ANαliZA

časopis za kritično misel

Letnik 17, številka 1-2, november 2013 · ISSN 1408-2969

KOGNITIVNE ZNANOSTI: ZAVEST, IZKUSTVO IN NEVROFENOMENOLOGIJA

David J. Chalmers

Kaj pomeni soočiti se s problemom zavesti?

prevedel Sebastjan Vörös

Citiranje: David J. Chalmers, »Kaj pomeni soočiti se s problemom zavesti?«, *Analiza: časopis za kritično misel* 17, št. 1-2 (2013): 61–84.

Prevod. Licenca Creative Commons, ki velja za arhiv revije, za prevode ne velja; pravice izvirnika pripadajo njegovim imetnikom.

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

Elektronska različica je bila leta 2026 rekonstruirana iz skena tiskane izdaje (OCR, ročno pregledano) in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

David J. Chalmers
Prevedel Sebastjan Vörös

KAJ POMENI SOOČITI SE S PROBLEMOM ZAVESTI?

1. Uvod

Zavest velja na področju znanosti duha za največjo uganko. Z zavestnim izkustvom smo seznanjeni intimneje kakor s čimerkoli drugim, a ga obenem pojasnimo težje kakor karkoli drugega. V zadnjih letih je znanstvenim raziskavam podleglo ničkoliko duševnih pojavov, le zavest se trmoglavo upira. Videti je, da razlage kljub številnim poskusom vselej zgrešijo tarčo. Nekateri so celo zaključili, da je problem nerešljiv in da ustrezne razlage preprosto ni mogoče najti.

Če se želimo pri problemu zavesti pomakniti naprej, se moramo z njim soočiti na neposreden način. V pričujočem prispevku najprej izpostavim res težek problem, kar storim tako, da ga razločim od obvladljivejših problemov in skušam pojasniti, zakaj nam povzroča toliko preglavic. Nato podam kritiko nekaterih sodobnih pristopov, ki se problema zavesti lotevajo z reduktivnimi metodami, in dokazujem, da so se takšne metode že v načelu nezmožne spoprijeti z najtežjim delom problema. Pripoznanje te nezmožnosti odpira vrata k nadaljnjemu napredku. V drugi polovici članka dokazujem, da nam prestop k novemu tipu nereduktivne razlage omogoča izdelati naturalistično teorijo zavesti, in ponudim tudi svojega lastnega kandidata za takšno razlago: nereduktivno teorijo, osnovano na načelih strukturne koherence, organizacijske nespremenljivosti in hipoteze o dveh vidikih informacije.

2. Lahek problem in težek problem

Ne obstaja en sam problem zavesti. »Zavest« je dvoumen izraz, ki se nanaša na različne pojave. Čeprav je treba pojasniti vsakega od njih, storimo to pri nekaterih lažje kakor pri drugih. Praktično je, če z zavestjo povezane probleme uvodoma razdelimo na »težke« in »lahke«. Za lahke probleme je značilno, da so neposredno dovzetni za standardne metode kognitivne znanosti in jih je potemtakem mogoče pojasniti s komputacijskimi ali živčnimi mehanizmi. Težki problemi se tem metodam upirajo.

Med lahke probleme zavesti uvrščamo tiste, ki se ukvarjajo z naslednjimi vprašanji:

- sposobnost prepoznavanja in razvrščanja okoljskih dražljajev ter odzivanja nanje;

- integriranje informacij v kognitivnem sistemu;

- sporočljivost duševnih stanj;

- sposobnost sistema, da dostopa do lastnih notranjih stanj;

- usmerjanje pozornosti;

- hoteni nadzor nad vedênjem;

- razlika med budnostjo in snom.

Vsi ti pojavi so povezani s pojmom zavesti. Tako denimo duševno stanje včasih označimo kot zavestno, če je verbalno sporočljivo ali notranje dosegljivo. Za sistem včasih pravimo, da se informacije zaveda, če jo je zmožen uporabiti v svojih odzivih, ali natančneje, če je nanjo pozoren oz. jo je zmožen integrirati in jo uporabiti v kompleksnem procesu nadzora vedênja. Dejanje včasih velja za zavestno, če je hoteno, izjava, da je nek organizem zavesten, pa pogosto ne pomeni nič drugega kakor to, da je buden.

Nobene resne dileme ni, ali je mogoče znanstveno pojasniti *te* pojave. Vsi so namreč neposredno dovzetni za komputacijske ali nevrofiziološke razlage. Če želimo npr. pojasniti dostop in sporočljivost, moramo le specificirati mehanizem, ki poskrbi za dostop do informacij o notranjih stanjih in jih naredi razpoložljive za verbalna poročila. Če želimo pojasniti integracijo informacij, moramo zgolj prikazati mehanizme, ki informacije združijo, in jih pripravijo za uporabo v nadaljnjih procesih. Če potrebujemo razlago spanja in budnosti, je dovolj, da navedemo nevrofiziološko razlago procesov, ki so odgovorni za vedênjske razlike v teh dveh nasprotujočih si stanjih. Očitno je, da lahko razlagalno vlogo v vseh teh primerih opravi ustrezen kognitivni ali nevrofiziološki model. Če bi zavest označevala samo navedene pojave, ne bi bila kaj prida problematična. Čeprav se trenutno nismo niti približali njihovi popolni razlagi, imamo vsaj dobro predstavo, kako naj bi potekala. Izraz »lahek« je seveda relativen. Razlaga vseh podrobnosti bo najbrž zahtevala stoletje ali dve zahtevnega empiričnega dela. Toda kljub temu obstajajo dobri razlogi za prepričanje, da bodo metode kognitivne znanosti in nevroznanosti pri tem uspešne.

Zares težek problem zavesti je problem *izkustva*. Kadar mislimo in zaznavamo, ni prisoten le trušč procesiranja informacij, temveč tudi subjektivni vidik. Kot je izpostavil Nagel (1990), obstaja nekaj takega, kot je *biti zavestni organizem*.

Ta subjektivni vidik je izkustvo. Kadar npr. nekaj gledamo, *izkušamo* vidne zaznave: občuteno takšnost rdečosti, izkustvo tem in svetlobe, kakovost globine v vidnem polju. Podobno velja za izkustva, ki se manifestirajo v drugih modalitetah (zvok klarineta, vonj naftalina), telesna občutja (od bolečine do orgazma), v notranjosti nastale duševne podobe, občuteno takšnost čustev in doživljanje toka zavestnih misli. Vsa navedena stanja združuje to, da obstaja nekaj takega, kot je biti v njih – vsa so torej izkustvena.

