opisuje zakonolike regularnosti u asocijativnim odnosima UP-a i NP-a koje su eksperimentalno potvrđene u paradigmi klasičnog i operantnog uvjetovanja (za kritičku raspravu, vidi Houwer i Hughes 2020, 101–9).

U terminima nomološkog modela objašnjenja, činjenicu da određeni organizam *O* pokazuje slabije povećanje stope u snazi naučenih asocijacija između UP-a i NP-a u kasnijoj fazi učenja u usporedbi sa snagom asocijacija koje su uspostavljene u ranijim fazama učenja mogli bismo objasniti pozivajući se na psihološki zakon kako je formuliran u (RW). To objašnjenje bi moglo imati sljedeću formu:

1) Ako je odnos asocijacije između UP-a i NP-a zahvaćen jednadžbom (RW) onda, ako je u trenutku t_1 učenja asocijacija UP-a i NP-a bila neočekivanija ili više iznenađujuća nego u trenutku t_2, rezultirajuće će povećanje snage asocijacije između UP-a i NP-a u t_2 biti manje nego u t_1 kako je to određeno jednadžbom (RW). (zakonolika regularnost)
2) Asocijacija između UP-a i NP-a bila je za organizam *O* neočekivanija ili više iznenađujuća nego u t_2. (početni uvjeti)

Dakle,

3) Rezultirajuće povećanje snage naučene asocijacije između UP-a i NP-a kod *O*-a u t_2 manje je nego u t_1 te prati dinamiku ažuriranja kako je to određeno u jednadžbi (RW).

Ako uzmemo u obzir ovaj fragment psihološke teorije učenja koji je zahvaćen sa (RW), možemo razmotriti što bi značilo da je psihologija reducibilna na neuroznanost.

Općenito možemo reći da je fragment psihološke teorije učenja koji se temelji na (RW) reducibilan na neku neuroznanstvenu teoriju ako se može pokazati da postoje zakoni premošćivanja između termina kojima referiramo na psihološke aspekte učenja i termina iz relevantne neuroznanstvene teorije.

Može li se to stvarno napraviti ostaje otvoreno empirijsko pitanje. Međutim, vrijedi istaknuti da postoje istraživanja koja se bave neurološkim temeljima asocijativnog učenja (vidi Roesch i ostali 2012). U tom smislu mogućnost takve redukcije nije empirijski nevjerojatna. U svakom slučaju, nama je bitno razmotriti postoje li pojmovni razlozi koji bi pokazivali da se takav projekt u principu ne može provesti.

Može se primijetiti da je teorija identiteta tipova u načelu kompatibilna s interteorijskom redukcijom psihologije na neuroznanost. Vidjeli smo da teoretičari identiteta tipova smatraju da nema načelnih razloga protiv prihvaćanja hipoteze da se mentalni predikati odnose na svojstva mozga. Dakle, prema ovom gledištu, nema načelnog razloga koji bi nas onemogućio u traženju neuralnih svojstava ili procesa koji bi bili identični procesu asocijativnog učenja i drugih psiholoških procesa opisanih jednadžbom (RW). Štoviše, iako to nije bio dio izvorne formulacije ovog gledišta, prema teoriji identiteta tipova nema apriornih razloga da se isključi da su upravo ta neuralna svojstva ona koja se spominju u zakonima neuroznanstvene teorije. Pretpostavka postojanja tih identiteta između neuralnih i mentalnih svojstava kako ih koncipira teorija identiteta tipova nudi nam potrebne zakone premošćivanja za interteorijsku redukciju psihologije na neuroznanost. Dakle, ako je određeni proces ili svojstvo M_1 koje se pojavljuje u zakonima psihologije identično svojstvu F_1 koje igra određene uloge u zakonima neuroznanosti, onda slijedi da se predikati koji se odnose na ta svojstva mogu pojaviti u bikondicionalnim iskazima sljedeće vrste:

(PP1) $M_1(x)$ ako i samo ako $F_1(x)$.

Ovakvi bikondicionali bi tada u sklopu neuroznanstvene teorije omogućili dedukciju regularnosti opisanu jednadžbom (RW) i drugim psihološkim zakonima čime bismo ujedno uspjeli derivirati neuroznanstvena objašnjenja psiholoških pojava koje se objašnjavaju pozivanjem na psihološke zakone.

Međutim, mnogi fizikalisti u filozofiji uma smatraju da se takva vrsta redukcije psihologije na neuroznanost ne može provesti ni u principu. U nastavku ćemo razmotriti utjecajan argument kojim se to nastoji pokazati.

6.4 Fodorov argument za fizikalistički antiredukcionizam

Jerry Fodor (1974; 1997) dao je utjecajan argument u prilog tvrdnji da se psihologija treba smatrati autonomnom znanošću u odnosu na neuroznanost. Njegova tvrdnja je da psihologija ima respektabilan znanstveni status čak i ako se ne može interteorijski reducirati na znanosti koje se bave fizičkim svojstvima mozga. Istodobno, Fodor tvrdi da se unatoč nemogućnosti redukcije na neuroznanost psihološka istraživanja u suštini odnose na fizičke događaje u mozgu. U tom smislu, možemo reći da zagovara nereduktivnu varijantu fizikalizma s kakvom smo se već susreli pri razmatranju kompjutacijske varijante funkcionalizma (vidi poglavlje 5). Prije nego što detaljnije razmotrimo Fodorov argument, vrijedi se osvrnuti na vrstu znanstvenog istraživanja uma koju je Fodor nastojao razjasniti u svojim filozofskim radovima.

Metodološka ideja da se um može istraživati neovisno o istraživanju mozga predstavlja srž kognitivnih znanosti koje su se razvile krajem šezdesetih. Nadahnuće za ovaj pristup bili su razvoj prvih računala i programskih jezika. Kao što smo vidjeli u Putnamovom (1995) slučaju, kao radnu hipotezu uzimalo se da se mentalne funkcije mogu proučavati u terminima obrade informacija ili točnije kao kompjutacijski procesi koji barataju mentalnim reprezentacijama (Miščević i Smokrović 2001). Ova pretpostavka je promicala povezivanje istraživanja umjetne inteligencije, koja je prema pretpostavci neovisna o istraživanjima hardvera, sa psihološkim istraživanjima koja su prema analogiji trebala biti neovisna o neuroznanstvenim istraživanjima mozga (Fodor 1968; Newell i Simon 1976; Pylyshyn i Pylyshyn 1984).

Kao što smo spomenuli, Fodor (1974) nastoji formulirati argument kojim će pokazati da su pristupi istraživanja u kognitivnim znanostima kompatibilni s fizikalističkom slikom svijeta, no da oni ne zahtijevaju prihvaćanje interteorijske redukcije psihologije na neuroznanost.

Jedan od glavnih ciljeva toga rada bio je otkloniti zabunu koja je prema Fodoru negativno utjecala na razmišljanja filozofa i znanstvenika, a odnosi se na sljedeće dvije teze:

1) Ako događaj *d* potpada pod neki znanstveni zakon, onda je to fizički događaj i potpada pod zakone fizike. (teza općenitosti fizike)
2) U konačnici će se sve znanosti, uključujući sociologiju, ekonomiju, psihologiju itd., reducirati na fizikalne teorije. (hipoteza jedinstvenosti znanosti) (Fodor 1974)

Fodor ustanovljuje svoju varijantu nereduktivnog fizikalizma prihvaćanjem 1) teze općenitosti fizike, i istovremenim odbacivanjem redukcionističke teze koja se odnosi na 2) hipotezu jedinstvenosti znanosti. Redukcionisti bi, s druge strane, bili oni koji prihvaćaju 1) i 2).

Fodor pokazuje kako bi se teza 1) mogla formulirati u pogledu posebnih znanosti na temelju primjera iz ekonomije. Uzmimo kao primjer ekonomsku transakciju. Svaka ekonomska transakcija može se smatrati događajem koji uključuje instancijaciju nekog fizičkog događaja. Na primjer, isplata novca u nekoj trgovini predstavlja ekonomski događaj koji će u Hrvatskoj najvjerojatnije uključiti prijenos određene svote kuna od kupca prema prodavaču. Isplata iste količine novca može se postići upotrebom kreditne kartice. U nekoj drugoj zemlji će se umjesto kuna koristiti druga novčana valuta. Intuitivno je uvjerljivo da se isplata novca ne može dogoditi, a da se ne ostvari neki fizički događaj. Na primjer, novac se neće prenijeti ako ne dođe do fizičkog prijenosa komada papira iz ruke u ruku, ili pokreta ruke kojom se obavljaju radnje vezane uz internetsko bankarstvo ili daje verbalna zapovijed bankaru da premjesti određena novčana sredstava i tako dalje. Slično prema Fodoru vrijedi za svaki mentalni događaj. Na primjer, imati neko vjerovanje pretpostavlja neki događaj u mozgu ili u slučaju umjetne inteligencije neki događaj u strujnom krugu računala. Stoga ima smisla tvrditi da ne može biti mentalnih događaja koji nisu instancirani u nekom fizičkom događaju. Budući da su ti događaji fizički, moraju se moći opisati kao da potpadaju pod neke fizičke zakone.

Unatoč tome što Fodor smatra da se psihologija ne može reducirati na neuroznanost, svejedno tvrdi da u psihologiji postoje prirodne vrste. Pod prirodnim vrstama u psihologiji misli se na to da postoje kategorije koje dobro razvrstavaju neke entitete prema njihovim psihološkim svojstvima i tipovima mentalnih procesa, kao što biološke vrste poput lavova i tigrova dobro razvrstavaju tipove životinja (vidi Brzović 2018). Iako psihološke vrste nisu koekstenzivne s fizičkim prirodnim vrstama, one prema Fodoru igraju značajne uloge u pravim psihološkim zakonima. Da nisu koekstenzivne znači da skup stvari na koje referiramo kada koristimo psihološke predikate neće nužno biti identičan nekom skupu stvari na koji referiramo fizikalnim predikatima. U tom smislu, psihologija je prema Fodoru prava znanost koja je nezavisna od neuroznanosti ili općenito fizikalnih znanosti.

Prema Fodoru, činjenica da svaka ekonomska transakcija ili psihološki događaj implicira da postoji određeni fizički događaj ne povlači da se ekonomija ili psihologija mogu reducirati na neuroznanost ili kombinaciju neuroznanosti i drugih temeljnijih fizikalnih znanosti.

Fodorov argument oslanja se na objašnjenje interteorijske redukcije kako smo je uveli u odjeljku 6.2. U tom kontekstu Fodorovo antiredukcionističko zaključivanje može se pojednostavljeno prikazati na sljedeći način:

1) Ako je psihologija reducibilna na neuroznanost, onda bi trebali postojati zakoni premošćivanja koji povezuju psihološke vrste i neuralne vrste.
2) Psihološke vrste su definirane funkcionalno. (pretpostavka funkcionalizma)

3) Ako su psihološke vrste definirane funkcionalno, onda se mogu višestruko realizirati. (teza o mogućnosti višestruke realizacije)
4) Ako se psihološke vrste mogu višestruko realizirati, onda ne postoje zakoni premošćivanja koji bi ih povezali s neuralnim vrstama.

Dakle,

5) Psihologija se ne može reducirati na neuroznanost.

U nastavku ćemo pojasniti premise ovog argumenta te navesti razloge zašto bismo ih prema Fodoru trebali prihvatiti.

U prvoj premisi spominje se princip premošćivanja koji bi trebao povezati psihološke i neuralne vrste. Stoga bismo trebali reći kako Fodor shvaća pojam vrste u ovom kontekstu. Rasprava o znanstvenim vrstama i njihovom odnosu sa znanstvenim zakonima zauzima značajno mjesto u filozofiji znanosti (za pregled rasprave, vidi Brzović 2018). Za naše potrebe znanstvena ili prirodna vrsta može se karakterizirati kao skup entiteta koji dijele određene značajke koje im omogućuju pojavljivanje u znanstvenim zakonima koji pak istraživačima omogućuju objašnjenje i predviđanje pojava. Klasičan primjer prirodne vrste je zlato. Ako je nešto zlato, onda možemo s izrazitom sigurnošću predvidjeti da će imati određena svojstva te da će se u određenim procesima ponašati na određeni način. S druge strane, svojstvo biti udaljen jedan kilometar od središta Rijeke nije prirodna vrsta. Ne postoji prirodni zakon koji bi omogućio da predvidimo ili objasnimo daljnje karakteristike svih entiteta koji su udaljeni jedan kilometar od središta Rijeke. Čini se jasnim da su sve te stvari koje su udaljene jedan kilometar od središta Rijeke previše heterogene te da stoga ne mogu igrati relevantne uloge u znanstvenim objašnjenjima i predviđanjima. Prema Fodoru, upravo principi premošćivanja, potrebni da se jedna teorija reducira na drugu, omogućuju da se termini kojima referiramo na prirodne vrste i zakone u reduciranoj teoriji prevedu u termine kojima referiramo na prirodne vrste i zakone reducirajuće teorije.

Druga premisa Fodorova argumenta predstavlja prihvaćanje jedne varijante funkcionalizma. Kao što znamo, funkcionalisti objašnjavaju prirodu mentalnih stanja u terminima funkcionalnih stanja. Funkcionalna stanja općenito se karakteriziraju u terminima input-output uzročnih relacija, koji mogu uključivati osjetilne podražaje, ponašanja, ali i druga mentalna stanja.

Treća premisa Fodorova argumenta oslanja se na pojam višestruke realizacije funkcionalnih stanja. Kao što smo vidjeli u prošlim poglavljima, prema toj tezi, isto funkcionalno stanje može se instancirati kroz različita stvarna i potencijalna/moguća fizička stanja.

Treća premisa može se interpretirati kao da izražava empirijsku ili pojmovnu tvrdnju. Kao empirijska tvrdnja temelji se na ideji da organizmi koji pripadaju različitim stvarnim vrstama mogu imati isti tip mentalnog

stanja unatoč različitoj neuroanatomiji ili neurofiziologiji. Na primjer, ljudi i hobotnice mogu doživjeti bol unatoč razlikama u živčanim sustavima. Ako se uzme kao pojmovna tvrdnja, onda hipoteza višestruke realizacije ukazuje na to da se psihološke vrste i mentalna svojstva općenitije mogu istraživati neovisno o fizičkim realizacijama. To nam omogućuje da zamislimo da postoje vrlo sofisticirani roboti ili izvanzemaljci koji su sastavljeni od potpuno različitih fizičkih materijala, no svejedno posjeduju mentalna svojstva poput ljudi.

Četvrta premisa predstavlja središnju tvrdnju u Fodorovu argumentu, stoga ćemo je malo detaljnije razmotriti. S obzirom na mogućnost višestruke realizacije mentalnih stanja možemo ustanoviti sljedeće identitete između mentalnih i fizičkih događaja:

(VR) $M_1(x)$ ako i samo ako $F_1(x)$ ili $F_2(x)$ ili $F_3(x)$ … ili $F_n(x)$.

Prema tvrdnji (VR) svaki mentalni događaj koji instancira mentalno svojstvo M identičan je fizičkom događaju koji instancira jedno od fizičkih svojstava mozga F_1, …, F_n. Međutim, ova svojstva mogu biti izuzetno različita iz perspektive neuroznanosti, neurobiologije, neurokemije i ostalih fizikalnih znanosti. Mentalni događaj koji uključuje osjećaj boli, na primjer, kod ljudi će biti realiziran fizičkim događajem $F_1(x)$, dok će kod hobotnica biti realiziran događajem $F_2(x)$, a kod neke treće vrste događajem $F_n(x)$ i tako dalje. Ako se radi o izvanzemaljcima ili sofisticiranim robotima, onda će fizičke realizacije mentalnih stanja kod njih biti radikalno drukčije jer će vjerojatno biti napravljeni od vrlo različitih materijala.

Međutim, možemo se zapitati zašto točno mogućnost višestruke realizacije mentalnih stanja implicira da ne postoji neuroznanstvena prirodna vrsta koja bi odgovarala pojedinim psihološkim vrstama. Na primjer, redukcionist bi mogao tvrditi da je mentalno svojstvo M_1 identično svojstvu koje dobijemo disjunkcijom cijelog niza fizičkih svojstava F_1,…, F_n koja instanciraju M_1 te da ta disjunkcija svojstava predstavlja reduktivnu bazu za mentalno svojstvo. Kako bismo uvidjeli zašto identifikacija mentalnog svojstva s disjunkcijom vrlo različitih fizičkih svojstava predstavlja loš temelj za interteorijsku redukciju, razmotrit ćemo zašto disjunktivna svojstva općenito predstavljaju loš temelj za odabir prirodnih vrsta i formulaciju zakona prirode.

Ranije smo spomenuli da je važno odrediti i kategorizirati entitete prema prirodnoj vrsti kojoj pripadaju jer nam one omogućuju da formuliramo zakone prirode, koji su nam pak važni za formulaciju kvalitetnih objašnjenja i predviđanja pojava. U tom pogledu, Kim (1993, 318) ističe da neko svojstvo može biti temelj za određivanje prirodne vrste kada su njegovi primjerci dovoljno *slični* da podržavaju induktivne projekcije općih iskaza pomoću kojih formuliramo zakonolike generalizacije. Kako bismo pojasnili tu ideju objasnit ćemo što bi značilo da je iskaz induktivno projicibilan (engl.

projectible). Na primjer, možemo primijetiti da su svi gavrani koje smo dosad vidjeli crni. Na temelju tih opažanja možemo donijeti induktivni zaključak da su svi gavrani crni. Međutim, opći iskaz „Svi gavrani su crni" odnosi se na sve gavrane koje smo opazili i na temelju njega projiciramo ili stvaramo očekivanje da će i svi budući gavrani koje još nismo opazili biti crni. U tom smislu, ako sljedeći gavran kojeg primijetimo bude crn dobivamo dodatnu dokaznu građu da je točna naša induktivna projekcija „Svi gavrani su crni". Općenito možemo reći da su opći iskazi kojima se izražava empirijska generalizacija poput „Svi F su G" projektibilni kada se mogu potvrditi pozitivnim instancama koje još nismo opazili. Dakle, iskaz „Svi gavrani su crni" projicibilan je jer se može potvrditi opažanjem novih instanci crnih gavrana. Kada proširimo tu ideju na svojstva ili predikate kojima ih označujemo možemo reći da je svojstvo ili predikat projicibilan kada se pojavljuje u iskazu koji se može induktivno potvrditi. Dakle, rekli bismo da su svojstva biti gavran i biti crn induktivno projektibilni jer se mogu koristiti u iskazima koji su induktivno projektibilni. U tom kontekstu, možemo reći da prirodne vrste zahvaćaju upravo ona svojstva koja su induktivno projektibilna te to objašnjava zašto igraju važne uloge u općim iskazima kojima izražavamo zakone prirode.

Sada se možemo vratiti pitanju mogu li disjunktivna svojstva ili predikati koji ih označuju predstavljati dobar temelj za formulaciju prirodnih zakona. Drugim riječima pitanje je mogu li disjunktivna svojstva predstavljati prirodne vrste koje zahvaćaju induktivno projektibilna svojstva? Općeniti odgovor je da u velikom broju slučajeva disjunktivni predikati neće biti projektibilni te stoga neće predstavljati dobre temelje za formulaciju zakona prirode. Kako bismo to vidjeli, razmotrimo sljedeći primjer.

Kim (1993, 319–20) daje primjer minerala koji se naziva žad. Iako se nekad smatralo da žad predstavlja jedinstvenu mineraloško vrstu, kasnije se otkrilo da je zapravo sastavljen od dva minerala: jadeita i nefrita. Dakle, svojstvo ili predikat biti žad možemo definirati na sljedeći način:

(1) *x* je žad ako i samo ako *x* jest jadeit ili *x* jest nefrit.

Postavlja se pitanje je li žad prirodna vrsta koja bi mogla igrati ulogu u formulaciji zakona prirode? Pretpostavimo da, slično kao i u slučaju višestruke realizacije psiholoških vrsta, netko želi tvrditi da je žad zapravo disjunktivna vrsta minerala koja se stoji od jadeita ili nefrita. Kim smatra da bi takvo gledište bilo neuvjerljivo te da se žad ne bi trebao smatrati prirodnom vrstom.

Kako bismo to uvidjeli razmotrimo sljedeći iskaz:

(2) Žad je zelen.

Budući da „žad“ zapravo označuje jadeit ili nefrit, onda je (2) ekvivalentan sljedećem iskazu:

(3) Jadeit ili nefrit su zeleni.

Međutim, (3) ne može konstituirati prirodni zakon. Vidjeli smo da opći iskazi koji izražavaju zakone prirode trebaju zadovoljiti kriterij projektibilnosti, tj. moramo ih moći potvrditi kroz opažanja pozitivnih instanci. Međutim, u ovom slučaju problem je što je moguće da sva opažanja koja potvrđuju iskaz (3) budu o jadeitu, a ne o nefritu. U tom slučaju (3) nije u cijelosti potvrđen. Štoviše, iako nađemo potvrdu da je jadeit zelen, svejedno nećemo imati nikakvu potvrdu da je nefrit zelen. Stoga možemo zaključiti da žad ne može biti prirodna vrsta pomoću koje možemo formulirati zakone prirode. Razlog tome je upravo činjenica da žad predstavlja preheterogenu vrstu na temelju koje ne možemo formulirati projektibilne zakonolike iskaze.

Sada možemo vidjeti zašto bi sličan zaključak trebao slijediti u slučaju odnosa psiholoških i neuroznanstvenih teorija unutar kojih formuliramo kategorije i zakone jedne i druge discipline. Ako je mentalno svojstvo, na primjer bol, ekvivalentno ili realizirano kroz disjunkciju heterogenih fizičkih svojstava, onda bol ne bi mogla biti psihološka vrsta pomoću koje možemo formulirati induktivno projektibilne zakone psihologije. Naime, prema analogiji sa žadom, kada bismo bol u iskazima psiholoških zakona htjeli zamijeniti sa svim mogućim fizičkim procesima (od kojih su neki radikalno drukčiji od stvarnih) koji mogu igrati ulogu boli te time pokušali formulirati neki neuroznanstveni zakon dobili bismo induktivno neprojicibilan opći iskaz. Međutim, ako je Fodor u pravu, u psihologiji postoje zakoni koji su induktivno projektibilni, a to znači da postoje i predikati ili svojstva koja su induktivno projektibilna. Iz toga slijedi da se ta svojstva ne mogu reducirati na disjunkciju fizičkih svojstava koja se nalaze u njihovoj podlozi. Zaključak je da, ako želimo sačuvati psihologiju i znanstvene zakone koje formuliramo unutar nje, psihologiju moramo smatrati disciplinom koja će biti neovisna i autonomna u odnosu na neuroznanost i ostale discipline koje se bave fizičkim aspektima mentalnih procesa.

Fodorov argument za nesvodljivost psihologije kao znanosti na neuroznanost i ostale temeljne discipline privukao je dosta veliku pažnju te se u literaturi i danas raspravljaju mnogi njegovi aspekti. U nastavku ćemo se usredotočiti na neke od značajnijih kritika tog argumenta.

6.5 Pojmovni prigovori Fodorovu argumentu

Kao što smo vidjeli, Fodorov argument sastoji se od premisa koje uključuju osnovne pojmove iz filozofije znanosti, poput pojma prirodne vrste, znanstvenog zakona, objašnjenja i interteorijske redukcije. U tom pogledu nije čudno da se mnoge reakcije na njegov argument odnose na to kako

bismo trebali shvatiti te pojmove i njihove logičke pretpostavke. U nastavku ćemo razmotriti neke od njih.

6.6 Redukcija ne zahtijeva bikondicionale

Fodorov antiredukcionistički argument može se kritizirati tako da se odbaci koncepcija redukcije koja se podrazumijeva u prvoj premisi. U vrijeme kada Fodor (1974) objavljuje svoj članak, vladalo je mišljenje da ispravni principi premošćivanja pretpostavljaju povezivanje termina iz reducirane i reducirajuće teorije na temelju ekvivalencija koje se izražavaju pomoću bikondicionalnih iskaza. Na primjer, osim Fodora i Putnam je smatrao da:

> se redukcija […] treba shvatiti kao dedukcija reducirane teorije iz reducirajuće teorije, uz pomoć bikondicionala koji izražavaju „koordinirajuće definicije". (Putnam 1975b, 117–18)

Međutim, Robert Richardson (1979) ukazuje na to da Nagelova izvorna karakterizacija interteorijske redukcije ne pretpostavlja postojanje takvih bikondicionala. Štoviše, Nagel navodi da je

> […] redukcija izvršena kada se za eksperimentalne zakone sekundarne nauke (a ako ova ima adekvatnu teoriju, onda to vrijedi i za njezinu teoriju), pokaže da predstavljaju logičke posljedice teorijskih pretpostavki (uključujući koordinirajuće definicije) primarne nauke (E. Nagel 1974, 312–13).[53]

U nastavku dodaje da:

> […] veza između A [pojma u sekundarnoj znanosti] i B [pojma u primarnoj znanosti] nema nužno oblik ekvivalencije [bikondicionala] i može, na primjer, imati oblik implikacije: ako B onda A. (E. Nagel 1974, 315, fusnota 5)

Richardson (1979) kao primjer koristi redukciju genetike na molekularnu biologiju te pokazuje da se u tom slučaju principi premošćivanja mogu izraziti kondicionalnim, a ne nužno bikondicionalnim iskazima. Dakle, ako prihvatimo da se interteorijska redukcija u znanosti može postići oslanjanjem na principe premošćivanja koji podrazumijevaju kondicionalne iskaze, onda nemamo razloga negirati njezinu mogućnost u slučaju psihologije i neuroznanosti.

---

[53] Nagel (1974, 300) naziva „primarna nauka" skup teorija koje predstavljaju bazu redukcije, dok „sekundarna nauka" naziva skup teorija koje se reduciraju na ove prve. Dakle, primarne znanosti predstavljaju ono što smo mi nazvali reducirajuće teorije, dok sekundarne znanosti predstavljaju ono što smo mi nazvali reducirane teorije.

Prema ovom objašnjenju redukcije, psihologija bi se mogla reducirati na neuroznanost kada bi postojali kondicionalni principi premošćivanja koji imaju sljedeću formu:

$$\forall x\, \mathrm{F}(x) \rightarrow \mathrm{M}(x)$$

Kao i prije, ovdje F označuje fizikalne predikate, dok M označuje mentalne predikate koji su karakterizirani u funkcionalnim terminima. Ovakva vrsta odnosa gdje je instancijacija određenog fizičkog svojstva *dovoljan* uvjet za instancijaciju određenog mentalnog svojstva je nešto što funkcionalisti obično prihvaćaju.[54]

Štoviše, funkcionalisti često prihvaćaju jaču tvrdnju koja isključuje dualizam, prema kojoj samo fizička svojstva mogu biti temelji za pripisivanje mentalnih predikata. U tom smislu Richardson zaključuje da:

> […] čak i ako su psihološka stanja funkcionalna stanja te čak i ako se psihološka stanja i fiziološka stanja ne mogu dovesti u jedan na jedan zakonoliki tip korespondencije, to ne stvara prepreku za redukciju u psihologiji. (Richardson 1979, 549)

Ova reakcija na Fodorov argument je zanimljiva i ukazuje na određene pretpostavke u njegovom argumentu koje ne moraju svi nužno prihvatiti. Međutim, kada govorimo o najpoznatijoj redukcionističko-fizikalističkoj teoriji, naime teoriji identiteta tipova, čini se da njezini zastupnici ne mogu prihvatiti formulaciju principa premošćivanja koji podrazumijeva samo kondicionalne iskaze. Naime, teoretičari identiteta tipova prihvaćaju ontološku pretpostavku da su mentalna svojstva *identična* fizičkim svojstvima. Ideja identiteta pretpostavlja da postoje bikondicionalni iskazi koji povezuju tvrdnje o mentalnim stanjima i svojstvima te tvrdnje o njihovim fizičkim realizatorima. Stoga se čini da prihvaćanje višestruke realizacije i pojma redukcije koji sugerira Richardson nije kompatibilno s ontologijom koju prihvaćaju zastupnici teorije identiteta tipova.

6.7 Redukcije koje su specifične za pojedine vrste ili individue

Drugi važan odgovor na Fodorov antiredukcionistički argument fokusira se na njegovo gledište da se funkcionalna analiza mentalnih predikata ne može pomiriti s redukcijom psihologije na neuroznanost. Vidjeli smo da se ova pretpostavka eksplicira u premisama 3) prema kojoj se mentalna svojstva

[54] Važno je primijetiti da u slučaju jakog funkcionalizma, gdje je svaki mentalni pojam uveden putem implicitnih Ramseyjevih definicija, sustav ima to fizičko svojstvo kada u njemu možemo pronaći dovoljno složenu uzročnu strukturu da ovaj sustav učini modelom funkcionalno-psihološke teorije T, u kojoj je definirano mentalno svojstvo (vidi poglavlje 5).

mogu višestruko realizirati kroz različite vrste i 4) gdje se tvrdi da mogućnost višestruke realizacije nije kompatibilna s redukcionizmom.

Kim (1993; vidi, također Lewis 1994) ukazuje na to da mogućnost višestruke realizacije ne predstavlja problem za redukcionizam, već daje razlog da se redukcija ograniči na određene skupove organizama. Na primjer, ako je bol višestruko ostvarljiva kod ljudi i hobotnica, onda trebamo zaključiti da se teorija boli kod ljudi reducira na teoriju o neuralnim stanjima koja realiziraju bol kod ljudi, dok se teorija boli kod hobotnica reducira na neuroznanstvene teorije koje se bave njihovim živčanim sustavima. Ova ideja se prema Kimu može formalnije zahvatiti na sljedeći način:

(1) $M(a) \leftrightarrow F_1(a) \vee F_2(a) \vee F_3(a)$

Iz te tvrdnje možemo zaključiti da:

(i) mentalni predikat M ne referira na prirodnu vrstu.

(ii) Ako je S skup nužnih i dovoljnih uvjeta koji omogućuju nekom primjerku da ga se dodijeli određenoj vrsti organizma, onda za svaki mentalni predikat M vrijedi:

(2) $(S(a) \wedge M(a) \leftrightarrow F_1(a))$

Ovaj bikondicional omogućuje redukciju tako da se psihološke teorije ograniče na određene *vrste* organizama. Stoga se psihologija može reducirati na neuroznanost ako nam nije stalo da imamo psihološke teorije koje uključuju generalizacije koje nadilaze pojedine vrste organizama. Međutim, ova vrsta odgovora na Fodorov argument susrela se s mnogo kritika. U nastavku ćemo razmotriti one koje se najčešće spominju.

Može se prigovoriti da je ideja o kompatibilnosti funkcionalne analize mentalnih svojstava i lokalnih redukcija koje su specifične za određenu vrstu previše *ad hoc*. Osim toga moglo bi se tvrditi da se tom idejom radi neprincipijelna razlika između psiho-neuralnih redukcija u odnosu na sve ostale slučajeve redukcije u znanosti. Na primjer, u slučaju redukcije teorije topline na mehaniku kako smo je predstavili u odjeljku 6.3 ne spominje se redukcija koja bi bila specifična za samo neku domenu termodinamike kao što se spominje kada govorimo o lokalnim redukcijama psiholoških svojstva na pojedine vrste organizama.

Na ove prigovore bi se moglo odgovoriti ukazivanjem na to da unatoč prividima čak i u klasičnom slučaju redukcije termodinamike postoji lokalna redukcija koja je specifična za stanja niske energije. Malo konkretnije, temperatura plazme, tj. ionizirane tvari koja nastaje pri visokim temperaturama, ne može se identificirati s prosječnom kinetičkom energijom molekula (vidi P. S. Churchland 1986, 356–57). No, to svejedno ne

implicira da treba odbaciti lokalnu redukciju temperature kada govorimo o domenama termodinamike gdje se ona može identificirati s prosječnom kinetičkom energijom molekula. Dakle, sama ideja lokalne redukcije ne izgleda *ad hoc*, već ima pandan u paradigmatskim znanstvenim područjima poput termodinamike. Unatoč tome, ideja da se redukcija treba provoditi lokalno prema domenama u znanosti, vrstama organizama i sličnome susreće se s dodatnim poteškoćama.

Iako lokalne redukcije mogu zahvatiti slučajeve višestruke realizacije kod različitih vrsta, neki autori ukazuju na to da postoje problematičniji slučajevi višestruke realizacije kada govorimo o primjercima iste vrste (Block 1978; Tye 1983; Endicott 1993). Možemo zamisliti, na primjer, da je kod određene manjine pojedinaca unutar ljudske vrste bol realizirana kroz neuralne procese koji su različiti od onih kod većine ljudi (vidi Tye 1983, 165). Mogućnost takve vrste višestruke realizacije ukazuje na to da, čak i kada govorimo o primjercima iste vrste, ista fizička stanja neće nužno realizirati iste uzročne uloge (vidi također poglavlje 4).

Neki na ovaj prigovor odgovaraju da se redukcija treba još više lokalizirati. Mogli bismo reći da su mentalna stanja identična fizičkim stanjima određene osobe u određenom trenutku (Jackson, Pargetter, i Prior 1982; Braddon-Mitchell i Jackson 2007, 101–3). Ideja je da se vratimo originalnoj formulaciji teorije identiteta tipova, gdje se tvrdilo da će točna priroda identifikacija između mentalnih i fizičkih stanja, a time i točna restrikcija domene na koje primjenjujemo principe premošćivanja i redukcije koje omogućuju, biti određena empirijskim istraživanjima. U tom smislu, Frank Jackson, Robert Pargetter i Elizabeth Prior predlažu da se funkcionalistička analiza kombinira s teorijom identiteta tipova i redukcionizmom na sljedeći način:

> Ukratko, teoriju identiteta tipova interpretiramo na način da je svaka (vrsta) mentalnog stanja (vrsta) stanja mozga, posebice (vrsta) stanja mozga koje za organizam u to vrijeme realizira funkcionalnu ulogu koja određuje što znači biti u tom mentalnom stanju. A pitanje možemo li ili u kojoj mjeri možemo odbaciti ograničenja na određene organizme i vremena stvar je empirijskog istraživanja. (Jackson, Pargetter, i Prior 1982, 212)

Prema Jacksonu i kolegama (1982, 210) ono što čini neko stanje organizma O stanjem boli u trenutku *t* je da okupira određenu uzročnu ulogu u trenutku *t.* Tu ulogu možemo nazvati bol-uloga. Iz toga zaključuju da ako struktura u mozgu, nazovimo je S, igra ili okupira bol-ulogu u trenutku *t* kod organizma O, onda slijedi da struktura mozga S = stanje boli organizma O u trenutku *t*. Struktura S može biti struktura u mozgu organizma koja je ista kroz cijeli životni vijek tog organizma. U tom slučaju mogli bismo ispustiti referencu na vrijeme i jednostavno reći da bol za organizam O nije ništa drugo nego biti u stanju mozga S. Međutim, Jackson i kolege (1982, 211–12) smatraju da čak i

ako se struktura mozga koja igra bol-ulogu mijenja od trenutka do trenutka to ne narušava redukcionističku tvrdnju da se mentalna svojstva svode na neka fizička svojstva.

Dakle, čini se da redukcionisti imaju dovoljno pojmovnih resursa da se obrane od Fodorova antiredukcionističkog argumenta. Kao što smo vidjeli, jedna moguća opcija uključuje prihvaćanje ideje da se mentalna svojstva, ovisno o domeni i vremenu istrage, mogu reducirati na fizička svojstva pojedinog primjerka određene vrste. U tom pogledu, ispostavlja se da je prihvaćanje funkcionalne analize mentalnih svojstava kompatibilno s prihvaćanjem interteorijskog redukcionizma. Međutim, ovakvi kompatibilistički pristupi suočavaju se sa sljedećim problemom.

Ako se mentalna stanja reduciraju pomoću lokalno određenih principa premošćivanja možemo se zapitati kakve implikacije to gledište ima za postojanje mentalnih svojstava (Block 1978). Ako vjerujemo da postoji *generički* pojam boli koji referira na svojstvo koje dijele, na primjer, ljudi i delfini, onda moramo zaključiti da to ne može biti njihova fizička realizacija niti uzročna uloga koju fizička stanja igraju kod tih vrsta. Štoviše, ako se prihvati prijedlog koji daju Jackson, Pargetter i Prior (1982), izgleda da više nemamo načina za reći koje bi to bilo mentalno stanje koje različite vrste dijele. Ako u trenutku *t* ulogu-boli kod ljudi igraju C-vlakna dok kod delfina igraju hipotetička D-vlakna, onda izgleda da bol kao generalno svojstvo koje mogu dijeliti ljudi i delfini ne postoji. Izgleda da postoje samo bolovi koji su individuirani s obzirom na pojedinu vrstu ili, još gore, postoji samo bol koja se manifestira u pojedinim primjercima unutar određene vrste u određenom trenutku. Iako možda ova razmatranja ne daju konkluzivan prigovor interteorijskoj redukciji, ona u najmanju ruku pokazuju da pobornici radikalnog lokalnog redukcionizma moraju ponuditi jasnije objašnjenje pojma mentalnog stanja i uzročnih uloga koje ga realiziraju.

Jedan od načina na koji se može probati odgovoriti na prethodni prigovor jest sljedeći. Možemo primijetiti da iako dva predmeta ne dijele nužno svojstva koja realiziraju ili instanciraju neko drugo svojstvo, svejedno mogu dijeliti to drugo instancirano svojstvo. Na primjer, zamislimo dva različita predmeta od kojih je jedan u potpunosti crven, dok je drugi u potpunosti plav. To su svojstva prvog reda koja naši zamišljeni predmeti imaju. No, iako nemaju istu boju te time ne dijele svojstva prvog reda, svejedno imaju zajedničko svojstvo koje je određeno činjenicom da su obojeni. U ovom je slučaju svojstvo biti obojen svojstvo drugog reda. Dakle, možemo primijetiti da se svojstvo drugog reda definira u odnosu na svojstva prvog reda. Ako je predmet crven, onda možemo reći da ima svojstvo biti obojen. Dakle, ako predmet ima svojstvo prvog reda, a to je da je crven, onda znači da ima svojstvo drugog reda biti obojen.

Na sličan način može se pokušati osmisliti objašnjenje prema kojemu iako mentalna svojstva imaju različite realizacije kroz različite vrste, te se prema

tome razlikuju u pogledu svojstava prvog reda, svejedno mogu dijeliti svojstva drugog reda koja se definiraju prema uzročnim ulogama koje svojstva prvog reda igraju kod različitih pojedinaca i vrsta organizama. Prema ovoj koncepciji bitno je razlikovati bol kao određeno *stanje* od boli kao *svojstva* koja određena stanja mogu instancirati. Pretpostavljamo da je kod ljudi bol identična s aktivacijom C-vlakana, dok je kod delfina identična s aktivacijom D-vlakana. Međutim, svojstvo biti bol nije identično s aktivacijom C ili D ili nekih drugih vlakana. Niti je svojstvo biti bol identično s disjunktivnim svojstvom biti aktivacija C vlakana ili biti aktivacija D vlakana. Biti bol je funkcionalno svojstvo koje sva ta stanja (zbog svojstava prvog reda koja imaju) dijele s obzirom na to da igraju relevantne uzročne uloge kod različitih vrsta organizama. S obzirom na tu razliku, svojstvo biti bol možemo definirati na sljedeći način:

> Individua *x* je u stanju boli ako i samo ako je *x* u stanju koje igra ili okupira bol-ulogu u vrsti kojoj *x* pripada.

Dakle, bol je kod delfina realizirana u stanju mozga koji smo nazvali aktivacija D-vlakana, dok se kod ljudi radi o aktivaciji C-vlakana. U tom smislu, biti aktivacija C-vlakana i biti aktivacija D-vlakana različita su svojstva prvog reda koja karakteriziraju mozgove ljudi i delfina. Ono što je zajedničko ljudima i delfinima je činjenica da se, kada su u boli, nalaze u stanju koje igra prikladnu uzročnu ulogu, tj. bol-ulogu koju povezujemo sa svojstvom biti bol. Drugim riječima, ono što ljudi i delfini dijele, unatoč njihovim neurofiziološkim razlikama koje određuju svojstva prvog reda, je funkcionalno svojstvo drugog reda, a to je da njihova posebna neurofiziološka svojstva instanciraju funkcionalno svojstvo drugog reda. Stoga se čini da postoji pojmovni prostor koji omogućuje redukcionistima da protiv Fodorova argumenta brane kompatibilizam između funkcionalne analize mentalnih svojstava i lokaliziranu verziju interteorijskog redukcionizma.

Vrijedi spomenuti da je dosadašnja rasprava pretpostavljala da su mentalna stanja višestruko ostvarljiva. Dugo se vremena ta pretpostavka jednostavno uzimala kao uvjerljiva hipoteza koja predstavlja kamen spoticanja za bilo koju verziju redukcionizma. Ona je bila toliko uvriježena da je Kim jednom prilikom spomenuo da je „konvencionalna mudrost u filozofiji uma da su mentalna stanja »višestruko realizirana« (…)“ (Kim 1993, 309). Međutim, u novijim se raspravama ta pretpostavka počela osporavati. Stoga ćemo se u nastavku osvrnuti na novije rasprave u kojima se kritički preispituje koji uvjeti moraju biti zadovoljeni kako bismo rekli da se neka vrsta ili svojstvo mogu višestruko realizirati.

6.8 Kriteriji određivanja mogućnosti višestruke realizacije i Fodorov argument

U nizu radova koji su rezultirali objavom knjige *The multiple realization book*, Larry Shapiro i Tom Polger (2016) kritiziraju Fodorov antiredukcionistički argument. Smatraju da prihvaćanje tvrdnje da su mentalna svojstva funkcionalna svojstva ne povlači tvrdnju da su mentalna svojstva višestruko ostvarljiva.[55] Dakle, napadaju premisu 3) u našoj rekonstrukciji Fodorova argumenta. Njihov se prigovor sastoji od dva koraka. Prvo, pokazuju da postoji pojmovni problem određivanja razlika i sličnosti koje moraju postojati između različitih događaja, stanja ili procesa koji konstituiraju moguće realizatore psiholoških svojstava. Drugo, oslanjajući se na empirijska razmatranja argumentiraju da nije tako da se mentalna svojstva mogu višestruko realizirati jer funkcionalna karakterizacija mentalnih svojstava, ili priroda njihovih pretpostavljenih realizatora, ne zadovoljava uvjete sličnosti i različitosti koji su utvrđeni u prethodnom pojmovnom koraku. To ih dovodi do zaključka da se argument višestruke realizacije često nelegitimno koristi protiv redukcionizma u filozofiji uma. U nastavku ćemo malo detaljnije razmotriti ova dva koraka njihove argumentacije.

Shapiro i Polger daju primjer otvarača za boce kako bi ilustrirali pojmovnu tvrdnju koja se odnosi na kriterije za odgovor na pitanje predstavlja li neki proces ili događaj jedan od mogućih realizatora određenog svojstva ili vrste. Svojstvo biti otvarač za boce ili vadičep može se smatrati funkcionalnom vrstom. Postavlja se pitanje kada smijemo reći da je otvarač za boce vrsta stvari koja se može višestruko realizirati? Zamislimo da imamo jednostavne vadičepove koji imaju samo svrdlo i držač, pomoću kojeg se svrdlo okreće i poteže, i koji su napravljeni od različitih materijala i imaju različite boje. Svi ti otvarači predstavljaju instance sličnog mehanizma koji obavlja funkciju koja definira vrstu stvari kojoj otvarači pripadaju. Ova vrsta varijacije ne opravdava zaključak da je vrsta ili svojstvo biti vadičep višestruko ostvarljiv. Dakle, iako se ovi otvarači razlikuju prema formi i materijalu od kojeg su napravljeni, svejedno instanciraju sličan mehanizam koji se sastoji od malog svrdla, polugice i ručice za povlačenje. U tom pogledu nije jasno da su realizatori ovih otvarača dovoljno različiti da bismo rekli da su višestruko realizirani. Ovaj primjer nam govori da nisu sve *varijacije* u sustavima koje ostvaruju istu funkciju primjeri višestruke realizacije. Često se nedovoljno pažnje pridaje ovoj tvrdnji kada se raspravlja o postojanju višestrukih realizatora određenog mentalnog svojstva.

Primjer vadičepa se također može koristiti za određivanje varijacija u realizatoru koje zaista predstavljaju višestruku realizaciju. Zamislimo dva vadičepa, od kojih jedan ima samo svrdlo i držač, dok drugi ima sofisticiraniji

[55] Za predstavljanje njihovih gledišta oslanjamo se na rad od Joséa Bermúdeza i Arnona Cahena (2020).

mehanizam koji se, osim svrdla, sastoji od poluge i zupčanika. Zupčanik koji drži svrdlo se uz pomoć drške okreće te kada se svrdlo dovoljno spusti, čep se vadi uz pomoć poluge. U ovom slučaju Shapiro i Polger smatraju da ovi primjeri otvarača opravdavaju zaključak da se vadičep kao vrsta stvari može višestruko realizirati jer „[…] ova dva uređaja koriste različite mehaničke principe“ (Polger i Shapiro 2016, 65).

U ovom primjeru višestruka ostvarljivost pretpostavlja realizatore koji su sastavljeni od različitih mehanizama koji izvode istu funkciju. Sada se postavlja pitanje kako odlučiti u pojedinom slučaju radi li se o varijacijama u instancijaciji funkcije koje podrazumijevaju višestruku realizaciju? Ili, još važnije za naš kontekst, pitanje je što određuje je li određeno mentalno svojstvo višestruko realizirano u različitim neuralnim procesima?

Polger i Shapiro smatraju da je to otvoreno empirijsko pitanje na koje bi nam znanstvena istraživanja trebala dati odgovor. Točnije, smatraju da se u odnosu na određeni taksonomski sustav koji pripada određenoj znanstvenoj disciplini specificiraju kriteriji koji određuju koje varijacije smijemo smatrati primjerima funkcionalne vrste koja se može višestruko realizirati. U tom pogledu navode sljedeće:

> Znanosti koriste mnoge metode za vrednovanje istosti i različitosti između stvari; grupiraju svoje teme prema različitim sličnostima i razlikama, često uzimajući u obzir razmatranja koja mogu iznenaditi autsajdere. Posljedično, nema prečaca u usporedbi i suprotstavljanju taksonomija stvarnih znanosti ili njihovih uzoraka, poput onih uključenih u pojedinačna objašnjenja ili modele […]. Pitanje višestruke realizacije je pitanje koje se odnosi na stvarne znanosti i uvijek je specifično i kontrastivno. Pitanje nije: „Jesu li oči višestruko realizirane?“, već je: „Je li vrsta biti oko u znanosti A višestruko realizirana kroz vrste K_1–K_n u znanosti B?“ (Polger i Shapiro 2016, 64)

U nastavku ćemo taksonomski sustav označiti slovom „S“. Polger i Shapiro navode sljedeće uvjete koje primjerci A i B trebaju zadovoljiti kako bismo bili opravdani tvrditi da predstavljaju višestruku realizaciju iste vrste u odnosu na određeni taksonomski sustav:

(i) A i B pripadaju istoj vrsti u modelu ili taksonomskom sustavu S_1. U našem slučaju, S_1 će biti neka psihološka teorija i njezini sustavi klasifikacija.

(ii) A i B pripadaju različitim vrstama u modelu ili taksonomskom sustavu S_2. U našem slučaju, S_2 će biti neka neuroznanstvena teorija.

(iii) Faktori koji upućuju na to da se A i B trebaju različito klasificirati u S_2 moraju biti među onima koji ih upućuju prema zajedničkoj klasifikaciji u S_1.

(iv) Relevantne varijacije u S_2 između A i B moraju biti različite u odnosu na unutar vrsne S_1 varijacije između A i B. (Polger i Shapiro 2016, 67)

Nakon što su ustanovili kriterije koji određuju je li neka vrsta višestruko ostvarljiva, Polger i Shapiro argumentiraju da empirijski dokazi pokazuju da psihološka svojstva zapravo nisu višestruko ostvarljiva kroz različite vrste organizama.

Pojašnjavaju svoju tvrdnju na primjeru oči i vida za koje se često uzima da mogu biti višestruko realizirani kod različitih vrsta organizama. Ako vid shvatimo na vrlo apstraktan način, kao nešto čija je funkcija da skuplja određenu vrstu informacije iz okoline, mogli bismo reći da su oči višestruko ostvarljive jer, na primjer, kod ljudi i hobotnica različiti mehanizmi realiziraju funkciju vida. Međutim, jednom kada promotrimo malo detaljnije kako oko funkcionira kod različitih vrsta i stoga kako funkcioniraju mehanizmi koji realiziraju te funkcije, možemo doći do zaključka da oni ipak ne realiziraju istu perceptivnu funkciju. Na primjer, dok u ljudskom oku postoje receptori koji nam omogućuju da vidimo boje, u očima hobotnica nema takvih receptora pa su one slijepe na boje. Kada se te činjenice uzmu u obzir, onda prestaje biti jasno da ove različite vrste očiju zaista realiziraju *istu* perceptivnu funkciju. Štoviše, izgleda da oči kod ljudi i hobotnica omogućuju različite perceptivne funkcije jer u jednom slučaju omogućuju vid boja dok u drugom slučaju omogućuju monokromatski vid.

Slično tome, Polger i Shapiro argumentiraju da ćemo, jednom kada uzmemo u obzir zahtjeve (i) – (iv) za mogućnosti višestruke realizacije u odnosu na stvarne prakse znanstvene kategorizacije, uvidjeti da nema empirijskih dokaza za smatrati da su mentalna svojstva višestruko realizirana kroz različite vrste. Na primjer, na visokoj razini apstrakcije bol se može funkcionalno karakterizirati kao ono stanje koje aktivira ponašanje izbjegavanja u odnosu na štetan podražaj. S obzirom na takvu općenitu funkcionalnu karakterizaciju, tvrdilo se da se bol može višestruko realizirati u živčanim sustavima ljudi i hobotnica. Međutim, ako prihvatimo teoriju višestruke realizacije koja je zahvaćena u zahtjevima (i) – (iv) i pažljivije razmotrimo dostupnu empirijsku dokaznu građu može se tvrditi da zapravo bol nije višestruko ostvarljiva.

Da bi bol bila višestruko ostvarljiva, mora se pokazati u skladu sa zahtjevom (i) da su stanje A kod ljudi i stanje B kod hobotnica, budući da su stanja boli, isto stanje iz perspektive taksonomije (S_1) neke psihološke teorije. Istovremeno, kako se traži zahtjevom (ii), A i B moraju biti različiti iz perspektive neke neuroznanstvene teorije koja daje taksonomiju (S_2). Ta taksonomija bi trebala kategorizirati neuralne mehanizme koji realiziraju

psihološka svojstava. Međutim, ljudi i hobotnice dijele slične neurofiziološke mehanizme poput nociceptora (osjetilni živčani sustavi koji registriraju štetne podražaje). Ta činjenica dovodi u pitanje pretpostavku da su, iz perspektive taksonomije neuroznanosti (S_2), stanja A i B toliko različita da ih možemo smatrati *različitim* realizacijama iste psihološke funkcije. Nadalje, ljudi imaju toplinske nociceptore koje hobotnice nemaju. Dakle, u tom pogledu ljudi i hobotnice će se razlikovati jer ti toplinski nociceptori omogućuju ljudima da izbjegavaju podražaje koji ukazuju na povišene temperature, dok hobotnice neće pokazivati slično ponašanje u odnosu na toplinske podražaje. Međutim, nasuprot prvotnoj pretpostavci ova činjenica nam pokazuje da stanja A i B kod ljudi i hobotnica zapravo ne realiziraju istu psihološku funkciju. Stoga se može argumentirati da ljudi i hobotnice ne dijele bol kao psihološku vrstu jer imaju različite neuralne sustave koji omogućuju različite psihološke funkcije. S druge strane, u mjeri u kojoj imaju slične neuralne sustave, oni također omogućuju slična ponašanja koja povezujemo uz bol. Stoga se čini da unatoč početnim pretpostavkama bol kao psihološka funkcija nije višestruko realizirana kod ljudi i hobotnica.

Razlog zašto pobornici argumenta višestruke realizacije nisu bili osjetljivi na ovu vrstu razmatranja je vjerojatno taj što nisu dovoljno pažnje obraćali na istraživanja fizičke podloge mentalnih procesa i njihove funkcije. Kao što smo ranije istaknuli, Fodorovo i Putnamovo razmišljanje bilo je inspirirano razvojem kognitivnih znanosti koje se temeljilo na računalnoj metafori koja je u osnovi imala za cilj elaboraciju apstraktnih modela kognitivnog funkcioniranja koji se mogu implementirati kao računalni program. Međutim, u novijim radovima na ovu temu ističe se suprotan trend. Razvoj kognitivne neuroznanosti se u velikoj mjeri razvija na temelju usporedbe ljudskih i životinjskih modela koji pretpostavljaju značajne funkcionalne i fiziološke sličnosti između ljudi i drugih životinja (Call i ostali 2017). Na primjer, istraživanja procesiranja vizualnih podražaja kod majmuna rhesus macaque bila su vrlo važna za istraživanje relevantnih područja za procesiranje vizualnih podražaja kod ljudi (za raspravu, vidi Bechtel i Mundale 1999). Stoga, mnogi smatraju da ćemo pažljivijim praćenjem razvoja kognitivne neuroznanosti uvidjeti da, u suprotnosti s apriornim intuicijama koje se nalaze u podlozi argumenta višestruke realizacije, postoje značajne psihološke i neurološke sličnosti između različitih vrsta organizama (vidi Bechtel i Mundale 1999; usp. Figdor 2010; Barrett 2013).

6.9 Zaključak

U ovom smo se poglavlju bavili idejom redukcije u filozofiji uma. Kao relevantan pojam redukcije preuzeli smo ideju interteorijske redukcije kako ju je formulirao Ernest Nagel (1974). Vidjeli smo da je pod pretpostavkom takvog shvaćanja redukcije, Jerry Fodor (1974) argumentirao da se psihologija ne može reducirati na temeljnije fizikalne znanosti te je u tom

smislu zastupao varijantu nereduktivnog fizikalizma. Nakon toga smo razmotrili različite pojmovne i empirijski utemeljene odgovore na Fodorov utjecajni antiredukcionistički argument. U oba slučaja, vjerujemo da nije ponuđen konkluzivan argument protiv Fodora. Međutim, smatramo da argumenti koje smo razmotrili u ovom poglavlju u najmanju ruku pokazuju da Fodorov argument unatoč velikom utjecaju među antiredukcionistima u filozofiji uma pretpostavlja tvrdnje za čije odbacivanje imamo dobrih razloga.

Međutim, čak i ako prihvatimo antiredukcionističku konkluziju Fodorova argumenta, otvaraju se nova filozofska pitanja koja moramo razmotriti. Pod pretpostavkom da je nereduktivni fizikalizam uvjerljivo gledište, postavlja se pitanje kako bi se zapravo trebao shvatiti odnos između uma i tijela. Budući da je to fizikalistička pozicija, tvrdi se da su svojstva uma na neki način povezana s fizičkim svojstvima. No, budući da se istodobno radi o antiredukcionizmu, tvrdi se da se mentalna svojstva ne mogu reducirati na fizička svojstva. Kao odgovor na to pitanje, mnogi tvrde da se ontološki odnos između uma i tijela treba shvatiti u terminima odnosa supervenijencije. Stoga ćemo se u sljedećem poglavlju baviti pojmom supervenijencije te ćemo razmotriti s kakvim poteškoćama se susreće nereduktivni fizikalizam koji se temelji na njemu.

7 Supervenijencija i antiredukcionizam

7.1 Uvod

Pod pretpostavkom plauzibilnosti nereduktivnog fizikalizma, postavlja se pitanje kako bismo trebali shvatiti ontološki odnos između uma i tijela. U ovom poglavlju razmotrit ćemo gledište prema kojemu se odnos uma i tijela, ili mentalnih i fizičkih svojstava, treba shvatiti u terminima relacije supervenijencije. Ovaj pojam smatra se ključnim u mnogim antiredukcionističkim gledištima u suvremenoj filozofiji uma (McLaughlin i Bennett 2018).

Izraz „supervenijencija“ označava filozofski pojam kojim se karakterizira vrsta metafizičke ili pojmovne relacije između skupova entiteta, poput stanja stvari ili svojstava. Filozofi koriste pojam supervenijencija u različitim kontekstima kako bi riješili različite filozofske probleme. U trenutnom filozofskom značenju, izraz „supervenijencija“ među prvima koristi moralni filozof Richard Hare (1998). On je uveo izraz „supervenijencija“ kako bi imenovao pojam koji je već od ranije u moralnoj filozofiji koristio poznati filozof G. E. Moore (1922). Naime, Moore je smatrao da se moralna svojstva ne mogu reducirati na fizička svojstva, no također je smatrao da su moralna svojstva na neki način određena fizičkim svojstvima. U smislu da, ako imamo dvije osobe koje postupaju na isti način, i u ostalim relevantnim aspektima su slične, onda bismo i njihove radnje trebali jednako moralno vrednovati (usp. McLaughlin i Bennett 2018). Pojam supervenijencije u filozofiji uma popularizira Donald Davidson kako bi opisao odnos između mentalnih i fizičkih događaja. Taj odnos opisuje na sljedeći način:

> mentalne karakteristike su u određenom smislu ovisne, ili supervenijentne, o fizičkim karakteristikama. [… To] znači da ne mogu postojati dva događaja koja su ista u svim fizičkim aspektima, ali različita u nekom mentalnom aspektu, ili da se predmet ne može promijeniti u nekom mentalnom aspektu bez promjene u nekom fizičkom aspektu. (Davidson 2001b, 214)

U filozofiji uma supervenijencija se koristi kako bi se na nereduktivan način objasnio odnos između mentalnih i fizičkih svojstava. Vidjeli smo da antiredukcionisti obično smatraju da su mentalna svojstva prava svojstva koja se ne mogu reducirati na fizička svojstva. Međutim, neki autori smatraju da se nereducibilnost mentalnosti na fizičko treba shvatiti na razini opisa, teorija ili objašnjenja te u tom smislu smatraju da fizički svijet, unatoč tome što je samo jedan, možemo opisivati na različite načine. Na primjer, možemo ga opisivati koristeći psihološke teorije ili fizikalne teorije koje se ne mogu reducirati jedne na druge (vidi, npr. Davidson 2001b). Stoga, ovisno o tome koju varijantu antiredukcionizma uzimamo u obzir, postoje dva načina na koji se supervenijencija može shvatiti.

U tom pogledu, korisno je podijeliti pojam supervenijencije na *askriptivnu* (tj. *pripisujuću*) i *ontološku* supervenijenciju (Klagge 1988; Horgan 1993). Askriptivna supervenijencija je pojmovna ili logička relacija između sudova ili rečenica pomoću kojih ih izražavamo. S druge strane, ontološka supervenijencija vezana je uz odnos između stvari po sebi bez obzira kako ih opisujemo. U nastavku rasprave govorit ćemo samo o ontološkom shvaćanju supervenijencije. Ovakvo je gledište u skladu s nereduktivnim gledištem na psihologiju kakvo su zastupali Fodor i drugi funkcionalisti čija smo gledišta razmatrali u prijašnjim poglavljima.

Ostatak poglavlja je podijeljen na sljedeći način. Prvo ćemo objasniti kako se shvaća pojam supervenijencije u filozofiji uma. Razlikovat ćemo slabiju i jaču tezu supervenijencije. Pokazat ćemo da antiredukcionisti moraju prihvatiti jaču tezu supervenijencije. Nakon toga ćemo preći na probleme s kojima se susreću antiredukcionisti fizikalisti ako prihvate jaču tezu supervenijencije. U zadnjem dijelu poglavlja, razmotrit ćemo poznati argument uzročne isključivosti koji daje Jaegwon Kim (1993). Kim pokazuje da se čini da, bez obzira na to kako se shvati pojam supervenijencije, antiredukcionistički fizikalizam nije stabilna pozicija. Naime, ako se pretpostavi da mentalna svojstva imaju uzročne moći, onda se čini da ih jedino mogu imati ako ih poistovjetimo s uzročnim moćima njihovih fizičkih realizatora. Taj argument nas stoga ponovo navodi na prihvaćanje neke varijante redukcionizma u filozofiji uma.

7.2 Teze antiredukcionizma i supervenijencija

Tri pretpostavke karakteriziraju varijantu nereduktivnog realizma u pogledu mentalnih svojstava o kojoj ćemo govoriti u nastavku. Prvo, smatra se da mentalna svojstva postoje (teza mentalnog realizma). Drugo, smatra se da su sva fizička svojstva ontološki primarna u odnosu na mentalna svojstva te ih na neki način određuju (teza ovisnosti). Treće, mentalna svojstva nisu fizička svojstva (teza ontološkog antiredukcionizma). U skladu s tim pretpostavkama potrebno je formulirati pojam supervenijencije koji će

zahvatiti intuiciju izraženu u tezi ovisnosti tako da bude konzistentna s prihvaćanjem mentalnog realizma i antiredukcionizma.

Široki spektar relacija supervenijencije može se odrediti s obzirom na dvije dimenzije. Prvo, filozofi imaju suprotna mišljenja o prirodi stvari koje stoje u odnosu supervenijencije. Prema tradiciji u filozofiji uma koju uspostavlja Davidson (2001b), pojam supervenijencija bi trebao zahvatiti intuitivnu ideju da „nema mentalnih razlika bez fizičkih razlika". Kao što ćemo vidjeti, utjecajan način tumačenja ove tvrdnje uključuje shvaćanje supervenijencije kao odnosa između skupova svojstava (Kim 1994; McLaughlin i Bennett 2018).

Drugo, postoje razlike među filozofima u pogledu toga kako shvatiti modalni odnos između predmeta koji stoje u relaciji supervenijencije. Obično se u obzir uzimaju tri vrste modalnih odnosa. Prema fizičkoj ili nomološkoj nužnosti rečenica A implicira rečenicu B ako postoji neki zakon prirode koji povezuje stvari opisane rečenicom A i stvari opisane rečenicom B. Primjer fizičke ili nomološke nužnosti bio bi iskaz „Ako se voda zagrije na 100 °C, onda će proključati". Ovaj iskaz je istinit jer je jedan od zakona prirode taj da voda u normalnim uvjetima ključa na 100 °C. Međutim, nije kontradiktorno zamisliti mogući svijet u kojem u normalnim uvjetima voda ne proključa na 100 °C nego na nekoj drugoj temperaturi.

Prema logičkom shvaćanju nužnosti, ako A implicira B, onda je istinitost te tvrdnje ovisna samo o zakonima logike. Na primjer, iskaz „Svi trokuti imaju tri kuta" istinita je rečenica čija istinitost ne ovisi o slučajnim ili akcidentalnim činjenicama ili zakonitostima koje vrijede u našem ili nekom drugom mogućem svijetu. Taj je iskaz istinit u svim mogućim svjetovima.

Konačno, postoji treće, metafizičko shvaćanje nužnosti koje bi po jačini trebalo biti između fizičke i logičke nužnosti. Prema metafizičkom poimanju nužnosti iskazi su istiniti kada određuju nužna svojstva predmeta o kojima govorimo. Drugim riječima, pomoću metafizičke nužnosti možemo karakterizirati esencijalna svojstava stvari, tj. ona svojstva koja odražavaju samu bit stvari (za više, vidi Kripke 1997). Na primjer, neki smatraju da genetski materijal koji smo dobili od roditelja predstavlja jedno od naših esencijalnih svojstava, u smislu da kada se ne bismo rodili s genetskim kodom koji stvarno imamo, to ne bismo bili mi nego neko drugo biće. U tom smislu, iskaz „DNA koju smo naslijedili od roditelja barem djelomično određuje naš osobni identitet" predstavljao bi metafizički nužnu istinu. Taj iskaz bi trebao biti istinit u svim mogućim svjetovima u kojima mi postojimo. Slično, ako je esencijalno svojstvo vode da ima molekularnu strukturu H_2O, onda je metafizički nužna istina da je voda H_2O. Dakle, u svakom mogućem svijetu u kojem postoji voda imat će molekularnu strukturu H_2O.

Odnos jačine ovih vrsta modalnosti bi trebao biti sljedeći. Budući da je metafizička nužnost između logičke i fizičke nužnosti, onda to povlači da su sve metafizički nužno istinite tvrdnje ujedno i nomološki nužne. No, obrnuto

ne vrijedi. Na primjer, iako u našem svijetu vrijedi zakon gravitacije, možemo zamisliti da u nekom drugom mogućem svijetu ne postoji zakon gravitacije kakav je kod nas. Međutim, ako postoji svijet u kojem se nalazi voda, onda u tom svijetu postoji nešto što ima molekularnu strukturu H_2O. Kada ta stvar ne bi imala strukturu H_2O onda to ne bi bila voda nego neka druga stvar. Slično prethodnom odnosu, sve logičke istine su metafizički nužne, no obrnuto ne bi trebalo vrijediti. Međutim, treba istaknuti da je manje jasno kako razlikovati logičke od metafizičke nužnosti. Naime, metafizička i logička nužnost se određuju u odnosu na tvrdnje koje su istinite u svim mogućim svjetovima. Jedan način na koji bismo ih mogli razlikovati odnosi se na temelje koji određuju njihovu istinitost. S jedne strane, iskaz „Trokut ima tri stranice“ nužno je istinit zato što naš pojam trokuta uključuje pojam predmeta koji ima tri stranice te u tom smislu logičke istine predstavljaju pojmovne istine. S druge strane, iskaz „Voda je H_2O“ nužno je istinit zbog empirijske prirode vode, ne zbog načina na koji koristimo pojmove VODA i H_2O. Stoga možemo reći da logičke istine vrijede u svim mogućim svjetovima bez obzira na to koji predmeti postoje u njima, dok metafizičke istine vrijede u svim mogućim svjetovima s obzirom na prirodu predmeta koji se nalaze u njima.

U kontekstu naše rasprave obično se uzima metafizičko shvaćanje nužnosti kao relevantno. Naime, odnos supervenijencije trebao bi zahvatiti odnos između fizičkih i mentalnih entiteta, a ne odnos između pojmova koje koristimo kada govorimo o njima. Stoga ćemo u nastavku pretpostaviti da odnos supervenijencije podrazumijeva metafizičku vrstu nužnosti.

Prije nego uvedemo različite pojmove supervenijencije, objasnit ćemo tehničke pojmove pomoću kojih ćemo ih definirati. Smatrat ćemo da $\mathbf{M}$ označuje skup mentalnih svojstava, tako da $\mathbf{M} = \{M_1, \ldots, M_n\}$. Smatrat ćemo da **F** označuje skup fizičkih svojstava, tako da $\mathbf{F} = \{F_1, \ldots, F_n\}$. Reći ćemo da su dva predmeta *x* i *y* fizički nerazlučivi (ili mentalno nerazlučivi) kada vrijedi odnos $x =_{\mathbf{F}} y$ (ili $x =_{\mathbf{M}} y$). S obzirom na $\mathbf{M}$ i **F**, reći ćemo da su dva predmeta *x* i *y* fizički nerazlučivi ($x =_{\mathbf{F}} y$) (ili mentalno nerazlučivi ($x =_{\mathbf{M}} y$)), ako za svaki *i*, gdje $1 \leq i \leq n$, $F_i(x) \leftrightarrow F_i(y)$ (ili $M_i(x) \leftrightarrow M_i(y)$). Sada možemo preći na formulaciju pojma supervenijencije (za detaljniji pregled, vidi McLaughlin i Bennett 2018).

Razmotrimo sljedeću definiciju supervenijencije koju možemo nazvati teza *slabe supervenijencije*:

> **M** *slabo supervenira* nad **F**, ako i samo ako za bilo koji mogući svijet *s* i za sve predmete *x* i *y* koji se nalaze u *s*-u, vrijedi da ako $x =_{\mathbf{F}} y$ u *s* onda $x =_{\mathbf{M}} y$ u *s*.

Tezom slabe supervenijencije tvrdi se da ne postoji mogući svijet u kojem bi predmeti ili pojedinci bili nerazlučivi u pogledu fizičkih svojstava, a razlučivi u pogledu mentalnih svojstava.

Postavlja se pitanje je li slaba supervenijencija dovoljna kako bi se analizirao ontološki odnos ovisnosti mentalnog o fizičkom koji je potreban za uvjerljivu formulaciju antiredukcionističkog mentalnog realizma? Kako bismo odgovorili na to pitanje moramo se pomnije usredotočiti na prirodu relacije ovisnosti koja se podrazumijeva u ovom kontekstu.

Relacija ovisnosti između mentalnih i fizičkih svojstava trebala bi podržavati odgovarajuće kontrafaktičke kondicionale (vidi Beckermann 1992, 12). Na primjer, ako je svojstvo biti bol ovisno o svojstvu biti aktivacija C-vlakana, onda bi iskazi poput „Da se u Marijinom živčanom sustavu nisu aktivirala C-vlakna ona ne bi osjećala bol" trebali biti istiniti. U tom smislu, čini se da relevantna relacija ovisnosti podrazumijeva istinitost kontrafaktičkih kondicionala koji govore o odnosima između mentalnih i fizičkih svojstava. Govor o kontrafaktičkim situacijama može se shvatiti kao govor o tome kako je neki mogući svijet mogao izgledati nasuprot tome kako zaista izgleda. U tom smislu, istinitost i neistinitost kontrafaktičkih kondicionala određujemo u odnosu na moguće svjetove koji su identični našem svijetu, osim što se razlikuju u pogledu onoga što se tvrdi u antecedensu kontrafaktičkog kondicionala. Na primjer, kada kažemo da Marija ne bi osjećala bol da se u njezinom živčanom sustavu nisu aktivirala C-vlakna, taj iskaz vrednujemo tako da gledamo mogući svijet koji je identičan našem osim što kod Marije nisu aktivirana C-vlakna. Ako je taj kontrafaktički kondicional istinit, onda će u tom mogućem svijetu biti istina da Marija ne osjeća bol kada u njezinom živčanom sustavu C-vlakna nisu aktivirana.

Međutim, tezom slabe supervenijencije tvrdi se samo da je nužno da ako osoba ima određeno fizičko svojstvo u nekom mogućem svijetu ima i određeno mentalno svojstvo u istom svijetu, a ne govori ništa o odnosima mentalnih i fizičkih svojstava u kontrafaktičkim situacijama. Stoga ostavlja otvorenom mogućnost da u nekom mogućem svijetu postoji osoba koja ima sva fizička svojstva koja u našem aktualnom svijetu *de facto* određuju neki skup mentalnih svojstava, a da u tom mogućem svijetu ne posjeduje ta nego neka druga mentalna svojstva. Na primjer, dopušta se da, iako u našem aktualnom svijetu aktivacija C-vlakana uvijek popraćena osjećajem boli, postoji mogući svijet u kojem C-vlakna mogu biti aktivirana, a da osoba ne osjeća bol. Dopušta se, na primjer, da ta osoba osjeća nešto drugo poput blagog svrbeža, škakljanja ili nečeg sasvim trećeg. Drugim riječima, slaba supervenijencija ne podržava relevantne kontrafaktičke kondicionale jer se dopušta da kontrafaktički kondicional poput „Da se C-vlakna nisu aktivirala, onda osoba ne bi osjećala bol" ne izražavaju istinite tvrdnje. To nam pokazuje da bi modalna veza između mentalnih svojstava koja superveniraju nad fizičkim svojstvima trebala biti snažnija nego ona koja se podrazumijeva u tezi slabe supervenijencije.

Prema tezi jake supervenijencije fizička nerazlučivost implicira mentalnu nerazlučivosti kroz sve moguće svjetove. Teza se može formulirati na sljedeći način:

> **M** jako supervenira nad **F** ako i samo ako za bilo koji mogući svijet s_i i s_j, te za bilo koji predmet *x* koji se nalazi u s_i i za bilo koji predmet *y* koji se nalazi u s_j, ako je $x =_{\mathbf{F}} y$ onda $x =_{\mathbf{M}} y$.

Tezom jake supervenijencije tvrdi se da bez obzira u kojem mogućem svijetu se nalazili, ne postoje predmeti ili pojedinci koji bi bili nerazlučivi u fizičkim aspektima, a razlučivi u mentalnim aspektima. Radi ilustracije ove teze pretpostavimo da Ivana živi u našem aktualnom svijetu i posjeduje svojstva **F** koja uključuju aktivaciju C-vlakana. Kada su, u našem aktualnom svijetu, kod Ivane aktivirana C-vlakna, onda ona osjeća bol koja pripada skupu **M**. Pretpostavimo da Marija živi u nekom drugom mogućem svijetu i posjeduje ista svojstva **F**. Prema jakoj supervenijenciji, ako su kod Marije aktivirana C-vlakna, onda ona također osjeća bol, tj. ima ista mentalna svojstva iz skupa **M**. Možemo primijetiti da jaka supervenijencija zadovoljava uvjet relacije ovisnosti koja podržava relevantne kontrafaktičke kondicionale. Kontrafaktički kondicional „Da kod Ivane nisu aktivirana C-vlakna, onda ne bi osjećala bol“ istinit je jer ne postoji neki drugi mogući svijet u kojem aktivacija C-vlakana ne bi bila popraćena osjećajem boli, tj. u kojem bi osoba bila fizička nerazlučiva od Ivane, a mentalno razlučiva.

Stoga se čini da jaka supervenijencija bolje zahvaća modalnu snagu relacije ovisnosti koju antiredukcionistički fizikalisti pretpostavljaju. Međutim, određene poteškoće pojavljuju se kada pomoću relacije supervenijencije pokušamo objasniti odnos ovisnosti koji podrazumijevaju antiredukcionisti (McLaughlin i Bennett 2018). Prvo što se može primijetiti jest da je relacija ovisnosti koju fizikalisti pretpostavljaju asimetrična, u smislu da fizička svojstva imaju temeljnu ontološku ulogu u odnosu na mentalna svojstva. Dakle, prema fizikalizmu nije moguće da fizička svojstva superveniraju nad mentalnim svojstvima. Međutim, kako je trenutno definiran odnos jake supervenijencije on po sebi ne uspijeva zahvatiti tu asimetričnost relacije ovisnosti. Trenutno je pojam supervenijencije određen kao neka vrsta kovarijacije između svojstava. Pretpostavljamo da promjene u fizičkim svojstvima kovariraju s promjenama u mentalnim svojstvima, a da se mentalna svojstva ne reduciraju na fizička svojstva. Shvaćanje supervenijencije kao određene vrste kovarijacije između svojstava nije dovoljno da se zahvati odnos ovisnosti koji podrazumijevaju pobornici antiredukcionističkog fizikalizma. Kako Kim ističe u sljedećem odlomku, relacija kovarijacije nije ni asimetrična, ni antisimetrična:

> […] zamislite domenu savršenih sfera. Površina svake sfere jako kovarira sa svojim obujmom, i obrnuto, obujam s površinom. I

> ne želimo reći da jedan određuje, ili ovisi o drugom, u bilo kojem smislu ovih pojmova koji implicira asimetriju. Postoji samo funkcionalno određenje, i ovisnost, u oba smjera; ali oklijevali bismo imputirati metafizičku ili ontološku ovisnost u bilo kojem smjeru. (Kim 1990, 144)

Nadalje, kovarijacija između svojstava ne implicira relaciju ovisnosti između tih svojstava jer ne možemo isključiti mogućnosti da svojstva koja kovariraju ovise o nekom trećem skupu svojstava. Kim ilustrira takvu mogućnost na sljedeći način:

> Čuo sam da postoji kovarijacija između inteligencije mjerene IQ-om i spretnosti s rukama. Moguće je da i spretnost s rukama i inteligencija ovise o određenim genetskim i razvojnim faktorima, te da inteligencija jako kovarira s vještim rukama, ali ne i obrnuto. Da je to slučaj, ne bismo smatrali da je inteligencija ovisna o ili određena posjedovanjem vještih ruku. (Kim 1990, 146)

Ova dva slučaja pokazuju da shvaćanje supervenijencije kao neke vrste kovarijacije između svojstava nije dovoljno za adekvatnu formulaciju pojma ontološke ovisnosti kakav podrazumijevaju pobornici antiredukcionističkog fizikalizma. Stoga adekvatna formulacija pojma supervenijencije mora uključiti dodatno pojašnjenje onoga u čemu se sastoji relacija ovisnosti i na koji način mentalna svojstva *ovise* o fizičkim svojstvima kako to podrazumijevaju antiredukcionisti.

Štoviše, ako antiredukcionisti ne ponude neko dodatno objašnjenje kako supervenijencija može objasniti ontološku ovisnost mentalnog o fizičkom, njihovom gledištu prijeti opasnost kolabiranja u neku varijantu redukcionizma. Pod pretpostavkom jake supervenijencije može se tvrditi da prihvaćanje supervenijencije obvezuje na zaključak da se mentalna svojstva reduciraju na fizička svojstva (vidi, npr. Kim 1990; 1993). Uvjerljivost ove tvrdnje postat će jasnija kada uzmemo u obzir jednu implikaciju teze jake supervenijencije. Ona se može formulirati na sljedeći način:

> **M** jako supervenira nad **F** ako i samo ako je nužno da za svaki predmet *x* i svako svojstvo M_i, ako *x* ima M_i, onda postoji fizičko svojstvo F_i koje je takvo da za svaki predmet *y* nužno je da ako *y* ima F_i, onda *y* ima M_i.

Ovom formulacijom jake teze tvrdi se da je posjedovanje fizičkog svojstva koje pripada supervenijentnoj bazi, tj. onome nad čime mentalno svojstvo supervenira, nomološki dovoljno za posjedovanje mentalnog supervenijentnog svojstva. Dakle, posjedovanje supervenijentne fizičke baze

nužno povlači posjedovanje mentalnog supervenijentnog svojstva. U tom su smislu, kondicionalne rečenice koje imaju formu „Ako F_i onda M_i" istinite te predstavljaju nomološki nužne korelacije između fizičkih i mentalnih svojstava. Takva vrsta korelacije sama po sebi ne implicira da se mentalna svojstva mogu reducirati na fizička svojstva. Sjetimo se da se često pretpostavlja da relacija redukcije zahtijeva postojanje bikondicionala kojima se tvrdi da su ekstenzije mentalnih predikata nomološki identične ekstenzijama fizikalnih predikata, tj. da su mentalna i fizička svojstva identična (za raspravu vidi poglavlje 7).

Međutim, Kim (1990, 151–52) argumentira da prihvaćanje jakog shvaćanja supervenijencije zapravo implicira da se mentalna svojstva mogu reducirati na fizička svojstva. Razmotrimo svojstvo M_i koje pripada skupu mentalnih svojstava **M**, koji jako supervenira nad skupom fizičkih svojstava **F**. Pod pretpostavkom jake supervenijencije slijedi da postoji podskup svojstava, koji možemo nazvati **F***, skupa fizičkih svojstava **F** koja su takva da, ako ih određena osoba posjeduje, onda posjeduje M_i. Kim tvrdi da je u tom slučaju M_i reducibilno na disjunkciju fizičkih svojstava koji čine **F***.[56] Ova tvrdnja može se dokazati tako da se pokaže da je M_i logički ekvivalentno disjunkciji **F***, tj. moramo pokazati istinitost bikondicionala: M_i ako i samo ako **F***. Kondicional s desna na lijevo (tj. **F*** $\rightarrow M_i$) relativno je lako dokazati. Naime, logička je istina da disjunktivni iskaz implicira sve što impliciraju njegovi pojedini disjunkti.[57] Budući da jedan od disjunkata iz **F*** predstavlja dovoljan uvjet da se pojavi svojstvo M_i, slijedi da cijela disjunkcija **F*** implicira M_i. Stoga ono što preostaje za pokazati jest da posjedovanje svojstva M_i također predstavlja dovoljan uvjet da se pokaže istinitost disjunkcije **F***, tj. da $M_i \rightarrow$ **F***.

U prilog te tvrdnje Kim daje sljedeći argument. Pretpostavimo da posjedovanje M_i nije dovoljan uvjet za posjedovanje nekog od svojstava iz **F***. Možemo pokazati da ta pretpostavka nije točna na sljedeći način. Ako osoba *a* ima svojstvo M_i, teza jake supervenijencije implicira da nužno postoji fizičko svojstvo F_i koje *a* posjeduje i dovoljno je za posjedovanje M_i. Budući da je F_i dovoljno za posjedovanje M_i, to znači da F_i pripada disjunkciji **F*** koja se sastoji od uvjeta koji su dovoljni za posjedovanje M_i. Ovo zaključivanje nam ukazuje na to da, suprotno inicijalnoj pretpostavci, za bilo koju osobu *a* vrijedi da je posjedovanje mentalnog svojstva M_i dovoljno da posjeduje fizičko svojstvo F_i te da se stoga disjunkcija **F*** odnosi na tu osobu (vidi Kim 1990, 152).

Ovaj je zaključak značajan jer pokazuje da prihvaćanje teze jake supervenijencije vodi u neku vrstu redukcionizma. Naime, pokazali smo da posjedovanje mentalnih svojstava pretpostavlja posjedovanje određenih

---

[56] **F*** se formalno može izraziti kao disjunkcija F_i v, …, v F_n.

[57] Ovaj se princip odnosi na sljedeću logičku istinu: $((P \rightarrow Q) \text{ v } (R \rightarrow Q)) \rightarrow (R \text{ v } P) \rightarrow Q)$, gdje P, Q i R predstavljaju propozicije ili iskaze.

fizičkih svojstava te posjedovanje tih fizičkih svojstava pretpostavlja posjedovanje mentalnih svojstava. Ako je pripisivanje mentalnih predikata u svim mogućim svjetovima ovisno o pripisivanju točno određenih fizičkih predikata, onda možemo zaključiti da oni zapravo referiraju na istovjetna svojstva. Ako je ovaj argument uvjerljiv, onda on pokazuje da antiredukcionisti ne mogu prihvatiti tezu jake supervenijencije.

Međutim, mnogi antiredukcionisti smatraju da ovaj način zaključivanja nije opravdan jer se temelji na neuvjerljivoj premisi. Naime, kako smo vidjeli u prošlom poglavlju, nije jasno da disjunktivna svojstva predstavljaju dobru reduktivnu bazu za druga svojstva (vidi poglavlje 6). Stoga se može tvrditi da skup **F*** ne predstavlja dobru fizičku bazu za ontološku redukciju mentalnih svojstava. Međutim, čini se da održivost ove vrste prigovora uključuje pretpostavke koje neće biti prihvatljive svim pobornicima antiredukcionističkog fizikalizma.[58] Kako bismo to uvidjeli trebamo se prisjetiti da se pojam redukcije može razumjeti na više načina (vidi poglavlje 6).

Formalni pristup redukciji kako ga je formulirao Nagel (1987), podrazumijeva da se relacija redukcije odnosi na teorije shvaćene kao lingvističke entitete, tj. skupove rečenica. U tom slučaju, svi su zahtjevi koje redukcija mora ispuniti sintaktički uvjeti koji povezuju predikate reducirajuće teorije s predikatima reducirane teorije. S druge strane, ako uzmemo u obzir ontološku perspektivu, ideja redukcije uma na tijelo podrazumijeva da se mentalne činjenice, svojstva ili događaji mogu identificirati s fizičkim entitetima istog tipa.

Antiredukcionisti obično argumentiraju protiv reduciranja mentalnih svojstava na beskonačne disjunkcije fizičkih svojstava na temelju praktične nemogućnosti izvođenja sintaktičke redukcije psiholoških teorija na fizikalne teorije. Mnogi od tih autora dopuštaju da mogu postojati istiniti zakoni premošćivanja koji povezuju mentalne predikate s disjunkcijama fizikalnih predikata (za raspravu, vidi Kim 1993). Međutim, tvrde da su, unatoč tome, takve disjunkcije fizikalnih predikata previše kompleksne da bi se koristile kada se stvarno bavimo znanošću i formuliramo znanstvene teorije. Drugi, kao što smo vidjeli u poglavlju 6, tvrde da disjunkcije fizičkih svojstava ne mogu podržavati znanstvene zakone jer takva vrsta predikata nije nužno projektibilna. Međutim, ovdje treba primijetiti da ova vrsta prigovora protiv Nagelova tipa redukcije, ako je uvjerljiva, samo pokazuje da redukcije mentalnih na fizička svojstva nisu moguće zbog našeg ograničenog znanja ili praktične nemogućnosti da prevedemo mentalne predikate u cijeli (potencijalno beskonačan) niz disjunkcija fizikalnih predikata. Stoga ova razmatranja ne podržavaju ontološki antiredukcionizam, već se samo

[58] Fodor (1974), Teller (1985) i Kim (1989) iznose različite razloge koji objašnjavaju ove nedostatnosti.

ukazuje na ograničenje naših sposobnosti da izvedemo redukciju teorija iz psihologije na teorije iz fizikalnih znanosti.

Antiredukcionisti mogu pokušati izbjeći urušavanje jake supervenijencije u redukcionizam oslanjajući se na razmatranja koja se odnose na jednostavnost jezika i eksplanatornih moći psiholoških teorija koje pretpostavljaju postojanje mentalnih svojstava. Naime, može se tvrditi da nema smisla reducirati psihološke teorije jer su one jednostavnije od fizikalnih teorija te da ćemo, ako stvari objašnjavamo koristeći psihološke predikate, imati jednostavnija i razumljivija objašnjenja nego da psihološke fenomene pokušamo objasniti pomoću disjunkcija različitih fizičkih svojstava.

Međutim, ovakva vrsta razmatranja ne podupire nešto više od instrumentalističkog gledišta na prirodu mentalnih stanja. Instrumentalizam u filozofiji uma bilo bi gledište prema kojemu su mentalna stanja korisna sredstva za predviđanje ponašanja, ali nije nužno da ona zaista postoje kao zasebni i nesvodljivi entiteti (za raspravu, vidi Dennett 1981). Ova pozicija nije privlačna realistima jer dopušta da su mentalna stanja i njihova svojstva samo korisni, i potencijalno fikcijski, predmeti koje koristimo kako bismo objasnili i predvidjeli tuđa ponašanja. Dakle, to gledište ne podrazumijeva da mentalna stanja zaista postoje kao ontološki posebni entiteti te ova vrsta razmatranja ne podupire realističku verziju antiredukcionističkog fizikalizma. Stoga, ako antiredukcionisti fizikalisti žele izbjeći prihvaćanje ovako oslabljene pozicije, moraju ponuditi neko ontološko objašnjenje toga zašto mentalna svojstva nisu reducibilna na fizička svojstva. Kako bi to postigli čini se da je potrebno eksplicitno formulirati relaciju ovisnosti koja bi mogla iz metafizičke perspektive objasniti u čemu se sastoji relacija supervenijencije (vidi Stenwall 2021).