Nobenega dvoma ni, da so nekateri organizmi deležni izkustev. To, kar nas bega, je, kako je to mogoče. Zakaj imamo, kadar v naših kognitivnih sistemih poteka

vidno in slušno procesiranje informacij, ustrezno vidno ali slušno izkustvo: takšnost temno modre, občutje srednjega C-ja? Kako naj pojasnimo, da obstaja nekaj takega, kot je imeti duševno podobo ali izkušati čustvo? Čeprav obstaja splošni konsenz, da se izkustva porajajo iz fizične osnove, nimamo nobene razlage, zakaj in kako do tega pride. Zakaj naj bi fizično procesiranje informacij vzbudilo bogato notranje življenje? Z objektivnega vidika se zdi to nerazumno, in vendar vidimo, da je tako.

Če je kateri problem upravičen do naziva *najtežji* problem zavesti, je to problem izkustva. Po tem osrednjem pomenu »zavesti« je nek organizem zavesten, če obstaja nekaj takega, kot je biti ta organizem, neko duševno stanje pa je zavestno, če obstaja nekaj takega, kot je biti v tem stanju. Včasih naletimo tudi na izraze, kakršna sta »fenomenalna zavest« ali »kvalije«, toda osebno se mi zdi naravneje govoriti o »zavestnem izkustvu« ali preprosto o »izkustvu«. Drugi učinkovit način, da se ognemo zmedi (uporabljen npr. v Newell, 1990; Chalmers, 1996), je, da izraz »zavest« omejimo na pojave, povezane z izkustvom, za bolj neposredne pojave, opisane zgoraj, pa uporabimo manj obtežen izraz »zavedanje«. Če bi se tega dogovora držali vsi, bi bila komunikacija veliko lažja; udeleženci v trenutnih razpravah o »zavesti« namreč pogosto govorijo drug mimo drugega.

Z dvoumnostjo izraza »zavest« se pogosto okoriščajo tako filozofi kakor znanstveniki, ki o njej pišejo. Tipična razprava o zavesti se nemalokrat začne z zaklinjanjem na njeno skrivnostnost: s čudenjem nad neotipljivostjo in neizrekljivostjo subjektivnosti ter izražanjem zaskrbljenosti nad pomanjkanjem vsakršne teorije o njej. Očitno je, da imamo na tej točki opravka s težkim problemom – problemom izkustva. V drugi polovici razprave postane razpoloženje bolj optimistično in avtor oriše svojo lastno teorijo zavesti. Natančnejša analiza pokaže, da gre za teorijo o enem od obvladljivih pojavov – o sporočljivosti, introspektivnem dostopu ali čem drugem. Avtor zaključi, da zavest vendarle ni neotipljiva, a bralec ima občutek, da so ga potegnili za nos. Težek problem ostaja nedotaknjen.

3. Funkcionalna razlaga

Zakaj so lahki problemi lahki in zakaj je težek problem težek? Lahki problemi so lahki natanko zato, ker se nanašajo na razlago kognitivnih *zmožnosti* in *funkcij*. Če želimo pojasniti kognitivno funkcijo, moramo le specificirati mehanizem, ki jo izvaja. Ker so takšnim razlagam metode kognitivne znanosti pisane na kožo, so pisane na kožo tudi lahkim problemom zavesti. Nasprotno pa je težek problem težek natanko zato, ker se ne nanaša na izvedbo funkcij. Problem bi namreč vztrajal, tudi če bi nam uspelo pojasniti izvedbo vseh relevantnih funkcij. (Izraza »funkcija« tukaj ne uporabljam v ozkem teleološkem pomenu namere, ki naj bi jo tak sistem uresničeval, temveč v širšem pomenu kakršnekoli vzročne vloge, ki jo tak sistem opravlja pri sprožanju vedênja).

Če želimo npr. pojasniti sporočljivost, moramo zgolj pojasniti, kako sistem izvaja funkcijo izdelovanja poročil o notranjih stanjih. Če želimo pojasniti notranji dostop, moramo pojasniti, kako na sistem učinkujejo njegova notranja stanja in

kako z informacijami o teh stanjih usmerja nadaljnje procese. Če želimo pojasniti integracijo in nadzor, moramo pojasniti, kako osrednjim sistemskim procesom uspe združiti informacijske vsebine in jih uporabiti pri izvedbi različnih dejanj. Vse to so problemi, ki se nanašajo na razlago funkcij.

Kako pojasnimo izvedbo neke funkcije? Tako, da specificiramo *mehanizem*, ki to funkcijo izvaja. Nevrofiziološko in kognitivno modeliranje je v teh primerih odlična izbira. Če nas zanima natančna razlaga z nižje ravni, bomo specificirali ustrezni živčni mehanizem; če nas zanima abstraktnejša razlaga, bomo mehanizem specificirali v komputacijskem kontekstu. Najsi izberemo prvo ali drugo možnost, pridemo na koncu do popolne in zadovoljive razlage. Ko enkrat specificiramo živčni ali komputacijski mehanizem, ki npr. opravlja funkcijo verbalnega poročila, je glavni del iskanja razlage sporočljivosti opravljen.

V nekem smislu je vse to trivialno. Ker so omenjeni pojavi *funkcionalno opredeljivi*, je namreč to, da je njihova razlaga odvisna od razlage različnih funkcij, *pojmovno* dejstvo. Trditev, da je sporočljivost uprimerjena v nekem sistemu, ne *pomeni* nič drugega kakor to, da je sistem zmožen izdelovati verbalna poročila o notranjih informacijah. In trditev, da je sistem buden, ne pomeni nič drugega kakor to, da je na ustrezen način dovzeten za informacije iz okolja in je zmožen te informacije na ustrezen način uporabiti za usmerjanje svojega vedênja. Da so tovrstne trditve pojmovna dejstva, je razvidno iz tega, da gre pri izjavi »Pojasnili ste izvedbo funkcije za izdelavo verbalnih poročil, niste pa pojasnili sporočljivosti« za trivialno pojmovno napako glede sporočljivosti. Pojasniti sporočljivost *ne more* pomeniti nič drugega kakor pojasniti, kako je relevantna funkcija izvedena; enako velja tudi za druge pojave, ki jih obravnavamo.