Možda bi se moglo tvrditi da već imamo razloga smatrati da prihvaćanje teze jake supervenijencije ne implicira neku formu redukcionizma (vidi Marras 1993, 221–23). Takvu bi se tvrdnju moglo braniti na sljedeći način. Čini se uvjerljivim pretpostaviti da postoji ontološka relacija ovisnosti koja vrijedi u svakom mogućem svijetu i koja je takva da asimetrično povezuje mentalna i fizička svojstva. Smatramo da mora postojati takva vrsta ovisnosti jer ona zahvaća intuiciju da mentalna svojstva asimetrično ovise o fizičkim svojstvima. Dakle, intuitivna razmatranja opravdavaju tvrdnju da postoji relacija prema kojoj su mentalna svojstva u svakom mogućem svijetu u kojem postoje ovisna o postojanju određenih fizičkih svojstava. No, obrnuto ne vrijedi, u smislu da mogu postojati mogući svjetovi u kojima fizička svojstva postoje, a ne postoje mentalna svojstva. Pojam supervenijencije moramo interpretirati u tom svjetlu. Naime, takva vrsta asimetrične relacije daje ispravnu interpretaciju pojma supervenijencije koji pretpostavljaju antiredukcionisti realisti u filozofiji uma. Jednom kada prihvatimo takav pojam supervenijencije koji asimetrično povezuje mentalna i fizička svojstva,

postaje jasno da se mentalna svojstva ne mogu reducirati na fizička svojstva. Naime, sama ideja asimetrične ontološke ovisnosti pretpostavlja da u toj relaciji stoje *različiti* entiteti. Ako su entiteti različiti, onda se ne mogu reducirati jedni na druge. Nadalje, ako ovakvo shvaćanje relacije ovisnosti implicira tezu jake supervenijencije te daje način na koji bi je trebalo interpretirati, onda nasuprot tvrdnjama Kima, jaka supervenijencija ne može implicirati mogućnost redukcije mentalnih svojstava na fizička svojstva. U suprotnom bi slijedilo da i pretpostavljena asimetrična relacija ovisnosti implicira mogućnost redukcije.

Ovaj odgovor na prvu izgleda uvjerljiv. Međutim, nije jasno da pomaže antiredukcionistima da objasne relaciju ovisnosti između mentalnog i fizičkog. Prisjetimo se da je pojam „supervenijencija" uveden kako bi se objasnila asimetrična ovisnost mentalnog nad fizičkim. Vidjeli smo da, ako se supervenijencija tumači prema tezi jake supervenijencije, prihvaćanje te relacije vodi u redukcionizam. Kako bi se spriječila ta implikacija formulirali smo relaciju ovisnosti koja bi nam trebala dati namjeravanu interpretaciju pojma supervenijencije. No, to ujedno znači da ova vrsta odgovora podrazumijeva da moramo imati pojam ovisnosti koji je neovisan i prethodi pojmu jake supervenijencije. To nas dovodi do zaključka da, unatoč prvotnom cilju, supervenijencija nije pojam koji može objasniti antiredukcionističko shvaćanje odnosa između mentalnih i fizičkih svojstava. U tom pogledu, Kim navodi da je supervenijencija vrsta relacije koja sama po sebi „šuti o prirodi relacije ovisnosti koja bi mogla objasniti zašto mentalno supervenira nad fizičkim" (Kim 1998, 14). To ne implicira da je nereduktivni realizam u pogledu mentalnih svojstava pogrešna teorija; no ukazuje na daljnju potrebu specificiranja neovisnog metafizičkog temelja koji bi objasnio relaciju asimetrične ovisnosti mentalnih svojstava o fizičkim svojstvima.[59]

Da sumiramo, dosad smo razmatrali kako antiredukcionisti i realisti u pogledu mentalnih stanja mogu koristiti pojam supervenijencije u svrhu objašnjenja ovisnosti mentalnih svojstava o fizičkim svojstvima. Kako bi se pokazala održivost takvog gledišta potrebno je formulirati pojam supervenijencije koji će uspjeti zahvatiti relevantni pojam ontološke ovisnosti, a da ne kolabira u ontološki redukcionizam. Dakle, održivost takve vrste antiredukcionističkog projekta zahtijeva opravdanje pretpostavke da postoji asimetrična relacija ontološke ovisnosti koja implicira relaciju supervenijencije, ali nije ekvivalentna s njom.

U nastavku ćemo razmotriti jedan drugi problem s kojim se susreću zastupnici antiredukcionizma koji pretpostavljaju relaciju supervenijencije

[59] Upravo u tom smjeru se počinje razvijati novija rasprava o ovom problemu. U svom recentnom radu, Robin Stenwall (2021) argumentira da bi antiredukcionisti fizikalisti trebali objasniti ovisnost mentalnih o fizičkim svojstvima koristeći tehnički filozofski pojam „utemeljenje". Za općenitu raspravu o pojmu utemeljenja, vidi Dasgupta (2014).

mentalnog nad fizičkim. Bez obzira na to kako relaciju supervenijencije shvatimo u modalnom smislu, čini se da prihvaćanje ideje da mentalna stanja superveniraju nad fizičkim stanjima nije kompatibilno s intuitivnom idejom da mentalna stanja posjeduju uzročne moći.

7.3 Argument uzročne isključivosti

Kim je formulirao još jedan utjecajan argument protiv nereduktivnog fizikalizma koji se naziva argument uzročne isključivosti. Njime se tvrdi da antiredukcionistički fizikalizam koji pretpostavlja relaciju supervenijencije ne može objasniti kako to da mentalna stanja imaju uzročne moći.[60] Argument uzročne isključivosti oslanja se na dva principa koja karakteriziraju fizikalističko gledište. Njih smo već spomenuli u poglavlju 2 kada smo se bavili prigovorima kartezijanskom dualizmu. Prvi se princip odnosi na uzročnu zatvorenost ili potpunost fizičkog svijeta koji možemo formulirati na sljedeći način:

> Ako fizički događaj ima dovoljan uzrok u vremenu *t*, onda ima dovoljan *fizički* uzrok u vremenu *t*. (Kim 2005, 15, kurziv dodan)

Drugi se princip odnosi na uzročnu isključivost kojom se negira mogućnost sistematične uzročne preodređenosti u nekoj domeni objašnjenja. Drugim riječima, a važno za naš kontekst, njome se isključuje mogućnost da svaki tjelesni pokret ima dva dovoljna istodobna uzroka. Kim uzročnu isključivost formulira na sljedeći način:

> Ako događaj *d* ima dovoljan uzrok *u* u *t*, nijedan događaj u *t* koji je različit od *u* ne može biti uzrok od *d* (osim ako se ne radi o pravom slučaju preodređenosti. (Kim 2005, 17)

Kim u argumentu koristi generalizaciju principa uzročne isključivosti, koji naziva princip određenja ili generativne isključivosti. Njega definira na sljedeći način:

> Ako je događaj *d*, ili instanca svojstva P, određen/generiran događajem *u* – uzročno ili na neki drugi način – tada događaj *d* nije određen/generiran od strane nekog događaja koji se u potpunosti razlikuje ili je neovisan o događaju *u* – osim ako se ne radi o pravom slučaju preodređenosti. (Kim 2005, 17)

Prema ovom principu, relacija uzročnosti samo je jedan element općenitijeg skupa relacija koji Kim naziva relacija određivanja. Ovim se principom tvrdi da, ako želimo izbjeći sistematsku preodređenost, onda moramo

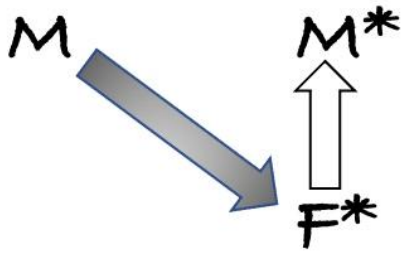
[60] Prezentacija argumenta u ovom dijelu se oslanja na rad Malatesti i Malatesti (2013).

pretpostaviti da, ako neki događaj *u* općenito određuje događaj *d*, onda što god drugo određuje *d* mora biti u relevantnim aspektima slično ili istovjetno događaju *u*. Važno je primijetiti da se ovaj princip općenito odnosi na relaciju određivanja ili determinacije. Budući da relacija supervenijencije predstavlja jedan element općenitijeg skupa relacija određivanja, onda se ovaj princip odnosi i na relaciju supervenijencije.

Pretpostavimo sada da instancijacija mentalnog svojstva M uzrokuje instancijaciju drugog svojstva M*. Možemo zamisliti da M predstavlja osjet boli prouzrokovan opeklinom, dok M* predstavlja želju da se odmaknemo od uzroka boli. Nadalje pretpostavimo da je fizičko svojstvo F supervenijentna baza za M, dok F* predstavlja supervenijentnu bazu za M*. Možemo pretpostaviti da F i F* predstavljaju određene obrasce aktivacije u našem živčanom sustavu.

Zadržimo se ovdje na svojstvu F*. Budući da mentalno svojstvo M* supervenira nad fizičkim svojstvom F*, aktivacija u živčanom sustavu koju F* predstavlja *određuje* pojavljivanje želje koju M* predstavlja. Također, budući da se radi o relaciji supervenijencije, znači da je F* dovoljan uvjet koji određuje pojavljivanje M*. Sada dolazi ključan trenutak u argumentu. Budući da je F* dovoljan uvjet koji generira M*, prema principu generativne isključivosti slijedi da prethodno mentalno stanje M ne može biti dovoljan uzrok za M*. Kada bi to bio slučaj, onda bismo imali neuvjerljivi slučaj preodređenja gdje su različita svojstva F* i M zasebno dovoljna da generiraju M*. Time dolazimo do neintuitivnog zaključka kod kojeg se čini da M (npr. osjet boli) ne može direktno uzrokovati M* (npr. želju da se odmaknemo od uzroka boli). Međutim, do ovog zaključka vode pretpostavka da mentalno supervenira nad fizičkim i princip isključivosti koji nam kaže da jedan događaj ne može biti u potpunosti određen s dva ili više različitih i neovisnih događaja.

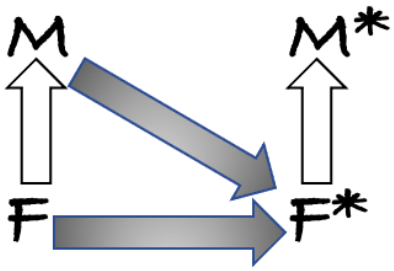
Moglo bi se odgovoriti da je, iako M ne uzrokuje M* direktno, moguće da ga uzrokuje indirektno. Na primjer, mogli bismo reći da M uzrokuje moždanu aktivnost F* koja onda određuje putem relacije supervenijencije M*. Ova



Slika 4

mogućnost je predstavljena na slici 4 gdje bijela strelica predstavlja relaciju supervenijencije između F* i M*, a siva uzročni odnos između M i F*.

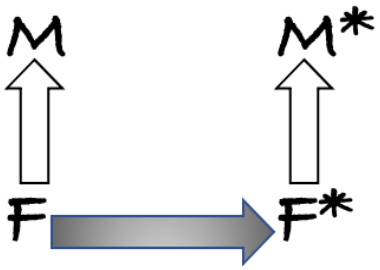
Kim smatra da ovaj odgovor nije uvjerljiv iz sljedećih razloga. Primijetimo da, s obzirom na relaciju supervenijencije, moramo pretpostaviti da i svojstvo M ima svoju supervenijentnu bazu u obliku moždane aktivnosti F. Iz toga slijedi da je instancijacija svojstva F dovoljan uvjet kako bi se instanciralo svojstvo M. Budući da smo pretpostavili da je instancijacija svojstva M dovoljan uzrok svojstva F*, slijedi da je također sama instancija F-a koji predstavlja supervenijentnu bazu za M dovoljan uzrok za instancijaciju svojstva F*. Ili koristeći manje apstraktnu terminologiju, pretpostavimo da aktivacija C-vlakana predstavlja supervenijentnu bazu osjeta boli (u ovom slučaju F). Ako osjet boli uzrokuje aktivaciju dijelova mozga koji se nalaze u supervenijentnoj bazi želje da se odmaknemo od



Slika 5

uzroka boli (ovdje F*), onda je zapravo sama aktivacija C-vlakana dovoljna da se aktiviraju dijelovi mozga koji instanciraju to svojstvo F*. Trenutna situacija je ilustrirana na slici 5. Pretpostavili smo da M uzrokuje F*. Međutim, F također predstavlja dovoljan uzrok za F*. Stoga se čini da F* ima dva u potpunosti različita, ali dovoljna uzroka.

Jasno je da se antiredukcionisti fizikalisti ovdje opet nalaze u problemima. Prema principu određenja/generativnosti, moramo ili F ili M isključiti kao uzrok za F*. Čini se da nijedna opcija nije povoljna za antiredukcionistički fizikalizam. Ako isključimo F kao uzrok za F*, onda slijedi da F* ima samo mentalni uzrok M. Međutim, time bismo prekršili princip uzročne zatvorenosti ili potpunost fizičkog svijeta. Ovaj princip nam nalaže da moramo pretpostaviti da F* ima dovoljan fizički uzrok F. Stoga slijedi da



Slika 6

moramo isključiti M kao dovoljan uzrok za F*. To nas dovodi do situacije koja je ilustrirana na slici 6.

Pod uvjetom da je Kimov argument uvjerljiv, slijedi da mentalna stanja ne mogu uzrokovati druga mentalna stanja. Jedino što može stajati u uzročnim odnosima su supervenijentne fizičke baze tih mentalnih stanja i svojstava. Dakle, ako razmotrimo naš primjer gdje nas boli ruka jer smo je opekli, taj osjećaj boli ne može biti uzrok želje da odmaknemo ruku. Ono što igra uzročne uloge su procesi u mozgu i tijelu koji se nalaze u podlozi tih mentalnih stanja. Najviše što možemo reći jest da mentalna stanja poput M i M* stoje u regularnim vezama na temelju kojih možda možemo donositi dobre zaključke i predviđanja o tome kako će se ljudi i drugi organizmi ponašati. U tom pogledu Kim navodi:

> Ove uočene korelacije ostavljaju dojam uzročnosti; međutim, to je samo privid, i ovdje nema više uzročnosti nego između dvije uzastopne sjene koje baca isti automobil u pokretu, ili dva uzastopna simptoma patologije u razvoju. (Kim 2005, 21)

Dakle, pod pretpostavkom da mentalna stanja superveniraju nad fizičkim stanjima, dolazimo do vrlo neintuitivnog zaključka da mentalna stanja nemaju uzročne moći već samo na neki način zrcale prave uzročne veze koje se nalaze u njihovim fizičkim podlogama.

7.4 Moguće reakcije na argument uzročne isključivosti

Posljedice Kimovog argumenta općenite su te se odnose na cijeli niz fizikalističkih gledišta koja zastupaju neku varijantu antiredukcionizma u filozofiji uma. Osnovna poruka njegovog argumenta je da se antiredukcionisti fizikalisti susreću s problem mentalnog uzrokovanja, tj. pitanjem kako objasniti zdravorazumsku tvrdnju da mentalna stanja imaju uzročne moći, a da se ona ne svode na svoju reduktivnu fizičku bazu. Kim ovaj generalni problem formulira na sljedeći način:

> *Problem mentalne uzročnosti*. Uzročna učinkovitost mentalnih svojstava nije konzistentna s istodobnim prihvaćanjem sljedeće četiri tvrdnje (i) uzročna zatvorenost fizičkog, (ii) uzročna isključivost, (iii) supervenijencija uma nad tijelom i (iv) dualizam mentalnih/fizičkih svojstava, gledište prema kojemu su mentalna svojstva nesvodiva na fizička svojstva. (Kim 2005, 21)

S obzirom na Kimov argument i opis situacije s problemom uzročnosti, antiredukcionisti koji prihvaćaju relaciju supervenijencije mogu pokušati odgovoriti na ovaj argument tako da odbace neku od njegovih pretpostavki. Na primjer, mogu odbaciti (i), (ii) ili čak pretpostavku da mentalna stanja imaju uzročne moći. U nastavku ćemo razmotriti koje bi bile implikacije tih odabira.

Odbacivanje (i) principa uzročne zatvorenosti fizičkog svijeta djeluje kao najmanje prihvatljiva opcija. Naime, mnogi smatraju da je upravo taj princip, koji je utemeljen na uspješnom razvoju znanosti, glavni temelj za prihvaćanje fizikalističkog gledišta (vidi Papineau 2002, 232–56). Nije jasno zašto bismo uopće smatrali da trebamo prihvatiti ili braniti neku varijantu fizikalizma kada bismo ga odbacili. Stoga se čini da napuštanjem ovog principa na neki način napuštamo i fizikalizam.

Druga opcija je ona da se odbaci pretpostavka uzročne djelotvornosti mentalnih stanja. U tom bi slučaju antiredukcionisti trebali prihvatiti neku varijantu epifenomenalizma. Međutim, epifenomenalizam ne djeluje kao uvjerljiva pozicija za dominantne varijante fizikalističkog antiredukcionizma. Naročito se ne čini spojiv s funkcionalizmom. Prisjetimo se da funkcionalizam pretpostavlja da su mentalna stanja određena uzročnim ulogama koje ona igraju u povezivanju podražaja i ponašanja. Kada bi se odbacila pretpostavka da mentalna stanja imaju uzročne moći, onda se standardna varijanta funkcionalizma uopće ne bi mogla formulirati.

Treća opcija je da odbacimo (ii) princip određenja/generativne isključivosti. Međutim, kako smo već istaknuli u raspravi Descartesovog dualizma (vidi poglavlje 2), ova opcija znači prihvaćanje opće preodređenosti. Iako nije logički nemoguće da u nekoj domeni stvarnosti postoji sustavna preodređenost uzroka, svejedno se čini da je takva mogućnost malo vjerojatna. Naime, ona zahtijeva da pretpostavimo da svako tjelesno ponašanje i aktivnost u mozgu koju naša mentalna stanja mogu prouzročiti podrazumijevaju da istodobno postoji mentalni i fizički uzrok od kojih je svaki za sebe dovoljan da proizvede tjelesnu posljedicu. Čini se da takvo objašnjenje fizičkih događaja krši princip Ockhamove britve. U ovom kontekstu taj nam princip nalaže shvaćanje da, ako pretpostavka mentalnosti ne doprinosi objašnjenju fizičkih događaja, jednostavno ni nema potrebe pretpostaviti da mentalna stanja i svojstva postoje neovisno o fizičkim stanjima i svojstvima kako ih zamišljaju antiredukcionisti.

Zaključno možemo reći da se čini kako Kimov argument uzročne isključivosti pokazuje da prihvaćanje supervenijentnog fizikalizma nije kompatibilno s prihvaćanjem antiredukcionizma u pogledu mentalnih svojstava te da bi fizikalisti trebali odbaciti (iv) i prihvatiti reduktivni fizikalizam.

7.5 Zaključak

U ovom poglavlju razmotrili smo utjecajnu varijantu antiredukcionističkog fizikalizma koja nastoji odnos mentalnog i fizičkog objasniti koristeći relaciju supervenijencije. Ako se želi sačuvati ontološka neovisnost mentalne od fizičke domene, takva vrsta fizikalizma susreće se s problemom objašnjenja ovisnost mentalnog nad fizičkim. Mnogi antiredukcionisti fizikalisti smatraju da se objašnjenje takve vrste ovisnosti može dati u terminima relacije

supervenijencije. U ovomu poglavlju razmotrili smo argumente koji pokazuju da relacija supervenijencije nije dovoljna kako bi se pružila uvjerljiva teorija antiredukcionističkog fizikalizma. Prvo smo razmotrili nekoliko shvaćanja pojma supervenijencije. Kao uvjerljivu interpretaciju prihvatili smo ono što se naziva teza jake supervenijencije. Međutim, kada se supervenijencija shvati na taj način, onda se čini da relacija supervenijencije kolabira u redukcionizam.

Drugi prigovor antiredukcionističkom fizikalizmu odnosi se na argument uzročne isključivosti. Ovdje smo vidjeli da, kada ekspliciramo pretpostavke koje karakteriziraju fizikalistička gledišta, onda bez obzira na različite modalne interpretacije, prihvaćanje supervenijencije vodi u redukcionizam. Naime, čini se da je najbolje objašnjenje toga da mentalno ovisi o fizičkom i da mentalna stanja imaju uzročne moći činjenica da mentalna stanja jesu jedna vrsta fizičkih stanja. U suprotnom bismo trebali prihvatiti da fizička stanja isključuju mentalna stanja iz domene uzročnih veza i objašnjenja. Zaključak do kojeg nas ova rasprava dovodi jest da antiredukcionizam nije lako pomiriti sa supervenijentnim fizikalizmom te da ima više smisla smatrati da se, ako je fizikalizam istinit, mentalna stanja u nekom smislu moraju moći reducirati na fizička stanja. Druga opcija bila bi pokušati ponuditi formulaciju antiredukcionističkog fizikalizma koja se neće oslanjati na pojam supervenijencije (za raspravu, vidi Stenwall 2021).

U sljedećem poglavlju bavit ćemo se nizom argumenata kojima se nastoji pokazati da fizikalizam općenito nije uvjerljiva pozicija. U tom smislu tvrdnja će biti da neovisno o tome prihvaćamo li redukcionističku ili antiredukcionističku varijantu, fizikalizam kao opće gledište na prirodu naših mentalnih života i odnosa s prirodnim svijetom nije uvjerljiv.

8 Svijest i fizički svijet

8.1 Uvod

U prethodnom smo poglavlju vidjeli da se fizikalizam može formulirati pomoću relacije supervenijencije. Ovo gledište temelji se na tezi da je nužno da mentalna svojstva superveniraju nad fizičkim svojstvima te se smatra da je to minimalna pretpostavka koju dijele redukcionističke i antiredukcionističke varijante fizikalizma.

U ovom poglavlju bavit ćemo se dvama utjecajnim prigovorima fizikalizmu kojima se nastoji pokazati da svjesna iskustva ne superveniraju nužno nad fizičkim svojstvima. Radi se o argumentu iz znanja koji je formulirao Frank Jackson (1982; također vidi Pećnjak i Špiljak 2014) i varijanti kartezijanskog argumenta pojmljivosti koji u suvremenim raspravama brani David Chalmers (1996; također vidi Pećnjak i Janović 2016, pogl. 9). Prvo ćemo objasniti pojam svijesti koji će nam biti važan za formulaciju ovih argumenata. Nakon toga ćemo izložiti argument znanja kako ga je Jackson formulirao te ćemo se baviti utjecajnim fizikalističkim odgovorima na njega. U zadnjem dijelu poglavlja bavit ćemo se argumentom pojmljivosti kako ga je formulirao Chalmers te ćemo također razmotriti utjecajne fizikalističke odgovore. Poseban naglasak ćemo staviti na one odgovore koji dovode u pitanje pouzdanost intuicija o prirodi fizičkog svijeta i svjesnog iskustva na kojima se temelje ovi utjecajni antifizikalistički argumenti.

8.2 Pojam svjesnog iskustva

Ovo poglavlje posvećeno je filozofskim raspravama kojima se nastoji pokazati da fizikalizam ne može zahvatiti važne aspekte svjesnih iskustva. Postoje različiti fenomeni koji se podrazumijevaju pod pojmom svijesti pa time i različiti načini kako možemo shvatiti pojam iskustva (Van Gulick 2018; Block 1995; Hill 1991, pogl. 1). Ovdje nećemo razmatrati sve načine na koje se pojmovi svijesti i iskustva upotrebljavaju. Naš cilj je skromniji. Nastojat ćemo ponuditi minimalnu karakterizaciju pojma svijesti koja će biti korisna i relevantna za predstavljanje nekih od središnjih rasprava u suvremenoj filozofiji uma. Stoga ćemo se u nastavku usredotočiti na pojam fenomenalne

ili pojavne svijesti, točnije na pojam fenomenalnog ili pojavnog karaktera iskustva.

Općenito govoreći, fenomenalni karakter iskustva je obilježje koje karakterizira, iz perspektive prvog lica osobe koja ima to iskustvo, prirodu tog iskustva. Razmotrimo, na primjer, vizualni doživljaj koji imamo kada gledamo crvenu površinu. Kad imamo to iskustvo, na poseban smo način svjesni boje površine. Drugim riječima, svjesni smo crvenosti koja karakterizira tu površinu. Fenomenalni karakter iskustva gledanja crvene boje je način na koji nam je boja dana (engl. *given*). Drugim riječima, fenomenalni karakter iskustva je u ovom slučaju određen svojstvom crvenosti. Slično tome, bolnost boli, svrbež kože, karakterističan okus određene hrane ili pića, karakterističan miris cvijeća i tako dalje, predstavljaju uobičajene primjere fenomenalnog karaktera različitih iskustava.

Prije nego što nastavimo dalje, vrijedi usporediti ovu karakterizaciju pojma fenomenalnog karaktera iskustva s drugim pojmovima koji se spominju u suvremenim raspravama. Prvo treba istaknuti da neki autori ističu da je fenomenalni karakter iskustva ono što određuje „kako je to" doživjeti određeno iskustvo. Thomas Nagel je, u svom poznatom radu *Kako je to biti šišmiš?* (1974), popularizirao termin „kako je to biti ili doživjeti" nešto te odredio daljnje rasprave o fenomenalnom karakteru iskustva kao posebnom svojstvu koje svjesna mentalna stanja imaju. Međutim, već su ranije o sličnom fenomenu govorili, koristeći slične izraze, Ludwig Wittgenstein i drugi (vidi Wittgenstein 1980, odjeljak 19; Farrell 1950). U novije vrijeme, Ned Block (1995) vrstu svijesti koja određuje kako je to doživjeti neko iskustvo naziva *fenomenalna svijest*. Dakle, ideja je da su svjesna iskustva ona za koja se možemo pitati postoji li nešto kako je to doživjeti neko iskustvo. Da se vratimo na primjer s bojom, ukazati na svjesni aspekt iskustva gledanja boje možemo tako da se pitamo *kako je to vidjeti neku boju*. Slično tome, osjećaj boli predstavlja vrstu iskustva čiji karakter možemo odrediti tako da se pitamo kako je to doživjeti bol. Dakle, općenito možemo reći da fenomenalni karakter iskustva predstavlja ono obilježje koje određuje kako je to doživjeti neko iskustvo.

Međutim, Nagel koristi pojam kako je to doživjeti iskustvo u inkluzivnijem, i više teorijskom, smislu od pojma fenomenalni karakter kako smo ga ranije obrazložili. To se vidi u sljedećem odlomku gdje Nagel navodi:

> [...] organizam ima svjesna mentalna stanja ako i samo ako postoji nešto kako je to *biti* taj organizam - nešto kako je to *za* taj organizam.
> To možemo nazvati subjektivnim karakterom iskustva. (T. Nagel 1974, 436)

Ovdje Nagel koristi pojam „kako je to biti“ na način koji karakterizira iskustvo bivanja određenim *tipom* ili *vrstom* organizma. Premda ne isključujemo da fenomenalni karakter iskustva nekog organizma može pridonijeti tome kako je to biti taj organizam, u ostatku poglavlja ćemo koristiti uže shvaćanje ovog pojma. Pojam fenomenalni karakter iskustva shvaćamo kao ono što karakterizira, za određenu osobu koja ga proživljava, to iskustvo.

Nadalje, Nagel u gornjem citatu povezuje pojam kako je to biti ili imati neko iskustvo sa subjektivnošću. Ovo potonje je obilježje iskustva koje je pak povezano s posjedovanjem određene perspektive na svijet. Međutim, iako se subjektivnost iskustva i pojam perspektive ili točke gledište (engl. *point of view*) mogu povezati s fenomenalnim karakterom iskustva, čini se da su to logički neovisna obilježja iskustvenih doživljaja. Naime, netko ili nešto može imati perspektivu na svijet i time subjektivnu točku gledišta, a da nema iskustva s fenomenalnim karakterom. Na primjer, možemo zamisliti da postoji autonomni robot koji je sposoban za vlastito odlučivanje i djelovanje u svijetu, čime pokazuje da ima određenu perspektivu na svijet. Unatoč tome, nije nužno da će unutarnja stanja tog autonomnog robota imati fenomenalni karakter koji bi odredio kako je to imati ili doživjeti ta unutarnja stanja. S obzirom na tu *prima facie* odvojenost između posjedovanja perspektive i imanja iskustva s fenomenalnim karakterom, u ostatku poglavlja nećemo se opredjeljivati ili raspravljati koja bi točno bila priroda odnosa između fenomenalnog karaktera i posjedovanje perspektive na svijet.

U ovom kontekstu mnogi filozofi koriste latinizam *quale* (mn. *qualia*) kako bi označili fenomenalni karakter iskustva. Međutim, smatramo da je bolje ne koristiti taj termin kada se određuje općeniti pojam fenomenalnog karaktera. Naime, *quale* za mnoge autore predstavlja tehnički filozofski termin koji je opterećen teorijskim pretpostavkama koje nisu svima prihvatljive (na primjer, vidi Stoljar 2006, 23 i odjeljak 9.5 u sljedećem poglavlju). Jedna od tih pretpostavki je da su *qualia* nesvodivo mentalna svojstva, te kao takva ne mogu biti fizička svojstva. Kasnije ćemo razmotriti neke argumente kojima se nastoji pokazati da fenomenalni karakter nije fizičko svojstvo iskustva. S obzirom na to da je upravo stvar rasprave predstavlja li fenomenalni karakter fizičko ili nereducibilno mentalno svojstvo iskustva, bolje je izbjegavati terminologiju koja bi prejudicirala odgovor na to pitanje. Drugim riječima, bolje je izbjegavati pojam *qualia* kada nastojimo dati teorijski neutralno tumačenje pojma fenomenalni karakter.

Drugi važan razlog za izbjegavanje korištenja pojma *qualia* je taj što su neki autori skloni poistovjetiti *qualia* s nereprezentacijskim svojstvima iskustva. Svi će se složiti da neki aspekti naših iskustava reprezentiraju ili predstavljaju svijet na određeni način te, stoga, posjeduju intencionalnost (vidi poglavlje 1). Na primjer, kada vidimo crvenu površinu imamo iskustvo

koje reprezentira površinu kao crvenu. Dakle, možemo reći da to iskustvo ima reprezentacijski sadržaj *da* je površina crvena. Pristaše reprezentacionalizma smatraju da se fenomenalni karakter iskustva sastoji od njega ili barem supervenira nad njegovim reprezentacijskim sadržajem (vidi Tye 1995; Dretske 1995; vidi, također Sarıhan 2020). Drugi smatraju da se fenomenalni karakteri iskustva zapravo svode na posjedovanje *qualia* koje, budući da su prema pretpostavci čisto kvalitativna svojstva iskustva, ne mogu biti reprezentacijska ili intencionalna svojstva (vidi Pećnjak i Janović 2016, 41–62; Block 1996; Peacocke 1983, pogl. 1). Iako rasprava između reprezentacionalista i nereprezentacionalista u pogledu *qualia* ima važno mjesto u suvremenoj filozofiji uma (Lycan 2019), ovdje se nećemo direktno njome baviti. U svakom slučaju, smatramo da je bolje ne interpretirati pojam fenomenalni karakter u terminima *qualia* koje sa sobom nose određene teorijske pretpostavke koje bi mogle prejudicirati našu raspravu o tome predstavljaju li fenomenalni karakteri nefizička svojstva iskustvenih doživljaja.[61]

Sada kada smo razjasnili pojam fenomenalnog karaktera koji ćemo podrazumijevati u ostatku poglavlja, možemo preći na razmatranje prvog argumenta kojim se nastoji pokazati da postojanje fenomenalnog karaktera stvara probleme za fizikalističku tezu da mentalna svojstva nužno superveniraju nad fizičkim svojstvima.

8.3 Argument iz znanja

Frank Jackson formulirao je argument iz znanja (AZ) u dva seminalna rada (vidi Jackson 1982; 1986). Iako ga se u kasnijim radovima odrekao (vidi Jackson 1998; 2003; 2006), njegov se argument pokazao izuzetno utjecajnim te je potaknuo jednu od najširih rasprava u suvremenoj filozofiji uma (vidi, npr. Alter 2021; Nida-Rümelin i O Conaill 2021; Ludlow, Nagasawa, i Stoljar 2004; Malatesti 2012; Pećnjak i Špiljak 2014). Srž Jacksonova izvornog argumenta je da:

> Ništa što možete reći o fizičkoj vrsti ne zahvaća miris ruže […]. Stoga, fizikalizam je lažan. (Jackson 1982, 469)

Jacksonov argument za ovu tvrdnju se temelji na poznatom misaonom eksperimentu koji uključuje Mary, svjetski poznatu neuroznanstvenicu koja posjeduje potpuno znanje o svim fizičkim svojstvima boja i ljudskom

[61] Tvrdnju da se fenomenalni karakteri iskustva mogu shvatiti kao *qualia* razmatramo u poglavlju 9.

perceptivnom aparatu.[62] Jackson na sljedeći način opisuje situaciju u kojoj zamišljamo da se Mary nalazi:

> Mary je znanstvenica koja je, iz nekog razloga, prisiljena istraživati svijet iz crno-bijele sobe putem crno-bijelog televizijskog monitora. Ona je specijalizirala neurofiziologiju vida i pretpostavimo da stječe sve fizičke informacije o tome što se događa kada vidimo zrele rajčice ili nebo, te koristimo izraze poput »crveno«, »plavo« itd. Ona otkriva, na primjer, koje kombinacije valnih duljina s neba stimuliraju mrežnicu, te kako točno to preko središnjeg živčanog sustava proizvodi kontrakciju glasnica i izbacivanje zraka iz pluća što rezultira izricanjem rečenice »Nebo je plavo«. (Jackson 1982, 471)

Dakle, ovdje zamišljamo da Mary, s obzirom da od rođenja živi u crno-bijeloj sobi, nikada nije imala pristup obojenim stvarima niti ih je negdje osobno percipirala. Unatoč tome što odrasta u crno-bijeloj sobi studirala je prirodu boje i ljudski perceptivni aparat te je postala svjetski stručnjak koji zna sve fizičke činjenice koje se mogu znati o prirodi boja, percepciji boja, kako ta percepcija proizvodi određene učinke u ljudskom mozgu, što ljudi kažu kada vide predmete u boji, koji se procesi odvijaju u njihovim mozgovima za to vrijeme i tako dalje.

U sljedećem koraku misaonog eksperimenta zamišljamo da je jednog dana Mary puštena izvan kuće te po prvi put vlastitim očima vidi crvenu ružu te kaže: „Aha, znači tako izgleda crvena ruža!“. Postavlja se pitanje je li Mary izlaskom iz kuće naučila neku novu činjenicu o svojem iskustvu crvene ruže ili nije? Jackson (1982) daje potvrdan odgovor na to pitanje. Štoviše, budući da je prije izlaska iz kuće Mary znala sve fizičke činjenice o osjetilnim iskustvima boja, Jackson tvrdi da ono što Mary spoznaje gledanjem crvene ruže mora biti neka nefizička činjenica o tome kako je to vidjeti crvenu ružu. Slijedi li taj zaključak iz misaonog eksperimenta s Mary, razmotrit ćemo u nastavku. No, prije toga, izložit ćemo malo detaljnije sam AZ.

U argumentu pretpostavljamo da Mary posjeduje sveobuhvatno fizičko znanje o prirodi boja i njihovoj percepciji. Dakle, prva premisa Jacksonova

---

[62] Jackson (1982, 470–71) također razmatra slučaj Freda. Mary zamišlja kao normalno ljudsko biće, barem što se tiče njezina vizualnog sustava, koje se nalazi u vrlo neobičnoj situaciji, naime veći dio života provodi u crno-bijeloj sobi. Nasuprot tome, Freda zamišlja kao jedinog čovjeka koji može percipirati određenu nijansu crvene boje. Intuicija na koju se u tom slučaju oslanja kako bi opovrgnuo fizikalizam jest da bez obzira na znanstveno znanje koje o njemu prikupimo, nikada nećemo saznati nešto o fenomenalnom karakteru njegovog doživljaja ove nijanse crvene boje. Međutim, ovaj misaoni eksperiment je u literaturi u potpunosti zanemaren u usporedbi s misaonim eksperimentom o neuroznanstvenici Mary. S obzirom na taj trend u literaturi, mi ćemo se također usredotočiti na potonji misaoni eksperiment.

argumenta iz znanja je da je moguće da Mary posjeduje *potpuno znanje* o svim fizikalnim informacijama koje se odnose na boje i percepciju boja, unatoč tome što ona nikada nije osobno imala iskustvo viđenja boja.

Jackson koristi pojam fizikalne informacije kao sinonim za fizičke činjenice. U tom smislu, ideja je da se Maryjino potpuno znanje odnosi na onu vrstu činjenica za koju fizikalisti, bilo reduktivnog ili nereduktivnog tipa, smatraju da se nalazi u fizičkoj bazi nad kojom mentalna svojstva superveniraju. U tom pogledu, Jackson pojašnjava da je:

> Nepobitno da su fizičke, kemijske i biološke znanosti omogućile veliki broj informacija o svijetu u kojem živimo i o nama samima. Koristit ću termin »fizikalna informacija« za tu vrstu informacija. (Jackson 1982, 469)

Nadalje, treba istaknuti da među fizikalne informacije Jackson (1982, 469) također uključuje one koje se odnose na funkcionalna svojstva mentalnih stanja i uloge koje igraju u mentalnoj i ponašajnoj ekonomiji.

Druga premisa je da Mary jednom kada izađe iz kuće nauči nešto novo o doživljaju boje što nije znala na temelju potpunog znanja fizičkih činjenica:

> Što će se dogoditi kada Mary puste da izađe iz svoje crno-bijele sobe ili joj daju televizijski monitor u boji? Hoće li naučiti nešto ili neće? Jednostavno se čini očitim da će naučiti nešto o svijetu i kako ga vizualno doživljavamo. (Jackson 1982, 471)

Jackson smatra da Mary izlaskom iz crno-bijele sobe stječe novo znanje koje se odnosi na spoznaju fenomenalnog karaktera određenog iskustva. Štoviše, Jackson zaključuje da je „njezino prethodno znanje bilo nepotpuno“, budući da je „imala sve fizikalne informacije“, iz čega slijedi da fizikalizam mora biti neistinit (Jackson 1982, 471).

Ako je fizikalizam neistinit, onda to znači da fenomenalni karakter iskustva nije svojstvo koje pripada fizici, kemiji, biologiji ili neuroznanosti niti je funkcionalno svojstvo koje supervenira nad takvim vrstama svojstava. U tom smislu, mogli bismo reći da AZ nastoji pokazati da postoje nefizičke *qualia*.

Premise AZ-a možemo prikazati na sljedeći način:

1. Ako je fizikalizam istinit i osoba zna sve fizičke (i funkcionalne) činjenice o iskustvu, onda nije moguće da imanjem određenog iskustva osoba nauči nešto novo o tom iskustvu.
2. Moguće je da Mary zna sve fizičke (i funkcionalne) činjenice o iskustvu.
3. Moguće je da Mary imanjem nekog iskustva nauči nešto o njegovom fenomenalnom karakteru.

4. Moguće je da Mary zna sve fizičke (i funkcionalne) činjenice o iskustvima i da imajući neko iskustvo nauči nešto novo o njegovom fenomenalnom karakteru.

Dakle:

5. Fizikalizam nije istinit.

Nakon što smo detaljnije razmotrili premise AZ-a, u nastavku ćemo razmotriti neke od značajnijih prigovora koji mu se upućuju.

8.4 Fizikalizam i argument iz znanja

Argumentom iz znanja napada se fizikalizam pod pretpostavkom da je fizikalizam gledište prema kojemu je nemoguća situacija da Mary ima potpuno znanje fizičkih činjenica te da nauči nešto novo o iskustvima tako da ih doživi. Međutim, teza da mentalna svojstva nužno superveniraju nad fizičkim svojstvima ne odnosi se na to što pojedine osobe znaju niti se njome eksplicitno negira postojanje mogućnosti opisane u misaonom eksperimentu s Mary. Pod pretpostavkom da je AZ uvjerljiv, on samo pokazuje da je moguće da netko zna sve o fizičkim svojstvima i mentalnim svojstvima koja superveniraju nad njima, a da ne zna sve o njihovim fenomenalnim karakterima. Ta mogućnost ne opovrgava tezu supervenijencije koju minimalne forme fizikalizma pretpostavljaju. Stoga je potrebno uvesti dodatne pretpostavke kako bismo mogli zaključiti da AZ dovodi u pitanje takvu minimalnu varijantu fizikalizma.

Jackson (1982; 2007) smatra da teza supervenijencije, osim što predstavlja metafizički nužan odnos između mentalnog i fizičkog, također pretpostavlja apriornu mogućnost spoznaje odnosa između mentalnog i fizičkog. To znači da će osoba, pod pretpostavkom da ima besprijekornu sposobnost za logičko zaključivanje te zna sve relevantne fizikalne informacije, biti sposobna na temelju postojećeg znanja o fizičkim činjenicama logički derivirati znanje o fenomenalnim karakterima iskustva, bez da nužno i sama doživi ta iskustva. Dakle, ako pretpostavimo da je Mary osoba koja posjeduje sposobnost za logičko zaključivanje te da u svom zaključivanju ne radi nikakve greške, slijedi da bi Mary trebala znati kako je to imati iskustvo viđenja crvene boje i prije nego što izađe iz crno-bijele sobe. Ako imamo intuiciju da Mary to ne može znati prije nego izađe iz crno-bijele sobe, onda to podržava zaključak da ovakva vrsta apriornog fizikalizma ne može biti točna.

Međutim, neki fizikalisti prihvaćaju aposteriorni fizikalizam (vidi McLaughlin 2007). To je gledište prema kojemu se ne može apriorno izvesti znanje o mentalnom samo na temelju znanja o fizičkom, a da se ne doživi određeno iskustvo. Međutim, moglo bi se tvrditi da AZ također opovrgava aposteriorni fizikalizam time što podriva tezu da mentalno nužno supervenira nad fizičkim (vidi Stoljar 2006, 40–41). Štoviše, ako Mary ne

može logički izvesti iz fizikalnih informacija koje posjeduje znanje o fenomenalnom karakteru tog iskustva, onda bi slijedilo da se odnos supervenijencije otkriva *a posteriori*. Međutim, u tom slučaju imali bismo razloga smatrati da teza supervenijencije uopće ne predstavlja *metafizički nužan* odnos između mentalnog i fizičkog te nam stoga daje razloga smatrati da fizikalizam nije istinit.

Uzmimo kao primjer da netko tvrdi da je metafizički nužno da je predsjednik Sjedinjenih Američkih Država muškarac. Pretpostavimo da je ta tvrdnja utemeljena na aposteriornom opažanju činjenice da su dosad svi predsjednici bili muškarci. Međutim, jasno je da ova vrsta aposteriornog znanja sama po sebi nije dovoljna da isključimo mogućnost da predsjednik SAD-a može biti žena. Štoviše, činjenica da do tvrdnje da su svi dosadašnji predsjednici bili muškarci dolazimo aposteriornim promatranjem daje nam razloga smatrati da uopće nije nužna tvrdnja da je predsjednik SAD-a muškarac. Čini se da se pristaše aposteriornog fizikalizma nalaze u sličnoj poziciji. Ako tvrde da je odnos supervenijencije između mentalnog i fizičkog nužan, a na temelju *a posteriori* znanja o odnosu mentalnog i fizičkog ne mogu isključiti mogućnost da fenomenalni karakteri ne superveniraju nad fizičkim svojstvima, onda izgleda da gube osnove za tvrditi da je ipak nužno da fenomenalni karakteri iskustva superveniraju nad nekim fizičkim svojstvima. Dakle, barem *prima facie*, izgleda da AZ dovodi u sumnju apriorni i aposteriorni fizikalizam jednom kada ta gledišta dodatno ekspliciramo. U nastavku ćemo razmotriti druge prigovore koji se upućuju argumentu iz znanja.

8.5 Imamo ograničeno shvaćanje Maryjinog znanja o fizikalnim činjenicama

Kako bismo odredili jesu li uvjerljive premise argumenta iz znanja moramo razmotriti što uopće znači da Mary posjeduje potpuno znanje o svim relevantnim fizičkim činjenicama. Naime, to je jedna od ključnih pretpostavki AZ-a na temelju koje se dolazi do zaključka da jednom kada Mary doživi određeno iskustvo spoznaje nešto novo što nije bilo zahvaćeno korpusom znanja koje je prethodno imala.

Neki autori smatraju da mi zapravo nemamo jasno razumijevanje što Maryjino znanje fizikalnih informacija podrazumijeva te bi nas to trebalo onemogućiti u donošenju pouzdanih zaključaka o tome što ona može naučiti određenim vizualnim iskustvom. Na primjer, Daniel Dennett (1991; 2006; 2013) u više je radova tvrdio da je AZ „pumpa za intuicije“, tj. filozofsko sredstvo koje nas tjera da zamislimo nešto što nadilazi naše kognitivne sposobnosti. Prema Dennettu, AZ se oslanja na našu nemogućnost shvaćanja ili zamišljanja Maryjinog potpunog znanja svih fizičkih činjenica koje se odnose na prirodu i percepciju boja (vidi Dennett 1991, 398–401; također vidi Foss 1989; Stemmer 1989). Dennett naglašava da, iako AZ ne zahtijeva

da je Mary sveznajuća u pogledu fizičkih činjenica, tvrdnja da zna sve fizičke činjenice o prirodi i percepciji boja zahtijeva od nas da zamislimo takav kompleksan i široki skup informacija da jednostavno ne možemo u tome uspjeti. Kao što osoba koja zamišlja krug, s ciljem da zamisli tijelo s tisuću stranica, ne može na temelju svog zamišljanja izvoditi zaključke o tijelu s tisuću stranica, tako i mi s obzirom na ograničenu mogućnost poimanja Maryjinog znanja o fizičkim činjenicama, ne možemo izvoditi zaključke o tome što će se dogoditi kada Mary vidi boju po prvi put (vidi Dennett 2013).

S obzirom na to da ne možemo pojmiti kakvo točno znanje o fizičkim činjenicama Mary posjeduje, prema Dennettu (2006) opravdani smo tvrditi da možda ona ne nauči ništa novo kada po prvi puta vidi boje. Štoviše, mogli bismo ići još dalje i reći da Mary na temelju svojeg znanja o fizičkim činjenicama zapravo može logički derivirati znanje o fenomenalnom karakteru iskustva kojeg nikada nije imala. Na primjer, zamislimo da neki istraživači testiraju Maryjino znanje o izgledu banana tako da joj pokažu plavu bananu očekujući da će misliti da tako izgleda normalna žuta banana. Dennett (1991, 399–400) kaže da s obzirom na našu nejasnu ideju što sve uključuje Maryjino znanje o fizičkim činjenicama prije nego izađe iz sobe, nije neuvjerljivo tvrditi da bi ona bila u stanju prepoznati pokušaj prevare istraživača i reći im da ljudi nemaju iskustvo s takvim fenomenalnim karakterom kada vide žutu bananu.

Unatoč mišljenjima nekih koji su kritizirali ovakvu vrstu odgovora na AZ (vidi, npr. H. M. Robinson 1993; Jacquette 1995), cilj Dennettovog prigovora nije tvrditi da Mary stvarno može logički derivirati znanje o fenomenalnom karakteru iskustva na temelju poznavanja fizičkih činjenica. Njegov cilj je pokazati da misaoni eksperiment s Mary zapravo ne isključuje tu mogućnost (Dennett 1991, 400). Međutim, ako ne isključuje tu mogućnost, onda AZ ne pokazuje da iz njegovih premisa slijedi zaključak da Mary izlaskom iz sobe spozna nešto novo. Stoga AZ ne može biti valjan argument protiv fizikalizma.

Dennettu bi se moglo odgovoriti da zapravo imamo dobro razumijevanje Maryjinog znanja o fizičkim činjenicama. Vidjeli smo da Dennett prigovara da ne možemo znati svaki detalj o fizičkim činjenicama koji Mary zna. Međutim, za potrebe AZ-a dovoljno je dati filozofsku karakterizaciju znanja o fizičkim činjenicama koja neće podrazumijevati da razumijemo svaki detalj o tim fizičkim činjenicama. Jedna opcija je da se AZ formulira s obzirom na trenutno dostupna znanstvena istraživanja percepcije boja. Barem njih bismo trebali biti u stanju pojmiti. Na primjer, u suvremenim istraživanjima percepcije boja koriste se statističke metode koje kao ulaznu informaciju uzimaju ljudske sudove koji opisuju sličnosti i razlike u doživljaju boja i kao izlaznu informaciju daju višedimenzionalne modele koji kategoriziraju ljudske fenomenalne karaktere doživljaja boja (vidi, npr. Clark 2000, 19–22; Malatesti 2008). Kada znanje o fizičkim činjenicama percepcije boja shvatimo kao znanje o takvim statističkim modelima, onda jasno razumijemo

pretpostavke pitanja može li Mary samo na temelju znanja o fizičkim i funkcionalnim činjenicama otkriti kako je to vidjeti predmet određene boje.

Međutim, problem s ovakvom karakterizacijom fizičkog znanja je to što previše ovisi o trenutno dostupnim spoznajama. Moramo imati na umu da u misaonom eksperimentu Mary posjeduje sve moguće znanje o fizičkom, što bi onda trebalo uključivati i sve buduće znanje koje bismo mogli steći metodama istraživanja koje nam trenutno možda nisu dostupne. Stoga bi fizikalist mogao prigovoriti da čak i ako AZ uspješno osporava mogućnost deriviranja znanja o fenomenalnim karakterima iskustva boja na temelju trenutnog znanja o fizičkoj podlozi ljudske percepcije, svejedno ne slijedi da fizikalizam nije istinit.

Druga opcija je razmotriti postoje li filozofske koncepcije koje zahvaćaju neko opće svojstvo pojma fizičkog koje bi bilo neovisno o specifičnim i trenutno dostupnim istraživačkim programima, ali opet dovoljno određeno za suvislu raspravu o AZ-u. U nastavku ćemo razmotriti prijedlog takve karakterizacije fizičkog koji daje David Lewis (1988).

U Lewisov prijedlog ćemo se uvesti indirektnim putem, tako da slijedimo njegovo zaključivanje koje ukazuje na to da AZ nije problematičan samo za fizikalizam, nego, pomalo neočekivano, i za dualizam u pogledu uma i tijela. U tom pogledu, Lewis navodi sljedeće:

> Argument iz znanja usmjeren je protiv Vas [dualista] ništa manje nego protiv samog Materijalizma. Neka parapsihologija bude znanost o svim nefizičkim stvarima, svojstvima, uzročnim procesima, zakonima prirode i tako dalje, koje mogu biti potrebne kako bismo objasnili stvari koje radimo. Pretpostavimo da u potpunosti naučimo parapsihologiju. To neće napraviti nikakvu razliku. Crno-bijela Mary može proučiti svu parapsihologiju kao i svu psihofiziku percepcije boja, ali još uvijek neće znati kako je to [imati iskustvo]. [...] Naše intuitivno polazište nije bilo samo to da predavanja iz fizike ne mogu pomoći neiskusnima da saznaju kako je to [imati iskustvo], već da predavanja ne mogu pomoći. Ako postoji takva stvar kao što je fenomenalna informacija, ona nije samo neovisna o fizikalnim informacijama. Nezavisna je od svih vrsta informacija koje bi se mogle upotrijebiti u predavanjima za neiskusne. (Lewis 1988, 588–89)

Ako je ova Lewisova primjedba točna, onda AZ od nas uopće ne zahtijeva da shvatimo sve fizičke činjenice nad kojima navodno superveniraju činjenice o mentalnim stanjima. Nasuprot tome, AZ je usmjeren protiv fizikalizma ako ga shvatimo kao gledište da se fizikalne informacije mogu razumjeti i prenositi putem „predavanja“, tj. opisa koji ne pretpostavljaju da je osoba

sama imala određeno iskustvo. Takvo poimanje fizikalne informacije moglo bi biti razumljivije i pristupačnije od onog koje pretpostavlja da u ovom trenutku možemo razumjeti sve fizičke činjenice o bojama i percepciji boja. U sljedećem odjeljku ćemo vidjeti da Lewis odgovara na AZ oslanjajući se na takvo shvaćanje pojma fizikalne informacije. No prije nego se osvrnemo na taj odgovor, razmotrit ćemo još neka poimanja fizičkog koja su dovoljno općenita da izbjegnu Dennettov prigovor.

Thomas Nagel (1974; 1986) karakterizirao je fizičko kao ono što je *objektivno*. Prema njemu, to znači da se fizičke činjenice mogu spoznati neovisno o bilo čijoj perspektivi ili točki gledišta. AZ navodno pokazuje da Mary može jedino spoznati fenomenalni karakter doživljaja boje tako da ima to iskustvo. Stoga bi se moglo reći da mogućnost takve spoznaje ovisi o perspektivi povezanoj s takvim tipom iskustva. Daniel Stoljar (2006, 61–62) umjesto toga karakterizira fizikalističku tezu relevantnu za ovaj kontekst kao tezu da iskustvo, shvaćeno kao ono što ima fenomenalni karakter, supervenira nad onim što nije iskustveno. Stoga se, prema ovom objašnjenju, pojam fizičkog razumije kao ono što nema fenomenalni karakter.

Međutim, čak i ako prihvatimo ovu karakterizaciju fizičkog, AZ i dalje ostaje otvoren prigovoru da, umjesto predstavljanja prave mogućnosti, predstavlja samo simptom našeg neznanja. Stoljar upravo to pokazuje prihvaćanjem onoga što naziva *epistemičko gledište* (Stoljar 2006). Argumentira da nam se primjer s Mary čini mogućim jer ignoriramo neiskustvene činjenice koje nužno određuju supervenijentnu bazu iskustava i njihovih fenomenalnih karaktera (Stoljar 2006, 139–40). Ističe da trenutno nije u stanju opisati te neiskustvene činjenice, no nudi nekoliko razloga koji indirektno podržavaju pretpostavku da je vrlo vjerojatno da uvjerljivost AZ-a ovisi o našem nepoznavanju tih činjenica. U nastavku ćemo razmotriti neke od tih razloga koji se temelje na povijesti filozofije i znanosti.

Stoljar razmatra dva povijesna slučaja gdje se činilo da su argumenti poput AZ-a uvjerljivi, no naknadnim se otkrivanjem činjenica određenog tipa pokazalo da nisu pouzdani (vidi P. S. Churchland 1986). Prvi slučaj odnosi se na argument Renéa Descartesa, kojim smo se bavili u poglavlju 2, a odnosi se na nemogućnost izgradnje stroja koji bi se mogao služiti prirodnim jezikom. Descartes je tvrdio da nije moguće zamisliti stroj, napravljen od čiste materije, koji bi mogao koristiti jezik i prikladno odgovarati na potencijalno beskonačan niz pitanja kako su to u stanju činiti ljudi. Stoga je zaključio da je ljudski um najvjerojatnije nematerijalan jer ima tu lingvističku sposobnost.

Netko bi mogao pokušati opravdati isti taj zaključak na temelju argumenta iz znanja. Na primjer, moglo bi se tvrditi da netko sa savršenim

znanjem fizičkih činjenica, tj. svim znanjem o protežnoj stvari, ne bi mogao derivirati tvrdnju da neki stroj može koristiti jezik kako ga ljudi koriste. Međutim, Stoljar (2006) smatra, i kako smo već isticali u drugom poglavlju, da bi se takva verzija AZ-a prije trebala uzeti kao znak našeg neznanja kakve je sve kompjutacije fizička materija u stanju provesti nego kao argument da je ljudski um nematerijalan. Suvremena znanstvena istraživanja pokazuju da vrsta jezičke sposobnosti o kakvoj je razmišljao Descartes može supervenirati nad fizičkim i kompjutacijskim činjenicama (vidi, npr. Miščević 1990, 31–34). To bi bile upravo one činjenice koje su uključene u opis funkcioniranja prikladno programiranog računala, poput onih na kojima rade različiti *chatbotovi*.

Drugi primjer koji Stoljar (2006, 135) razmatra odnosi se na argument koji navodi C. D. Broad u prilog tvrdnji da je nemoguće derivirati kemijske činjenice iz fizičkih činjenica. Broad (1925, 63) je tvrdio da na temelju znanja, na primjer, fizičkih svojstava kisika i vodika nije moguće derivirati znanje o kemijskoj činjenici da kisik i vodik u kombinaciji s omjerom jedan na prema dva tvore vodu. Stoljar primjećuje da se ovaj argument temelji na našem nepoznavanju činjenica iz kvantne mehanike, koje su Broadu bile nedostupne. Upravo poznavanje tih fizičkih činjenica omogućuje derivaciju kemijskih činjenica, barem onih o kojima je govorio Broad (za raspravu, vidi Malatesti i Malatesti 2013; Weisberg, Needham, i Hendry 2019).

Da zaključimo, čini se da postoje dobri razlozi za smatrati da se intuitivna uvjerljivost AZ-a temelji na našem nepoznavanju određenih fizičkih činjenica koje bi mogle biti ključne za otkrivanje da fenomenalni karakter iskustva supervenira nad fizičkim svojstvima.

Iako smo skloni mišljenju da AZ nije pouzdan jer ovisi o našem nepoznavanju relevantnih činjenica o fizičkom svijetu, važno je istaknuti da je ovaj argument inspirirao izuzetno veliki broj kvalitetnih filozofskih radova čija vrijednost često nadilazi njihovu specifičnu ulogu u raspravi o uvjerljivosti AZ-a. U sljedećim odjeljcima, razmotrit ćemo različita gledišta na pitanje nauči li Mary nešto novo i, ako je odgovor potvrdan, što točno nauči kada ima vizualno iskustvo boja. Kao što ćemo vidjeti, filozofi koji se bave ovim pitanjima ponudili su zanimljive teorije o prirodi pojmova, misli, vjerovanja i znanja o svjesnim iskustvima čija primjena nadilazi samu raspravu AZ-a. Stoga ćemo se raspravom ovih prijedloga dotaknuti nekih važnih tema koje se nalaze na razmeđu filozofije jezika, metafizike i epistemologije.

8.6 Mary ne stječe novo činjenično znanje

Uvjerljivost AZ-a ovisi o pretpostavci da Mary, imajući određena iskustva, stječe znanje o činjenicama koje nisu bile zahvaćene njezinim prethodnim

korpusom znanja o fizičkim činjenicama. Nasuprot tome, mnogi autori tvrde da Mary ne stječe, kada po prvi put ima vizualno iskustvo boje, novo znanje o činjenicama, već stječe novu *sposobnost* (engl. *ability*) (Lewis 1988; Nemirow 1990; Mellor 1993; Meyer 2001). Sličnu vrstu odgovora daje i sam Jackson (2003) u kasnijim radovima u kojima osporava uvjerljivost AZ-a. U nastavku ćemo se usredotočiti na utjecajnu varijantu odgovora sposobnosti (engl. *the ability reply*) koju je ponudio Lewis (1988).

Lewis smatra da AZ nije valjan jer u sebi sadrži pogrešku ekvivokacije. Pogreška ekvivokacije javlja se kada argument sadrži dvosmislene termine. Kako bismo pojasnili ovu vrstu pogreške, razmotrimo sljedeći argument:

1) Postojanje zakona implicira da postoji zakonodavac.
2) Postoje zakoni prirode.

Dakle:

3) Postoji kozmički zakonodavac.

Ovaj argument nije valjan jer se izraz „zakon“ u premisi 1) i 2) koristi u dva različita značenja. U 1) izraz „zakon“ koristi se u smislu pravnih normi koje donose ljudi, dok se u premisi 2) „zakon“ odnosi na prirodne regularnosti koje su neovisne o odlukama ljudi.

Prema Lewisu, slična pogreška se javlja u AZ kada se govori o znanju. Treba razlikovati barem dva smisla riječi „znati“: „znanje da“ i „znanje kako“ (engl. *knowledge that* i *knowing how*). „Znanje da“ vrsta je deklarativnog znanja koje izražavamo uz pomoć propozicija da je nešto slučaj. Na primjer, znamo *da* je snijeg bijel, znamo *da* se nogomet igra s loptom, znamo *da* je Pariz u Francuskoj i tako dalje. Dakle, deklarativno znanje se izražava uz pomoć propozicija koje mogu biti istinite ili neistinite. „Znanje kako“ odnosi se na sposobnosti i vještine koje ne moramo nužno moći izraziti kroz deklarativne rečenice. Na primjer, možemo znati voziti bicikl, iako ne znamo nužno opisati što uključuje to znanje. Drugi su primjeri „znanja kako“ znati *kako* plivati, znati *kako* voziti automobil, znati upravljati dizalicom, znati koristiti računalo i tome slično. Dakle, to nije vrsta znanja čiji se sadržaj izražava putem propozicija koje mogu biti istinite ili neistinite, već znanje koje se manifestira primjenom određenih sposobnosti.

Prema Lewisu, AZ nije valjan jer se u jednom dijelu argumenta izraz „znanje“ koristi u deklarativnom smislu, dok se u drugom dijelu koristi u smislu posjedovanja sposobnosti. Kada se govori o Maryjinom poznavanju fizičkih činjenica onda govorimo o njezinom deklarativnom znanju. Ova vrsta znanja uključuje vjerovanja o činjenicama koje možemo vrednovati kao istinite i neistinite. Međutim, kada se u argumentu spominje da Mary stječe znanje o fenomenalnom karakteru iskustva, onda Lewis smatra da ona stječe znanje koje se odnosi na stjecanje određenih sposobnosti. Dakle, središnja tvrdnja u odgovoru sposobnosti upravo je hipoteza da Mary izlaskom iz crno-bijele sobe stječe novo znanje, ali ne deklarativno znanje o novim činjenicama, već stječe novi skup sposobnosti. Koje bi to sposobnosti bile?

Prema Lewisu, kada Mary po prvi put vidi obojene predmete onda stječe nove sposobnosti da (i) diskriminira, (ii) prepoznaje, (iii) pamti i (iv) zamišlja to iskustvo. No, ta vrsta znanja ne implicira da ona ujedno stječe i nova vjerovanja o (fizičkim ili nefizičkim) činjenicama koje ranije nije imala.

Ova vrsta odgovora na AZ izazvala je dosta polemika među filozofima koji podražavaju AZ, ali i među onima koji ga nastoje opovrgnuti (vidi Jackson 1986; Bigelow i Pargetter 2006; Loar 2004; Conee 1994; Tye 2000, pogl. 1; Nemirow 1990; Malatesti 2004; Papineau 2002). Na primjer, iako prihvaćanje razlike između znanja kako i znanja da ima dosta dugu filozofsku tradiciju (u suvremene rasprave ovo razlikovanje uvodi Ryle 1949, pogl. 2), neki autori osporavaju da ona predstavlja dvije nesvodivo različite vrste znanja (vidi, npr. Crane 2001, 94–95). Dakle, kada bi se moglo pokazati da se znanje kako može reducirati na znanje da, onda bi ipak govorili o jednoj vrsti znanja te prigovor ekvivokacije ne bi bio uvjerljiv.

No, osim Lewisovog izvornog odgovora, u literaturi o AZ-u postoje različite varijante prigovora kojima se negira da Mary novim iskustvom stječe propozicijsko znanje činjenica. Neki od tih odgovora temelje se na tezi da Mary, imajući vizualno iskustvo, postaje *upoznata* (engl. *acquainted*) s njima i njihovim fenomenalnim karakterom na način koji se ne može reducirati na vrstu znanja koja imamo o fizičkim činjenicama (Conee 1994; Bigelow i Pargetter 1990). U filozofskim raspravama pojam „upoznatost“ koristi se kao tehnički termin koji predstavlja jednu formu nepojmovne (engl. *non-conceptual*) i nepropozicijske svjesnosti predmeta (Hasan i Fumerton 2020). U tom smislu riječi „upoznatost“ među prvima koristi Bertrand Russell (1911), koji razlikuje između znanja putem upoznatosti i znanja putem opisa kako bi adresirao određene probleme koji se javljaju u raspravama o temeljima lingvističkog značenja. U ovom kontekstu razlika se može ilustrirati sljedećim primjerom. Možemo znati identitet neke osobe tako da nam netko opiše njezine karakteristike. No isto tako možemo identificirati tu osobu kroz direktnu upoznatost, tj. tako da sami vidimo obilježja koja je karakteriziraju. Ako se prihvati ova razlika između znanja putem upoznatosti i znanja putem opisa, onda se može tvrditi da AZ nije valjan jer implicitno koristi dva različita pojma znanja. U crno-bijeloj sobi Mary ima propozicijsko znanje o fizičkim činjenicama koje stječe putem opisa. Kada jednom izađe iz sobe onda kroz direktnu upoznatost sa svojim iskustvima vanjskih predmeta stječe novu vrstu nepropozicijskog znanja.

Međutim, ova vrsta odgovora na AZ nije jako popularna u suvremenoj literaturi (Alter 1998; Gertler 1999; Papineau 2002). Naime, iako razlika između znanja putem upoznatosti i znanja putem opisa djeluje intuitivno, ne treba smetnuti s uma da je „upoznatost“ filozofski termin čiju se valjanost, naročito među fizikalistima, osporava. Prema Russellu (1911) spoznaja putem upoznatosti podrazumijeva da imamo *direktan* i *neposredan* dodir s predmetom spoznaje, dok je znanje putem opisa inferencijalno i omogućuje

samo *posrednu* spoznaju predmeta. Na primjer, prema Russellu znanje o postojanju vanjskih predmeta, poput vjerovanja da vani pada kiša, moguće je samo putem opisa. Razlog tome je što uvijek možemo sumnjati u istinitost tvrdnje da vani pada kiša, tj. uvijek ostaje otvorena mogućnost da sanjamo, imamo nekakve iluzije i tome slično. No, ono oko čega se ne možemo varati je ono s čime smo *direktno* upoznati. U ovom slučaju to bi bilo osjetilno iskustvo koje uključuje *percepciju* padanja kiše. Upravo ta pretpostavka da postoje predmeti iskustva kojih smo neposredno svjesni i o kojima imamo nepobitno znanje za mnoge je autore sporna. Jedan od razloga je taj što nije jasno kako bi se takva vrsta mentalnih predmeta mogla uklopiti u znanstvenu sliku svijeta. Čini se da takva iskustva nisu vani u svijetu jer bi ih onda mogli istraživati kao ostale fizičke pojave. Također, čini se da se ne nalaze u našim glavama jer prema našim moždanim stanjima nemamo epistemički neopovrgljiv pristup. S obzirom na te probleme, mnogi autori odbacuju postojanje takvog znanja i predmeta koje ono pretpostavlja (vidi, npr. Sellars 1956; Fumerton 2005).

Paul Churchland ponudio je zanimljivu varijantu odgovora na AZ koja se također oslanja na razlikovanje između deklarativnog i drugih vrsta znanja. Churchland (1985) nudi rekonstrukciju AZ-a oslanjajući se na neurokompjutacijske modele i neuroznanstvene podatke kojima nastoji ukazati na činjenicu da su različiti dijelovi mozga zaduženi za obradu informacija koje povezujemo s deklarativnim znanjem i za one informacije koje dobivamo kroz percepciju. Na primjer, kada Mary uči znanstvene činjenice o prirodi boja i ljudskom perceptivnom aparatu i kada drugima izlaže te informacije, služi se lingvističkim formatom reprezentacija koje se obrađuju u dijelovima mozga zaduženima za obradu jezika (poput Wernickeovog područja u temporalnom režnju). Međutim, kada Mary direktno percipira obojani predmet, njezin mozak reprezentira tu informaciju u nelingvističkoj formi koja se obrađuje u vizualnom režnju mozga. Stoga Churchland tvrdi da ono što Mary nauči kada ima vizualno iskustvo boja nije *nova* činjenica, koja se može analizirati u terminima deklarativnog ili propozicijskog znanja, već znanje o već poznatim činjenicama predstavljeno na novi način, tj. u novom nepropozicijskom ili nelingvističkom formatu.

Prethodnim odgovorima kojima se Maryjino novo znanje interpretira u terminima nepropozicijskog znanja prigovara se da ipak ne mogu isključiti da s novim iskustvom Mary stječe i nova vjerovanja, tj. informacije koje su predstavljena u propozicijskom formatu. Tu se posebice ističe odgovor Briana Loara (2004) koji ukazuje na to da jednom kada Mary vidi crvenu boju, onda dolazi u poziciju da formulira misli koje joj ranije nisu bile dostupne. Na primjer, nakon izlaska iz crno-bijele sobe može imati sljedeću misao:

> Ako je ovo fenomenalni karakter doživljaja crvene boje, onda mi se to ne sviđa.

Logički veznici poput kondicionala koji izražavamo u rečenicama s „ako … onda", povezuju *propozicijski* sadržaj naših vjerovanja. Dakle, antecedens gornje kondicionalne rečenice izražava propozicijski sadržaj koji možemo opisati na sljedeći način:

> […] ovo je fenomenalni karakter doživljaja crvene boje.

Teško je objasniti kako bi Mary mogla, jednom kada ima doživljaj boje, razmatrati takvu misao, ako doživljajem boje stječe samo sposobnosti da diskriminira, zapamti i tome slično, a ne i nove propozicijske stavove poput vjerovanja.

Razmatranja koja smo dosad iznijeli zasigurno ne predstavljaju konačnu riječ o uvjerljivosti odgovora sposobnosti. No, u nastavku se nećemo više njime baviti. U sljedećem odjeljku prelazimo na drugu vrstu utjecajnih odgovora na AZ kojima se dopušta da Mary izlaskom iz crno-bijele sobe stječe nova vjerovanja o doživljajima boja, no negira se da iz toga slijedi neistinitost fizikalizma.

8.7 Mary stječe nova vjerovanja o fizičkim činjenicama

Neki autori odgovaraju na AZ prihvaćajući ono što se ponekad naziva odgovor stare činjenice/nove reprezentacije (engl. *the old-fact/new-representation reply*). Prema ovom gledištu, Mary izlaskom iz crno-bijele sobe uči nove načine razmišljanja o fizičkim činjenicama s kojima je već ranije bila upoznata (Loar 1990; 2004; Horgan 1984; McMullen 1985; Pereboom 1994; Tye 2002; Malatesti 2012; Papineau 2002; Carruthers 2000). Obično ova vrsta odgovora pretpostavlja mogućnost imanja dva različita vjerovanja o istoj činjenici, gdje se vjerovanje shvaća kao *način* razmišljanja o nekom mogućem ili aktualnom stanju stvari. U tom smislu, iako možemo reći da se vjerovanje V_1 čiji je sadržaj „Superman je visok" razlikuje od vjerovanja V_2, čiji sadržaj je „Clark Kent je visok", oba vjerovanja referiraju na istu činjenicu da je osoba na koju referiramo izrazima „Superman" i „Clark Kent" visoka. Dakle, možemo reći da svako vjerovanje uključuje reprezentaciju određenog *aspekta* mogućeg ili aktualnog stanja stvari.

Odgovor stare činjenice/nove reprezentacije također se u literaturi naziva *strategija fenomenalnih pojmova*. Ova vrsta odgovora pretpostavlja da se mentalna stanja poput vjerovanja individuiraju na temelju pojmova koji čine njihov sadržaj. Da uzmemo prethodni primjer, vjerovanja V_1 i V_2 smatraju se različitima jer je njihov sadržaj sastavljen od različitih pojmova koje izražavamo riječima „Superman" i „Clark Kent". Stoga možemo reći da, iako vjerovanja V_1 i V_2 referiraju na istu činjenicu (da je Superman, tj. Clark Kent visok), njihovi *sadržaji* se razlikuju jer tu istu činjenicu predstavljaju na drugačiji način, tj. predstavljaju je pod različitim pojmovima.

Kako bi osporili konkluziju na koju upućuje AZ, pristaša ove strategije mora argumentirati da je vjerovanje koje Mary stječe o fenomenalnom karakteru njezinog iskustva uključuje pojam koji, iako referira na neko fizičko svojstvo o kojem je već ranije imala znanje, različito od pojmova koje je usvojila pri učenju fizičkih činjenica u crno-bijeloj sobi. Taj se pojam naziva *fenomenalni pojam*. Upravo zbog tog pojma Mary može imati novo vjerovanje, tj. stječe novi način razmišljanja o već poznatim fizičkim činjenicama o fenomenalnom karakteru iskustva.

Općenito smatra se da su fenomenalni pojmovi uključeni u misli, vjerovanja i spoznaje koje se odnose na fenomenalni karakter iskustva. Ti pojmovi se uče kroz razmišljanje iz perspektive prvog lica o fenomenalnom karakteru iskustava koje proživljavamo. Stoga prema pristašama strategije fenomenalnih pojmova, kada razmišljamo o tome kako nam je u iskustvu dana, na primjer, plava boja mora koje promatramo ili zamišljamo i kada izražavamo vjerovanja da imamo doživljaj s tim fenomenalnim karakterom, onda koristimo fenomenalne pojmove. U nastavku ćemo shematski objasniti kako bi strategija fenomenalnih pojmova trebala funkcionirati.

Prema pristašama strategije fenomenalnih pojmova, Mary prije izlaska iz crno-bijele sobe posjeduje pojam **FC** (u nastavku otisnutim slovima označujemo pojmove) koji referira, na primjer, na fenomenalni karakter iskustva viđenja crvene boje. Referent **FC**-a, tj. ono što **FC** označuje, može biti neka složena fizička ili funkcionalna relacija između mozga i okoline u kojoj se nalazi. Kada Mary izađe iz sobe i po prvi put vidi predmet crvene boje, stječe i počinje koristiti fenomenalni pojam **FFC**, koji prema pretpostavci ima istog referenta kao i **FC**. Dakle, ideja je da Mary još dok se nalazi u crno-bijeloj sobi može imati vjerovanje V_3 čiji se sadržaj može izraziti na sljedeći način:

FC je fenomenalni karakter iskustva crvene boje.

Jednom kada zaista vidi crvenu boju onda može formirati novo vjerovanje V_4 čiji se sadržaj može izraziti ovako:

FFC je fenomenalni karakter doživljaja crvene boje.

Prema pobornicima strategije fenomenalnih pojmova, V_3 i V_4 predstavljaju različita vjerovanja jer im je sadržaj određen različitim pojmovima **FC** i **FFC**. Međutim, oba vjerovanja se odnose na istu fizičku činjenicu koju je Mary već znala prije nego je stekla novo svjesno iskustvo gledanja crvene boje. Kao što netko može naučiti da je Clark Kent ista osoba kao i Superman, tako i Mary može naučiti da se pojmovi **FFC** i **FC** odnose na isto fizičko svojstvo ili relaciju.

Ovo je shematski prikaz načina na koji strategija fenomenalnih pojmova nastoji blokirati AZ. Daljnja elaboracija ove strategije zahtijeva suočavanje s određenim izazovima.

Jedan od izazova odnosi se na odabir varijante fizikalizma koji se želi obraniti. Neki pobornici strategije fenomenalnih pojmova prihvaćaju apriorni fizikalizam (Tye 2000; Hellie 2004). Prema njima, Mary ne može logički derivirati znanje o fenomenalnom karakteru iskustva na temelju znanja o fizičkim činjenicama jer prije izlaska iz crno-bijele sobe ne posjeduje fenomenalne pojmove. Međutim, prema njima to ne implicira da tezu supervenijencije treba shvatiti kao aposteriornu relaciju. Dapače, smatraju da osoba, jednom kada stekne određena iskustva te stoga posjeduje relevantne fenomenalne pojmove, može derivirati znanje o fenomenalnom karakteru iskustva iz potpunog znanja o fizičkim činjenicama.

Međutim, mnogi pobornici strategije fenomenalnih pojmova prihvaćaju aposteriorni fizikalizam. Prema njima, iako AZ ne dokazuje da je fizikalizam neistinit, potvrđuje da se teza supervenijencije treba shvatiti kao aposteriorna relacija. Međutim, kao što smo istaknuli ranije u odjeljku 7.1.1., pobornici aposteriornog fizikalizma moraju moći pokazati da unatoč tome što se relacija supervenijencije spoznaje *a posteriori*, mentalno svejedno *nužno* supervenira nad fizičkim. Problem za aposteriorni fizikalizam je taj da, ako se neke činjenice o odnosu mentalnog i fizičkog spoznaju *a posteriori*, onda nam to daje razlog za sumnju da relacija supervenijencije predstavlja nužan odnos između mentalnog i fizičkog. Međutim, aposteriorni fizikalisti mogu na ovaj problem odgovoriti na više načina. U nastavku ćemo razmotriti neke od tih odgovora.

Fizikalisti mogu prihvatiti ovu nepoželjnu posljedicu (engl. *bite the bullet*) prihvaćajući da je teza supervenijencije kontingentna. Ovaj potez bi uključivao svojevrsni povratak izvornoj verziji teorije identiteta tipova o kojoj smo govorili u poglavlju 4. Neki autori i dalje prihvaćaju ovu verziju fizikalizma (vidi Berčić 2012, pogl. Um; Carruthers 2000). Alternativno, fizikalisti bi mogli tvrditi da, iako je supervenijencija mentalnog nad fizičkim aposteriorna, to ne implicira da je njihov odnos kontingentan. Ako je Saul Kripke (1997) u pravu da postoje nužne istine koje se spoznaju *a posteriori*, poput toga da je voda istovjetna s H_2O, onda bi se moglo tvrditi da je relacija supervenijencije mentalnog nad fizičkim nužna iako je spoznajemo kroz aposteriorna istraživanja. Ovo gledište prihvaćaju neki od autora čija ćemo stajališta razmotriti kasnije, kada se budemo bavili argumentom pojmljivosti. U nastavku ćemo se vratiti razmatranju načina na koji se dalje može razviti strategija fenomenalnih pojmova.

Daljnja elaboracija strategije fenomenalnih pojmova zahtijeva adresiranje određenih poteškoća koje se javljaju s obzirom na razinu općenitosti ove rasprave. Na najopćenitijoj razini javljaju se problemi oko shvaćanja prirode i uvjeta individuacije mentalnih stanja poput vjerovanja, pojmova i njihovih međusobnih odnosa. To su sve stvari oko kojih ne postoji jasan konsenzus te se po tim pitanjima i dalje vode rasprave u filozofiji i kognitivnim znanostima (vidi Margolis 2000; Machery 2009). Nadalje, fizikalisti koji koriste strategiju

fenomenalnih pojmova moraju ponuditi teoriju fenomenalnih pojmova koja će biti kompatibilna s fizikalizmom i općenitijim teorijama o tome kako pojmovi inače funkcioniraju. Žustre rasprave oko ovih pitanja ukazuju na to da i daljnje ostaje otvoreno pitanje mogu li fizikalisti ponuditi zadovoljavajuća rješenja u tom pogledu (Schiffer 1987).

Osim tih općenitih izazova, postoje i specifičniji problemi koji proizlaze iz određenih pretpostavki koje fenomenalni pojmovi moraju zadovoljiti kako bi ova strategija uspješno funkcionirala. Prvo što je važno istaknuti jest da bi uspješna analiza Maryjinih novih vjerovanja u terminima fenomenalnih pojmova trebala zahvatiti točno onaj sadržaj novih misli o fenomenalnom karakteru iskustva koji postaje dostupan tek kada Mary izađe iz crno-bijele sobe. Drugo, fenomenalni pojmovi se na neki način moraju razlikovati od fizikalnih pojmova koje Mary od ranije posjeduje. Dakle, strategija fenomenalnih pojmova pretpostavlja teoriju pojmova koja dopušta da koreferencijalni pojmovi mogu biti različiti, u smislu da imaju različite uvjete individuacije. U tu svrhu neki autori prihvaćaju kriterije za individuaciju pojmova inspiriranu teorijom Gottloba Fregea (1995). Takvu je teoriju razvio Christopher Peacocke, prema kojemu se pojmovi mogu razlikovati na sljedeći način:

> Različitost pojmova: Pojmovi **C** i **D** različiti su ako i samo ako postoje dva cjelovita propozicijska sadržaja koja se razlikuju jedino u tome da u jednome **D**, na jednom ili više mjesta, zamjenjuje **C**, i od kojih je jedan potencijalno informativan dok drugi nije. (Peacocke 1983, 2)

Na primjer, pojmovi **SUPERMAN** i **CLARK KENT** različiti su jer su sadržaji propozicija u kojima se pojavljuju različiti. Kada kažemo „Superman = Superman“ onda nismo dali baš informativni iskaz. Međutim, ako kažemo „Superman = Clark Kent“ onda smo dali informativniju propoziciju od prethodne iako jedna i druga referiraju na isto stanje stvari. Dakle, budući da zamjenom pojma **SUPERMAN** pojmom **CLARK KENT** unutar neke propozicije dobivamo novu propoziciju koja je potencijalno informativnija od prethodne, radi se o različitim pojmovima. Taj uvjet, na primjer, nije zadovoljen kod sljedećih propozicija. Uzmimo rečenicu „Neženja je neženja“ i zamijenimo pojam **NEŽENJA** s **NEOŽENJENI MUŠAKARAC.** Budući da propozicija „Neženja je neženja“ nije informativnija od propozicije „Neženja je neoženjeni muškarac“ onda slijedi, ne samo da pojmovi **NEŽENJA** i **NEOŽENJENI MUŠAKARAC** referiraju na istu stvar, tj. osobu, već predstavljaju *isti* pojam.

Pobornici strategije fenomenalnih pojmova smatraju da, ako se prihvati Peacockeovo gledište na razlikovanje pojmova, onda AZ samo pokazuje da su **FC** i **FFC** različiti pojmovi (vidi, npr. Papineau 2002). Štoviše, prema

njihovoj interpretaciji Maryjinog novog vjerovanja, iako Mary neće smatrati informativnim iskaz:

FC je **FC**,

ipak će smatrati informativnim iskaz:

FC je **FFC**.

To slijedi upravo iz intuicije na kojoj se temelji misaoni eksperiment da Mary ne može derivirati vjerovanje o fenomenalnom karakteru iskustva na temelju potpunog znanja fizičkih činjenica.

Vidjet ćemo, međutim, da Peacockeov princip za razlikovanje pojmova nije univerzalno prihvaćen među filozofima te da neki autori osporavaju uspješnost strategije fenomenalnih pojmova upravo propitivanjem uvjerljivosti tog principa. No, prije toga, u nastavku ćemo razmotriti druga ograničenja koja fenomenalni pojmovi moraju zadovoljiti kako bi ih se moglo koristiti u argumentu protiv AZ-a.

Treće ograničenje je da djelatnik mora imati iskustvo određenog tipa kako bi posjedovao određeni fenomenalni pojam. Dakle, Mary ne može imati ili naučiti pojam **FFC**, a da ne proživi iskustvo kako je to vidjeti crvenu boju. To objašnjava zašto Mary jednom kada vidi crvenu boju stječe sposobnost novog načina razmišljanja o fenomenalnom karakteru tog iskustva koje prema fizikalistima supervenira nad fizičkim činjenicama.

Četvrto ograničenje je da fenomenalni pojmovi ne smiju referirati na svojstva koja ne superveniraju nad fizičkim. Ovo se ograničenje odnosi na opći zahtjev da se pojmovi moraju moći objasniti na način koji će biti u skladu s minimalističkom verzijom fizikalizma, a to je supervenijentni fizikalizam. Jednom kada smo naveli ograničenja koje fenomenalni pojmovi moraju zadovoljiti, razmotrit ćemo glavne načine na koje se strategija fenomenalnih pojmova koristi kako bi se odgovorilo na AZ. Kao što ćemo vidjeti, pobornici strategije fenomenalnih pojmova mogu se razlikovati prema tome kako shvaćaju prirodu fenomenalnih pojmova.

Neki pobornici ove strategije smatraju da se fenomenalni pojmovi trebaju shvatiti kao indeksikalni ili demonstrativni pojmovi (Perry 2001; McMullen 1985). Ovi pojmovi izražavaju se indeksikalnim i demonstrativnim izrazima poput „ja”, „ti“ ili „to“, „ovo“ ili „ono“.

Ovdje ćemo razmotriti gledište koje je razvio John Perry (2001), budući da je ono integrirano sa širom raspravom o semantici indeksikalnih i demonstrativnih izraza. Iako se ovdje nećemo specifično baviti semantikom koju koristi Perry, važno je istaknuti da Perry razvija svoju varijantu strategije fenomenalnih pojmova na temelju semantike indeksikala i demonstrativa koja nije formulirana *ad hoc* kako bi se odgovorilo na AZ, već se ona neovisno razvija kroz dugogodišnju tradiciju u filozofiji jezika.

Osnovna misao Perryjevog prijedloga je da se Mary, imajući svjesno iskustvo boje i dodatne kolateralne informacije o boji koju gleda, dovodi u situaciju u kojoj može referirati na fenomenalni karakter tog iskustva pomoću demonstrativnih pojmova. Budući da je prema Perryju fenomenalni karakter svojstvo koje supervenira nad fizičkim, slijedi da Mary na to isto svojstvo može referirati pomoću pojma koji pripada korpusu znanja o fizičkim činjenicama. Označimo ponovo taj pojam sa **FC**. Dakle, prema Perryju, Mary izlaskom iz crno-bijele sobe i opažanjem crvene ruže nauči sljedeće:

Fenomenalni karakter **FC** doživljaja crvene ruže je *ovo*.[63]

Prema Perryju, izraz „ovo“ iz prethodne rečenice izražava fenomenalni pojam koji je Mary stekla jednom kada je vidjela crvenu ružu. To je poseban tip demonstrativnog pojma koji, uz pomoć introspekcije, referira na fenomenalni karakter iskustva koje osoba proživljava. S obzirom na njegovu važnost za našu raspravu, označit ćemo taj pojam s **Ovo$_{FC}$**. Sada ćemo razmotriti zadovoljava li ovaj pojam prethodno navedena ograničenja i pretpostavke koje fenomenalni pojmovi moraju zadovoljiti kako bi uspješno odgovorili na AZ.