Na ta način delujejo reduktivne razlage v vseh znanostih višjega reda. Če želimo npr. pojasniti gen, moramo specificirati mehanizem, ki shranjuje in prenaša dedne informacije z ene generacije na drugo. Izkaže se, da to funkcijo izvaja DNK – ko smo enkrat pojasnili izvedbo funkcije, smo pojasnili tudi gen. Če želimo pojasniti življenje, moramo v zadnji instanci pojasniti, kako se nek sistem razmnožuje in prilagaja na okolje, kakšno presnovo ima itd. Ker se vsa ta vprašanja nanašajo na izvedbo funkcij, so dovzetna za reduktivno razlago. Podobno velja za večino problemov s področja kognitivne znanosti. Če želimo pojasniti učenje, moramo pojasniti način, kako se vedênjske zmožnosti sistema spreminjajo v luči okoljskih informacij in kako nove informacije vplivajo na prilagajanje delovanja sistema glede na njegovo okolje. Podobne razlage lahko najdemo tudi za druge kognitivne pojave, denimo zaznavo, spomin in jezik. Čeprav zahteva opredeljevanje relevantnih funkcij včasih precejšnjo mero subtilnosti, je jasno, da so razlage kognitivne znanosti uspešne ravno toliko, kolikor temeljijo na razlagi izvedbe funkcij.

Ko v igro vstopijo zavestna izkustva, pa tovrstnim razlagam spodleti. Kar dela težek problem težek in domala edinstven, je to, da *presega* probleme, povezane z izvedbo funkcij. To postane lepo razvidno, če pomislimo, da kljub razlagi izvedbe vseh kognitivnih in vedênjskih funkcij, ki spremljajo neko izkustvo – zaznavno prepoznavanje, razvrščanje, notranji dostop, verbalno poročanje —,

ostaja dodatno vprašanje *»Zakaj izvedbo teh funkcij sploh spremlja izkustvo?«* brez odgovora. Preprosta razlaga funkcij pušča to vprašanje odprto.

Pri razlagi genov, življenja ali učenja analognega dodatnega vprašanja ni. Izjava »Vidim, da si pojasnil, kako DNK shranjuje in prenaša dedne informacije z ene generacije na drugo, toda nisi pojasnil, v kakšnem smislu naj bi bila istovetna *genu«* je pojmovno napačna. Biti gen namreč ne pomeni nič drugega kakor biti neka entiteta, ki opravlja relevantno funkcijo shranjevanja in prenašanja informacij. Izjava »Vidim, da si pojasnil, kako je informacija prepoznana, integrirana in poročana, toda nisi pojasnil, v kakšnem smislu naj bi bila *izkušena«* pa ni pojmovno napačna. Tukaj gre za netrivialno dodatno vprašanje.

To dodatno vprašanje je pri problemu zavesti ključnega pomena. Zakaj vse to procesiranje informacij ne poteka »v temi«, prosto vsakršnega notranjega občutka? Zakaj sta, ko elektromagnetna valovanja trčijo ob mrežnico ter jih vidni sistem prepozna in razvrsti, funkciji prepoznavanja in razvrščanja doživeti kot izkustvo živo rdeče barve? Čeprav vemo, da se izkustvo ob izvedbi funkcij *poraja*, je največja skrivnost, zakaj se poraja. Med funkcijami in izkustvom zeva *razlagalna vrzel* (termin je leta 1983 uvedel Levine), ki jo je mogoče prečkati samo po razlagalnem mostu. Ker ostaja razlaga funkcij na eni strani vrzeli, bo treba material za izdelavo mostu poiskati drugod.

To pa ne pomeni, da izkustvo *nima* nobenih funkcij. Morda se izkaže, da igra pomembno kognitivno vlogo. Toda ne glede na vlogo, ki bi jo lahko imelo, bo razlaga izkustva presegala preprosto razlago njegovih funkcij. Morda se celo izkaže, da bomo med pojasnjevanjem funkcij deležni ključnega uvida, s katerim bomo nato uspeli pojasniti smo izkustvo. Toda v tem primeru bo takšno odkritje le *dodatna* razlagalna nagrada. Za nobeno kognitivno funkcijo namreč ne moremo vnaprej zatrditi, da bo njena razlaga *avtomatično* razložila tudi izkustvo smo.

Za razlago izkustva potrebujemo nov pristop. Običajne razlagalne metode kognitivne znanosti in nevroznanosti niso dovolj. Te metode so se razvile z namenom, da pojasnijo izvedbo kognitivnih funkcij, in pri tem so zelo uspešne. Toda v svoji trenutni postavitvi so primerne *zgolj in samo* za razlago izvedbe funkcij. Ko imamo opravka s težkim problemom, nam standardni pristop ne pove ničesar.

4. Nekaj študij primerov

V zadnjih nekaj letih je nastalo več del, ki skušajo problem zavesti rešiti v okviru kognitivne znanosti in nevroznanosti. To nemara vzbuja vtis, da je zgornja analiza napačna, a jo natančnejši pretres relevantnih študij le dodatno učvrsti. Ko si namreč ogledamo, h katerim vidikom so tovrstne študije usmerjene in katere vidike naposled pojasnijo, ugotovimo, da si za svoj končni cilj vselej izberejo enega od lahkih problemov. Dovolite, da to ponazorim na dveh reprezentativnih primerih.

Prvi je »nevrobiološka teorija zavesti«, ki sta jo očrtala Crick in Koch (1990; glej

tudi Crick, 1994), in se osredotoča na živčna nihanja s frekvenco od 35 do 75 Hz v možganski skorji. Crick in Koch domnevata, da so ta nihanja osnova zavesti, saj v številnih modalitetah – denimo v vidnih in vohalnih sistemih – sovpadajo z zavedanjem, obenem pa ponujajo razlago *povezovanja* informacijskih vsebin. Povezovanje je proces, pri katerem pride do združevanja ločeno reprezentiranih informacij o eni sami entiteti, ki tako postanejo razpoložljive za nadaljnjo obdelavo. En tak primer je denimo integracija informacij o barvi in obliki zaznanega predmeta, ki so nastale v različnih vidnih predelih. Izhajajoč iz dela drugih raziskovalcev (npr. Eckhorn et aL, 1988), sta Crick in Koch postavila hipotezo, da je povezovanje posledica sinhronega nihanja živčnih skupin, ki reprezentirajo relevantne vsebine. Povezovanje dveh informacij bi po tej hipotezi pomenilo, da relevantne živčne skupine nihajo z isto frekvenco in so usklajene.

Čeprav so podrobnosti o povezovalnem procesu zaenkrat slabo poznane, predpostavimo, da nam jih bo sčasoma uspelo razvozlati. Kaj bi takšna teorija pojasnila? Očitno je, da bi pojasnila povezovanje informacijskih vsebin in morda ponudila splošnejšo razlago integracije informacij v možganih. Ker sta Crick in Koch prepričana, da omenjena nihanja aktivirajo mehanizme delovnega spomina, bi nemara pojasnila tudi slednjega in v daljnji prihodnosti še ostale oblike spomina. Morda bi sčasoma prerasla celo v splošno teorijo o tem, kako pride do povezovanja in shranjevanja zaznanih informacij v spominu, kjer čakajo na nadaljnjo obdelavo.

Takšna teorija bi bila sicer dragocena, a nam ne bi povedala ničesar o tem, zakaj relevantne vsebine izkušamo. Crick in Koch menita, da so nihanja živčni *korelati* izkustva. Takšna trditev je vprašljiva – mar do povezovanja ne pride tudi pri procesiranju nezavednih informacij? —, a tudi če jo sprejmemo, ostane še vedno brez odgovora *razlagalno* vprašanje: zakaj nihanja sploh porajajo izkustvo? Edina osnova za razlagalno povezavo je vloga, ki jo nihanje igra pri povezovanju in shranjevanju, pri čemer vprašanje, zakaj povezovanje in shranjevanje sploh spremlja izkustvo, ostaja nedotaknjeno. Če ne vemo, zakaj povezovanje in shranjevanje porajata izkustvo, nam zgodba o nihanju ni v nobeno pomoč. In nasprotno, če bi *vedeli*, zakaj povezovanje in shranjevanje porajata izkustvo, bi bile nevrofiziološke podrobnosti le češnja na torti. Ker teorija Cricka in Kocha svojo prepričljivost črpa iz *predpostavljene* zveze med povezovanjem in izkustvom, ji zveze same ne uspe pojasniti.

Sam za razliko od nekaterih drugih interpretov nisem mnenja, da si Crick in Koch domišljata, da sta pojasnila težki problem. Domet njune teorije je zelo jasno orisan v objavljenem intervjuju s Kochom:

»Najprej je treba pozabiti res težka vprašanja, recimo subjektivne občutke, saj obstaja možnost, da teh po znanstveni poti ne bomo rešili. Subjektivna stanja, povezana z igro, bolečino, ugodjem, modro barvo, vonjem vrtnice – zdi se, da je med materialistično ravnjo, ki pojasnjuje molekule in nevrone, in subjektivno ravnjo ogromen preskok. Osredotočimo se raje na stvari, ki jih lažje preiskujemo, denimo vidno zavedanje. Ta hip se npr. pogovarjate z mano, a ne gledate vame, temveč v svoj kapučino, kar pomeni, da se ga zavedate. Poveste lahko: ‘Pred

mano je skodelica z nekaj tekočine'«. Če vam jo izročim, boste svojo roko premaknili in jo vzeli, tj. odzvali se boste na smiseln način. Temu pravim zavedanje« (»What is Consciousness«, *Discover*, nov. 1992: 96).

Drugi primer je vzet iz kognitivne psihologije. Gre za Baarsovo teorijo zavesti, t.i. teorijo globalnega prostora, ki jo je predstavil v knjigi *A Cognitive Theory of Consciousness*. Po tej teoriji so vsebine zavesti vsebovane v *globalnem delovnem prostoru*, osrednjem procesorju, ki deluje kot nekakšen posrednik v komunikaciji med množico specializiranih nezavednih procesorjev. Ko morajo ti specializirani procesorji neko informacijo dostaviti ostalim delom sistema, to storijo tako, da jo pošljejo v delovni prostor, ki služi kot nekakšna skupinska tabla, dostopna vsem ostalim procesorjem.

Baars v svojem modelu naslavlja različne vidike človeške kognicije in pojasnjuje številna nasprotja med zavestnim in nezavednim kognitivnim delovanjem. Toda v zadnji instanci gre za teorijo *kognitivne dostopnosti*, ki pojasnjuje, na kakšen način so določene informacijske vsebine v sistemu širše dostopne, ter za teorijo informacijske integracije in sporočljivosti. Teorija je obetaven kandidat za teorijo zavedanja, tj. za teorijo funkcionalnega korelata zavestnega izkustva, ne pojasnjuje pa izkustva kot takega.

Morda je videti, da so po Baarsovi teoriji izkustvene vsebine istovetne s vsebinami globalnega delovnega prostora. Toda tudi če bi bilo to res, nam teorija sama po sebi ne *pojasni*, zakaj naj bi bile informacije v globalnem prostoru izkušene. Največ, kar lahko naredi, je, da pove, da so informacije izkušene, ker so *globalno dostopne*. Toda sedaj se vprašanje zastavi v drugačni obliki: zakaj naj bi globalna dostopnost porajala zavestno izkustvo? Kakor vedno ostaja problem premostitve brez odgovora.

Enaki kritiki lahko podvržemo skoraj vse ostale kognitivne ali nevroznanstvene pristope zadnjih nekaj letih. Edelmanov »živčni darvinizem« (1989) se denimo loteva vprašanj zaznavnega zavedanja in pojma jaza, a ne pove ničesar o tem, zakaj je prisotno tudi izkustvo. Dennettov model »mnogoterih osnutkov« (2012) se v glavnem osredotoča na sporočljivost nekaterih duševnih vsebin. Jackendoffova teorija »vmesne ravni« (1988) pa pojasnjuje komputacijske procese, na katerih počiva zavest, a ob tem poudarja, da vprašanje, kako so ti procesi »projicirani« v zavestno izkustvo, ostaja nepojasnjeno.

Čeprav so raziskovalci, ki te metode uporabljajo, glede vprašanja zavestnega izkustva pogosto dvoumni, jih nekaj vendarle zastopa jasno izdelana stališča. A tudi pri teh se mnenja zelo razlikujejo. Če se vprašanja izkustva lotimo s tovrstnimi pristopi, imamo na voljo več različnih strategij. Zelo koristno bi bilo, če bi bile te strateške možnosti večkrat eksplicirane.