Perry smatra da Maryjino znanstveno znanje ne sadrži indeksikalne ili demonstrativne pojmove koji zahtijevaju imanje iskustva boje kako bi referirali na fenomenalni karakter tog iskustva. Ali to nije zato što fenomenalni karakter ne supervenira nad fizičkim svojstvima, već je to instanca općenite činjenice da se indeksikalni i demonstrativni pojmovi razlikuju od neindeksikalnih i nedemonstrativnih pojmova, čak i kada referiraju na istu stvar. Razliku između indeksikalnih i ostalih vrsta pojmova možemo pokazati na primjeru koji koristi Perry (1979). Zamislimo da se John nalazi u trgovini i prati trag šećera koji ispada iz kolica koja gura neka osoba. John može razmišljati o osobi koja ostavlja trag šećera pod opisom „Osoba koja radi nered u trgovini“ ili „Osoba koja će biti vrlo razočarana kada otkrije da nema šećera u kolicima“ i tako dalje. Pretpostavimo sada da je John upravo ta osoba kojoj ispada šećer iz kolica te da zapravo prati svoj trag, a da toga nije svjestan. Prema tome, misao koja je izražena u sljedećem iskazu:

Osoba koja radi nered je osoba koja radi nered,

je manje informativna za Johna nego misao:

Ja sam osoba koja radi nered.

[63] Možemo zamisliti da Mary ujedno i prstom pokazuje na cvijet, ali ne nužno. Fenomenalni pojmovi bi prema Perryju trebali biti neka vrsta unutrašnjeg, introspektivnog „pokazivanja“ koje nije nužno popraćeno javno opažljivom tjelesnom gestikulacijom.

Dakle, iako je misao izražena s rečenicom „Osoba koja radi nered je ostala bez šećera“ i ona koju izražavamo s „Ja sam ostao bez šećera“ o istoj osobi i istoj činjenici, one pretpostavljaju dva različita načina razmišljanja o njima.

Perry (2001) tvrdi nešto slično za demonstrativne pojmove. Tu daje primjer osobe koja se divi djelima filozofa Freda Dretskea, ali ne zna kako on izgleda. Kada ga sretne na zabavi onda shvati da iskaz „Htio bih se rukovati s Fredom Dretskeom“ ima istu referenciju kao i iskaz „Htio bih se rukovati s ovom osobom“ jer se odnosi na Dretskea koji stoji ispred nje. Ovdje opet imamo slučaj da je sadržaj „Fred Dretske je Fred Dretske“ manje informativan za ovu osobu od iskaza „Ova osoba je Fred Dretske“.

Ova analiza koristi pojmove koji zadovoljavaju nekoliko ograničenja na fenomenalne pojmove koje smo razmotrili ranije. Fenomenalni pojam **Ovo$_{FC}$**, budući da je demonstrativni pojam, različit je i ne može se derivirati na temelju pojmova koje Mary mora posjedovati da bi imala potpuno znanje o fizičkim činjenicama. Osim toga, Mary ne može naučiti koristiti **Ovo$_{FC}$** i primijeniti ga, a da nema relevantno iskustvo. Nadalje, iz Perryjevog objašnjenja semantičkih svojstava demonstrativnih pojmova poput „ovaj FC“ jasno je da ne moramo pretpostaviti da referira na neka nefizička svojstva ili relacije. Dakle, čini se da objašnjenje fenomenalnih pojmova u terminima demonstrativa može zadovoljiti većinu kriterija potrebnih za uspješno odgovaranje na AZ. Ipak, neki autori propituju nudi li ovo objašnjenje adekvatnu karakterizaciju svega što Mary nauči imanjem određenog svjesnog iskustva.

Mnogi pobornici strategije fenomenalnih pojmova, naročito oni koji smatraju da fenomenalni pojmovi podrazumijevaju sposobnost specifičnog načina diskriminacije i prepoznavanja iskustva, složit će se da Mary, jednom kada ima iskustvo boje, stječe sposobnost korištenja demonstrativa kojima može na novi način referirati na fenomenalni karakter iskustva (Tye 2000; Carruthers 2000; Malatesti 2012). Međutim, dodat će da Mary ujedno stječe dodatne sposobnosti koje treba istaknuti u adekvatnoj analizi onoga što je uključeno u stjecanje novog vjerovanja o fenomenalnom karakteru iskustva. Prema njima, kako bi savladala demonstrativni fenomenalni pojam **Ovo$_{FC}$**, Mary mora biti sposobna diskriminirati i izdvojiti fenomenalni karakter iskustva od drugih obilježja koje to iskustvo može imati. Nadalje, ova sposobnost diskriminacije će u nekim slučajevima biti nužna za prepoznavanje radi li se o istom ili različitom fenomenalnom karakteru iskustva kada se ono pojavi u nekom kasnijem vremenu. Pobornici teorije prema kojoj su fenomenalni pojmovi *sposobnosti* za prepoznavanje ili diskriminiranje ne smatraju samo da takve sposobnosti individuiraju određene nedemonstrativne fenomenalne pojmove, već da one same predstavljaju posebnu vrstu fenomenalnih pojmova. Dakle, pretpostavljaju da se neki fenomenalni pojmovi reduciraju na sposobnosti diskriminacije i prepoznavanja fenomenalnih karaktera iskustva. U nastavku ćemo

razmotriti uspijeva li ova teorija zadovoljiti relevantna ograničenja na fenomenalne pojmove.

Čini se da zadovoljava. Prvo, fenomenalni pojmovi shvaćeni kao sposobnosti za diskriminaciju ili prepoznavanje omogućuju nove misli koje nisu bile dostupne prije nego se proživi iskustvo određenog tipa. Drugo, ovako shvaćeni fenomenalni pojmovi ne mogu se derivirati iz fizikalnih pojmova jer se temelje na sposobnosti diskriminacije i prepoznavanja. Treće te se sposobnosti stječu imanjem iskustva određene vrste. Četvrto, kada pripisujemo Mary sposobnosti za diskriminaciju i prepoznavanje onda ne moramo pretpostaviti da one referiraju na neka nefizička svojstva i relacije.

Međutim, ni ovo gledište nije imuno na prigovore. Na primjer, Fodor (1998) je argumentirao, neovisno o raspravi o strategiji fenomenalnih pojmova, da ne može postojati vrsta pojmova koja se svodi na imanje nekih sposobnosti. Razlog tome je što su pojmovi prema svojoj prirodi kompozicionalni. U smislu da se pojmovi mogu kombinirati kako bi se formirali kompleksniji pojmovi. Na primjer, pojmovi pas i crveno mogu se povezati kako bi se formirao složeniji pojam crveni pas. Međutim, nije jasno da se sposobnosti mogu kombinirati na sličan način. Kombinacijom sposobnosti za vožnju bicikla i sposobnosti za vožnju automobila nećemo dobiti novu smislenu kompleksniju sposobnost vožnje bicikl-automobila. Stoga, ako su fenomenalni pojmovi neka vrsta sposobnosti, onda se čini da zapravo ne govorimo više o *pravim* pojmovima.

Također vrijedi spomenuti gledište Davida Papineaua (2002) prema kojemu su fenomenalni pojmovi vrsta citirajućih pojmova (engl. *quotational concepts*). Ti pojmovi, prema analogiji s tekstualnim citatima koji u sebi sadrže dio citiranog teksta, u sebi sadrže iskustvo ili fenomenalni karakter na koji referiraju. Ovo gledište također može objasniti zašto Mary u crno-bijeloj sobi nema fenomenalne pojmove. Citirajući pojmovi su upravo oni koje stječemo imanjem određenog svjesnog iskustva. Nadalje, budući da znanje fizičkih činjenica ne uključuje imanje svjesnih iskustava, ta iskustva ne mogu biti sadržana u fizikalnim pojmovima, niti se na temelju njih mogu naučiti citirajući pojmovi.

Shvaćanje fenomenalnih pojmova kao citirajućih pojmova također ima određene probleme. Čini se da pomoću citirajućih pojmova ne možemo zahvatiti razmišljanje o fenomenalnim iskustvima u trenucima kad ih ne proživljavamo (vidi Crane 2005). Naime, rekli smo da ovo gledište pretpostavlja da je, kada koristimo citirajuće pojmove, samo iskustvo na koje referiramo, poput citiranog teksta, već sadržano u tim pojmovima. Drugim riječima, ispada da nije moguće razmišljati o tim iskustvima kroz prizmu fenomenalnih pojmova ako ih ujedno ne proživljavamo. Dakle, ako smatramo da korištenjem fenomenalnih pojmova možemo razmišljati o svjesnim karakterima iskustva čak i kada ih ne proživljavamo, onda se gledište citirajućih pojmova neće činiti uvjerljivim. Međutim, u nastavku

nećemo raspravljati uvjerljivost pojedinih varijanti strategije fenomenalnih pojmova. Umjesto toga, osvrnut ćemo se na jedan općeniti prigovor protiv korištenja te strategije.

8.8 Kritika strategije fenomenalnih pojmova

Vidjeli smo da je jedna od temeljnih postavki strategije fenomenalnih pojmova da se pojmovi mogu individuirati pomoću propozicija koje ljudi smatraju informativnima. Ovo se može nazvati *internalističkim* gledištem na individuaciju pojmova. Prema ovom gledištu, identitet pojma ovisi o onome što djelatnik koji kompetentno koristi taj pojam smatra informativnim iz svoje perspektive. Međutim, eksternalisti u pogledu sadržaja misli smatraju da su psihološka svojstva, poput informativnosti za pojedinu osobu, nebitni za individuaciju pojmova. Nasuprot tome, eksternalisti smatraju da je referencija pojma ono što je bitno za njegovu individuaciju. Stoga prema eksternalistima ne mogu postojati dva različita pojma koja referiraju na istu stvar. Eksternalizam se u filozofiji jezika i uma branio vrlo utjecajnim misaonim eksperimentima koje su osmislili Hilary Putnam (1975g) i Tyler Burge (1979). Bez ulaženja u detaljnu diskusiju (za pregled rasprave, vidi Rowlands, Lau, i Deutsch 2020), u nastavku ćemo eksternalističko gledište objasniti koristeći jedan Putnamov (1975g) primjer.

Uvjerljivom se čini tvrdnja da se ljudi mogu razlikovati u stupnju razumijevanja određenog pojma. Na primjer, osoba može znati da pojmovi „bukva“ i „brijest“ označuju različite vrste stabla. Međutim, možemo zamisliti da prosječna osoba koja posjeduje te pojmove neće, kada vidi stablo brijesta, nužno moći odrediti radi li se o bukvi ili brijestu. Za razliku od nje, botaničar može jednostavno na temelju promatranja stabla reći da se zapravo radi o brijestu. Ovaj primjer nam pokazuje da se dvije osobe mogu razlikovati u pogledu toga što će smatrati informativnim. Međutim, iz toga ne slijedi da prosječna osoba i stručnjak imaju različite *pojmove* bukve i brijesta. Ono po čemu se oni razlikuju odnosi se na kompetentnost služenja tim istim pojmovima.

Štoviše, eksternalisti poput Tylera Burgea (1979) tvrde da čak i osoba koja povezuje u potpunosti pogrešna vjerovanja s određenim pojmom svejedno može posjedovati taj pojam. Burge kao primjer daje slučaj Alfa koji misli da je artritis, između ostalog, i upala mišića. Pa tako kada osjeća bol u bedru misli da je to zato što ima artritis. Prema Burgeu, može se reći da Alf pogrešno vjeruje da zbog artritisa osjeća bol u bedru. Kada mu doktor kaže da to što osjeća nije posljedica artritisa jer je artritis poglavito upala zglobova, očekujemo da će Alf prihvatiti da je pogriješio kada je mislio da ima artritis. To očekujemo jer zapravo smatramo da Alf i doktor posjeduju isti pojam artritisa koji Alf pogrešno koristi. Kada to ne bismo smatrali onda bi imalo smisla očekivati da će Alf nastaviti tvrditi da je zapravo imao točno

vjerovanje o poremećaju u bedrima koji ona naziva „artritis“, a razlikuje se od onoga što doktora naziva „artritis“.

Dakle, prema eksternalistima, posjeduje li osoba određeni pojam ne ovisi isključivo o njezinim psihološkim sposobnostima ili o tome što smatra informativnim kada koristi taj pojam. Osim znanja i sposobnosti koje osoba može imati, ono što je važno za određivanje i individuiranje pojmova je jezična zajednica kojoj pripada pojedina osoba.

Derek Ball (2009) i Michael Tye (2009) argumentirali su da jednom kada prihvatimo eksternalizam više ne možemo smatrati da postoje fenomenalni pojmovi te se ne možemo oslanjati na njih kako bismo odgovorili na AZ. Štoviše, ako prihvatimo fizikalističku pretpostavku da fenomenalni karakter supervenira nad fizičkim svojstvima, onda možemo reći da Mary već prije izlaska iz crno-bijele sobe posjeduje pojam koji se odnosi na fenomenalni karakter iskustva. Prema eksternalističkom gledištu, imanjem vizualnog iskustva boje Mary ujedno stječe sposobnost prepoznavanja fenomenalnog karaktera iskustva. Međutim, kako nam pokazuje Putnamov primjer s pojmovima bukve i brijesta, ova sposobnost ne omogućuje stjecanje novog pojma. Ova razmatranja naizgled ukazuju na to da uspješna primjena strategije fenomenalnih pojmova zahtijeva obranu internalizma u pogledu pojmova. To znači da pobornici ove strategije moraju pokazati zašto eksternalizam nije uvjerljivo gledište ili argumentirati da uvjerljiv oblik eksternalizma nije nespojiv s postojanjem fenomenalnih pojmova.

Nakon što smo razmotrili različite moguće odgovore na argument iz znanja i probleme s kojima se suočava, u nastavku ćemo se baviti jednim drugim utjecajnim argumentom protiv fizikalizma. Razmotrit ćemo suvremenu varijantu argumenta pojmljivosti kako ga je formulirao David Chalmers.

8.9 Chalmersov argument pojmljivosti

Chalmers (1996, 93–171; 2010b) argumentom pojmljivosti nastoji pokazati da naša svjesna iskustva imaju obilježja koja ne superveniraju nad fizičkim svojstvima. Kao što ćemo vidjeti, njegov argument je dosta kompleksan, stoga nije naodmet na prvom koraku ga predstaviti u simplificiranom obliku te postepeno dodavati složenije detalje. Osnovna struktura argumenta pojmljivosti (skraćeno AP) može se iskazati na sljedeći način:

1) Zombiji su pojmljivi.
2) Štogod je pojmljivo, moguće je.
3) Dakle, zombiji su mogući.
4) Ako su zombiji mogući, onda je fizikalizam neistinit.

Dakle:

5) Fizikalizam je neistinit.

Kako bismo provjerili valjanost ovog argumenta, moramo provjeriti jesu li mu premise istinite. Međutim, kako bismo to učinili prvo moramo pojasniti tehničke pojmove koje Chalmers koristi u AP.

Pojam zombija iz premise (1) odnosi se na entitet koji s ljudima dijeli sva fizička, funkcionalna i biološka svojstva, uključujući i bihevioralne dispozicije. Jedino u čemu se zombiji razlikuju od ljudi je činjenica da njihova iskustva nemaju fenomenalni karakter. Dakle, ne postoji nešto kako je to biti zombi ili kako je to za zombija imati određeno iskustvo. U tom smislu, ova vrsta filozofskog zombija treba se razlikovati od holivudskog zombija. Filozofski zombiji su na van (u fizičkom smislu) identični ljudima, jedino po čemu se razlikuju je što nemaju svjesna mentalna stanja. Dakle, kada ih nastojimo zamisliti, ne smijemo zamišljati ljude koji su se vratili iz mrtvih i kojima upravljaju bazični instinkti za prežderavanjem, već bismo trebali zamisliti ljude koji su poput nas, jedino što nemaju osjetilna iskustva i druga *svjesna* mentalna stanja. Važne aspekte filozofskih zombija možemo malo preciznije formulirati oslanjajući se na pojam *zombi-svijeta* koji koristi Chalmers (2010b).

Pretpostavimo da je „F konjunkcija svih mikrofizičkih činjenica o svijetu“ i da je „Q istinita tvrdnja o fenomenalnom karakteru nekog iskustva“. Na primjer, Q se može odnositi na činjenicu da osoba osjeća probadajuću bol. U tom smislu iskaz F&¬Q opisuje zombi-svijet. Dakle, to bi trebao biti svijet koji je u svim fizičkim aspektima identičan našem svijetu, uključujući zakone prirode, ali se razlikuje u tome što u njemu ne postoje iskustva s fenomenalnim karakterom Q. Na primjer, pretpostavimo da se udarimo u prst čekićem te osjetimo bol koja ima određeni *bolni* fenomenalni karakter. Zombi-svijet je fizički duplikat našeg svijeta sve do posljednjeg detalja, koji sadrži naš fizički duplikat (našeg zombija) koji kada se udari u prst čekićem (kao što smo se mi udarili) počinje izvoditi identične radnje koje i mi izvodimo (na primjer, baca čekić na pod, hvata se za udareni prst, trlja ga i jauče), ali nema svjesno iskustvo boli s karakterističnim fenomenalnim karakterom. Dakle, prva premisa ovog argumenta preciznije se može izraziti na sljedeći način:

1) Pojmljivo je da F&¬Q.

Međutim, da bismo dali potpunu karakterizaciju ove premise potrebno je pojasniti pojam pojmljivosti koji se u njoj koristi.

Chalmers (2010b) tvrdi da se ovdje pojmljivost treba shvatiti kao *idealna pojmljivost.* Općenito možemo reći da je sadržaj S neke rečenice *idealno pojmljiv* kada osoba može zamisliti da je S istinit i iz pozicije *idealne* refleksije, uzimajući u obzir sve moguće zaključke, ne bi došla do konkluzije da S sadrži nekakvu kontradikciju. Na primjer, možemo idealno pojmiti da trokut ima tri stranice. Međutim, ova vrsta pojmljivosti nije zadovoljena u slučaju osobe koja poima da je 683*6 = 4101 jer se ta osoba ne nalazi u poziciji u kojoj može

idealno reflektirati. Kada bi napravila potrebne kalkulacije otkrila bi da ono što poima nije istinito, već da je 683,5*6 = 4101.

Chalmers (2010b) još razlikuje pozitivnu od negativne pojmljivosti. Pod pozitivnim poimanjem Chalmers misli na zamišljanje konkretne situacije koja na neki način egzemplificira ono što je idealno pojmljivo. Na primjer, čini se da možemo pozitivno idealno zamisliti zombije jer si možemo predočiti situaciju u kojoj postoje nama fizički identični ljudi koji nemaju svjesna iskustva. Ako osoba u poziciji idealne refleksije ne može pronaći razlog da odbaci određenu hipotezu, a ujedno ne može *pozitivno* zamisliti scenarij u kojem je hipoteza istinita, onda se radi samo o negativnom poimanju hipoteze. U nastavku ćemo govoriti samo o pozitivnoj idealnoj pojmljivosti.

Dakle, prva premisa u AP je da možemo pozitivno pojmiti zombije, tj. da je slučaj F&¬Q, i da iz pozicije idealne refleksije ne možemo pronaći razlog za smatrati da postoji kontradikcija u sadržaju koji zamišljamo. No ovdje nije kraj. Chalmers dalje uvodi razliku između *primarne* i *sekundarne* pojmljivosti koje se temelje na razlici između primarne i sekundarne mogućnosti (usp. Pećnjak i Janović 2016, pogl. 9).

Prema Chalmersu, sadržaj S određene misli ili iskaza je primarno ili epistemički moguć ako je, s obzirom na ono što *a priori* znamo, S istinit u mogućem svijetu za koji se uzima da je aktualni svijet. Na primjer, pretpostavimo da je aktualni svijet onaj koji sadrži vodu čija je molekularna struktura XYZ. Dakle, sadrži tvar s molekularnom strukturom XYZ koja zadovoljava sva uobičajena svojstva koja povezujemo s vodom na Zemlji, poput toga da je prozirna, tekuća, pitka, da se nalazi u rijekama, jezerima i tako dalje.[64] U tom svijetu je iskaz „Voda = XYZ" istinit, a iskaz „Voda = H_2O" neistinit. Budući da se čini mogućim, s obzirom na sve što *a priori* znamo o vodi, da postoji svijet u kojem je voda identična s nečim što ima molekularnu strukturu XYZ, onda je prema Chalmersu u primarnom ili epistemičkom smislu *moguće* da voda nije identična sa H_2O.

S druge strane, S je sekundarno ili metafizički moguć kada je C istinit u nekom mogućem svijetu koji je kontrafaktički u odnosu na aktualni svijet. S

[64] Primjer dolazi od Putnama (1975g) koji ga koristi u argumentu za semantički eksternalizam. Argument se temelji na misaonom eksperimentu u kojem zamišljamo Oskara koji živi u 17. stoljeću na Zemlji prije nego je otkrivena molekularna struktura vode. Toskar je psihofizički identičan Oskaru te živi na Zemlji blizanki, kontrafaktičkom svijetu koji je identičan našoj Zemlji osim što je tamo molekularna struktura vode XYZ. Toskar poput Oskara ne zna koji je kemijski sastav vode na Zemlji blizanki. Međutim, kada Oskar razmišlja da bi popio čašu vode njegova misao se odnosi na tvar koja ima molekularnu strukturu H_2O. Kada Toskar razmišlja o vodi njegova misao se odnosi na tvar koja ima molekularnu strukturu XYZ. Budući da su Oskar i Toskar prema pretpostavci psihofizički identični, onda ono što određuje sadržaj njihovih misli o vodi ne mogu biti psihološke činjenica o njima. Cilj ovog misaonog eksperimenta je podržati zaključak da je referencija naših misli i iskaza kojima ih izražavamo, barem kada se radi o pojmovima i terminima kojima označujemo prirodne vrste, određena objektivnim činjenicama i našim uzročnim odnosima sa svijetom.

obzirom na to da u našem svijetu voda jest H_2O, i da iskazi identiteta o supstancijama predstavljaju nužne istine (Kripke 1997), onda u XYZ svijetu ne može biti istinita tvrdnja „Voda je XYZ“, kao što nije istinita tvrdnja „Voda nije H_2O“. Drugim riječima, iskaz „Voda nije H_2O“ u sekundarnom smislu izražava metafizičku nemogućnost.[65]

Da sumiramo, ako se sadržaj S može pojmiti kao istinit u nekom svijetu za koji uzimamo da je aktualan, tada je S primarno pojmljiv. Dakle, procedura za određivanje primarne pojmljivosti izvodi se tako da uzmemo neki mogući svijet kao aktualan i onda iz njegove perspektive gledamo što je istinito u njemu. S druge strane, ako se S može pojmiti kao istinit samo u nekom svijetu za koji se uzima da je kontrafaktičan našem aktualnom svijetu, tada je sekundarno pojmljiv. Na primjer, možemo zamisliti da je prvi predsjednik Sjedinjenih Američkih Država bio Abraham Lincoln. Budući da to možemo pojmiti kao istinu u nekom svijetu koji je kontrafaktičan našem aktualnom, onda se radi o sekundarnoj pojmljivosti. S pojašnjenjem pojmova primarne i sekundarne pojmljivosti možemo razmotriti još precizniju formulaciju prve premise:

1) F&¬Q je primarno, idealno i pozitivno pojmljiv.

Dakle, prva premisa AP-a zahtijeva da na temelju onoga što nam je *a priori* dostupno iz pozicije idealne refleksije možemo pozitivno pojmiti situaciju u kojoj je aktualni svijet zombi-svijet. U nastavku ćemo razmotriti detaljnije drugu premisu.

U drugoj premisi prelazi se od poimanja zombi-svijeta na zaključak da je takav svijet zaista moguć. Slično karakterizaciji prve premise, druga premisa koristi pojam primarne, idealne i pozitivne pojmljivosti. Ali sada, zbog razlike između primarne (epistemičke) i sekundarne (metafizičke) mogućnosti, legitimno je pitati o kojoj se vrsti mogućnosti radi. Chalmers tvrdi da u ovom slučaju o pojmljivosti treba razmišljati kao primarnoj, idealnoj i pozitivnoj pojmljivosti koja je povezana i na neki način implicira sekundarnu, tj. metafizičku mogućnost. Bez ulaženja u detalje, možemo reći da se takvo shvaćanje mogućnosti čini uvjerljivim ako se fizikalistička teza o supervenijenciji uzme kao da izražava metafizički nužnu tvrdnju. U prethodnom poglavlju vidjeli smo da je tome tako jer se očekuje od relacije supervenijencije da bude dovoljno robusna da može podržavati kontrafaktičku ovisnost, što se može ostvariti ako njihov odnos shvatimo kao metafizički nužan. S obzirom na to, drugu premisu možemo formulirati na sljedeći način:

[65] Kada smo u poglavlju 4 govorili o Kripkeovoj tvrdnji da iskazi identiteta izražavaju nužne istine i da nije moguće da identitet između mentalnog i fizičkog predstavlja samo kontingentnu istinu, onda smo prema Chalmersovoj terminologiji govorili o mogućnosti i nemogućnosti u sekundarnom smislu.

2) Ako je F&¬Q primarno, idealno i pozitivno pojmljivo, onda je F&¬Q metafizički moguće.

Sada se treća premisa može preformulirati na sljedeći način:

3) Ako je F&¬Q metafizički moguće, onda je fizikalizam neistinit.

Dakle, zaključak je:

4) Fizikalizam je neistinit.

Nakon što smo detaljnije obrazložili premise AP-a, u nastavku ćemo se osvrnuti na značajnije prigovore koji se upućuju ovom argumentu.

8.10 Zombiji nisu pojmljivi

Chalmersov argument izazvao je dosta kritika te se razvila široka rasprava oko svake premise njegovog argumenta (Kirk 2019). Mnogi napadaju prvu premisu kojom se tvrdi da možemo pojmiti zombije. Dennett (1995) je, u skladu sa svojom dijagnozom problema s AZ, argumentirao da zombiji ili zombi-svijet, iako *prima facie* djeluje pojmljiv, zapravo nije pojmljiv u smislu koji to zahtijeva AP. Kada nastojimo pojmiti zombije moramo zamisliti duplikate pojedinaca ili čak duplikate cijelih mogućih svjetova koji su identični nekom pojedincu ili svijetu s obzirom na svaki fizički, ponašajni i funkcionalni detalj. Kao i u slučaju s Mary, nije jasno da smo sposobni pojmiti sve što je uključeno u zamišljanje tih fizičkih činjenica i činjenica koje superveniraju nad njima (vidi Dennett 1991, pogl. 10-12; 1995; 2005; Marcus 2004).

Osim Dennetta, i neki apriorni fizikalisti negiraju pojmljivost zombija. Smatraju da ispravno razumijevanje iskustva i fizičkog svijeta pokazuje, barem u slučaju idealne refleksije, da racionalna osoba može kroz *a priori* refleksiju pronaći kontradikciju u pretpostavci da postoji zombi-svijet. Na primjer, Perry (2001) tvrdi da, osim ako AP ne pretpostavlja epifenomenalizam, tj. da fenomenalni karakter iskustva ne može imati uzročne učinke u fizičkom svijetu, pojam zombi-svijeta sadrži kontradikciju. Ako fenomenalni karakteri imaju uzročne moći, onda mogu utjecati na ponašanje pojedinaca koji ih doživljavaju. Perry argumentira da u tom slučaju pak ne možemo pojmiti zombi-svijet. Naime, ako u zombi-svijetu nema fenomenalnih karaktera, onda oni neće moći uzrokovati ponašanja i druge fizičke učinke. Međutim, ako se taj mogući svijet razlikuje od našeg prema fizičkim učincima, onda iz toga slijedi da se ipak ne radi o fizičkom duplikatu našeg svijeta (također vidi Tye 2007). S druge strane, ako AP ne pretpostavlja epifenomenalizam, može se argumentirati da zombiji također

nisu pojmljivi pod pretpostavkom da su fenomenalni karakteri uzrokovani fizičkim događajima (vidi Berčić 2012, 2:181–82). Stoga, ako je svijet fizički duplikat našeg svijeta onda bi isto tako trebao imati fenomenalni karakter, pod pretpostavkom da potpuni fizički uzroci nekog događaja uvijek imaju taj događaj kao učinak (za raspravu oba argumenta, vidi Malatesti 2013). Ako su ovi prigovori dobri onda slijedi da AP ne dokazuje da fenomenalni karakter ne supervenira nad fizičkim te bi se prava rasprava trebala fokusirati na pitanje kako točno fenomenalni karakter i fizička svojstva uzročno interagiraju.

Dosad smo razmatrali prigovor AP-u koji se temelji na negiranju uvjerljivosti prve premise. Međutim, neki fizikalisti dopuštaju da možemo pojmiti zombije čak i u jakom smislu koji pretpostavlja Chalmersov argument te nastoje pokazati što ne valja s drugom premisom njegovog argumenta. Drugim riječima, neki fizikalisti osporavaju tvrdnju da pojmljivost zombija implicira mogućnost njihovog postojanja. U sljedećem pododjeljku ćemo razmotriti tu vrstu prigovora.

8.11 Zombiji nisu mogući

Neki filozofi tvrde da su zombi-svjetovi pojmljivi, ali nisu mogući. Kako bi opravdali tu tvrdnju oslanjaju se na razmatranja o prirodi fenomenalnih pojmova kojima referiramo na fenomenalne karaktere iskustva (Loar 1990; Carruthers 2000). U slučaju AP-a, fenomenalni pojmovi moraju objasniti dvije stvari. Zašto su zombiji pojmljivi i zašto njihova pojmljivost ne implicira da su zombiji mogući. Pristaše ove strategije tvrde da možemo pojmiti zombije jer imamo fenomenalne pojmove, koji su, zato što demonstrativni (ili sposobnosti za prepoznavanje ili sredstva citiranja), različiti od pojmova koji se koriste za davanje potpunog opisa fizičkog svijeta F. Zbog toga možemo zamisliti scenarij gdje je F istinito, a Q neistinito. Također ne možemo pronaći kontradikciju kada zamišljamo zombi-svijet jer kada razmišljamo o fenomenalnim karakterima koji su predstavljeni sa Q koristimo fenomenalne pojmove, dok kada razmišljamo o opisima koji su predstavljeni sa F onda koristimo fizikalne pojmove. Budući da nema *a priori* poveznice između pojmova kojima opisujemo F i Q, onda ne možemo doći do zaključka da tvrdnja F&¬Q uključuje kontradikciju.

Nadalje, apriorni fizikalisti smatraju da ova vrsta pojmljivosti ne implicira da je F&¬Q moguće jer pripada poznatoj skupini slučajeva gdje pojmljivost ne implicira metafizičku mogućnost. Već smo ranije vidjeli da se jedan od osnovnih pojmova u AP odnosi na primarnu mogućnost. Međutim, primarna mogućnost općenito ne implicira sekundarnu, tj. metafizičku mogućnost. Na primjer, vidjeli smo da, iako možemo primarno pojmiti da voda nije H_2O, svejedno nije moguće u sekundarnom smislu da voda nije H_2O. Drugim riječima, iako je iz onoga što *a priori* znamo o pojmu vode (da je prozirna, pitka, bez mirisa, stvar koja se nalazi u rijekama, jezerima itd.) moglo ispasti

da voda ima molekularnu strukturu XYZ, to ne znači da je metafizički moguće da voda nije H_2O. Apriorni fizikalisti smatraju da slično vrijedi za odnos mentalnog i fizičkog. Iako na temelju pojmova možemo u primarnom smislu smatrati mogućim F&¬Q, svejedno iz toga ne slijedi da je ta tvrdnja moguća u sekundarnom, tj. metafizičkom smislu.

Međutim, autori koji podržavaju ovaj odgovor moraju moći odgovoriti na izazov prema kojemu primarna i sekundarna pojmljivost koincidira kada se radi o svjesnim iskustvima. Ako primarna i sekundarna pojmljivost koincidira, onda slijedi da će primarna i sekundarna mogućnost također koincidirati. Drugim riječima, iz epistemičke mogućnosti da pojmimo zombi-svijet slijedit će metafizička mogućnost zombi-svijeta. Vidjeli smo da kod primjera znanstvenih *a posteriori* nužnih identiteta primarna pojmljivost ne implicira sekundarnu pojmljivost. Zašto isto ne vrijedi u slučaju fenomenalnih karaktera iskustva? Chalmers (2010b), slijedeći Kripkea (1997) koji u suvremenoj filozofiji ponovo dovodi u fokus ovu vrstu argumenta, ukazuje na sljedeće; kada govorimo o znanstvenim *a posteriori* nužnim istinama, onda govorimo o stvarima kod kojih postoji razlika između izgleda (kako nam se stvari čine) i stvarnosti (kakve stvari zaista jesu). Međutim, kada govorimo o fenomenalnim karakterima iskustva, onda nema razlike između izgleda i stvarnosti, tj. stvari su onakve kakvima nam se čine. Iz toga slijedi da, ako nam se čini da fenomenalna svojstva nisu identična ili ne superveniraju nad fizičkim svojstvima, onda nije nužno da je to slučaj. Ovu razliku između *a posteriori* nužnih istina i istina o fenomenalnim karakterima iskustva možemo ilustrirati koristeći sljedeći primjer.

Znanost nam otkriva da je toplina identična s prosječnom kinetičkom energijom molekula. Nama se naravno može činiti da taj odnos nije nužan, već da je toplina mogla biti identična s nekim drugim svojstvom. Međutim, u ovom slučaju imamo dostupno objašnjenje zašto nam se to čini. Prema Kripkeu (1997, 111–12), značenje termina „toplina“ fiksiramo putem *osjeta* topline. Dakle, kada razmišljamo o toplini kao objektivnoj veličini, na nju referiramo terminom „toplina“ čiju smo referenciju fiksirali posredstvom našeg osjeta topline. Zbog toga je osjet topline samo kontingentno svojstvo topline shvaćene kao objektivne veličine. Naime, jasno je da toplina ne bi prestala postojati kada mi i naši osjeti ne bismo postojali. To nam također objašnjava zašto nam se čini da toplina može biti nešto drugo osim prosječne kinetičke energije molekula. Naime, ono što mi zapravo poimamo u tom slučaju nije negacija tvrdnje da je toplina prosječna kinetička energija molekula, već poimamo da je *osjet* topline moglo uzrokovati nešto što nije

gibanje molekula. Na primjer, možemo zamisliti da u nama osjet topline uzrokuju fotoni koji čine vidljivu svjetlost.[66]

Međutim, kada govorimo o samim fenomenalnim karakterima iskustva ne možemo napraviti sličnu razliku između načina na koji fiksiramo referenciju izraza i onoga na što oni referiraju u svijetu. Uzmimo kao primjer riječ „bol". Prema Kripkeu (1997, 120) i autorima koji ga slijede, referenciju izraza „bol" fiksiramo pozivanjem na esencijalni karakter boli, tj. način na koji je doživljavamo. Iz toga slijedi da fenomenalni karakter boli kojim fiksiramo značenje izraza „bol" nije samo kontingentno svojstvo boli, kao što je osjet topline samo kontingentno svojstvo gibanja molekula i njihove prosječne kinetičke energije. Stoga, ako nam se čini da bol nije identična ili da ne supervenira nad aktivacijom C-vlakana, onda slijedi da nije *nužno* da su one na taj način povezane. Ili da preformuliramo ovaj način razmišljanja: ako fizikalisti žele tvrditi da iz činjenice što možemo primarno pojmiti da fenomenalni karakteri ne superveniraju nad fizikalnim svojstvima ne slijedi da je metafizički moguće da oni ne superveniraju, onda fizikalisti moraju moći objasniti zašto iz primarne pojmljivost ne slijedi sekundarna pojmljivost. Kao što smo vidjeli u slučaju drugih znanstvenih *a posteriori* nužnih istina imamo za to dostupno objašnjenje. Međutim, u slučaju fenomenalnih karaktera iskustva slično objašnjenje nije nam dostupno jer se čini da ne postoji razlika između toga kako nam se stvari čine i kakve one zaista jesu.

Ovdje možemo primijetiti da ovaj način razmišljanja, koji u suvremenu raspravu ponovo u fokus dovodi Kripke (1997), podrazumijeva da se pristup svjesnim iskustvima temelji na neposrednom *a priori* pristupu. Ideja je da nam se, kada razmišljamo o vlastitom svjesnom iskustvu, njegova esencijalna priroda otkriva u samom doživljaju tog iskustva. Drugim riječima, jedino kada razmišljamo o fenomenalnom karakteru naših iskustava iz perspektive prvog lica direktno zahvaćamo njegovu esencijalnu prirodu. Ovaj način argumentacije temelji se na kartezijanskim intuicijama da apriornim razmišljanjem ili načinom na koji ih shvaćamo ujedno utvrđujemo njihovu ontologiju (vidi Kripke 1997, 119–21).

Fizikalisti su dakako ponudili različite odgovore kojima se nastoji pokazati da su Kripke (1997) i njegovi sljedbenici poput Chalmersa (2010b) u krivu kada ističu razlike između prirode fenomenalnog karaktera iskustva i ostalih znanstvenih *a posteriori* nužnih istina (za kritičku raspravu, vidi Stoljar 2006, pogl. 9). Međutim, rasprava tih odgovora nadilazi ciljeve ovog poglavlja.

[66] Slično objašnjenje možemo dati u slučaju vode. Referenciju izraza „voda" fiksiramo putem načina na koji doživljavamo vodu. Dakle, vodu doživljavamo kao prozirnu, tekuću stvar koja je bez mirisa i okusa. A ti doživljaji su naravno samo kontingentno povezani s fizikalnim svojstvima vode. Ako nam se čini da je molekularna struktura vode mogla biti XYZ, umjesto H_2O, onda je to zato što je XYZ mogao proizvesti doživljaje posredstvom kojih fiksiramo referenciju izraza „voda".

Ovdje smo poglavito htjeli ukazati na izazove i prepreke koje fizikalisti moraju savladati kako bi opravdali svoje gledište. Također smo htjeli pokazati kako nefizikalisti kao osnovnu filozofsku metodologiju koriste apriorne intuicije koje, slično kao i kod Descartesovih argumenata, često podrazumijevaju da se priroda mentalnih stanja otkriva posebnim epistemičkim pristupom koji čini naše znanje mentalnih stanja neposrednim i neopovrgljivim te time i kategorički različitim od spoznaje drugih nementalnih svojstva svijeta. U sljedećem poglavlju ćemo vidjeti do kakvih radikalnih posljedica može dovesti takvo gledište te kako fizikalisti mogu odgovoriti na njih.

8.12 Zaključna razmatranja

U ovom poglavlju kritički smo razmotrili argument iz znanja i argument iz pojmljivosti kojima se nastoji pokazati neodrživost redukcionističkog ili supervenijentnog fizikalizma. Vidjeli smo da fizikalisti imaju različite mogućnosti odgovaranja na ove argumente. Ono što treba istaknuti jest da većina razmotrenih odgovara podrazumijeva da ovi argumenti ukazuju na neodrživost ontološke teze da se u podlozi svjesnih iskustava nalaze fizikalna svojstva. Međutim, ovdje treba istaknuti jedan drugi izazov na koji nas navode argumenti iz znanja i pojmljivosti. To nije toliko upućivanje na ontološku konkluziju, već na suočavanje s onim što se naziva eksplanatorni jaz. Čini se da trenutno, ili zbog naše ograničene kognitivne konstitucije, u principu (McGinn 1978) nismo sposobni shvatiti kako fizikalizam u pogledu fenomenalnih karaktera iskustva može biti istinit. Stoga mnogi autori smatraju da čak i ako argumenti poput argumenta pojmljivosti ne opravdavaju ontološku konkluziju da fizikalizam nije istinit, on u najmanju ruku predstavlja značajan epistemološki prigovor na koji svaka fizikalistička pozicija mora ponuditi nekakav odgovor. U sljedećem poglavlju ćemo se baviti upravo tom vrstom epistemoloških prigovora fizikalističkim gledištima.

9 Eksplanatorni jaz i težak problem svijesti

9.1 Uvod

U prošlom poglavlju vidjeli smo da postoji više načina na koje fizikalisti mogu odgovoriti na argument iz znanja i argument pojmljivosti. Tim odgovorima se obično nastoji pokazati da iz pojmljivosti ili mogućeg zamišljanja neke situacije ne slijedi zaključak da je fenomenalni karakter iskustva nefizičko svojstvo ili svojstvo koje ne supervenira nad fizičkim svojstvima (Levine 2004).

Međutim, čini se da argument iz znanja, argument pojmljivosti i slični argumenti u najmanju ruku pokazuju da se fizikalisti suočavaju s pojmovnim jazom kada nastoje objasniti odnos fenomenalnih karaktera iskustava i fizičkih svojstava. Na primjer, Nagel u svom poznatom radu „Kako je to biti šišmiš“ (1974) nije tvrdio da je fizikalizam neistinit. U njemu je Nagel ukazao na to da iz perspektive prvog lica, razmišljajući o fenomenalnom karakteru iskustva, mi zapravo ne možemo shvatiti kako fizikalizam u pogledu prirode svjesnog iskustva može biti istinit. Drugim riječima, zaključak njegove rasprave je da trenutno ne možemo shvatiti kako je moguće da svjesna iskustva imaju objektivnu prirodu koja se može istraživati dostupnim znanstvenim metodama. Sličan argument razvija Joseph Levine (1983) te tvrdi da je prava poruka argumenta pojmljivosti da fizikalizam ne može objasniti kako to da fizička materija poput mozga može imati svjesna iskustva. Na temelju sličnih je razmatranja Chalmers (2010a) formulirao takozvani teški problem svijesti koji je po njemu poguban za fizikalizam te daje osnove za rehabilitaciju dualističkih gledišta.

U ovom poglavlju prvo ćemo razmotriti prigovor eksplanatornog jaza kako ga je formulirao Levine (1983). Nakon toga ćemo se nadovezati na njegovu općenitiju formulaciju u terminima teškog problema svijesti (Chalmers 2010a). U ovom dijelu vidjet ćemo na koji način Chalmers razlikuje lake od teških problema svijesti i zašto potonji prema njemu predstavljaju nepremostivi problem za fizikalizam. Chalmers smatra da postojanje teškog problema svijesti opravdava povratak dualističkim gledištima. Međutim, način na koji Chalmers formulira svoju varijantu dualizma izrazito podsjeća na panpsihizam u pogledu mentalnih svojstava. I doista, u suvremeno doba

veći broj autora smatra da teški problem svijesti ukazuje na plauzibilnost panpsihizma. Stoga ćemo se u nastavku baviti suvremenim varijantama panpsihizma te ćemo razmotriti nekoliko argumenata kojima ga se nastoji opravdati.

Poglavlje ćemo zaključiti tako da se malo odmaknemo od trenutne rasprave i preispitamo mogućnost da sve strane u njoj samouvjereno pretpostavljaju da postoji problem svijesti te da razumijemo u čemu se taj problem sastoji. U tom pogledu razmotrit ćemo Denettove (1988) argumente kojima nastoji pokazati da zapravo ne postoji problem objašnjenja svijesti za fizikaliste jer kada shvatimo kako se pojam „svijest“ shvaća u ovim raspravama uvidjet ćemo da u tom smislu svijest ni ne postoji.