Prva strategija je, da preprosto *pojasnimo nekaj drugega*. Nekateri raziskovalci dajejo jasno vedeti, da je problem izkustva zaenkrat pretežek oziroma da morda povsem presega znanstvene zmogljivosti, zato se osredotočajo na enega od oprijemljivejših problemov, kakršna sta sporočljivost ali pojem jaza. Čeprav sem tovrstne probleme imenoval »lahke«, sodijo med najzanimivejše nerešene

probleme v kognitivni znanosti, kar pomeni, da je delo na njih zelo koristno. Slaba stran te strategije je, da je v kontekstu raziskovanja zavesti razmeroma neambiciozna in so njene ugotovitve včasih podvržene napačnim interpretacijam.

Druga možnost je, da uberemo tršo taktiko in *izkustvo zanikamo*. (Variacije na to temo srečamo pri Allportu, 1988; Dennettu, 2012; in Wilkes, 1988). Ta strategija uči, da po tem, ko smo pojasnili funkcije, kakršne so dostopnost, sporočljivost in tako dalje, ni treba iskati še razlage za neki dodaten pojav, imenovan »izkustvo«. Nekateri avtorji so v svojem zanikanju eksplicitni in npr. trdijo, da to, česar ne moremo verificirati z zunanjimi metodami, ne obstaja; spet drugi dosežejo isti učinek s tem, da obstoj izkustva sicer dopuščajo, ampak le, če ga poistovetimo z nečim takim, kot je zmožnost prepoznave in sporočanja.

Ti pristopi resda privedejo do enostavnejše teorije, a so v zadnji instanci nezadovoljivi. Izkustvo je najbolj osrednji in prepoznavni vidik našega duševnega življenja, zato je v znanosti duha nemara ključni eksplanandum. To je glavni razlog, zakaj izkustva ob pojavu nove teorije ne moremo preprosto ovreči, kakor smo ovrgli denimo vitalnega duha. Nasprotno, gre za osrednje dejstvo, ki ga mora pojasniti vsaka teorija zavesti. Teorija, ki izkustvo zanika, »reši« problem tako, da se mu ogne.

Zagovorniki tretje možnosti *trdijo, da so pojasnili zavest v* polnem pomenu. Za razliko od zagovornikov prejšnje možnosti skušajo ti raziskovalci jemati zavest skrajno resno: funkcionalni modeli ali teorije, ki jih ponujajo, naj bi pojasnili subjektivni značaj izkustva v celoti (npr. Flohr, 1992; Humphrey, 1992). Ker pa se ključnega koraka v razlagi komajda dotaknejo, je vse skupaj videti kot nekakšna čarovnija. Ko nanizajo nekaj podrobnosti o procesiranju podatkov, v zgodbo na vsem lepem vstopijo izkustva, ne da bi pojasnili, *zakaj* naj bi jih ti funkcionalni procesi sploh porajali. Nekaterim se to morda zdi samoumevno, toda v tem primeru razlaga ni popolna in imamo torej opravka z različico pete strategije, opisane spodaj.

Četrti, obetavnejši pristop skuša s temi metodami *pojasniti strukturo izkustva*.

Obstajajo denimo dokazi, da lahko z razlago prepoznav, do katerih pride v vidnem sistemu, pojasnimo tako strukturne odnose med različnimi barvnimi izkustvi kakor tudi geometrično strukturo vidnega sistema (glej npr. Clark, 1992 in Hardin, 1992). V splošnem velja, da se določena dejstva o strukturah procesiranja skladajo z dejstvi o strukturi izkustva in jih potemtakem pomagajo pojasniti. Takšna strategija je sprejemljiva, a omejena. V najboljšem primeru jemlje obstoj izkustva za nekaj samoumevnega in uspe pojasniti določena dejstva o njegovi strukturi, tj. ponudi nereduktivno razlago njegovih strukturnih vidikov (več o tem kasneje). To je v več ozirih sicer koristno, a nam ne pove, zakaj do izkustva sploh pride.

Peta in dokaj smotrna strategija je *izolacija substrata izkustva*. Ker se skoraj vsi strinjamo, da so možganski procesi tisti, ki na ta ali oni način *porajajo* izkustva, se zdi smiselno tovrstne procese identificirati. Crick in Koch denimo pravita, da skušata s svojimi raziskavami izolirati živčne korelate zavesti, podobne trditve pa

srečamo tudi pri Edelmanu (1989) in Jackendoffu (1988). Utemeljitev takšnih izjav kliče po nadrobni teoretski analizi – zlasti ker izkustev v eksperimentalnem kontekstu ni mogoče opazovati neposredno toda če smo v svojih analizah previdni, lahko s to strategijo problem izkustva osvetlimo na posreden način. Kljub temu pa je očitno, da je strategija nepopolna. Za zadovoljivo teorijo moramo vedeti več kot le to, *kateri* procesi porajajo izkustvo – potrebujemo tudi razlago, zakaj in kako ga porajajo. Popolna teorija zavesti mora zgraditi razlagalni most.

5. Dodatna sestavina

Videli smo, da obstajajo sistematični razlogi, zakaj običajne metode kognitivne znanosti in nevroznanosti pri razlagi zavestnega izkustva niso uspešne. Gre preprosto za napačne metode: nič od tega, kar nudijo, ne zadostuje za zadovoljivo razlago. Da bi pojasnili zavestno izkustvo, potrebujemo v svoji razlagi *dodatno sestavino*. To pa nas vse, ki jemljemo težki problem zavesti resno, postavlja pred naslednji izziv: predlagati ustrezno dodatno sestavino in utemeljiti, zakaj naj bi prav *ta* pojasnila zavestno izkustvo.

Takšnih sestavin je v obtoku veliko. Če nekateri predlagajo injekcijo kaosa in nelinearne dinamike, so drugi mnenja, da je ključ v nealgoritemskem procesiranju. In če se nekateri zanašajo na prihodnja odkritja v nevrofiziologiji, so drugi spet prepričani, da je rešitev v kvantni mehaniki. Ni težko sprevideti, v čem je čar teh predlogov. Ker so vse stare metode neučinkovite, mora biti rešitev v *nečem* novem. Žal pa vse te predloge pestijo enake tegobe.

Penrose (1989; 1994) je denimo nealgoritemsko procesiranje predlagal predvsem zaradi vloge, ki naj bi jo imelo pri zavestnem matematičnem uvidu. Argumenti o matematiki so sporni, a tudi če bi se izkazali za uspešne in bi nas pripeljali do razlage nealgoritemskega procesiranja v človeških možganih, bi šlo še vedno le za razlago *funkcij*, povezanih z matematičnim sklepanjem itd.

Tako pri nealgoritemskem kakor pri algoritemskem procesu ostaja namreč vprašanje »Zakaj naj bi ta proces porajal izkustvo?« neodgovorjeno; in pri iskanju odgovora na *to* vprašanje nima nealgoritemsko procesiranje nobene privilegirane vloge.

Enako velja za nelinearno in kaotično dinamiko. Lahko nam sicer ponudi svežo razlago dinamike kognitivnega delovanja, ki se bo močno razlikovala od dinamike, dobljene s standardnimi metodami kognitivne znanosti, a pri tem ostaja nesporno, da z dinamiko dobimo le več dinamike. Vprašanje izkustva ni zaradi tega nič manj skrivnostno. To je še toliko bolj očitno pri novih nevrofizioloških odkritjih. Čeprav nam takšna odkritja pomagajo izboljšati razumevanje delovanja možganov, pa vsak živčni proces, ki ga izoliramo, sproži isto vprašanje. Težko si je predstavljati, kaj zagovorniki nove nevrofiziologije mislijo, da bi poleg razlage nadaljnjih kognitivnih funkcij še lahko ugotovili. Saj vendar ne bomo na vsem lepem odkrili fenomenalnega sija v notranjosti živčne celice!

Nemara najbolj priljubljena »dodatna sestavina« je kvantna mehanika (npr. Hameroff, 1994). Razlog za njeno priljubljenost se morda skriva v zakonu o minimizaciji skrivnostnosti: ker sta skrivnostni tako zavest kakor kvantna mehanika, imata morda obe skrivnosti skupen izvor. Kljub temu pa se kvantne teorije zavesti spopadajo z istimi težavami kakor živčne in komputacijske. Kvantni pojavi imajo nekaj nenavadnih funkcionalnih lastnosti, denimo nedeterminiranost in nelokaliziranost. Hipoteza, da so te lastnosti soudeležene pri razlagi kognitivnih funkcij (npr. pri naključni izbiri in integraciji informacij), je smotrna in je ne gre *a prion* izključiti. Toda ko imamo opravka z razlago izkustva, so kvantni procesi v isti barki kakor vsi ostali. Vprašanje, zakaj naj bi ti procesi porajali izkustvo, ostaja v celoti brez odgovora.

(Eden od razlogov za privlačnost kvantnih teorij je ta, da igra zavest po nekaterih interpretacijah kvantne mehanike dejavno vlogo pri »kolapsu« kvantne valovne funkcije. Toda takšne interpretacije ne le, da so sporne, temveč ne vzbujajo prav nobenega upanja, da jim bo s kvantnimi procesi zavest uspelo *pojasniti*. Nasprotno, te teorije obstoj zavesti *predpostavljajo* in z njo pojasnjujejo kvantne procese. Največ, kar nam takšne teorije povedo, je, kakšno fizikalno vlogo bi lahko imela zavest. Ničesar pa nam ne povedo o tem, kako do nje pride).

Enaka kritika velja za *vse* povsem fizikalne razlage zavesti. Ne glede na to, kateri fizični proces specificiramo, bo vprašanje »Zakaj naj bi ta proces porajal izkustvo?« ostalo neodgovorjeno. Trditev, da bi lahko vsak tak proces potekal brez izkustva, je namreč pojmovno povsem koherentna. Iz tega sledi, da nam nobena razlaga izključno fizičnih procesov ne bo povedala, zakaj pride do zavesti. Izkustvo presega vse, kar je mogoče izpeljati iz fizikalne teorije.

Povsem fizikalne razlage so primerne za pojasnjevanje fizičnih *struktur*, tj. za razlago makroskopskih struktur v kontekstu njihovih mikroskopskih sestavnih delov, in za pojasnjevanje izvedbe *funkcij*, tj. za razlago funkcij v kontekstu fizičnih mehanizmov, ki jih izvajajo. To je tako, ker so dejstva o strukturi in funkciji v fizikalni razlagi že *zajeta*: ko smo enkrat razdelali notranje podrobnosti fizikalne razlage, se nam samodejno razjasnijo še strukturne in funkcionalne lastnosti. Toda iz strukture in dinamike fizičnih procesov sledi le več strukture in dinamike, zato tovrstni procesi ne morejo pojasniti ničesar drugega kakor strukturo in dinamiko. Lastnosti izkustva ne morejo biti avtomatična posledica katerekoli fizikalne razlage, saj je trditev, da lahko poljubni fizični proces poteka brez izkustva, pojmovno povsem koherentna. Izkustvo se iz fizičnega substrata *poraja*, a v njem ni *zajeto*.

Nauk vsega tega je, da *razlaga zavesti ne more biti poceni*. Prav osupljivo je, da so se reduktivne metode – metode, ki pojave z višjih ravni pojasnjujejo z osnovnejšimi fizičnimi procesi – izkazale za uspešne na tako številnih področjih. Večino bioloških in kognitivnih pojavov v nekem smislu *je* mogoče pojasniti poceni, namreč v smislu, da so pojmovani kot avtomatične posledice osnovnejših procesov. Bilo bi čudovito, če bi reduktivne metode lahko pojasnile tudi izkustvo – in dalj časa sem upal, da jim bo to uspelo. Žal pa obstajajo sistematični razlogi, zakaj so te metode obsojene na neuspeh. Reduktivne metode so v večini domen

uspešne samo zato, ker je v teh domenah treba pojasniti strukture in funkcije, to pa je nekaj, kar je v fizikalni razlagi zajeto. Toda ko imamo opravka s problemom, ki razlago struktur in funkcij presega, so te metode nemočne.

Morda to zveni podobno kakor vitalistična trditev, da življenja ni mogoče pojasniti z nobeno fizikalno razlago, toda podobnost je le navidezna. Gonilo vitalističnega skepticizma je bil dvom, ali so fizični mehanizmi zmožni opravljati številne osupljive funkcije, ki jih povezujemo z življenjem, denimo kompleksno prilagoditveno vedênje in razmnoževanje. Vitalisti so se implicitno strinjali s pojmovno trditvijo, da je za pojasnitev življenja potrebna funkcionalna razlaga, a so spričo pomanjkljivega poznavanja biokemičnih procesov bili v dvomih, ali je lahko katerikoli fizični proces tej razlagalni nalogi kos, in so zato kot alternativo ponudili hipotezo o vitalnem duhu. Ko se je enkrat izkazalo, da fizični procesi relevantne funkcije lahko opravljajo, so se vitalistični dvomi razblinili.