Argument eksplanatornog jaza

Kao što smo spomenuli ranije, Levine (1983) smatra da iako različite varijante argumenta pojmljivosti ne pokazuju da fizikalizam mora biti neistinit, ipak nam pokazuju da je fizikalizam nepotpun. Nepotpun je u eksplanatornom smislu jer nam ne može objasniti kako to da svjesna iskustva nastaju kao posljedica određene organizacije fizičke materije. Stoga smatra da se fizikalistička gledišta suočavaju s eksplanatornim ili objašnjavalačkim jazom. U nastavku ćemo pojasniti što se pod time misli.

U prijašnjim poglavljima vidjeli smo da se fizikalisti često pozivaju na uspješna objašnjenja u znanostima kako bi motivirali tvrdnju da će se i svojstva uma i njihova povezanost s fizičkim jednom moći objasniti na sličan način. Tu se najviše ističu teoretičari identiteta tipova koji su tvrdili da je najbolji istraživački program za otkrivanje prirode uma onaj koji pojedine vrste mentalnih stanja poistovjećuje s nekim stanjima u mozgu (vidi poglavlje 4). Tu se onda pozivaju na poznate znanstvene identifikacije koje su omogućile daljnji razvoj znanstvenih teorija i objašnjenja. Zadržimo se na primjeru identifikacije vode sa H_2O. Kada se otkrilo da voda ima molekularnu strukturu H_2O to je omogućilo različita objašnjenja svojstava vode. Na primjer, ta identifikacija nam objašnjava zašto se voda u procesu elektrolize pretvara u plinove vodika i kisika, zašto voda ključa na 100°C, zašto reflektira svjetlo određene boje, zašto može mijenjati agregatna stanja, provodi struju i tako dalje. Dakle, čini se da jednom kada znamo kemijsku strukture vode shvaćamo kako izgleda, ili u principu može izgledati, potpuno objašnjenje prirode vode te se čini da nema nekog dodatnog aspekta vode koji bi ostao nerazjašnjen ili neshvatljiv.

No, čini se da je situacija drukčija s poistovjećivanjem mentalnih stanja s fizičkim stanjima. Uzmimo opet klasičan primjer da je bol identična ili supervenira nad aktivacijom C-vlakana. Čini se da bismo tim poistovjećivanjem mogli objasniti mehanizme koji se nalaze u pozadini funkcionalnih svojstava boli. Na primjer, bol je ono što se pojavljuje kada imamo oštećenje tkiva te uzrokuje da se odmaknemo od izvora boli. Poznavanje biokemijskih svojstava C-vlakana bi zasigurno moglo objasniti

kako se te funkcije izvode. Međutim, Levine (1983, 357) ističe da bi ovakva vrsta objašnjenja ostavila značajnu „rupu“ u našem shvaćanju prirode boli; naime, ostaje neobjašnjeno zašto bol osjećamo i doživljavamo na određeni način. Drugim riječima, zašto bol ima određeni fenomenalni karakter. Budući da fenomenalni karakter boli nećemo moći objasniti otkrićem fizičke podloge boli, izgleda da ovakva vrsta poistovjećivanja ostavlja eksplanatorni jaz koji, čini se, ne postoji u drugim slučajevima znanstvenih poistovjećivanja i objašnjenja.

Kao što smo ranije spomenuli, u podlozi ovog argumenta nalaze se razmatranja koja motiviraju argument pojmljivosti. Naime, čini nam se da otkrivanje fizičke podloge boli ili drugih svjesnih mentalnih stanja neće biti dovoljno za objašnjenje prirode svjesnih mentalnih stanja jer se čini da uvijek možemo zamisliti da postoji to fizičko stanje, a da nije popraćeno nekim svjesnim mentalnim stanjem. Stoga, eksplanatornim jazom se postavlja pitanje zašto, na primjer, aktivaciju C-vlakana doživljavamo kao osjećaj boli, a ne primjerice kao osjećaj škakljanja ili nešto drugo? Ili zašto gledanje crvene rajčice u nama stvara doživljaj crvene boje, a ne plave? Ili zašto je oslobađanje dopamina popraćeno osjećajem ushićenja, a ne osjećajem strepnje? Još općenitije, možemo se pitati zašto uopće aktivaciju C-vlakana ili nekog drugog dijela mozga doživljavamo na određeni način, tj. zašto je popraćena određenim fenomenalnim karakterom, a ne unutrašnjom tamom kakvu zamišljamo kod filozofskih zombija?

Prigovorom eksplanatornog jaza dovode se u pitanje i slabije funkcionalističke varijante fizikalizma. Slično kao i ranije, možemo se pitati zašto je funkcionalna uloga koju C-vlakna igraju popraćena određenim fenomenalnim karakterom koji povezujemo s doživljajem boli? Ili ako pretpostavimo da je funkcija želja da nas navedu na određene radnje, zašto su želje obično povezane s određenim doživljajem privlačnosti prema predmetu želje? I ovdje se općenito možemo pitati zašto su uzročne uloge koja mentalna stanja igraju često povezane s određenim fenomenalnim karakterima, a ne fenomenalnom tamom? Čini se da bi se život u funkcionalno-fizičkom pogledu mogao normalno odvijati, čak i kada naša iskustva ne bi imala fenomenalne karaktere. Stoga Levine tvrdi da će, koliko god toga saznali o fizičkim mehanizmima koji se nalaze u podlogama svjesnih iskustava i njihovim funkcionalnim ulogama, ovaj eksplanatorni jaz ostati otvoren.

Prigovorom eksplanatornog jaza nastoji se pokazati da se fenomenalna svojstva iskustva ne mogu objasniti u sklopu objašnjenja fizičkih procesa koji se nalaze u podlozi naših mentalnih života. Iz toga bi trebalo slijediti da je fizikalizam metodološki nepotpuna teorija te da nam ne može omogućiti stvarno razumijevanje prirode svjesnih mentalnih stanja. To nas dovodi do zaključka da, čak i ako prihvatimo fizikalizam, svejedno ostaje svojevrsni misterij kako to da su fenomenalni karakteri mentalnih stanja identični s

nekim fizičkim ili funkcionalnim svojstvima ili pak zašto superveniraju nad njima.

Problem eksplanatornog jaza upućuje na potrebu da se ponudi neki odgovor na pitanje kako to da su određena fizička stanja popraćena određenim svjesnim mentalnim stanjima. Možemo primijetiti da bi pobornici strategije fenomenalnih pojmova objasnili postojanje tog jaza time što pri opisivanju fenomenalnih iskustava koristimo pojmove koji su kognitivno odijeljeni od pojmova kojima opisujemo fizičke procese u svijetu. Međutim, drugi autori smatraju da eksplanatorni jaz ukazuje na dublje probleme koji se ne mogu riješiti samo razlikovanjem pojmova koje koristimo pri opisivanju različitih događaja. Štoviše, neki smatraju da ovaj problem ukazuje na to da je potrebno prihvatiti radikalnija rješenja koja uključuju mijenjanje koncepcije toga čemu sve možemo pripisati svjesna iskustva i kako pojam fenomenalnog karaktera treba shvatiti. No, prije nego pređemo na te odgovore, u sljedećem odjeljku ćemo razmotriti suvremeniju verziju eksplanatornog jaza kako ga je formulirao Chalmers. To će nam biti važno jer je njegova formulacija u značajnoj mjeri odredila rasprave o odnosu duha i tijela u suvremenoj filozofiji uma i znanostima koje se bave istraživanjem svijesti.

9.2 Teški problem svijesti

Chalmers (2010a, pogl. 1) poslužio se sličnim razmatranjima koje nudi i Levine kako bi istaknuo probleme za fizikalizam koje skupno naziva teški problem svijesti. Chalmers razlikuje lake od teških problema svijesti. Laki problemi svijesti odnose se na one probleme za koje možemo očekivati da će se s vremenom riješiti korištenjem standardnih metoda istraživanja iz neurokognitivnih znanosti. Teški problemi su oni za koje imamo *a priori* razloge očekivati da se neće moći riješiti takvim metodama. Kada govori o „lakim“ problemima, Chalmers ne misli da ih je jednostavno riješiti. Štoviše, možda se laki problemi svijesti neće uspjeti riješiti tijekom naših života. Međutim, oni su laki jer barem znamo formu odgovora koju moraju zadovoljiti i načine na koje se mogu riješiti. Fenomeni koje Chalmers (2010a, 4) ubraja u skupinu lakih problema svijesti su:

- Sposobnost diskriminacije, kategorizacije i reakcije na okolišne podražaje
- Integracija informacija u kognitivnom sustavu
- Mogućnost izvještavanja o vlastitim mentalnim stanjima
- Sposobnost pristupa vlastitim mentalnim stanjima
- Fokusiranje pažnje
- Voljna kontrola ponašanja
- Razlika između budnosti i spavanja

Prema Chalmersu, ovi fenomeni podrazumijevaju posjedovanje svjesnih mentalnih stanja u nekom smislu riječi „svijest“. Međutim, smatra da svijest koja je potrebna za te sposobnosti nije ona vrsta svijesti koja ima fenomenalni karakter. Na primjer, Block (1995) razlikuje barem tri pojma svijesti: svijest višeg reda, pristupnu svijest i fenomenalnu svijest. Svijest višeg reda odnosi se na sposobnost nadziranja vlastitih mentalnih stanja. Pristupna svijest odnosi se na sposobnost diskriminacije, kategorizacije te upotrebe različitih vrsta informacija kako bi se modificirala pažnja, izvodili zaključci i usmjeravale radnje. Fenomenalna se svijest odnosi na iskustva i doživljaje koji imaju fenomenalni karakter.

Sposobnosti, poput ranije navedenih, koje podrazumijevaju da ljudi posjeduju pristupnu i svijest višeg reda upravo su one koje se mogu objasniti u funkcionalnim terminima. Pod time Chalmers misli na sposobnosti čije se funkcije mogu odrediti njihovim uzročnim ulogama; jednom kada ih odredimo možemo tražiti fizičke mehanizme koji implementiraju ili omogućuju te uzročne uloge. Na primjer, sposobnost diskriminacije, kategorizacije i reakcije na okolišne podražaje najčešće podrazumijeva da su nam u svijesti dostupne određene informacije koje nam omogućuju te kognitivne radnje. Međutim, jednom kada znamo uzročne uloge koje definiraju tu sposobnost, onda možemo tražiti mehanizme u mozgu koji povezuju percepciju podražaja s određenom ponašajnom reakcijom. Taj problem je „lako“ rješiv u smislu da iako možda nećemo znati sve detalje kako taj mehanizam u stvarnosti izgleda ipak znamo oblik koji će rješenje tog problema poprimiti. Slično vrijedi i za sve ostale ranije navedene sposobnosti. Na primjer, mogućnost verbalnog izvještavanja o svjesnim mentalnim stanjima često povezujemo sa sposobnošću imanja svijesti višeg reda, gdje su sadržaji naših mentalnih stanja druga mentalna stanja. Međutim, samu sposobnost da govorimo o mentalnim stanjima možemo objasniti povezivanjem s mehanicističkim objašnjenjima koja su već dostupna u neurokognitivnim znanostima (vidi, npr. Craver 2007).

Težak problem svijesti jest težak upravo zato što uopće nije jasno kakav oblik će njegovo rješenje poprimiti. Naime, težak problem svijesti podrazumijeva fenomenalni pojam svijesti koji se odnosi na objašnjenje prirode iskustva koja imaju fenomenalni karakter. Za tu vrstu svjesnih iskustava, ističe Chalmers, nije jasno kako ih možemo objasniti kroz funkcionalno-mehanicistička objašnjenja. Problem je upravo u tome što se čini da postoji eksplanatorni jaz između poznavanja činjenica o fizičkim stanjima mozga i kako na temelju njih nastaju mentalna stanja s fenomenalnim karakterima. Prema Chalmersu (2010a), teški problem svijesti odnosi se na pitanja kako i zašto su neka fizička stanja, procesi i događaji popraćeni iskustvima koja imaju određeni fenomenalni karakter. Ili još preciznije, zašto su određena stanja mozga uopće popraćena imanjem mentalnog stanja s određenim fenomenalnim karakterom i zašto baš tim, a

ne nekim drugim. Na primjer, možemo se pitati zašto je aktivacija C-vlakana *uopće* popraćena iskustvima koja imaju određeni fenomenalni karakter. Međutim, tu nije kraj jer čak i da znamo odgovor, ostaje daljnje pitanje zašto je aktivacija C-vlakana popraćena baš tim specifičnim osjećajem boli, a ne doživljajem s nekim drugim fenomenalnim karakterom. Prema Chalmersu, za razliku od lakih problema, za ove probleme nije jasno kako bismo ih mogli započeti objašnjavati pozivajući se na mehanizme i funkcije koje naši mozgovi izvode. Naime, uvijek se možemo pitati zašto izvođenje te funkcije ili tog mehanizma doživljavamo na određeni način. Čini se da će, koje god mehanicističko ili funkcionalno objašnjenje smislimo, ostati eksplanatorni jaz o tome zašto njihovim izvođenjem doživljavamo stvari na određeni način, tj. zašto postoji nešto kako je to doživjeti njihovu aktivaciju.

Budući da nam fizikalističko gledište naizgled ne pruža odgovor na ova pitanja, Chalmers (2010a) zaključuje da moramo napustiti fizikalizam kao znanstveno-metodološki okvir unutar kojeg ćemo objašnjavati odnos između mentalnog i fizičkog te da trebamo prihvatiti određenu varijantu dualizma prema kojoj fenomenalni karakteri postoje kao zasebna nefizička svojstva nekih fizičkih stvari.

Štoviše, Chalmers smatra da se svijest ne može objasniti u terminima nekih drugih jednostavnijih entiteta ili svojstava jer je svijest fundamentalni entitet koji karakterizira naš svijet. U tom pogledu, Chalmers (2010a, 16–17) uspoređuje svijest s nekim drugim fundamentalnim entitetima iz fizike. Kao primjer navodi slučaj elektromagnetskih sila. Do sredine 19. stoljeća, elektromagnetske procese nastojalo se objasniti pozivanjem na mehaničke procese i sile koje su dobro opisane teorijama Newtonove fizike. James Clerk Maxwell zaključio je da je to nemoguće te je pretpostavio postojanje elektromagnetske sile kao još jedne vrste fundamentalne sile koja se ne može dalje objašnjavati pozivanjem na neke druge entitete. Maxwellova je pretpostavka konačno potkrijepljena krajem 1880-tih kada je Heinrich Rudolph Hertz eksperimentalno dokazao postojanje elektromagnetskih valova (Poljak, Sokolić, i Jakić 2011, 22–23). Danas se smatra da postoje četiri fundamentalne sile pomoću kojih se mogu objasniti sve ostale sile, a to su jaka nuklearna, elektromagnetska, slaba nuklearna i gravitacijska sila. Chalmers smatra da nam pojmovna nemogućnost objašnjavanja načina na koji svijest nastaje iz fizičkih procesa sugerira da je možda i svijest još jedan takav fundamentalni entitet koji ne možemo objasniti pozivanjem na jednostavnije entitete. Stoga, Chalmers (2010a, 18) svoje gledište naziva naturalističkim dualizmom. Pod time misli na to da osim mase, električnog naboja i prostor-vremena, svijest predstavlja dodatni fundamentalni entitet. Slično kao što fundamentalne fizikalne veličine opisujemo fundamentalnim zakonima i procesima, tako bi se i ispravna teorija svijesti trebala izgraditi otkrivanjem i formulacijom fundamentalnih psihofizičkih zakona koji povezuju fizičke i mentalne događaje.

Možemo se zapitati kako izgledaju ti fundamentalni zakoni koji bi trebali opisivati svjesne događaje i povezati ih s fizičkim događajima. Ovdje Chalmers (vidi 2010a, 25–27) ne daje jasno definiranu teoriju, međutim sluti da bi se fundamentalni principi koji karakteriziraju svijest mogli formulirati uz pomoć pojma informacije na sljedeći način. Obično se u kognitivnim znanostima ljudske psihološke sposobnosti shvaća i objašnjava kao jednu vrstu sposobnosti za procesiranje informacija. U tom smislu, ljudi se, kao fizička bića, mogu promatrati kao entiteti koji utjelovljuju informacije i načine na koje se one povezuju. Međutim, Chalmers ovdje dodaje da informacija može biti fizički utjelovljena, ali isto tako može imati fenomenalni aspekt koji se očituje u fenomenalnim iskustvima i relacijama u kojima ona stoje. Na primjer, ljudsko fenomenalno iskustvo boje pojavljuje se u određenim relacijama. Pa tako crvenu boju vidimo kao sličniju ružičastoj nego zelenoj, ili plavu vidimo kao sličniju zelenoj nego narančastoj. Isti taj proces u terminima prijenosa informacija moći će se opisati u terminima fizičkih veličina koje možemo odrediti eksperimentalnim promatranjem vizualnog procesiranja boja iz perspektive trećeg lica. No, umjesto jednog fenomenalnog svojstva koje bi bilo identično nekom fizičkom svojstvu, kako su to smatrali teoretičari identiteta tipova, ovdje prema Chalmersu imamo niz fizičkih događaja koji se odnosi na procesiranje informacija, ali ta informacija ima dva aspekta, fizički i fenomenalni, čija je veza određena fundamentalnim zakonima koji se ne mogu dalje objašnjavati ili reducirati na neke druge elemente.

Chalmersov naturalistički dualizam ostavlja dosta otvorenih pitanja. Jedno se pitanje odnosi na mogućnost objašnjenja uzročnih moći koje pretpostavljamo da iskustva imaju. Čini se da se Chalmers slaže da je fizički svijet, shvaćen kao skup fundamentalnih entiteta koji ne obuhvaćaju svijest, uzročno zatvoren (vidi Chalmers 2010a, 17). Dakle, ako nam je dovoljno da fizičke događaje objasnimo pozivanjem samo na fizičke uzroke ili, u ovom slučaju, pozivanjem na fizikalne aspekte informacija koje ljudi procesiraju, čini se da onda ne ostaje eksplanatorni posao koji bi fenomenalni karakteri mogli igrati u psihološkim objašnjenjima. S jedne strane, to nas upućuje na epifenomenalizam za koji mnogi smatraju da je intuitivno neprihvatljiv. S druge strane, nekima to može predstavljati razlog za odbacivanje postojanja svijesti shvaćene kao nefizikalni aspekt informacija.

Drugo važno pitanje je što znači da informacija ima fenomenalni aspekt? Naime, ako bilo koja informacija ima fenomenalni aspekt znači li to da postoji nešto kako je to biti ta informacija? Chalmers ne daje jasno objašnjenje što bi to značilo da informacija može imati fenomenalni aspekt.

Treće važno pitanje odnosi se na određivanje vrste informacija koje posjeduju fenomenalni aspekt (Chalmers 2010a, 27). Dakle, čak i ako dopustimo da informacija ima svjesni aspekt, postavlja se pitanje imaju li sve informacije fenomenalni aspekt ili samo neke? I kako bismo mogli razlikovati

one informacije koje ga imaju od onih koje ga nemaju? Chalmers ova pitanja ostavlja otvorenima. Zapravo nije ni jasno kako bi mogao napraviti principijelnu razliku između informacija koje imaju i onih koje nemaju svjesna iskustva. Vjerojatno zato Chalmers tvrdi da čak i ako se prihvati da sve informacije posjeduju fenomenalni aspekt to nije toliko čudno koliko se na prvu čini. Kako bi ublažio neintuitivnost ideje da bilo koja informacija ima fenomenalni aspekt, Chalmers se poziva na slučajeve *procesiranja* informacija. Pa navodi da ima smisla smatrati da će, ako postoji jednostavno procesiranje informacija, ono biti popraćeno jednostavnijim iskustvom, dok će kompleksnije procesiranje informacija biti popraćeno kompleksnijim iskustvom. Kao primjer navodi da „miš ima jednostavniju strukturu za procesiranje informacija od čovjeka te u skladu s tim ima jednostavnije iskustvo (…)“ (Chalmers 2010a, 27).

Međutim, sada se postavlja pitanje što je s još jednostavnijim procesorima informacija koji ne moraju nužno biti živi; imaju li i oni fenomenalna iskustava? Ima li svjesna iskustva, na primjer, termostat koji bilježi temperaturu u prostoriji te uključuje ili isključuje grijanje? Chalmers prihvaća da, ako nema ograničenja za to koja vrsta informacija ima fenomenalni aspekt, onda slijedi da i termostati imaju svjesna iskustva (vidi Chalmers 2010a, 27). Štoviše, iz ovog gledišta slijedi da u bilo kojem slučaju gdje događaje možemo opisati u terminima informacija postoji svjesno iskustvo. Na primjer, u slučaju kvantnog sprezanja (engl. *quantum entanglement*) spin jedne subatomske čestice ovisi o spinu druge subatomske čestice. Ovaj fenomen pokazuje da, kada se utvrdi spin jedne čestice koja se nalazi u odnosu sprege, istodobno se dobiva informacija o spinu druge čestice. Dakle, ako se ovdje može govoriti o prijenosu informacija, onda bi slijedilo da subatomske čestice i veze među njima posjeduju svjesne aspekte.

Sada se možemo zapitati kakva je ovo varijanta dualizma? Krenuli smo od intuicije da postoji pojmovni jaz između znanja činjenica o strukturnim i funkcionalnim svojstvima fizičke materije i znanja o našim svjesnim iskustvima. To nas je uputilo prema zaključku da možda svjesna iskustva ni nisu svodiva na druga fizička svojstva predmeta nego da su ona, uz druge temeljne fizičke entitete i zakone, dodatni fundamentalni aspekti koji karakteriziraju informacijska stanja našeg svijeta. Na temelju tog zaključivanja dospjeli smo do gledišta da gdje god postoji prijenos informacija postoje i svjesna iskustva. Pa tako u principu ništa ne isključuje gledište da i subatomske čestice posjeduju svjesna iskustva. Drugim riječima, ne isključuje se da postoji nešto kako je to biti kvark ili foton (za raspravu, vidi Chalmers 2010a, 133–37).

Kad se stvari ovako ekspliciraju, Chalmersov naturalistički dualizam prestaje sličiti na varijantu klasične dualističke pozicije te počinje podsjećati na određenu varijantu panpsihizma, gledište prema kojemu sve što postoji

posjeduje mentalna svojstva (vidi Chalmers 2016a). Štoviše, navedene posljedice ovog gledišta čine se toliko neintuitivnima da bi ga se moglo smatrati redukcijom originalnog gledišta na apsurd (Berčić 2012, 2:209–11).

Međutim, iako bi se takva reakcija u ranija vremena smatrala opravdanom (Nagasawa 2021), u suvremenoj filozofiji uma panpsihizam u različitim varijantama postaje jedno od standardnih potencijalnih rješenja za problem uma i tijela. Dapače, neki autori smatraju da problemi s fizikalizmom koje smo ranije razmotrili pokazuju da jedino prihvaćanjem neke varijante panpsihizma možemo riješiti teški problem svijesti. Stoga ćemo u sljedećem odjeljku razmotriti kako se u novijim raspravama panpsihizam definira i kojim argumentima se brani.

9.3 Panpsihizam

Panpsihizam dolazi od riječi grčke riječi *pan* koja znači sve i *psiha*, koja znači um ili duša. U prijevodu, dakle, panpsihizam znači da sve ima um ili dušu. Ovo naizgled neintuitivno gledište dobiva na popularnosti među filozofima u novije doba te je sve prisutnije u generalnim raspravama o prirodi svjesnih mentalnih stanja (vidi, npr. Strawson 2008; Chalmers 2016a; Seager 2016; Goff 2017; 2019).[67] Suvremeni panpsihisti smatraju da su fizikalističke i dualističke teorije neuspješne u odgovaranju na pitanje kako to da postoji svijest i interakcija između mentalnih i fizičkih stanja te predlažu da moramo uzeti za ozbiljno alternativu da zapravo sve na neki način ima svijest. Panpsihisti imaju dugačku tradiciju u filozofiji. Neki među njih ubrajaju poznate filozofe poput Platona, Barucha de Spinoze i Gotfrieda Leibniza (za pregled, vidi Goff, Seager, i Allen-Hermanson 2020). Mi se nećemo baviti ovim povijesnim ličnostima, već ćemo se usredotočiti na suvremene varijante panpsihizma.

U novije vrijeme, Thomas Nagel (1979) daje jedan od ranijih impetusa panpsihizmu. On ga definira kao gledište prema kojemu temeljne tvari koje konstituiraju svemir imaju svjesna mentalna svojstva, bez obzira na to jesu li ili nisu dijelovi živih organizama. Kada se govori o svjesnim mentalnim stanjima misli se na stanja s fenomenalnim karakterima. Dakle, prema panpsihistima postoji nešto kako je to biti kvark ili elektron ili što god smatramo da čini osnovne sastojke našeg svemira. U tom pogledu, važno je istaknuti nekoliko stvari (vidi Goff 2019, pogl. 4). Prvo, suvremeni panpsihisti predlažu da se svjesna iskustva ljudi i životinja trebaju objasniti u terminima jednostavnijih tipova svijesti. No, nije tako da sve stvari imaju svijest iste kompleksnosti. Na primjer, ljudi će imati vrlo kompleksna svjesna iskustva,

[67] Neki smatraju da panpsihizam postaje popularan i u znanstvenim objašnjenjima svjesnog iskustva. Pa se tako *Teorija integrirane informacije* koju razvija Gulio Tononi s kolegama (vidi, npr. Tononi i ostali 2016) često spominje u kontekstu teorija koje su u najmanju ruku kompatibilne s panpsihizmom.

životinje će imati manje kompleksna svjesna iskustva, dok će kvarkovi i elektroni imati vrlo bazična svjesna iskustva. Dakle, postoji gradacija svjesnosti, no sve stvari će u nekom smislu posjedovati iskustva s fenomenalnim karakterima. Drugo što je važno za opis suvremenog panpsihizma jest da, iako bazične fizičke čestice posjeduju svjesna iskustva, neće sve kombinacije tih čestica nužno imati svjesna iskustva. Panpsihisti ostavljaju otvorenim mogućnost da recimo stvari poput mozga definitivno imaju svijest, no da kamenje, tj. kombinacije kvarkova i drugih mikrofizičkih čestica koje čine kamenje možda nemaju svijest.

Zaključivanje koje vodi suvremene panpsihiste prema ovom gledištu može se opisati na sljedeći način (Goff, Seager, i Allen-Hermanson 2020). Osnovni problem od kojeg se kreće jest problem zašto postoji svijest, tj. svjesna iskustva? Vidjeli smo da težak problem svijesti ukazuje na to da zapravo nemamo objašnjenje zašto mozak producira iskustva s fenomenalnim karakterima. Znamo svašta o mozgu, njegova funkcionalna svojstva, kako obrađuje informacije, kako reagira na fizičke podražaje, koliko mu treba da donese odluku i tome slično. Međutim, nije jasno kako se ta svojstva povezuju sa svijesti jer uvijek možemo zamisliti ili pojmiti da postoji fizički svijet bez svjesnih iskustava.

Fizikalisti smatraju da se ovaj problem može riješiti. U prošlom poglavlju smo vidjeli da prema pristašama strategije fenomenalnih pojmova problem objašnjenja svijesti nastaje zato što kada razmišljamo o fizičkim stvarima onda koristimo jedan skup pojmova, dok kada razmišljamo o svjesnim mentalnim stanjima onda koristimo drugi skup pojmova. Budući da nam fenomenalni pojmovi omogućuju da razmišljamo o iskustvima direktno i neovisno o tome kako razmišljamo o fizičkim stvarima, onda nam se čini da te dvije vrste pojmova ne referiraju na istu stvar. U tom pogledu, teški problem svijesti objašnjavaju kao iluziju koja je posljedica različitih pojmovnih shema koje koristimo kada razmišljamo o fizičkom svijetu. Možemo napraviti analogiju. Ista osoba može govoriti o Zornjači i Večernjači kao da se radi o dvije različite zvijezde, a da ne zna da se u stvarnosti radi o istoj stvari, naime o planeti Veneri. Štoviše, čak i nakon što sazna da pojmovi Zornjača i Večernjača referiraju na istu stvar, njoj se može činiti da one nisu ista stvar.

Drugi fizikalisti ne misle da se teški problem svijesti temelji na zabuni ili iluziji, nego smatraju da ćemo s vremenom uspjeti objasniti kako fizička stvar poput mozga može proizvesti svijest (vidi, npr. P. M. Churchland 1988).

Međutim, panpsihisti ukazuju na to da se obje varijante fizikalizma temelje na metodološkoj pretpostavci da se svijest treba objasniti u terminima procesa koji ne uključuju svijest. Zbog argumenata koje smo razmotrili u poglavlju 8, vidjeli smo zašto mnogi smatraju da to nije moguće. Stoga, panpsihisti predlažu alternativni pristup; smatraju da se ljudska i životinjska svijest može objasniti u terminima temeljnijih formi svijesti koje i

nežive tvari posjeduju. Ne samo to, već smatraju da uvjerljive varijante fizikalizma neće biti održive ako nismo spremni prihvatiti da temeljna materija može posjedovati svjesna iskustva. U nastavku ćemo razmotriti dva argumenta kojima se nastoji podržati ta tvrdnja.

Jedan argument u prilog tom gledištu je dao Nagel (1979). Zanimljivost tog argumenta je što kreće od premisa koje su *prima facie* uvjerljive iz perspektive nereduktivnog fizikalizma. Premise argumenta su sljedeće:

(1) Materijalna kompozicija: Nagel (1979, 181) smatra da je uvjerljiva tvrdnja prema kojoj su organizmi kompleksni materijalni sustavi. Prema ovoj tezi, svaki organizam, uključujući i ljude, sastavljen je od materijalnih komponenti koje po sebi nisu ništa posebno. Pod time Nagel misli na to da, kada bismo razlamali stvari u najsitnije moguće dijelove, vidjeli bismo da smo svi sastavljeni od istih mikrofizičkih čestica. Dakle, u principu ljudi i stolovi sastavljeni su od istih mikrofizičkih komponenti. Ono što nas razlikuje je kako su te čestice posložene i međusobno povezane. No sama priroda tih čestica se ne razlikuje od stvari do stvari.

(2) Antiredukcionizam: ovom premisom Nagel (1979, 181) tvrdi da postoje neka mentalna stanja, poput onih koje posjeduju fenomenalni karakter, koja ne spadaju među fizička svojstva organizma. Jasno je zašto Nagel to smatra. Anticipirajući argumente iz znanja i eksplanatornog jaza, Nagel (1974) tvrdi da iz znanja o objektivnim činjenicama o ljudima nikada nećemo moći derivirati znanje o subjektivnim karakterima iskustva. Na temelju tih razmatranja Nagel (1979, 188–89) opravdava ontološku konkluziju da svojstva svjesnih mentalnih stanja ne mogu biti fizička svojstva predmeta.

(3) Realizam u pogledu mentalnih svojstava: ovom premisom Nagel (1979, 181) ističe da je svjesnost svojstvo organizama. Nagel smatra da ne postoje bestjelesne duše, stoga mu je jedina uvjerljiva alternativa da mentalna svojstva karakteriziraju organizme kao jednu vrstu materijalnih predmeta.

(4) Ne postoje emergentna svojstva: prema Nagelu (1979, 182) ne postoje istinski emergentna svojstva, već se u ontološkom smislu sva svojstva kompleksnog sustava temelje na svojstvima komponenti tog sustava. Obično se smatra da „emergencija“ označuje svojstvo kompleksnih sustava koje ne posjeduju komponente koje sačinjavaju taj sustav. Na primjer, voda ima svojstvo likvidnosti, međutim, molekule od kojih je voda sastavljena nemaju to svojstvo. Dakle, mogli bismo reći da je likvidnost emergentno svojstvo vode. Međutim, prema Nagelu emergencija je epistemološki pojam, u smislu da odražava samo naše neznanje u pogledu toga kako od svojstava komponenti nekog sustava nastaju svojstva sustava kojeg su dio. Nagel smatra da, kada govorimo o emergentnim svojstvima, onda zapravo govorimo o svojstvima kompleksnog sustava koja još nismo u stanju objasniti na temelju poznavanja njegovih komponenti; što znači da dijelovi tog

sustava imaju komponente koje nam još nisu poznate ili već poznate komponente imaju svojstva koja nam trenutno nisu poznata. Stoga, prema Nagelu slijedi da likvidnost vode ne može biti istinski emergentno svojstvo jer nam je jasno kako ona nastaje na temelju poznavanja molekularne strukture vode (vidi također Goff, Seager, i Allen-Hermanson 2020). Naime, molekule H_2O se konstantno miču i sudaraju pri velikim brzinama te se vodikove veze brzo razbijaju i ponovo formiraju. Međutim, zbog slabijih van der Waalsovih sila molekule vode se odbijaju i skližu jedne od drugih te time proizvode pojavu likvidnosti.

Ako prihvatimo ove četiri premise, onda prema Nagelu slijedi da moramo prihvatiti panpsihizam. Do tog zaključka dolazi na sljedeći način:

> Ako mentalna svojstva organizma nisu implicirana nikakvim fizičkim svojstvima, već moraju proizaći iz svojstava sastavnih dijelova organizma, tada ti sastojci moraju imati nefizička svojstva iz kojih slijedi pojava mentalnih svojstava kada se radi o kombinaciji prave vrste. Budući da organizam može biti sastavljen od bilo koje tvari, sva tvar mora imati ta svojstva. (T. Nagel 1979, 182)

Dakle, ako pretpostavimo da postojanje svjesnih iskustava u principu ne možemo objasniti na temelju poznavanja fizičkih svojstava komponenti od kojih su organizmi sastavljeni, onda se čini da moramo prihvatiti da već te komponente od kojih je sve sastavljeno posjeduju neka nefizička svojstva na temelju kojih nastaju svjesna iskustva. Što znači da temeljna materija od koje je sav fizički svijet sastavljen mora posjedovati neka mentalna svojstva.

Iz fizikalističke perspektive premise (1) i (3) izgledaju prihvatljivo. Naime, premisom (1) ističe se fizikalistička tvrdnja da su u suštini sve stvari u našem svijetu sastavljene od nekakve materije. Premisom (3) uvjerljivo se tvrdi da su mentalna svojstva poput subjektivnih iskustava ono što karakterizira neke organizme. Čak i premisa (4) ne djeluje toliko sporno iz fizikalističke perspektive. Naime, ima smisla tvrditi da bi se, ako neki sustav ima određeno svojstvo, naročito ako se radi o fizičkom sustavu, ono trebalo barem u principu moći objasniti na temelju poznavanja dijelova tog sustava i principa prema kojima oni interagiraju. Stoga, ako je Nagelov argument valjan, čini se da iz fizikalističke perspektive ima najviše smisla osporavati istinitost premise (2).

Već nam je i jasno kako bi se to moglo činiti. Pristaše strategije fenomenalnih pojmova bi mogli tvrditi da, samo zato što na temelju korištenja fenomenalnih pojmova ne možemo derivirati znanje o fizičkim i funkcionalnim činjenicama koje se nalaze u podlozi subjektivnih iskustava, nije tako da ona predstavljaju nefizička ili nefunkcionalna svojstva organizama. Stoga pozivanje na to da na temelju pojmova o fizičkim činjenicama nećemo moći derivirati činjenice o subjektivnim iskustvima kako

ih opisujemo fenomenalnim pojmovima ne slijedi da subjektivna iskustva nisu identična nekim fizičkim ili funkcionalnim svojstvima. Moglo bi se inzistirati da je upravo snaga Nagelova i sličnih argumenata u tome što ističu da uopće nije jasno kako objasniti da ono na što referiramo fenomenalnim pojmovima može biti identično nekim fizičkim ili funkcionalnim svojstvima. Međutim, za očekivati je da iz fizikalističke perspektive to inzistiranje neće biti dovoljno uvjerljivo da se prihvate nefizikalistički zaključci, naročito ne ako je implikacija toga panpsihizam. Štoviše, ovdje bi pristaše strategije fenomenalnih pojmova mogli uzvratiti navođenjem da upravo oni imaju objašnjenje zašto se antiredukcionistima čini da se činjenice o mentalnim stanjima ne mogu objasniti pozivanjem na fizičke činjenice. Naime, upravo to što o nekim vlastitim stanjima razmišljaju koristeći fenomenalne pojmove je ono što ih sprečava da dođu do dobrih objašnjenja zašto su određena svjesna iskustva identična određenim fizičkim ili funkcionalnim svojstvima. Jednom kada se uvidi uloga pojmovnog okvira koji koristimo kada razmišljamo o svojim iskustvima, onda možemo objasniti otkuda se pojavljuju te antiredukcionističke intuicije.

Drugi argument za panpsihizam koji ćemo razmotriti temelji se na ideji intrinzične prirode materije. Među prvima ga je formulirao Bertrand Russell (1927) prema kojemu se gledište ponekad naziva russellijanski monizam (vidi Chalmers 2010a, 133). Kreće se od primjedbe da postoji određeni jaz u slici svijeta koju dobivamo od fizikalnih znanosti. Obično se smatra da će nam fizika dati potpuno objašnjenje fundamentalne prirode materijalnog svijeta jer je upravo fizika ona znanost koja se bavi prirodom prostora, vremena i materije. Međutim, panpsihisti upozoravaju da je ovakvo gledište pogrešno. Prema Philipu Goffu (2019) to možemo vidjeti kada se osvrnemo na vokabular i povijest fizikalnih znanosti. Goff ističe da je krucijalni trenutak u znanstvenoj revoluciji onaj kada je Galileo Galilei objavio da je knjiga svemira napisana jezikom matematike (vidi Goff 2019, 14–23). Od tada matematika predstavlja jezik kojim se služe fizičari. Taj jezik nije u cijelosti matematički, međutim, Galileo je odredio fiziku kao u suštini kvantitativnu znanost koja se bavi opisom stvarnosti u terminima zakona prirode koji se mogu matematički opisati. Time je, navodi Goff (2019, 21), Galileo izbacio iz fizike govor o kvalitativnim stanjima i uopće mogućnost njihova objašnjenja.

Ono što je problem s tim prijelazom na matematički vokabular koji se bavi samo kvantitativno-matematičkim odnosima je što nije jasno da se njime može zahvatiti u cijelosti priroda naše stvarnosti. Matematički opis situacije apstrahira od konkretne stvarnosti. Ono o čemu matematički jezik govori samo su odnosi među veličinama te u tom smislu može govoriti samo o strukturalnim svojstvima stvari. Na primjer, paradigmatski primjer zakona prirode ima sljedeću formu: F = m*a, tj. sila je jednaka masi puta ubrzanje. No, taj zakon nam ne kaže, na primjer, koja je priroda mase, tj. koja je priroda materije i sile. Samo nam kaže u kojim odnosima oni stoje. Na primjer, iz ove

jednadžbe znamo da je masa jednaka količini sile podijeljene s količinom ubrzanja (m = F/a) i da je ubrzanje jednako sili podijeljenoj s masom (a=F/m).[68] Dakle, ove jednadžbe nam ne govore koja je intrinzična priroda mase ili sile. Slično nam matematičke formulacije drugih fizikalnih zakona neće reći ništa o njihovoj prirodi. To nije problematično ako želimo, na primjer, predvidjeti kako će se elektroni ponašati. I naravno svi se slažu da nam matematički modeli ponašanja materije daju izuzetno korisne informacije. Na temelju tih informacija uspjeli smo manipulirati prirodom na različite načine, napraviti lasere, nuklearne elektrane i poslati ljude na mjesec.

Međutim, panpsihisti smatraju intuitivnim da elektroni osim ponašanja također posjeduju *intrinzičnu* prirodu; tj. da mora postojati odgovor na pitanje kakvi su elektroni i druge čestice sami po sebi, te koja su njihova kvalitativna, a ne samo kvantitativna svojstva. Ako je točno da nam fizika ne daje odgovor na ovo pitanje, onda nam čisto fizički opis svijeta nikada neće dati potpunu i adekvatnu teoriju prirode materijalnog svijeta.

Stoga, panpsihisti smatraju da filozofima ostaje otvorena mogućnost da spekuliraju koja bi mogla biti intrinzična priroda stvari u svijetu, uključujući prirodu fundamentalnih čestica od kojih je sav materijalni svijet sastavljen (vidi Chalmers 2010a, 133–34; Goff 2019, 132–33). Sam Russell je smatrao da se svijet u suštini ne sastoji od mentalnih ili fizičkih svojstava, već da je intrinzična priroda svijeta određena nekakvim trećim faktorom koji se nalazi u podlozi fizičkih i fenomenalnih svojstava. Stoga se njegova pozicija često naziva neutralni monizam (vidi Chalmers 2010a, 133–34; Goff 2019, 137). Kod Russella malo ostaje nejasno što bi točno bila ta neutralna intrinzična priroda koja na neki način prethodi fenomenalnim i fizičkim svojstvima. Suvremeni panpsihisti su u tom pogledu malo jasniji te predlažu da upravo svjesnost čini intrinzičnu prirode materije (vidi Goff 2019, 137). Stoga, ako pretpostavimo da elektroni čine dio fundamentalne konstituente prirode, onda prema panpsihistima slijedi da je elektron esencijalno stvar koja ima svijest (naravno misle na vrlo bazičan i primitivan pojam svijesti) te da postoji nešto kako je to biti elektron.

Goff (vidi, npr. 2019, 134) smatra da se takvo gledište, tj. jednostavnost takve hipoteze, može opravdati pozivanjem na Ockhamovu britvu. To je pojam s kojim smo se susreli kada smo govorili o teoriji identiteta tipova u poglavlju 4. Ideja je da je bolja ona teorija koja na jednostavniji način ili postuliranjem manjeg broja predmeta, principa, zakona itd. može bolje objasniti stvari od neke druge teorije.

Goff (2019, 132–33) argumentira da nije jasno koji bi bili alternativni prijedlozi za objasniti intrinzičnu prirodu materije. Sve što znamo o prirodi fundamentalnih čestica temelji se na posrednom znanju koje dobivamo kroz iskustvo ili oslanjanjem na instrumente koji se koriste za izvođenje fizikalnih

[68] Ili kako kaže pjesma *mala moja dajem ti na znanje sila masi daje ubrzanje*.

eksperimenata. Takav način spoznaje nam neće previše reći o intrinzičnoj prirodi fizičke materije. Stoga, panpsihisti poput Goffa navode da su nam jedine jasne opcije za objašnjenje intrinzične prirode fundamentalnih čestica prihvaćanje panpsihizma ili priznanje da zapravo ne znamo koja je intrinzična priroda materije.

Nadalje, ako pretpostavimo da je dualizam neistinit, znamo da postoji barem jedna vrsta tvari koja je svjesna, a to je mozak. Moguće je da mozak predstavlja jedini trag za općenito otkrivanje intrinzične prirode materije koja se nalazi izvan mozga. U tom pogledu, Goff (2019, 123) tvrdi da je, u odsutnosti razloga da formiramo drugačije mišljenje, najjednostavnija i najelegantnija hipoteza ta da je materija izvan našeg mozga u kontinuitetu s materijom koja čini mozak, pa čak i u pogledu toga da posjeduje svijest.