Po drugi strani pa pri izkustvu ni problematična fizikalna razlaga funkcij, temveč *pojmovna* ugotovitev, da razlaga funkcij za zadovoljivo razlago izkustev preprosto ne zadošča. Na to temeljno pojmovno ugotovitev nadaljnje nevroznanstvene raziskave ne bodo imele nobenega vpliva. V podobnem smislu se izkustvo razlikuje tudi od *član vital*. Ker so vitalisti vitalnega duha postulirali, da bi z njim pojasnili relevantne funkcije, so ga lahko kasneje, ko je biološki znanosti uspelo te funkcije pojasniti brez njega, preprosto zavrgli. Izkustvo pa ni razlagalni postulat, temveč avtonomni eksplanandum, in potemtakem za tovrstno eliminacijo ni primeren kandidat.

Morda bo kdo ugovarjal, da se je doslej že za prenekateri skrivnostni pojav izkazalo, da je razložljiv v fizikalističnem okviru. Toda pri vseh teh problemih je šlo za merljivo vedênje fizičnih predmetov, ki ga je bilo mogoče zvesti na probleme razlage struktur in funkcij. Zato so ti pojavi vseskozi sodili v kategorijo pojavov, ki jih je s fizikalno razlago *mogoče* pojasniti, tudi če so na neki točki obstajali tehtni razlogi, da do takšne razlage ne bo prišlo prav kmalu. To privlačno induktivno sklepanje pa se pri problemu zavesti konča, saj slednji ni odvisen od fizičnih struktur in funkcij. Problem zavesti je skrivnosten na popolne ma drugačen način. Analiza problema pokaže, da zavestnega izkustva s povse reduktivno razlago preprosto ni mogoče pojasniti.

6. Nereduktivna razlaga

Na tej točki je nekatere zamikalo vreči puško v koruzo, češ da je teorija zavestnega izkustva nedosegljiva. McGinn (1989) je denimo mnenja, da je problem zavesti za našega omejenega duha pretežek in da smo ljudje zanj »kognitivno zaprti«. Drugi avtorji dokazujejo, da zavestno izkustvo presega domeno znanstvenih teorij.

Osebno menim, da je tak pesimizem preuranjen. Sedaj ni čas za vdajo, ampak za marljivo nadaljnje delo. Ko je človek izključil vse preproste razlagalne metode, je napočil čas, da se razgleda po alternativah. Ker je reduktivnim razlagam spodletelo, so naravna izbira *nereduktivne* razlage.

Čeprav se je za presenetljivo veliko število pojavov izkazalo, da jih je mogoče pojasniti s preprostejšimi entitetami, pa to ni vselej tako. V fiziki se včasih zgodi, da moramo neko entiteto vzeti za *temeljno*. Temeljnih entitet ni mogoče pojasniti z ničimer preprostejšim, zato jih moramo vzeti za osnovne in izdelati teorijo o tem, v kakšnem odnosu so z drugimi pojavi v svetu. V devetnajstem stoletju se je npr. izkazalo, da elektromagnetnih procesov ni mogoče docela pojasniti z mehanskimi procesi, na katere so se sklicevale predhodne fizikalne teorije, zato so Maxwell in drugi v fizikalno teorijo vpeljali nove temeljne sestavine: elektromagnetni naboj in elektromagnetne silnice. Razlaga elektromagnetizma je zahtevala razširitev ontologije fizike, želja po izgradnji zadovoljive teorije je klicala po novih temeljnih lastnostih in zakonih.

Med lastnosti, ki jih fizikalna teorija pojmuje kot temeljne, sodita tudi masa in prostor-čas. Čeprav ju noben fizik ne skuša pojasniti s čim preprostejšim, iz tega ne sledi, da o njiju ne moremo izdelati nobene teorije. Nasprotno, na voljo imamo zapleteno teorijo o tem, kako sta masa in prostor-čas medsebojno prepletena in kateri temeljni zakoni vladajo med njima. S temi temeljnimi načeli lahko nato pojasnimo številne znane, z maso, prostorom in časom povezane pojave višjega reda.

Po mojem mnenju mora izkustvo v teoriji zavesti zasedati temeljno vlogo. Ker je celotna zgradba fizikalne teorije združljiva z odsotnostjo zavesti, bo teorija zavesti od nas zahtevala, da k ontologiji dodamo *nekaj* temeljnega. Možno je, da bo ta dodatek neka popolnoma nova nefizična lastnost, iz katere bo mogoče izpeljati izkustvo, a si je težko predstavljati, kakšna naj bi ta lastnost bila. Verjetneje je, da bomo poleg mase, naboja in prostor-časa med temeljne lastnosti sveta uvrstili kar izkustvo samo. To bi nam namreč omogočilo, da se lotimo dejanske izdelave teorije izkustva.

Kjer so temeljne lastnosti, tam so tudi temeljni zakoni. Nereduktivna teorija zavesti bo nabor osnovnih zakonov narave obogatila z novimi načeli in ta bodo v zadnji instanci nosila njeno razlagalno breme. Tako kot si pri razlagi za maso značilnih pojavov višjega reda pomagamo z za maso in druge entitete značilnimi osnovnejšimi načeli, si bomo tudi pri razlagi tipičnih izkustvenih pojavov pomagali z za izkustvo in druge entitete značilnimi osnovnejšimi načeli.

Nereduktivna teorija izkustva bo pomembna predvsem zaradi specifikacije osnovnih načel, ki nam bodo povedala, kako je izkustvo odvisno od fizičnih lastnosti sveta. Ta *psihofizična* načela se v fizikalne zakone, ki so najverjetneje povezani v zaprt sistem, ne bodo vmešavala, ampak jih bodo dopolnjevala. Pri fizikalni teoriji gre za teorijo fizikalnih procesov, pri psihofizični teoriji pa za razlago, kako ti procesi porajajo izkustvo. Vemo, da je izkustvo odvisno od fizičnih procesov, a obenem tudi vemo, da te odvisnosti ne moremo izpeljati samo iz fizikalnih zakonov. Nova temeljna načela, ki jih postulira nereduktivna teorija, tako zasedejo mesto dodatne sestavine, ki jo potrebujemo za izgradnjo razlagalnega mostu.