Ako prihvatimo ovakvo gledište, onda na neki način rješavamo teški problem svijesti jer se ne moramo pitati kako to da mozak proizvodi svjesna mentalna stanja. Naime, mozak ne proizvodi svijest, već je sva materija, uključujući i onu koja čini mozak, već svjesna. Mozak samo predstavlja jednu kompleksniju varijantu svjesne materije koju u jednostavnijoj formi posjeduju sve čestice koje sačinjavaju naš materijalni svijet.

Unatoč navedenim argumentima, panpsihizam se i dalje percipira kao vrlo kontraintuitivno gledište (Frankish 2021; Nagasawa 2021). Jedna od stvari koja odmah dolazi do izražaja jest činjenica da je teško zamisliti ili pojmiti kako je to biti elektron ili neka druga fundamentalna čestica te što bi uopće značilo pojmiti fenomenalni karakter iskustva elektrona ili neke druge fundamentalne čestice (Goff 2016; Miller 2017; za raspravu, vidi Siddharth 2021). No, unatoč tom nedostatku pozitivne koncepcije fenomenalnih svojstava bazične materije i koliko god bio kontraintuitivan, panpsihizam se nametnuo kao legitimna pozicija u suvremenim raspravama o odnosu uma i tijela (Chalmers 2016a; Seager 2016; Goff 2017). Štoviše, s obzirom na argumente koje nude, panpsihisti smatraju da je, ako se netko ne slaže s njihovim prijedlogom u pogledu intrinzične prirode materije, teret dokaza na drugima da pokažu što ne valja s panpsihičkim gledištem te da predlože alternativna rješenja (Goff 2017; 2019).

Međutim, također postoji određeni konsenzus da se panpsihizam suočava s unutrašnjim problemima koji utječu na njegovu koherentnost (vidi npr. Chalmers 2016b; Goff 2016; Miller 2017; Siddharth 2021). U nastavku ćemo se osvrnuti na jedan takav problem koji se naziva problem kombinacije fenomenalnih svojstava.

9.4 Problem kombinacije za panpsihiste

Problem kombinacije sastoji se u tome da nije jasno kako se fenomenalna svojstva mikrofizičkih entiteta poput kvarkova i fotona mogu kombinirati da nastanu makrofenomenalna iskustva koja karakteriziraju ljudske doživljaje (Chalmers 2016b, 179). Na primjer, ljudi koji osjećaju bol imaju doživljaj s

određenim fenomenalnim karakterom. Pitanje je kako se fenomenalna svojstva mikrofizičkih entiteta mogu povezati tako da proizvedu poznati osjećaj boli kakav ljudi mogu doživjeti. Smatra se da je među prvima ovaj problem formulirao William James. Referirajući se na osjećaje, James navodi sljedeće:

> Uzmite ih stotinu, promiješajte ih i spakirajte što bliže možete (što god to značilo); i dalje svaki ostaje isti osjećaj kakav je uvijek bio, zatvoren u vlastitu kožu, bez prozora, ne znajući što su i znače drugi osjećaji. Postojao bi sto i prvi osjećaj, ako (…) bi se pojavila svijest koja pripada grupi kao takvoj. I ovaj 101. osjećaj bio bi potpuno nova činjenica; 100 izvornih osjećaja moglo bi, prema nekom čudnom fizikalnom zakonu, biti signal za njegovo stvaranje, kada se oni zajedno spoje; ali oni ne bi bili supstancijalno istovjetni s njim, niti on s njima, i nikada se ne bi moglo derivirati jednog iz drugih, ili (u bilo kojem razumljivom smislu) reći da se razvio iz njih.
>
> Uzmite rečenicu koja ima tucet riječi, i uzmite dvanaest muškaraca i recite svakom po jednu riječ. Zatim postavite muškarce u red ili ih zgrčite u hrpu, i neka svaki razmišlja o svojoj riječi koliko hoće; nigdje neće biti svijesti o cijeloj rečenici. Govorimo o »duhu vremena« i »sentimentu naroda« i na razne načine hipostaziramo »javno mnijenje«. Ali znamo da je to simboličan govor i niti u snu ne smatramo da duh, mišljenje, osjećaj itd., sačinjavaju svijest različitu od, i koja nadilazi, svijest nekoliko pojedinaca koje riječi »dob«, »ljudi« ili »javnost« označuju. Privatni umovi se ne aglomeriraju u viši složeni um. (citirano u Chalmers 2016b, 179–80)

Ovdje James pokazuje da je problematična ideja prema kojoj se osjećaji, ili općenito svjesna iskustva, mogu kombinirati i aglomerirati kako bi od njih nastala složenija iskustva. Drugim riječima, ako panpsihisti smatraju da je ono što objašnjava ljudska svjesna iskustva činjenica da ona nastaju kao posljedica kombinacije svjesnih iskustava temeljnih mikrofizičkih čestica, onda moraju moći objasniti kako je to moguće. James ukazuje na to da nije jasno da možemo pojmiti nešto takvo jer se intuitivno čini da bi, čak i kada bi na temelju 100 postojećih osjećaja nastao novi 101. osjećaj, njegovo postojanje bilo još jedna temeljna činjenica koja se ne bi mogla derivirati iz iskustva prethodnih osjećaja. Jer, kako navodi James, svaki osjećaj je „zatvoren u vlastitu kožu, bez prozora, ne znajući što su i znače drugi osjećaji“.

Na temelju ove primjedbe mogu se formulirati različite varijante argumenta protiv panpsihizma ovisno o tome koji aspekt panpsihizma se stavlja u fokus (za pregled, vidi Chalmers 2016b). Mi ćemo se usredotočiti na dva aspekta panpsihizma kako bismo ilustrirali problem kombinacije.

Prvi aspekt odnosi se na činjenicu da postojanje fenomenalnog karaktera podrazumijeva postojanje subjekta koji ima ta iskustva. Dakle, prema panpsihizmu postoje barem neke mikrofizičke čestice koje su subjekti s određenim svjesnim iskustvima. Sada se postavlja pitanje kako je moguće da osoba koja ima iskustva može biti sastavljena od drugih subjekata koji imaju svoja iskustva. Naročito se javlja problem ako smatramo da su subjekti jedinstvena i nedjeljiva „polja" koja omogućuju postojanje iskustava (za raspravu, vidi Siddharth 2021). Dakle, postavlja se pitanje kako nizanjem i kombinacijom puno takvih malih subjekata koji predstavljaju jedinstvena i nedjeljiva polja iskustava mogu nastati kompleksni subjekti poput osobe sa svojim jedinstvenim iskustvima i doživljajima. Čini se da ništa u postojanju tih mikrosubjekata ne implicira nužno postojanje makrosubjekta.

Drugi aspekt panpsihizma koji ćemo istaknuti jest tvrdnja da mikrofizičke čestice posjeduju kvalitativna stanja. Ovdje se postavlja pitanje kako se te mikrokvalitete mogu kombinirati i proizvesti makrokvalitete koje karakteriziraju ljudska iskustva. Chalmers daje primjer:

> Ovdje su makrokvalitete specifične fenomenalne kvalitete poput fenomenalne crvenosti (kako je to vidjeti crveno), fenomenalne zelenosti i tako dalje. Prirodno je pretpostaviti da mikroiskustva uključuju mikrokvalitete, koje mogu biti primitivno analogne makrokvalitetama. Kako se oni kombiniraju? (Chalmers 2016b, 183)

Dakle, opet se postavlja pitanje kako i zašto određena kombinacija mikrokvaliteta može proizvesti makrokvalitete koje karakteriziraju ljudske doživljaje. Štoviše, ova razmatranja sugeriraju općenitiji problem s panpsihizmom.

9.5 Argument pojmljivosti protiv panpsihizma

Sjetimo se da panpsihisti smatraju svoje gledište rješenjem za teški problem svijesti koji se upućuje fizikalističkim gledištima. Objašnjenje koje nude jest to da već na mikrofizičkoj razini pronalazimo mentalna svojstva. Međutim, kako bi odgovorili na teški problem svijesti, panpsihisti moraju objasniti kako od svjesnosti na mikrorazini dolazimo do svjesnih iskustava koja karakteriziraju ljude na makrorazini. Ako se to ne može objasniti, onda zapravo ne nude bolje objašnjenje pojave svijesti kod ljudi nego što su to u stanju učiniti fizikalisti. Upravo nam na to ukazuje problem kombinacije koji proizlazi iz nejasnoće što na mikrorazini može objasniti ili čini nužnim

postojanje iskustva na makrorazini. Štoviše, čini se da možemo zamisliti da svaki subjekt sa svojim iskustvima postoji, a da ne postoji niti jedan drugi subjekt. Ako zamislivost ove vrste implicira mogućnost, onda slijedi da je moguće da postojanje nekog subjekta s iskustvima ne implicira postojanje drugog subjekta s iskustvima. Dakle, moguće je da postojanje subjekata na mikrorazini ne implicira postojanje subjekata na makrorazini.

Ovaj problem Chalmers poopćuje i formulira kao argument pojmljivosti protiv panpsihizma. Pretpostavimo da FP predstavlja konjunkciju svih mikrofizičkih i mikrofenomenalnih istina o našem svijetu (tj. istina o instancijaciji mikrofizičkih i mikrofenomenalnih svojstava). Pretpostavimo da je Q neka istinita makrofenomenalna tvrdnja, poput „Neki makroskopski entiteti su svjesni“. Koristeći ove simbole Chalmers formulira argument na sljedeći način:

1) FP&¬Q je pojmljivo.
2) Ako je FP&¬Q pojmljivo, onda je metafizički moguće.
3) Ako je FP&¬Q metafizički moguće, panpsihizam je neistinit.
4) Dakle, panpsihizam je neistinit. (Chalmers 2016b, 187)

Chalmers (2016b) ističe da problem kombinacije posebice zahvaća ono što naziva konstitutivni panpsihizam. To je gledište prema kojemu postojanje činjenica o mikrofenomenalnim svojstvima implicira postojanje činjenica o makrofenomenalnim svojstvima.

Međutim, postoje druge nekonstitutivne verzije panpsihizma. Na primjer, prema emergentnom panpsihizmu makrofenomenalna svojstva nastaju ili emergiraju iz mikrofenomenalnih svojstava, no ona ih ne konstituiraju. Ovakvo shvaćanje emergencije je u suprotnosti s Nagelovim (1979) jer se tvrdi da, ontološki gledano, na različitim razinama stvarnosti postoje svojstva koja se ne mogu derivirati iz svojstava s nižih razina stvarnosti. Ovakvom varijantom ontološkog emergentizma tvrdi se da postoje veze između mikro i makro činjenica, samo što su one povezane kontingentnim zakonima prirode (Chalmers 2016b, 192–93). Emergentni panpsihizam u principu zaobilazi problem kombinacije jer se njime niti ne tvrdi da ćemo iz mikrofenomenalnih činjenica moći derivirati ili uopće shvatiti kako nastaju makrofenomenalne činjenice. Štoviše, emergentisti mogu reći da su makrofenomenalne činjenice ontološki temeljne poput onih mikrofenomenalnih te da je jedino što nam preostaje za utvrditi prema kojim se kontingentnim zakonima one međusobno povezuju.

Međutim, problem s ovakvim gledištem je što dosta sliči na neku varijantu dualizma te je s obzirom na to podložno sličnim prigovorima (vidi Chalmers 2016b, 183). U poglavlju 2 vidjeli smo da se dualizam susreće s problemom uzročnosti pod pretpostavkom da je fizikalni svijet uzročno zatvoren. Sličan problem se javlja i ovdje. Ako pretpostavimo da je fizički svijet uzročno zatvoren, i da su u ontološkom smislu svi uzročni događaji utemeljeni na

uzročnim moćima mikrofizičkih svojstava, onda nije jasno kako makrosubjekti, uključujući njihova fenomenalna svojstva, mogu imati uzročne moći. Naime, čak i ako dopustimo da mikrofenomenalna svojstva imaju uzročne moći, one se ne mogu prenijeti na makrofenomenalna svojstva jer potonja nisu utemeljena na njima, već predstavljaju novu i neovisnu ontološku domenu postojanja. Stoga se čini da će emergentni panpsihizam imati kao posljedicu epifenomenalizam u pogledu makrofenomenalnih svojstava.

No, čak i ako ostavimo po strani probleme uzročnosti, čini se da će sve varijante nekonstitutivnog panpsihizma imati isti problem. Kako smo ranije istaknuli, panpsihizam se u suvremenim raspravama motivira antiredukcionističkim argumentima kojima se podupire postojanje teškog problema svijesti. Ako pretpostavka da mikrofizički konstituenti našeg svijeta imaju već svjesna stanja ne može (ili čak niti nema za cilj) objasniti kako to da makrofizički objekti poput ljudi posjeduju svjesna mentalna stanja, onda nije jasno zašto bismo uopće razmatrali panpsihizam kao ozbiljnu opciju za rješenje teškog problema svijesti.

Čini se da smo napravili puni krug. Antifizikalisti tvrde da postoji teški problem svijesti koji nam pokazuje da se trebamo okrenuti dualističkim i panpsihističkim gledištima. Potreba da se formulira naturalistički respektabilna dualistička teorija vodi filozofe poput Chalmersa da formuliraju gledišta koja su jako bliska panpsihizmu. Štoviše, s obzirom na probleme uzročnosti koji opterećuju dualiste, čak ih navodi da prihvate neku varijantu panpsihizma (Goff 2019). Sada vidimo da i panpsihizam pati od unutrašnjih problema. Za njih možda ne postoji teški problem svijesti u smislu u kojem se upućuje fizikalistima, ali postoji problem određenja prirode mikrofenomenalnih stanja i načina na koji ona zajedničkim djelovanjem stvaraju poznata makrofenomenalna svojstva koja karakteriziraju ljude. Ovi pojmovni problemi nas mogu navesti da se zapitamo jesmo li negdje pogriješili kad smo došli do zaključka da se pojmovi fenomenalnih karaktera odnose na nešto sasvim različito od pojmova kojima referiramo na fizička i funkcionalna svojstva svjesnih bića. Je li moguće da smo došli do ove situacije pogrešnom upotrebom pojmova i posljedičnog pogrešnog shvaćanja toga što je uopće svijest i što znači da imamo doživljaje s određenim kvalitativnim svojstvima? U posljednjem odjeljku koji slijedi ćemo malo detaljnije razmotriti tu mogućnost.

9.6 Dennettova kritička analiza pojma „qualia”

Dosad smo pretpostavljali da pojam fenomenalnog iskustva označuje pravu pojavu koja se odnosi na subjektivni karakter naših doživljaja, iskustva i drugih svjesnih mentalnih stanja. Vidjeli smo da, kada se prihvati da taj pojam označuje poseban aspekt našeg iskustva, mnogi autori smatraju da fenomenalni karakter iskustva ne može biti identičan ili supervenirati nad

fizičkim svojstvima u našem svijetu. Štoviše, neke autore je to dovelo do prihvaćanja ideje da možda i mikrofizički konstituenti našeg svijeta posjeduju subjektivna iskustva. Ovakav razvoj događaja možda i nije neočekivan ako se fenomenalne pojmove već na početku rasprave shvati kao da referiraju na stvari koje su nevezane uz psihološko-funkcionalne karakteristike ljudi koje se mogu istraživati pod egidom „lakog problema svijesti" (vidi Frankish 2021).

Međutim, ako je fizikalno objašnjenje svijesti nemoguće zbog teškog problema svijesti, a potonji je posljedica *a priori* nemogućnosti povezivanja fenomenalnih i fizikalnih pojmova, onda nam je to možda znak da trebamo ponovno razmotriti pojmove kojima referiramo na te stvari. Dakle, trebali bismo razmotriti kako nefizikalisti točno shvaćaju pojam fenomenalnog karaktera. Ako bismo to shvatili, možda bismo onda bolje mogli razumjeti otkud dolazi ta jaka intuicija da fenomenalni karakter iskustva ne može biti još jedno fizičko-funkcionalno svojstvo naših mentalnih stanja.

U ostatku ovog poglavlja razmotrit ćemo to pitanje. Vidjet ćemo da nefizikalističko shvaćanje fenomenalnog karaktera, više no što nam daje razlog da budemo nefizikalisti, možda predstavlja razlog da zaključimo da svjesna iskustva u tom smislu riječi zapravo ne postoje. Ovakvo gledište se u suvremenim raspravama naziva iluzionizam (Frankish 2016). Ideja je da se nama čini da postoje fenomenalni karakteri kako se na njih referira s fenomenalnim pojmovima. Međutim, to je zapravo samo iluzija; svjesna stanja postoje, ali ona nemaju svojstva koja bi podupirala postojanje teškog problema svijesti.[69]

Kako bismo bolje shvatili način na koji nefizikalisti shvaćaju pojam fenomenalnog iskustva, osvrnut ćemo se na pojam *qualia* koji često koriste kako bi zahvatili govor o subjektivnim aspektima iskustva. Dosad smo nastojali koristiti pojam fenomenalnog karaktera iskustva na neutralan način koji neće podrazumijevati nefizikalističke zaključke u pogledu prirode subjektivnog iskustva. Međutim, moguće je da se u pozadini intuicija koje podupiru teški problem svijesti nalazi karakteriziranje subjektivnog iskustva kakvo se obično povezuje s pojmom *qualia*. Stoga će nam u nastavku biti relevantna rasprava koju započinje Dennett u radu *Quining qualia* (1988).

U tom radu Dennett daje niz razmatranja kojima nastoji neutralizirati intuicije koje određuju karakterizaciju svjesnih iskustava kao entiteta koji se ne mogu akomodirati u prirodnom svijetu. To čini na način da prvo pokušava vidjeti što definira fenomene koji spadaju pod pojam *qualia*. Nakon toga

[69] Iako poticaj za iluzionizam dolazi od Dennetta, najistaknutiji suvremeni iluzionist je Keith Frankish (2016). Ponekad se u literaturi iluzionizam i slična gledišta nazivaju eliminativizmom u pogledu svijesti (za raspravu eliminativizma, vidi Pećnjak i Janović 2016, pogl. 6). Međutim, čini nam se da je taj termin nezgodan jer najčešće tvrdnja nije da *ne* postoje svjesna mentalna stanja čiju bi prirodu trebalo objasniti, već se tvrdi da *pojam* svjesnog iskustva kako ga nefizikalisti shvaćaju ne referira na pravi fenomen.

nastoji pokazati da se tako shvaćene *qualia* ne odnose na prave fenomene u svijetu. Posljedica toga je da, ako fenomenalni karakter iskustva ne postoji u smislu *qualia*, onda nije jasno da uopće ima smisla odgovarati na antifizikalističke argumente čija se intuitivna snaga temelji na ideji *qualia*. Pa da vidimo kako Dennett dolazi do tog zanimljivog, ali kontroverznog zaključka.

9.7 Qualia i njihova svojstva

Prema Dennettu (1988) pojam *qualia*, kako ga nefizikalisti shvaćaju, pretpostavlja da su to svojstva iskustva, dakle njihovi fenomenalni karakteri, koja mogu biti u potpunosti odvojena od dispozicija za ponašanje te vjerovanja i drugih funkcionalnih mentalnih stanja koja osoba može imati prema njima. Frankish (2021) je takvo shvaćanje nazvao u potpunosti depsihologiziranim pojmom fenomenalnog karaktera. Takvo je shvaćanje pojma fenomenalnog karaktera ili *qualia* jasno već iz podjele na teške i lake probleme svijesti, gdje su teški problemi upravo oni koji se odnose na pojmove svijesti koji ne referiraju na iskustva kao psihološko-funkcionalna stanja osobe. Dennett smatra da takvo depsihologizirano shvaćanje *qualia* proizlazi iz konfuznih predteorijskih pretpostavki o svjesnim isksutvima koje trebamo odbaciti. Drugim riječima, Dennett smatra da *qualia* ne postoje. No da bismo shvatili kako Dennett dolazi do tog zaključka, trebamo razmotriti koje su pretpostavljene karakteristike svjesnih iskustava na koje referiramo uz pomoć pojma *qualia*.

Prema Dennettu (1988, 229), autori koji smatraju da postoje *qualia* shvaćaju ih kao da imaju sljedeća svojstva:

1) Neiskazivost/neopisivost: pojedina *quale* je nešto što se ne može iskazati drugoj osobi. Na primjer, ideja je da nikada u potpunosti ne možemo opisati drugoj osobi kako izgleda subjektivni doživljaj nekog mirisa, okusa, izgleda stvari i tome slično.
2) Intrinzičnost: *qualia* bi trebala biti homogena svojstva iskustva čije postojanje ne ovisi o postojanju drugih predmeta. Ona su specifična za ljudska subjektivna iskustva te mogu postojati neovisno o bilo čemu drugome. Na primjer, *quale* crvene boje je svojstvo našeg vizualnog iskustva koje je jednostavno i ne može se dalje razlomiti u dijelove.
3) Privatnost: *qualia* su dostupna putem introspekcije samo osobi koja ih doživljava. Nitko drugi im ne može imati pristup poput samog subjekta tog doživljaja.
4) Direktno ili neposredno shvatljivo u svijesti: osoba koja ima *qualia* neposredno ih je svjesna te njihovu prirodu u potpunosti zahvaća

> samom činjenicom da ih je svjesna. Dakle, priroda *qualia* se u potpunosti otkriva u samom činu doživljavanja.

Dennett shvaća ova svojstva kao povezana, u smislu da će dio razloga zašto se prihvaća jedna karakteristika *qualia* biti objašnjen prihvaćanjem druge karakteristike. Na primjer, barem dio razloga zašto su *qualia* neiskazive je to što su one jednostavna i nesastavljena intrinzična svojstva iskustva. Također, ona se ne mogu uspoređivati među ljudima jer su privatna. Što povlači da su direktno dostupna samo osobi koja ih doživljava u određenom trenutku.

Takvo shvaćanje *qualia* podrazumijevaju mnogi antifizikalistički i antifunkcionalistički argumenti koje smo obrađivali kroz različita poglavlja. Uzmimo na primjer misaoni eksperiment obrnutog spektra (vidi poglavlje 5). Čini se da možemo zamisliti da postoje dvije osobe koje su u fizičkom i funkcionalnom smislu identične, ali imaju u potpunosti obrnuti doživljaj boja. Gdje jedna osoba vidi zelenu boju, druga će vidjeti crvenu, iako će obje osobe govoriti da vide predmet zelene boje te će imati iste dispozicije za korištenje termina „zelena boja“ i tome slično. Ako je to moguće onda slijedi da su fenomenalni karakteri iskustva boja neiskazivi, privatni i u potpunosti odvojeni od psihološko-ponašajnih dispozicija osoba.

Međutim, Dennettova argumentacija podsjeća da, iz fizikalističke perspektive, ako ne možemo pokazati da postoji razlika u iskustvima dvije osobe koja bi se mogla provjeriti na intersubjektivan način, onda nije jasno zašto bismo uopće pretpostavili da postoje takva svojstva koja se nikako ne mogu detektirati. Ovom načinu argumentacije moglo bi se prigovoriti da pretpostavlja neprihvatljiv oblik verifikacijsko-biheviorističke metodologije (vidi poglavlje 3), koju upravo intuicije u pozadini misaonih eksperimenata poput obrnutog spektra osporavaju.

Kako bi dodatno podržali tu vrstu intuicija pristaše *qualia* mogu se pozivati na *intra*personalnu varijantu obrnutog spektra (vidi, npr. Shoemaker 1975). Čini se da možemo zamisliti da osoba doživi promjenu *qualia* za koju nitko drugi osim nje neće primijetiti da se dogodila. Pretpostavimo da je zli neuroznanstvenik utjecao na funkcioniranje mozga osobe O te se ona jedno jutro probudi i primijeti da je trava crvena, nebo žuto, ruže zelene i tako dalje. O također primjećuje da se svi drugi ljudi ponašaju normalno kao i do sada što znači da nitko osim nje nije primijetio da su stvari u svijetu promijenile boju. Na temelju toga O zaključuje da mora biti da se samo kod nje dogodila inverzija *qualia* boja u odnosu na ono kako ih se sjeća od dana ranije.

Dennett (1988, 231) ukazuje na to da čak ni ovaj primjer ne pokazuje da postoje *qualia* koje bi bile direktno dostupne samo osobi koja ih doživljava. Naime, barem su dva načina na koja je zli neuroznanstvenik mogao utjecati na promjenu *qualia* kod O. Jedan način je da „preokrene“ optički živac tako

da jednom kada svjetlo padne na retinu oka onda mozak interpretira vizualne informacije suprotno od onoga kako ih inače interpretira. Drugi način je da utječe na dostupnost sjećanja i njihovu povezanost u pamćenju. Na primjer, može podesiti memoriju osobe O tako da ona misli da je jučer vidjela boje stvari na jedan način, a danas ih vidi kao da imaju drugu boju. U tom slučaju, ne bi se radilo o stvarnoj promjeni *qualia* obojanih predmeta, već o podešavanju dispozicija za reagiranje na iskustvo predmeta koje se temelje na promijenjenim sjećanjima o bojama predmeta.

Kada se misaoni eksperiment na ovaj način nadopuni, postaje jasnije da O nema direktan uvid u svoja kvalitativna stanja. Naime, Dennett (1988, 231) pokazuje da prava reakcija koju bi O trebala imati jest da je došlo do promjene u njezinim kvalitativnim doživljajima ili u reakcijama prema njima koje su posljedice promjena u pamćenju, a ne u samim doživljajima stvari. Iz same perspektive O nije jasno je li došlo do promjene u *qualia* ili u reakcijama prema nepromijenjenim subjektivnim iskustvima. Iz toga slijedi da O nema direktni i nepogrešivi uvid u svoja kvalitativna stanja. Štoviše, jedini način da otkrije je li došlo do promjene u *qualia* jest da to pokuša otkriti kroz neke objektivne metode (npr. da pita zlog neuroznanstvenika što je točno napravio).

Dennett nastoji poopćiti ovaj zaključak kako bi pokazao da je sama ideja intrinzičnih, privatnih i neposredno dostupnih *qualia* nekoherentna te u tu svrhu razmatra nekoliko misaonih eksperimenata.[70] Mi ćemo se ovdje usredotočiti na jedan koji pokazuje da pojam *qualia* ne može istodobno referirati na mentalna stanja koja bi bila intrinzična, neposredno dostupna i nepovezana s psihološko-ponašajnim funkcijama.

9.8 Argument za nepostojanje qualia

Zamislimo da su Amalija i Helena kušačice kave u prehrambenoj kompaniji Franck (usp. Dennett 1988, 231–32, intuicijska pumpa 7). Zajedno s ostalim kušačima kave, njihov zadatak je osigurati razinu kvalitete Franck kave, što uključuje provjeravanje da okus različitih vrsta kave ne varira značajno kroz različite periode proizvodnje. Nakon šest godina rada za Franck, Amalija u povjerenju kaže Heleni da više ne uživa u svom poslu kao kada je prije šest godina počela raditi te navodi:

[70] Kako smo ranije naveli, svojstvo neiskazivosti ovisi o shvaćanju *qualia* kao intrinzičnih svojstava iskustva. Stoga, ako se može pokazati da *qualia* nisu intrinzična svojstva, onda se dovodi u pitanje i sama ideja neiskazivosti. Međutim, u kojoj mjeri se dovodi u pitanje neiskazivost kao svojstva *qualia* nije toliko važno za općenitiju raspravu o tome može li fizikalizam dati točno objašnjenje prirode svjesnih iskustava. Stoga se u nastavku nećemo specifično referirati na argumente kojima se dovodi u pitanje neiskazivost *qualia*.

> Kada sam počela raditi za Franck mislila sam da je njihova kava najbolja na svijetu. Bila sam ponosna što sam dijelila odgovornost za očuvanje tog okusa kroz godine. I dobro smo obavljali svoj posao; kava ima isti okus danas kao što je imala kada sam došla. Ali znaš, više mi se ne sviđa! Moji okusni doživljaji su se promijenili. Postala sam sofisticiraniji kavopija. Više mi se uopće ne sviđa *taj okus*. (Prilagođeni primjer iz Dennett 1988, 232, kurziv je u originalu)

Heleni je bilo drago čuti Amalijinu ispovijest jer je i sama doživjela sličnu promjenu. Helena navodi da je, slično kao i Amalija, kada je počela raditi za Franck smatrala da se tu proizvodi najbolja kava na svijetu. Međutim, sada je promijenila svoj stav.

> Poput tebe, sada me više nije briga za kavu koju proizvodimo. Ali *moji* okusni doživljaji se nisu promijenili; *moji* […] kušači su se promijenili. Odnosno, mislim da je nešto pošlo po zlu s mojim okusnim pupoljcima ili nekim drugim dijelom moje perceptivne mašinerije za analizu okusa. Franck kava mi nema okus kao nekad; kada bi barem imala, i dalje bih je voljela, jer još uvijek mislim da je *taj okus* najbolji okus kave. Ne kažem da nismo dobro odrađivale svoj posao. Svi drugi kušači se slažu da je okus isti, a moram priznati da ni ja iz dana u dan ne mogu uočiti nikakvu promjenu. Dakle, mora da je to samo moj problem. Mislim da više nisam dobra za ovaj posao. (Prilagođeni primjer iz Dennett 1988, 232, kurziv je u originalu)

Amalija i Helena slične su u jednom pogledu i različite u drugom. Obje tvrde da su nekada voljele Franck kavu, a sada je više ne vole. Ono po čemu se razlikuju su razlozi zašto je više ne vole. Amalija tvrdi da joj Franck kava i dalje ima isti okus, no smatra da su joj se ukusi promijenili, postala je sofisticiraniji kušač, te više nema isti *stav* sviđanja prema Franck kavi. Helena nije promijenila ukus, u smislu stava sviđanja prema Franck kavi, već joj se čini da njihova kava više nema isti okus kakav je imala prije šest godina.

Dennett (1988, 232) pita se što se točno događa u ovim slučajevima, imaju li Amalija i Helena zaista pravo kada tvrde da se dogodila neka promjena u njihovim iskustvima? Moramo li uzeti njihove tvrdnje zdravo za gotovo? Je li moguće da se jedna od njih ili obje varaju u pogledu toga što im se dogodilo? Jesu li njihova iskustva zapravo ista, a samo se razlikuju u načinu na koji izvještavaju o njima? Oba slučaja uključuju intrigantne veze između davanja iskaza o trenutnim iskustvima i načina na koji su ona zabilježena u pamćenju. Stoga pouzdanost njihovih odgovora u značajnoj mjeri ovisi o pouzdanosti njihovih pamćenja. I naravno postavlja se pitanje kako bismo

mogli provjeriti njihovu pouzdanost. S obzirom na to Dennett (1988, 232) ukazuje na to da postoje tri mogućnosti koje karakteriziraju trenutne situacije. Zadržat ćemo se na slučaju Amalije.

Moguće je da je Amalijina *quale* okusa kave ostala ista, a da su se njezini reaktivni stavovi prema toj *quale* promijenili. Na primjer, moguće je da su se kod Amalije promijenili estetski standardi te je u tom smislu promijenila ukus i više ne smatra da Franck proizvodi dovoljno dobru kavu. Druga mogućnost je da se Amalija vara u pogledu nepromjenljivosti svojih *qualia* kave. Fenomenalni karakteri njezinog iskustva su se bez njezinoga znanja postepeno s godinama promijenili, dok su estetski standardi ili standardi ukusa ostali isti. U tom slučaju, rekli bismo da Amalija ima iluziju da je njezin ukus postao sofisticiraniji. Dapače, možemo reći da se ona nalazi u istom stanju poput Helene, samo joj nedostaje Helenina samospoznaja. Treća mogućnost je da se Amalija nalazi u situaciji koja bi bila negdje između prve i druge mogućnosti. Dakle, moguće je da su njezine *qualia* donekle pomaknute i da su se standardi ukusa promijenili. Slične mogućnosti vrijede i za slučaj Helene.

Koherentnost ovih mogućnosti dovodi u pitanje ideju neposredne spoznaje vlastitih svjesnih iskustava. Budući da Amalijine tvrdnje o promjeni standarda ukusa ovise o njezinom sjećanju kakav je okus kava imala kroz sve ove godine, onda je sasvim moguće da nju sjećanje vara i da se nije promijenio njezin ukus, već da sada okus te iste kave doživljava na drugačiji način. Iz subjektivne Amalijine perspektive to pitanje ne može dobiti konkluzivan odgovor. Drugim riječima, samo na temelju introspektivne spoznaje Amalija ne može riješiti ovaj problem. Amalija bi mogla potražiti pomoć neurofiziologa. Na primjer, neurofiziolog bi mogao pokušati isključiti barem jednu mogućnosti tako da provjeri je li se nešto promijenilo u okusnim pupoljcima kako to tvrdi za sebe Helena. Međutim, ako smatramo da je to dobar način utvrđivanja je li došlo do promjene u fenomenalnom karakteru iskustva, onda nam to ukazuje na to da *qualia* zapravo nisu privatne i neposredno dostupne samo osobi koja ih doživljava, već ih se može istraživati kao i druga objektivno dostupna fizička svojstva (vidi Dennett 1988, 235). Stoga, ako je neposredna spoznaja esencijalna karakteristika *qualia*, onda imamo razloga smatrati da one ne postoje.

Međutim, možda ova kritika pojma *qualia* pogrešno pretpostavlja mogućnost razlikovanja između *qualia* i mogućnosti imanja reaktivnih stavova prema njima. Štoviše, prethodna razmatranja pretpostavljaju shvaćanje pojma *qualia* koje je u potpunosti odvojeno od psiholoških reakcija koje osoba može imati prema njima. Vidjeli smo da ono što razlikuje Amaliju od Helene jest ideja da u prvom slučaju *quale* kave ostaje nepromijenjena, dok se mijenja stav sviđanja prema njoj. U drugom pak

slučaju stav sviđanja ostaje isti, a mijenja se *quale* kave. Pristaše *qualia* bi mogli tvrditi da ova razlika nije legitimna te da se promjenom stava sviđanja prema nekom iskustvu ujedno mijenja *quale* tog iskustva. Kako to na svoj duhovit način iznosi Dennett, Amalijin partner mogao bi reći „ne budi blesava! Jednom kada dodaš stav nesviđanja, promijenit će ti se iskustvo!" – – te nakon malo razmišljanja Amalija zaključuje da je on u pravu" (prilagođeno iz Dennett 1988, 236). U tom slučaju bismo bili opravdani reći da Amalija ipak ima neposredan uvid u svoja iskustva. Naime, budući da Amalija ima introspektivnu spoznaju da joj se Franck kava više ne sviđa, onda zna da je došlo do promijene u fenomenalnom karakteru tog iskustva.

Ako pristaša *qualia* prihvati ovakav način razmišljanja, opet se nalazi u problemima. Naime, ako su naše psihološke reakcije i stavovi konstitutivni za imanje *qualia*, slijedi da one ne mogu biti *intrinzična* svojstva iskustva. Ispada da su *qualia* ekstrinzična ili relacijska svojstva koja ovise o ljudskim psihološkim reakcijama koje u odnosu na vanjske ili unutrašnje podražaje proizvode određene doživljaje. U tom slučaju, fenomenalne karaktere iskustva mogli bismo opisati kao dispozicijska svojstva naših perceptivnih aparata koji u odnosu na određene podražaje proizvode psihološke ili ponašajne reakcije.

Ova dijalektika dovela nas je do sljedeće situacije. S jedne strane, pristašama *qualia* čini se da postoje privatna iskustva u koja imamo neposredan uvid. S druge strane, čini im se da su *qualia* intrinzična svojstva iskustva. Međutim, Izgleda da prihvaćanje jednog gledišta isključuje drugo. Sada se možemo zapitati kako bismo onda trebali shvatiti situaciju s Amalijom? Trebamo li je shvatiti kao da se radi o tome da je *quale* ostala ista dok se promijenio stav prema njemu ili se zajedno sa stavom promijenila i *quale*? S obzirom na ovaj konflikt u intuicijama nije jasno kako bi se pristaše *qualia* trebali opredijeliti. Štoviše, Dennett ističe da forsiranje odgovora na ovo pitanje upravo pokazuje da nema pravog odgovora:

> Ako priznate da odgovor nije očigledan, a pogotovo ako se požalite da ovaj prisilni izbor razdvaja dva aspekta za koje ste pretpostavljali da su sjedinjeni u vašem predteorijskom pojmu, podržavate moju tvrdnju da u svakodnevnoj „pučkoj psihologiji" nema sigurnog temelja za pojam *qualia*. Obično razmišljamo na zbunjen i potencijalno nekoherentan način kada razmišljamo o načinima na koje nam se stvari čine. (Dennett 1988, 237)

Slučaj Amalije sugerira nam da nije moguće da postoje svojstva koja bi istodobno bila intrinzična, privatna i neposredno dostupna. Stoga, Dennett zaključuje da je naš intuitivni pojam *qualia* nekoherentan i kao takav ne označuje pojavu vrijednu istraživanja. Prema Dennettu definitivno je da

svjesna iskustva postoje, no ona nemaju svojstva poput *qualia.* Ako netko misli da *qualia* postoje, onda pati od svojevrsne kognitivne iluzije (Frankish 2016). Prema Dennettu, nema previše smisla pokušati revidirati taj pojam zbog ontološkog tereta koji sa sobom nosi. Stoga kao jedino suvislo rješenje vidi odustajanje od korištenja termina „qualia" u istraživanju svijesti.

Ako je stvarno tako da *qualia* ne postoje, ima smisla tvrditi da su svojstva koja se manifestiraju aktivacijom psihološko-ponašajnih dispozicija jedino što postoji kada govorimo o svjesnim iskustvima. To su upravo svojstva koja karakteriziraju neke neurofiziološke sustave te se kao takva mogu istraživati koristeći znanstvene metode. Nadalje, ako se prihvati ovo gledište, onda je moguće da teški problem svijesti počiva na nekoherentnom pojmu. To bi značilo da su zapravo svi problemi s istraživanjem svijesti u principu ono što Chalmers naziva lakim problemima svijesti. Posljedica tog gledišta bi bila da problemi svijesti nisu u suštini filozofski, već predstavljaju fenomene čiju će nam pravu prirodu otkriti psihološke i neurokognitivne znanosti. Naravno, pristaše dualizma i panpsihizma ne slažu se s takvim gledištem te nastoje pokazati zašto zaključci koje sugeriraju Dennett i njegovi sljedbenici nisu valjani. O tome da rasprave među njima ne jenjavaju svjedoče nam brojne publikacije koje se bave prirodom svjesnih iskustava i njihovim mjestom u svijetu kako nam ga otkriva znanost.[71]

9.9 Umjesto zaključka

U raspravi o prirodi svijesti, koju smo započeli u poglavlju 8, razmotrili smo može li se svijest zahvatiti u fizikalističkim terminima. Vidjeli smo da mnogi autori smatraju da postoje intuitivni razlozi za smatrati da korištenje isključivo fizikalističkog vokabulara, teorija i objašnjenja izostavlja važan aspekt nekih mentalnih stanja, a to je ono kako je to doživjeti ta mentalna stanja. U ovom poglavlju raspravu smo započeli eksplanatornim jazom kojim se dodatno nastoji istaknuti taj nedostatak fizikalizma. Koliko god pokušavali smisliti fizikalističke teorije o odnosu svjesnih iskustava i njihovih neuralnih ili općenito fizikalnih podloga nikada nećemo objasniti zašto je neko iskustvo popraćeno određenim fenomenalnim karakterom; na primjer, zašto aktivaciju C i drugih vlakana doživljavamo kao osjećaj boli? Zašto aktivaciju amigdale doživljavamo kao osjećaj straha u nekim situacijama? I tako dalje.

Ukazivanje na eksplanatorna ograničenja fizikalizma dovelo nas je do formulacije teškog problema svijesti. Njime se želi pokazati da za mnoge probleme svijesti, koji se odnose na ljudsku sposobnost reagiranja na podražaje, njihovu individuaciju, kognitivne sposobnosti višeg reda i tome slično, znamo koji će oblik njihova objašnjenja poprimiti, iako nam trenutno

[71] Vidi, na primjer, posebna izdanja časopisa *Journal of Consciousness Studies* (Frankish 2016) i (Goff i Moran 2021).

možda nisu dostupna. Međutim, za objašnjenje svijesti u smislu fenomenalnih karaktera iskustva uopće ne razumijemo, ako se zadržimo u okviru fizikalističkih gledišta, kakvu formu bi njihovo objašnjenje trebalo poprimiti.

Intuitivna privlačnost ovih argumenata u suvremeno doba mnoge autore motivira na odbacivanje fizikalističkih gledišta i povratak klasičnim dualističkim teorijama ili razvoju novih. Štoviše, zbog nezadovoljstva s dualizmom na koji ukazuju fizikalisti, sve više autora okreće se panpsihizmu, prema kojemu i fundamentalni fizički sastojci našeg svijeta posjeduju subjektivnost i nekakvu bazičnu formu osjetilnosti. Panpsihisti smatraju da zaobilaze teški problem svijesti s kojim se susreće fizikalizam jer ne nastoje svjesna iskustva objasniti u terminima nesvjesnih fizikalnih činjenica. Nadalje, smatraju da zaobilaze probleme uzročnosti s kojima se dualizam susreće jer pretpostavljaju da fenomenalni karakteri utemeljuju intrinzične aspekte fizičkih čestica koji se nalaze u podlozi njihovih uzročnih i dispozicijskih svojstava. Međutim, panpsihizam zamjenjuje te probleme svojim specifičnim unutrašnjim tenzijama koje se manifestiraju u problemu kombinacije; kako se točno svjesna iskustva mikrofizičkih čestica mogu kombinirati da proizvedu makrofenomenalna iskustva koja prepoznajemo u svakodnevnom ljudskom životu? Kako se niz neovisnih mikrosubjekata može povezati i stvoriti ljudski makrosubjekt? Kako se povezivanjem niza jednostavnih i neovisnih kvalitativnih iskustava mikrofizičkih čestica mogu stvoriti jedinstvena i iskustva makrosubjekata? Nemogućnost, ili u najmanju ruku trenutni nedostatak, jasnog shvaćanja kako mikročesticama pripisati osjetilne doživljaje i razmišljati o njihovim ontološkim kombinacijama motivira nas da se zapitamo kako smo uopće došli do ove situacije. Koji su nas razlozi doveli do situacije gdje panpsihizam predstavlja uvjerljivo rješenje za problem odnosa uma i tijela? Imamo razloga smatrati da nas do te situacije dovode intuicije koje se nalaze u pozadini teškog problema svijesti. To nas upućuje na potrebu preispitivanja pouzdanosti tih intuicija.

Teški problem svijesti u suštini se temelji na određenom razumijevanju pojma fenomenalnog karaktera svjesnih iskustava. U tom pogledu, neki autori ističu da nešto ne valja s pojmom svijesti koji generira intuicije u pozadini teškog problema svijesti. Vidjeli smo da teški problem svijesti pretpostavlja depsihologizirani pojam fenomenalnog karaktera prema kojem su to svojstva iskustva koja ne ovise o javno dostupnim psihološko-funkcionalnim obilježjima osobe. Takav depsihologizirani pojam fenomenalnog karaktera obično se veže uz pojam *qualia*. One bi trebale biti svojstva osjetilnih iskustava koja su neiskaziva, intrinzična, privatna i neposredno dostupna. Međutim, slično kao što se misaonim eksperimentima nastoji pokazati da postoje fenomenalni karakteri iskustva koji su u potpunosti depsihologizirani, tako se drugim ili istim, ali detaljnije razrađenim misaonim eksperimentima, nastoji pokazati da se taj naizgled

intuitivan pojam kvalitativnog iskustva zapravo temelji na nekoherentnim predteorijskim sudovima o prirodi svjesnih mentalnih stanja. Takvim razmatranjima fizikalistima se otvara prostor za tvrditi da se intuicije u pozadini teškog problema svijesti temelje na kognitivnim iluzijama koje nastaju kao posljedica nedostatnog poznavanja empirijskih činjenica o funkcioniranju ljudskog perceptivnog aparata i ograničenog korištenja mašte. No, je li i koliko je to gledište zaista uvjerljivo, ostavljamo čitateljima da na temelju informiranog razmišljanja sami prosude.

Literatura

Alter, Torin. 1998. „A limited defense of the knowledge argument“. *Philosophical Studies* 90 (1): 35–56. https://doi.org/10.1023/A:1004290020847.

———. 2021. „Knowledge Argument against Physicalism“. The Internet Encyclopedia of Philosophy. 2021. https://iep.utm.edu/know-arg/.

Aranyosi, István. 2013. *The peripheral mind: philosophy of mind and the peripheral nervous system*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199989607.001.0001.

Armstrong, David M. 1968. *A materialist theory of the mind*. London; New York: Routledge.

———. 1995. „The causal theory of the mind“. U *Modern philosophy of mind*, uredio William Lyons, 175–90. London: Everyman.

Ball, Derek. 2009. „There are no phenomenal concepts“. *Mind* 118 (472): 935–62. https://doi.org/10.1093/mind/fzp134.

Barrett, David A. 2013. „Multiple realizability, identity theory, and the gradual reorganization principle“. *The British Journal for the Philosophy of Science* 64 (2): 325–46.

Bechtel, William i Jennifer Mundale. 1999. „Multiple realizability revisited: linking cognitive and neural states“. *Philosophy of Science* 66 (2): 175–207. https://doi.org/10.1086/392683.

Beckermann, Ansgar. 1992. „Introduction. Reductive and nonreductive materialism“. U *Emergence or reduction? Essays on the prospects of nonreductive physicalism*, uredili Ansgar Beckermann, Hans Flohr i Jaegwon Kim, 1–21. Berlin: deGruyter.

Bennett, M. R. i P. M. S. Hacker. 2003. *Philosophical foundations of neuroscience*. Malden, MA: Blackwell Publishing.

Ben-Yami, Hanoch. 2018. „The logical contingency of identity“. *European Journal of Analytic Philosophy* 14 (2): 5–10.

Berčić, Boran. 2002. *Filozofija Bečkog kruga*. Zagreb: KruZak.

———. 2012. *Filozofija*. Sv. 2. Zagreb: Ibis Grafika.

Bermúdez, José Luis. 2014. *Cognitive science: an introduction to the science of the mind*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107279889.

Bermúdez, José Luis i Arnon Cahen. 2020. „Fodor on multiple realizability and nonreductive physicalism: Why the argument does not work“. *Theoria* 35 (1): 59. https://doi.org/10.1387/theoria.20772.

Bigelow, John C. i Robert Pargetter. 1990. „Acquaintance with qualia“. *Theoria* 61 (3): 129–47. https://doi.org/10.1111/j.1755-2567.1990.tb00179.x .

———. 2006. „Re-acquaintance with qualia“. *Australasian Journal of Philosophy* 84 (3): 353–78. https://doi.org/10.1080/00048400600895847.

Biondić, Marin. 2017. „Pučka psihologija: znanstvene perspektive realizma, eliminativizma i instrumentalizma“. *Filozofska istraživanja* 37 (3): 559–78. https://doi.org/10.21464/fi37310.

Blair, R. J. R., D. G. V. Mitchell i Karina Blair. 2008. *Psihopat: emocije i mozak*. Jastrebarsko: Naklada Slap.

Block, Ned. 1978. „Troubles with functionalism“. *Minnesota Studies in the Philosophy of Science* 9: 261–325.

———. 1980a. „Are absent qualia impossible?“ *The Philosophical Review* 89 (2): 257–74. https://doi.org/10.2307/2184650.

———. 1980b. „Introduction: what is functionalism?“ U *Readings in philosophy of psychology*, 171–84. Cambridge Mass.: Harvard University Press.

———. 1995. „On a confusion about a function of consciousness“. *Behavioral and Brain Sciences* 18 (2): 227–47. https://doi.org/10.1017/S0140525X00038188.

———. 1996. „Mental paint and mental latex“. *Philosophical Issues* 7: 19–49. https://doi.org/10.2307/1522889.

Block, Ned i Jerry A. Fodor. 1972. „What psychological states are not“. *Philosophical Review* 81 (2): 159–81. https://doi.org/10.2307/2183991.

Boghossian, Paul A. 1989. „Content and self-knowledge“. *Philosophical Topics* 17 (1): 5–26.

Braddon-Mitchell, David i Frank Jackson. 2007. *The philosophy of mind and cognition: an introduction*. 2. ed,. Oxford: Blackwell.

Brentano, Franz. 1874. *Psychology from an empirical standpoint*. London: Routledge.

Broad, C. D. 1925. *The mind and its place in nature*. Tarner Lectures; 1923. London: Routledge & Kegan Paul.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan i ostali. 2020. „Language models are few-shot learners“. *arXiv:2005.14165 [cs]*. http://arxiv.org/abs/2005.14165.

Brzović, Zdenka. 2018. „Natural kinds“. U *Internet Encyclopedia of Philosophy*. https://www.iep.utm.edu/nat-kind/.

Burge, Tyler. 1979. „Individualism and the mental“. *Midwest Studies in Philosophy* 4 (1): 73–122.

Call, Josep, Gordon M. Burghardt, Irene M. Pepperberg, Charles T. Snowdon i Thomas Zentall, ur. 2017. *APA handbook of comparative psychology: perception, learning, and cognition.* Washington: American Psychological Association. https://doi.org/10.1037/0000012-000.

Carnap, Rudolf. 1931. „Die physikalische Sprache als Universalsprache der Wissenschaft“. *Erkenntnis* 2 (1): 432–65. https://doi.org/10.1007/BF02028172.

———. 1959. „The elimination of metaphysics through logical analysis of language“. U *Logical positivism*, uredio Alfred Jules Ayer. New York: The Free Press.

———. 1995. „Psychology in the language of physics“. U *Modern philosophy of mind*, uredio William Lyons, preveli Niamh Ni Bhleithin, Fionnula Meehan i Daniel Steur. London: Everyman.

Carroll, John W. 2020. „Laws of nature“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/laws-of-nature/.

Carruthers, Peter. 2000. *Phenomenal consciousness: a naturalistic theory*. Cambridge, UK; New York: Cambridge University Press.

Carus, A. W. 2009. *Carnap and twentieth-century thought: explication as enlightenment*. Cambridge: Cambridge University Press.

Cauman, L. 2004. *Uvod u logiku prvog reda*. Zagreb: Naklada Jesenski i Turk.

Chalmers, David J. 1996. *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press.

———. 2010a. *The character of consciousness*. New York: Oxford University Press.

———. 2010b. „The two-dimensional argument against materialism“. U *The character of consciousness*, 141–205. New York: Oxford University Press.

———. 2016a. „Panpsychism and panprotopsychism“. U *Panpsychism*, uredili Godehard Bruntrup i Ludwig Jaskolla, 19–47. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199359943.003.0002.

———. 2016b. „The combination problem for panpsychism“. U *Panpsychism: contemporary perspectives*, uredili Godehard Bruntrup i Ludwig Jaskolla, 179–214. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199359943.003.0008.

Child, William. 1996. *Causality, interpretation and the mind*. Oxford: Oxford University Press.

Chisholm, Roderick M. 1957. *Perceiving: a philosophical study*. Ithaca, N.Y.: Cornell University Press.

Churchland, Patricia S. 1986. *Neurophilosophy: toward a unified science of the mind-brain*. Cambridge, Mass.: MIT Press.

Churchland, Paul M. 1985. „Reduction, qualia and the direct introspection of brain states“. *Journal of Philosophy* 82: 8–28. https://doi.org/10.2307/2026509.

———. 1988. *Matter and consciousness: a contemporary introduction to the philosophy of mind*. Cambridge, Mass: MIT Press.

———. 1993. „Eliminativni materijalizam i propozicijski stavovi“. U *Filozofija psihologije*, uredili Nenad Miščević i Snježana Prijić, 45–63. Rijeka: Hrvatski kulturni dom.

Clark, Austen. 2000. *A theory of sentience*. New York: Oxford University Press.

Clarke, Desmond M. 2005. *Descartes's theory of mind*. Oxford: Clarendon Press.

———. 2006. *Descartes: A biography*. Cambridge: Cambridge University Press.

Cole, David. 2020. „The Chinese room argument“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/chinese-room/.

Conee, Earl. 1994. „Phenomenal knowledge“. *Australasian Journal of Philosophy* 72 (2): 136–50. https://doi.org/10.1080/00048409412345971.

Cottingham, John. 2005. „Cartesian dualism: theology, metaphysics, and science“. U *The cambridge companion to Descartes*, 236–57. Cambridge: Cambridge University Press.

Crane, Tim. 2001. *Elements of mind: an introduction to the philosophy of mind*. Oxford: Oxford University Press.
———. 2005. „Papineau on phenomenal concepts“. *Philosophy and Phenomenological Research* 71 (1): 155–62.
———. 2009. *Elements of mind: an introduction to the philosophy of mind*. Oxford: Oxford University Press.
Craver, Carl F. 2007. *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press, Clarendon Press.
Čuljak, Zvonimir. 2015. *Znanje i epistemičko opravdanje: uvod u epistemologiju*. Zagreb: Ibis Grafika.
Dancy, Jonathan. 2001. *Uvod u suvremenu epistemologiju*. Preveo Zvonimir Čuljak. Zagreb: Hrvatski studiji.
Dasgupta, Shamik. 2014. „The possibility of physicalism“. *The Journal of Philosophy* 111 (9/10): 557–92.
Davidson, Donald. 1970. „Mental events“. U *Essays on actions and events*, uredili L. Foster i J. W. Swanson, 207–24. Oxford: Clarendon Press.
———. 2001a. *Essays on actions and events*. Clarendon: Oxford University Press. https://doi.org/10.1093/0199246270.001.0001.
———. 2001b. „Mental events“. U *Essays on actions and events*, 207–24. Clarendon: Oxford University Press.
Davies, Brian. 2004. *An introduction to the philosophy of religion*. 3. izd. New York: Oxford University Press.
Dennett, Daniel C. 1980. „The milk of human intentionality“. *Behavioral and Brain Sciences* 3 (3): 428–30. https://doi.org/10.1017/s0140525x0000580x.
———. 1981. „True believers: the intentional strategy and why it works“. U *Scientific explanation: papers based on Herbert Spencer lectures given in the University of Oxford*, uredio A. F. Heath, 150–67. Oxford: Clarendon Press.
———. 1988. „Quining qualia“. U *Consciousness in modern science*, uredili A. Marcel i E. Bisiach. Oxford: Oxford University Press, Pretiskano u Chalmers D. (ur.). 2002. Philosophy of mind: contemporary and classical readings, New York: Oxford University Press. str. 226-246.
———. 1991. *Consciousness explained*. Boston: Little, Brown and Co.
———. 1995. „The unimagined preposterousness of zombies“. *Journal of Consciousness Studies* 2 (4): 322–26.
———. 2005. *Sweet dreams: philosophical obstacles to a science of consciousness*. The Jean Nicod Lectures. Cambridge, Mass: MIT Press.
———. 2006. „What Robomary knows“. U *Phenomenal concepts and phenomenal knowledge: new essays on consciousness and*

physicalism, uredili Torin Alter i Sven Walter. New York: Oxford University Press.
———. 2013. *Intuition pumps and other tools for thinking*. New York: W. W. Norton & Company.
Descartes, René. 1951. *Rasprava o metodi*. Preveo Niko Berus. Zagreb: Matica hrvatska.
———. 1985. *The philosophical writings of Descartes. Vol. I*. Preveo John Cottingham, Robert Stoothoff i Dugald Murdoch. Sv. 1. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805042.
———. 1993. *Razmišljanja o prvoj filozofiji*. Preveo Tomislav Ladan. Zagreb: Demetra.
———. 2007. „Descartes to Elisabeth, 21 May 1643, AT III 665–66“. U *The correspondence between Princess Elisabeth of Bohemia and René Descartes*, uredili Elisabeth von der Pfalz i Lisa Shapiro, 63–67. Chicago: University of Chicago Press.
———. 2014. *René Descartes, principia philosophiae. Načela filozofije*. Zagreb: KruZak.
———. 2015. *Meditacije o prvoj filozofiji*. Preveo Josip Talanga. Zagreb: KruZak.
Descartes, René i Princeza Elizabeta od Boemije. 2007. *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. Uredila Elisabeth von der Pfalz. Prevela Lisa Shapiro. Chicago: University of Chicago Press.
Dretske, Fred. 1995. *Naturalizing the mind*. MIT Press.
Elizabeta od Boemije. 2007. „Elisabeth to Descartes, 6 May 1643 (AT III 660)“. U *The correspondence between Princess Elisabeth of Bohemia and René Descartes*, uredili Elisabeth von der Pfalz i Lisa Shapiro, 61–62. Chicago: University of Chicago Press.
Endicott, Ronald P. 1993. „Species-specific properties and more narrow reductive strategies“. *Erkenntnis* 38 (3): 303–21. https://doi.org/10.1007/BF01128233.
Farkas, Katalin. 2009. „Not every feeling is intentional“. *European Journal of Analytic Philosophy* 5 (2): 39–52. https://hrcak.srce.hr/clanak/95134.
Farrell, B. A. 1950. „Experience“. *Mind* 59: 170–98. https://doi.org/10.1093/mind/LIX.234.170.
Feigl, Herbert. 1967. *The mental and the physical: the essay and a postscript*. Minneapolis: University of Minnesota Press. https://muse.jhu.edu/book/32829.

Figdor, Carrie. 2010. „Neuroscience and the multiple realization of cognitive functions“. *Philosophy of Science* 77 (3): 419–56. https://doi.org/10.1086/652964.

Fodor, Jerry A. 1968. *Psychological explanation: an introduction to the philosophy of psychology*. New York: Random House.

———. 1974. „Special sciences (or: the disunity of science as a working hypothesis)“. *Synthese* 28 (2): 97–115. https://doi.org/10.1007/BF00485230.

———. 1997. „Special sciences: still autonomous after all these years“. *Noûs* 31 (S11): 149–63. https://doi.org/10.1111/0029-4624.31.s11.7.

———. 1998. „There are no recognitional concepts, not even red“. *Philosophical Issues* 9: 1–14. https://doi.org/10.2307/1522954.

———. 2001. „Problem duha i tijela“. U *Računala, mozak i ljudski um*, preveli Nenad Miščević i Nenad Smokrović, 2. izd., 63–84. Rijeka: Izdavački centar Rijeka.

Fodor, Jerry A. i Zenon W. Pylyshyn. 1988. „Connectionism and cognitive architecture: a critical analysis“. *Cognition* 28 (1–2): 3–71. https://doi.org/10.1016/0010-0277(88)90031-5.

Foss, Jeff. 1989. „On the logic of what it is like to be a conscious subject“. *Australasian Journal of Philosophy* 67 (2): 305–20. https://doi.org/10.1080/00048408912343771.

Frankish, Keith. 2016. „Illusionism as a theory of consciousness“. *Journal of Consciousness Studies* 23 (11–12): 11–39.

———. 2021. „Panpsychism and the depsychologization of consciousness“. *Aristotelian Society Supplementary Volume* 95 (1): 51–70. https://doi.org/10.1093/arisup/akab012.

Frege, Gottlob. 1995. *Osnove aritmetike i drugi spisi*. Preveli Filip Grgić i Maja Hudoletnjak Grgić. Zagreb: KruZak.

Fumerton, Richard. 2005. „Speckled hens and objects of acquaintance“. *Philosophical Perspectives* 19 (1): 121–38. https://doi.org/10.1111/j.1520-8583.2005.00056.x.

Garber, Daniel. 2001. *Descartes embodied: reading Cartesian philosophy through Cartesian science*. Cambridge: Cambridge University Press.

Gertler, Brie. 1999. „A defense of the knowledge argument“. *Philosophical Studies* 93 (3): 317–36. https://doi.org/10.1023/A:1004216101557.

Gibbard, Allan. 1975. „Contingent identity“. *Journal of Philosophical Logic* 4 (2): 187–222. https://doi.org/10.1007/BF00693273.

Glüer, Kathrin. 2011. *Donald Davidson: a short introduction*. Oxford: Oxford University Press.

Goff, Philip. 2016. „The phenomenal bonding solution to the combination problem“. U *Panpsychism: contemporary perspectives*, uredili Godehard Brüntrup i Ludwig Jaskolla, 283–302. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199359943.003.0012.

———. 2017. *Consciousness and fundamental reality*. New York: Oxford University Press.

———. 2019. *Galileo's error: foundations for a new science of consciousness*. New York: Pantheon Books.

Goff, Philip i Alex Moran. 2021. „Is consciousness everywhere? Essays on panpsychism“. *Journal of Consciousness Studies* 28 (9–10): 9–15.

Goff, Philip, William Seager i Sean Allen-Hermanson. 2020. „Panpsychism“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/panpsychism/.

Graham, George. 2019. „Behaviorism“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2019. Stanford University. https://plato.stanford.edu/archives/spr2019/entries/behaviorism/.

Hacker, P. M. S. 2009. „Philosophy: a contribution, not to human knowledge, but to human understanding“. *Royal Institute of Philosophy Supplement* 65 (listopad): 129–53. https://doi.org/10.1017/S1358246109990087.

Hanlon, Robert T. 2020. *Block by block: the historical and theoretical foundations of thermodynamics*. Oxford: Oxford University Press.

Hanžek, Ljudevit. 2017. „Brentano on self-consciousness“. U *Perspectives on the self*, uredio Boran Berčić, 171–89. Rijeka: University of Rijeka.

Hare, Richard M. 1998. *Jezik morala*. Preveo Filip Grgić. Zagreb: Hrvatski studiji.

Harman, Gilbert H. 1965. „The inference to the best explanation“. *The Philosophical Review* 74 (1): 88–95. https://doi.org/10.2307/2183532.

Hasan, Ali i Richard Fumerton. 2020. „Knowledge by acquaintance vs. description“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/knowledge-acquaindescrip/.

Hatfield, Gary. 2014. *The Routledge guidebook to Descartes' Meditations*. London i New York: Routledge. https://doi.org/10.4324/9781315797878 .

Hellie, Benj. 2004. „Inexpressible truths and the allure of the knowledge argument". U *There's something about Mary*, uredili Yujin Nagasawa, Peter Ludlow i Daniel Stoljar, 333–64. MIT Press.

Hempel, Carl G. 1965. *Aspects of scientific explanation and other essays in the philosophy of science*. New York: The Free Press.

———. 1976. „Empiricist criteria of cognitive significance: problems and changes". U *Can theories be refuted? Essays on the Duhem-Quine thesis*, uredila Sandra G. Harding, 65–85. Synthese Library. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1863-0_3.

———. 1980. „The logical analysis of psychology". U *Readings in philosophy of psychology*, uredio Ned Block, 1–14. Cambridge: Harvard University Press.

Hill, Christopher S. 1991. *Sensations: a defense of type materialism*. Cambridge: Cambridge University Press.

Horgan, Terence E. 1984. „Jackson on physical information and qualia". *Philosophical Quarterly* 34: 147–52. https://doi.org/10.2307/2219508.

———. 1993. „From supervenience to superdupervenience: meeting the demands of a material world". *Mind* 102 (408): 555–86. https://doi.org/10.1093/mind/102.408.555.

de Houwer, Jan i Sean Hughes. 2020. *The psychology of learning: an introduction from a functional-cognitive perspective*. Cambridge, Massachusetts: The MIT Press.

Jackson, Frank. 1982. „Epiphenomenal qualia". *Philosophical Quarterly* 32: 127–36. https://doi.org/10.2307/2960077.

———. 1986. „What Mary didn't know". *Journal of Philosophy* 83 (5): 291–95. https://doi.org/jphil198683566.

———. 1998. „Postscript on qualia". U *Mind, method, and conditionals: selected essays*, 76–79. New York: Routledge.

———. 2003. „Mind and illusion". *Royal Institute of Philosophy Supplement* 53: 251–71. https://doi.org/10.1017/s1358246100008365 .

———. 2006. „The knowledge argument, diaphanousness, representationalism". U *Phenomenal concepts and phenomenal knowledge: new essays on consciousness and physicalism*, uredili Torin Alter i Sven Walter, 52–64. New York: Oxford University Press.

———. 2007. „A priori physicalism“. U *Contemporary debates in philosophy of mind*, uredili Brian P. McLaughlin i Jonathan D. Cohen, 185–99. Blackwell.

Jackson, Frank, Robert Pargetter i Elizabeth W. Prior. 1982. „Functionalism and type-type identity theories“. *Philosophical Studies* 42: 209–25. https://doi.org/10.1007/BF00374035.

Jacquette, Dale. 1995. „The blue banana trick: Dennett on Jackson's color scientist“. *Theoria* 61 (3): 217–30. https://doi.org/10.1111/j.1755-2567.1995.tb00498.x.

James, William. 1995. *The principles of psychology*. Sv. 1. New York: Dover.

Jurjako, Marko. 2020. „Samoobmana, namjere i pučko-psihološko objašnjenje djelovanja“. *Prolegomena* 19 (1): 91–117. https://doi.org/10.26362/20200106.

Jurjako, Marko i Zdenka Brzović. 2021. „Mora li identitet biti nužan?“ *Metodički ogledi: časopis za filozofiju odgoja* 28 (2): 54–76.

Kim, Jaegwon. 1989. „The myth of non-reductive materialism“. *Proceedings and addresses of the American Philosophical Association* 63 (3): 31–47. https://doi.org/10.2307/3130081.

———. 1990. „Supervenience as a philosophical concept“. *Metaphilosophy* 21 (1–2): 1–27. https://doi.org/10.1111/j.1467-9973.1990.tb00830.x.

———. 1993. „Multiple realization and the metaphysics of reduction“. U *Supervenience and mind*, 309–35. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511625220.017.

———. 1994. „Supervenience“. U *A companion to the philosophy of mind*, uredio Samuel Guttenplan, 575–84. Oxford: Blackwell.

———. 1996. *Philosophy of mind*. Cambridge, MA: Westview Press.

———. 1998. *Mind in a physical world: an essay on the mind-body problem and mental causation*. Cambridge, Mass: MIT Press.

———. 2005. *Physicalism, or something near enough*. Princeton, N.J. i Woodstock: Princeton University Press.

———. 2006. *Philosophy of mind*. 2. izd. Cambridge, MA: Westview Press.

Kirk, Robert. 2019. „Zombies“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/zombies/.

Klagge, James C. 1988. „Supervenience: ontological and ascriptive“. *Australasian Journal of Philosophy* 66 (4): 461–70. https://doi.org/10.1080/00048408812343521.

Kovač, Srećko i Berislav Žarnić. 2008. *Logička pitanja i postupci*. Zagreb: KruZak.

Kripke, Saul. 1971. „Identity and necessity“. U *Identity and individuation*, uredio M. K. Munitz, 135–64. New York: New York University Press.
———. 1997. *Imenovanje i nužnost*. Zagreb: KruZak.
Landucci, Sergio. 1997. „Introduzione“. U *Meditazioni metafisiche*, od Renéa Descartesa, preveo Sergio Landucci, V–lX. Roma; Bari: Laterza.
Leibniz, Gottfried Wilhelm. 1980. *Izabrani filozofski spisi*. Preveo Milivoj Mezulić. Zagreb: Naprijed.
Levine, Joseph. 1983. „Materialism and qualia: the explanatory gap“. *Pacific Philosophical Quarterly* 64 (4): 354–61. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x.
———. 2004. *Purple haze*. Oxford: Oxford University Press.
Lewis, David K. 1966. „An argument for the identity theory“: *Journal of Philosophy* 63 (1): 17–25. https://doi.org/10.2307/2024524.
———. 1971. „Counterparts of persons and their bodies“. *Journal of Philosophy* 68 (7): 203–11. https://doi.org/10.2307/2024902.
———. 1972. „Psychophysical and theoretical identifications“. *Australasian Journal of Philosophy* 50 (3): 249–58. https://doi.org/10.1080/00048407212341301.
———. 1988. „What experience teaches“. *Proceedings of the Russellian Society* 13: 29-57. Pretiskano u The Nature of Consciousness: Philosophical Debates, uredili N. Block et al., 579-96. Cambridge, MA: MIT Press.
———. 1994. „Reduction of mind“. U *Companion to the philosophy of mind*, uredio Samuel Guttenplan, 412–31. Blackwell.
Loar, Brian. 1990. „Phenomenal states“. *Philosophical Perspectives* 4: 81–108. https://doi.org/10.2307/2214188.
———. 2004. „Phenomenal states“. U *There's something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*, uredili Peter Ludlow, Yujin Nagasawa i Daniel Stoljar, 219–40. Cambridge, Mass.: MIT Press.
Lowe, Ernest J. 2002. *A survey of metaphysics*. Oxford i New York: Oxford University Press.
Ludlow, Peter, Yujin Nagasawa i Daniel Stoljar, ur. 2004. *There's something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*. Cambridge, Massachusetts: MIT Press.
Lycan, William G. 2019. „Representational theories of consciousness“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Fall 2019. Metaphysics Research Lab, Stanford University.

https://plato.stanford.edu/archives/fall2019/entries/consciousness-representational/.

Machery, Edouard. 2009. *Doing without Concepts*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195306880.001.0001.

Malatesti, Luca. 2004. „Knowing what it is like and knowing how“. *Mind and Causality* 55: 119.

———. 2008. „Mary's scientific knowledge“. *Prolegomena* 7 (1): 37–59.

———. 2012. *The knowledge argument and phenomenal concepts*. Cambridge Scholars Publishing.

———. 2013. „Zombies, the uniformity of nature, and contingent physicalism: a sympathetic response to Boran Berčić“. *Prolegomena* 12 (2): 245–59.

———. 2014. *Filozofija uma: intencionalnost u suvremenim filozofskim raspravama*. Rijeka: Filozofski fakultet u Rijeci. https://www.bib.irb.hr/704041.

Malatesti, Luca, Ana Gavran Miloš i Filip Čeč. 2015. *Filozofsko pisanje bez filozofiranja*. E-Izdanje. Rijeka: Filozofski fakultet u Rijeci. https://www.ffri.uniri.hr/files/izdavacka/L%20Malatesti%20-%20Filozofsko%20pisanje%20bez%20filozofiranja.pdf.

Malatesti, Luca i Nela Malatesti. 2013. „Supervenience, mind and chemistry“. U *Filozofija u dijalogu sa znanostima*, uredili Luka Boršić i Ivana Skuhala Karasman, 253–71. Institut za filozofiju.

Marcus, Eric. 2004. „Why zombies are inconceivable“. *Australasian Journal of Philosophy* 82 (3): 477–90. https://doi.org/10.1080/713659880.

Margolis, Eric, ur. 2000. *Concepts: Core Readings*. 2. izd. A Bradford Book. Cambridge, Mass.: MIT Press.

Marras, Ausonio. 1993. „Supervenience and reducibility: an odd couple“. *The Philosophical Quarterly* 43 (171): 215. https://doi.org/10.2307/2220371.

Maslin, Keith T. 2001. *An introduction to the philosophy of mind*. Cambridge, UK; Malden, MA: Polity.

McGinn, Colin. 1978. „Mental states, natural kinds and psychophysical laws“. *Proceedings of the Aristotelian Society, Supplementary Volumes* 52: 195–236.

McLaughlin, Brian P. 2007. „On the limits of a priori physicalism“. U *Contemporary debates in philosophy of mind*, uredili Brian P. McLaughlin i Jonathan D. Cohen. Blackwell.

McLaughlin, Brian P. i Karen Bennett. 2018. „Supervenience“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University.

https://plato.stanford.edu/archives/spr2018/entries/supervenience/.
McMullen, Carolyn. 1985. „'Knowing what it's like' and the essential indexical". *Philosophical Studies* 48 (rujan): 211–33. https://doi.org/10.1007/BF00356500.
Međedović, Janko. 2015. *Nomološka mreža psihopatije*. Beograd: Institut za kriminološka i sociološka istraživanja.
Mellor, D. H. 1993. „The presidential address: nothing like experience". *Proceedings of the Aristotelian Society* 93 (1): 1–16. https://doi.org/10.1093/aristotelian/93.1.1.
Meyer, Uwe. 2001. „The knowledge argument, abilities, and metalinguistic beliefs". *Erkenntnis* 55 (3): 325–47. https://doi.org/10.1023/A:1013308719802.
Milkov, Nikolay. 2003. *A hundred years of English philosophy*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-0177-8.
Miller, Gregory. 2017. „Forming a positive concept of the phenomenal bonding relation for constitutive panpsychism". *Dialectica* 71 (4): 541–62. https://doi.org/10.1111/1746-8361.12207.
Miščević, Nenad. 1988. *Radnja i objašnjenje*. Zagreb: Hrvatsko filozofsko društvo.
———. 1990. *Uvod u filozofiju psihologije*. Zagreb: Grafički zavod Hrvatske.
Miščević, Nenad i Nenad Smokrović, ur. 2001. *Računala, mozak i ljudski um: zbornik tekstova iz teorije umjetne inteligencije i kognitivne teorije*. 2. izd. Rijeka: Izdavački centar Rijeka.
Moore, George E. 1922. „The conception of intrinsic value". U njegovoj knjizi *Philosophical Studies*, 253–75. New York: Harcourt.
Moran, Richard. 2001. *Authority and estrangement: an essay on self-knowledge*. Princeton, NJ: Princeton University Press.
Nagasawa, Yujin. 2021. „A panpsychist dead end". *Aristotelian Society Supplementary Volume* 95 (1): 25–50. https://doi.org/10.1093/arisup/akab011.
Nagel, Ernest. 1974. *Struktura nauke*. Beograd: Nolit.
———. 1987. *The structure of science: problems in the logic of scientific explanation*. 2. izd. Indianapolis, Ind.: Hackett.
Nagel, Thomas. 1974. „What is it like to be a bat?" *The Philosophical Review* 83 (4): 435–50. https://doi.org/10.2307/2183914.
———. 1979. „Panpsychism". U *Mortal questions*, 181–95. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107341050.
———. 1986. *The view from nowhere*. New York: Oxford University Press.

Nemirow, Laurence. 1990. „Physicalism and the cognitive role of acquaintance“. U *Mind and cognition: a reader*, uredio William G. Lycan, 490–99.

Newell, Allen i Herbert A. Simon. 1976. „Computer science as empirical inquiry: symbols and search“. *ACM Turing Award lecture* 19. https://doi.org/10.1145/360018.360022.

Nida-Rümelin, Martine i Donnchadh O Conaill. 2021. „Qualia: the knowledge argument“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/qualia-knowledge/.

Noonan, Harold W. 1993. „Constitution is identity“. *Mind* 102 (405): 133–46.

Papineau, David. 2002. *Thinking about consciousness*. Oxford: Oxford University Press. https://doi.org/10.1093/0199243824.001.0001.

Peacocke, Christopher. 1983. *Sense and content: experience, thought, and their relations*. New York: Oxford University Press.

Pećnjak, Davor i Tomislav Janović. 2011. „Nefunkcionalnost funkcionalizma“. U *Aspekti uma*, uredili Maja Hudoletnjak Grgić, Filip Grgić i Davor Pećnjak, 1–18. Zagreb: Institut za filozofiju.

———. 2016. *Prema dualizmu: Ogledi iz filozofije uma*. Zagreb: Ibis grafika.

Pećnjak, Davor i Ivan Špiljak. 2014. „Argument iz znanja i modalni argument za dualizam u filozofiji uma“. *Bogoslovska smotra* 84 (1): 73–95.

Pereboom, Derk. 1994. „Bats, brain scientists, and the limitations of introspection“. *Philosophy and Phenomenological Research* 54 (2): 315–29. https://doi.org/10.2307/2108491.

Perry, John. 1979. „The problem of the essential indexical“. *Noûs* 13 (1): 3. https://doi.org/10.2307/2214792.

———. 2001. *Knowledge, possibility, and consciousness*. Cambridge: MIT Press.

Place, Ullin T. 1956. „Is consciousness a brain process“. *British Journal of Psychology* 47 (1): 44–50.

Polger, Thomas W. 2009. „Identity theories“. *Philosophy Compass* 4 (5): 822–34. https://doi.org/10.1111/j.1747-9991.2009.00227.x.

Polger, Thomas W. i Lawrence A. Shapiro. 2016. *The multiple realization book*. New York: Oxford University Press.

Poljak, Dragan, Franjo Sokolić i Mirko Jakić. 2011. „Znanstveno-filozofski aspekti Boškovićeva djela i utjecaj na razvoj klasične i moderne fizike“. *Metodički ogledi: časopis za filozofiju odgoja* 18 (1): 11–34.

Putnam, Hilary. 1965. „Brains and behavior“. U *Analytical philosophy: Second series*, uredio Ronald J. Butler. Oxford: Blackwell.

———. 1975a. „Brains and behavior“. U *Mind, Language and Reality*, 2:325–41. Philosophical papers. Cambridge: Cambridge University Press.

———, ur. 1975b. „How not to talk about meaning“. U *Mind, language and reality*, 117–31. Cambridge: Cambridge University Press.

———. 1975c. „Minds and machines“. U *Mind, language and reality*, 2:362–85. Philosophical papers. Cambridge: Cambridge University Press.

———. 1975d. „Other minds“. U *Mind, language and reality*, 2:342–61. Philosophical papers. Cambridge: Cambridge University Press.

———. 1975e. „Philosophy and our mental life“. U *Mind, language and reality*, 2:291–303. Cambridge: Cambridge University Press.

———. 1975f. „Robots: machines or artificially created life?“ U *Mind, language and reality*, 2:386–407. Philosophical papers. Cambridge: Cambridge University Press.

———. 1975g. „The meaning of ‚meaning‘“. *Mind language and reality Philosophical papers,* sv. 2: 215–71. https://doi.org/10.1088/1367-2630/13/7/073031.

———. 1995. „Priroda mentalnih stanja“. U *Filozofija psihologije: zbornik radova*, uredili Nenad Miščević i Snježana Prijić, prevela Vanda Božičević, 64–73. Rijeka: Izdavački centar.

Pylyshyn, Zenon W. i Zenon W. Pylyshyn. 1984. *Computation and cognition toward a foundation for cognitive science*. Cambridge, Mass.: MIT Press.

Ravenscroft, Ian. 2005. *Philosophy of mind: a beginner's guide*. Oxford and New York: Oxford University Press.

Rescorla, Robert A. i Allan R. Wagner. 1972. „A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement“. *Classical conditioning ii: current research and theory*, izd. 2: 64–99.

Richardson, Robert C. 1979. „Functionalism and reductionism“. *Philosophy of Science* 46 (4): 533–58. https://doi.org/10.1086/288895.

Robinson, Howard M. 1993. „Dennett on the knowledge argument“. *Analysis* 53 (3): 174–77. https://doi.org/10.1093/analys/53.3.169.

Robinson, William S. 2010. „Epiphenomenalism“. *WIREs Cognitive Science* 1 (4): 539–47. https://doi.org/10.1002/wcs.19.

Roesch, Matthew R., Guillem R. Esber, Jian Li, Nathaniel D. Daw i Geoffrey Schoenbaum. 2012. „Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain“. *The European Journal of*

Neuroscience 35 (7): 1190–1200. https://doi.org/10.1111/j.1460-9568.2011.07986.x.

Rowlands, Mark, Joe Lau i Max Deutsch. 2020. „Externalism about the mind“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/content-externalism/.

Rozemond, Marleen. 1998. *Descartes's dualism*. Cambridge, Mass.: Harvard University Press.

Russell, Bertrand. 1911. „Knowledge by acquaintance and knowledge by description“. *Proceedings of the Aristotelian Society* 11: 108–28. https://doi.org/10.1093/aristotelian/11.1.108.

———. 1927. *The analysis of matter*. London: Keegan.

Ryle, Gilbert. 1949. *The concept of mind*. Abingdon, Oxon: Routledge.

———. 2009. „Abstractions“. U njegovoj knjizi *Collected papers*, 2:448–58. London; New York: Routledge.

Salmon, Wesley C. 1989. *Four decades of scientific explanation*. Sv. 13. Minneapolis: University of Minnesota Pres.

Sarıhan, Işık. 2020. „Double vision, phosphenes and afterimages: non-endorsed representations rather than non-representational qualia“. *European Journal of Analytic Philosophy* 16 (1): 5–32. https://doi.org/10.31820/ejap.16.1.1.

Schiffer, Stephen R. 1987. *Remnants of meaning*. MIT Press.

Seager, William. 2016. „Panpsychist infusion“. U *Panpsychism*, uredili Godehard Bruntrup i Ludwig Jaskolla, 229–48. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199359943.003.0010.

Searle, John R. 1980. „Minds, brains, and programs“. *Behavioral and Brain Sciences* 3 (3): 417–24. https://doi.org/10.1017/S0140525X00005756.

———. 2001. „Umovi, mozgovi i programi“. U *Računala, mozak i ljudski um*, uredili Nenad Miščević i Nenad Smokrović, preveo Stipe Grgas, 134–53. Rijeka: Izdavački centar.

Sellars, Wilfrid S. 1956. „Empiricism and the philosophy of mind“. *Minnesota Studies in the Philosophy of Science* 1: 253–329.

Sesardić, Neven. 1984. *Fizikalizam*. Beograd: Istraživačko-izdavački centar SSO Srbije.

Shea, William R. 1991. *The magic of numbers and motion: the scientific career of René Descartes*. 1. izd. Canton, MA: Science History Publ.

Shoemaker, Sydney. 1975. „Functionalism and qualia“. *Philosophical Studies* 27: 291–315. https://doi.org/10.1007/BF01225748.

———. 1981. „Absent qualia are impossible – a reply to Block“. *Philosophical Review* 90: 581–99. https://doi.org/10.2307/2184608.

Siddharth, S. 2021. „Against phenomenal bonding“. *European Journal of Analytic Philosophy* 17 (1): (D1)5-16. https://doi.org/10.31820/ejap.17.1.3.

Skinner, B. F. 1953. *Science and human behavior*. New York: Macmillan.

Smart, J. J. C. 1959. „Sensations and brain processes“. *Philosophical Review* 68: 141–56. https://doi.org/10.2307/2182164. Prevedeno u Smart, Dž. Dž. 1993. „Oseti i moždani procesi“. *Theoria* 2: 79-91. Preveo Aleksandar Gordić.

———. 1963. *Philosophy and scientific realism*. London i New York: Routledge & Kegan Paul.

———. 2017. „The mind/brain identity theory“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/mind-identity/.

Smith, Peter i O. R. Jones. 1988. *The philosophy of mind: an introduction*. Cambridge: University Press.

Spinoza, Baruch de. 2000. *Etika: dokazana geometrijskim redom*. Preveo Ozren Žunec. Zagreb: Demetra.

Stemmer, Nathan. 1989. „Physicalism and the argument from knowledge“. *Australasian Journal of Philosophy* 67 (1): 84–91. https://doi.org/10.1080/00048408912343691.

Stenwall, Robin. 2021. „A grounding physicalist solution to the causal exclusion problem“. *Synthese* 198 (12): 11775–95. https://doi.org/10.1007/s11229-020-02829-3.

Stoljar, Daniel. 2006. *Ignorance and imagination: the epistemic origin of the problem of consciousness*. Oxford: Oxford University Press.

Strawson, Galen. 2008. „Realistic monism: why physicalism entails panpsychism“. U *Real materialism*, 53–74. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199267422.001.0001.

Sušnik, Matej. 2012. „Hjumovska teorija motivacije: u obranu dogme“. *Prolegomena* 11 (1): 83–105.

Tanney, Julia. 2009. „Rethinking Ryle: a critical discussion of the concept of mind“. U *The concept of mind*, od Gilbert Ryle, ix–lvii. Abingdon, Oxon: Routledge.

Teller, Paul. 1985. „Is supervenience just disguised reduction?“ *The Southern Journal of Philosophy* 23 (1): 93–99. https://doi.org/10.1111/j.2041-6962.1985.tb00379.x.

Teymoori, Ali, T. J. Perkins, Viorel Pâslaru, Daniel Cohnitz i Rose Trappes. 2020. „The online alternative: sustainability, justice, and conferencing in philosophy“. *European Journal of Analytic Philosophy* 16 (2): 145–71. https://doi.org/10.31820/ejap.16.2.7.

Tolman, E. C. i C. H. Honzik. 1930. „‚Insight‘ in rats“. *University of California Publications in Psychology* 4: 215–32.

Tononi, Giulio, Melanie Boly, Marcello Massimini i Christof Koch. 2016. „Integrated information theory: from consciousness to its physical substrate“. *Nature Reviews Neuroscience* 17 (7): 450–61. https://doi.org/10.1038/nrn.2016.44.

Turing, Alan M. 1950. „Computing machinery and intelligence“. *Mind* LIX (236): 433–60. https://doi.org/10.1093/mind/LIX.236.433.

Tye, Michael. 1983. „Functionalism and type physicalism“. *Philosophical Studies* 44: 161–74. https://doi.org/10.1007/BF00354097.

———. 1995. *Ten problems of consciousness: a representational theory of the phenomenal mind*. MIT Press.

———. 2000. *Consciousness, color, and content*. MIT Press.

———. 2002. „Representationalism and the transparency of experience“. *Noûs* 36 (1): 137–51. https://doi.org/10.1111/1468-0068.00365.

———. 2007. „New troubles for the qualia freak“. U *Contemporary debates in philosophy of mind*, uredili Brian P. McLaughlin i Jonathan D. Cohen, 303–18. Malden, MA: Blackwell.

———. 2009. *Consciousness revisited: materialism without phenomenal concepts*. The MIT Press.

Van Gulick, Robert. 2018. „Consciousness“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/consciousness/.

Weisberg, Michael, Paul Needham i Robin Hendry. 2019. „Philosophy of chemistry“. U *The Stanford Encyclopedia of Philosophy*, uredio Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/chemistry/.

Weiskopf, Daniel A. i Frederick Adams. 2015. *An introduction to the philosophy of psychology*. Cambridge: Cambridge University Press.

Wilson, Margaret Dauler. 1978. *Descartes*. The arguments of philosophers. London i New York: Routledge, Taylor & Francis Group.

———. 1999. *Descartes*. London: Routledge.
Wittgenstein, Ludwig. 1980. *Remarks on the philosophy of psychology*. Oxford: Basil Blackwell.
Wright, Georg H. von. 1975. *Objašnjenje i razumevanje*. Beograd: Nolit.