Seveda nam nereduktivni pristop s tem, da izkustvo okliče za temeljno, v nekem smislu ne pove, zakaj do njega sploh pride. Toda to velja za vsako temeljno

teorijo. Nič v fiziki nam npr. ne pove, zakaj sploh obstaja snov, a tega teorijam snovi ne štejemo v slabo. Vsaka znanstvena teorija jemlje nekatere lastnosti sveta za osnovne. Teorija snovi je zmožna vsakovrstna dejstva o snovi pojasniti s tem, da pokaže, kako ta dejstva izhajajo iz osnovnih zakonov. Enako velja za teorijo izkustva.

Ker osnovne lastnosti, ki jih postulira nereduktivna drža, presegajo lastnosti, na katere se sklicuje fizika, je takšno držo mogoče označiti za različico dualizma. Toda gre za nedolžno različico, ki je popolnoma združljiva z znanstvenim pogledom na svet in ni nikjer v navzkrižju s fizikalno teorijo: da bi pojasnili, kako fizični procesi porajajo izkustvo, moramo preprosto dodati nekaj nadaljnjih *premostitvenih* načel. Na tej teoriji ni ničesar pretirano duhovnega ali mističnega – po svoji splošni obliki spominja na fizikalno teorijo, kjer je nekaj temeljnih entitet povezanih s temeljnimi zakoni. Resda nekoliko razširi ontologijo, toda to velja tudi za Maxwellovo teorijo. Splošna struktura takšne drže je popolnoma naturalistična, saj trdi, da univerzum v zadnji instanci sestoji iz mreže osnovnih entitet, ki se pokoravajo preprostim zakonom, in da je s temi zakoni v zadnji instanci mogoče izraziti teorijo zavesti. Če bi ji moral nadeti ime, bi bila dobra izbira *naturalistični dualizem*.

Če je nereduktivna hipoteza pravilna, bo imela teorija zavesti v nekem smislu več skupnega s teorijami v fiziki kakor s teorijami v biologiji. Ker načela, ki nastopajo v bioloških teorijah, niso temeljna v opisanem smislu, se jih drži določena mera kompleksnosti in neurejenosti; teorije v fiziki – vsaj tiste, ki se ukvarjajo s temeljnimi načeli – pa stremijo k preprostosti in elegantnosti. Temeljni zakoni narave so del osnovne garniture sveta, o kateri fizikalne teorije učijo, da je izjemno preprosta. Če so v teoriji zavesti prisotna temeljna načela, lahko pričakujemo, da bodo načela preprostosti, elegantnosti in celo lepote, ki usmerjajo fizikovo iskanje temeljne teorije, veljala tudi zanjo.

(Tehnična opomba: Nekateri filozofi dokazujejo, da med fizičnimi procesi in izkustvom zeva *pojmovna*, a ne nujno metafizična vrzel, kar pomeni, da je izkustvo v nekem smislu vendarle lahko fizično (npr. Hill, 1991; Levine, 1983; Loar, 1990). Ta smer argumentacije je običajno podkrepljena s sklicevanjem na pojem *a posteriori* nujnosti (Kripke, 1980). Toda sam menim, da takšno stališče bodisi počiva na napačnem razumevanju *a posteriori* nujnosti bodisi zahteva novo vrsto nujnosti, za katero nimamo nobenih dobrih razlogov, da bi verjeli vanjo; za podrobnosti glej Chalmers, 1996 (tudi Jackson, 1994 in Lewis, 1994). V vsakem primeru pa takšno stališče še vedno priznava, da fizikalne procese in izkustvo loči *razlagalna* vrzel. Ker načel, ki povezujejo fizično in izkustveno raven, ni mogoče izpeljati iz fizikalnih zakonov, jih moramo jemati kot *razlagalno* temeljna. Iz tega pa sledi, da tudi to stališče priznava, da je razlagalna struktura teorije zavesti v veliki meri takšna, kot sem jo opisal zgoraj.)

7. Očrt teorije zavesti

Ni prezgodaj, da začnemo oblikovati teorijo. Na tej točki smo že zmožni razumeti nekatera ključna dejstva o odnosu med fizičnimi procesi in izkustvi ter

o pravilnostih, ki jih povezujejo. Ko reduktivno razlago postavimo na stran, lahko ta dejstva položimo na mizo in jih uporabimo kot začetne gradnike v nereduktivni teoriji zavesti in kot omejitve osnovnih zakonov, iz katerih bo zgrajena končna teorija.

Izdelavo teorije zavesti pesti očiten problem, namreč pomanjkanje objektivnih podatkov. Ker v eksperimentalnem kontekstu zavestnega izkustva ne moremo opazovati neposredno, je naš dostop do podatkov o odnosu med fizičnimi procesi in izkustvi omejen. Vsi pa imamo dostop do bogatega vira prvoosebnih podatkov in lahko številne pomembne pravilnosti med izkustvi in procesiranjem izpeljemo iz lastnih izkustev. Poleg tega imamo na voljo dobre posredne podatke iz situacij, ki jih je mogoče opazovati, denimo iz verbalnih poročil testirancev. Čeprav imajo vse te metode svoje omejitve, trenutno razpolagamo z več kot dovolj podatki, da lahko teorijo spravimo v pogon.

Filozofska analiza je koristna tudi zato, ker nam pomaga, da iz razpoložljivih podatkov iztržimo čim več: nudi namreč vpogled v številna načela, ki povezujejo zavest in kognicijo, in s tem močno omeji obliko končne teorije. Kot bomo videli, nas pri izdelavi teorije do pomembnih rezultatov privedeta tudi metoda miselnih poskusov in pa – glede na to, da iščemo *temeljno* teorijo – sklicevanje na neempirične omejitve, kot so preprostost, homogenost in podobno. Prizadevati si moramo, da razpoložljive informacije sistematiziramo, jih z natančno analizo čim bolj razširimo in nato na tej podlagi izdelamo najpreprostejšo teorijo, ki bi te informacije pojasnila, a obenem ohranila status sprejemljive kandidatke za sestavni del temeljne garniture sveta.

Ker pri raziskovanju zavesti ni mogoče izvesti dokončnih intersubjektivnih poskusov, bodo tovrstne teorije vselej ohranile spekulativni element, ki v drugih znanstvenih teorijah ni prisoten. Toda to še ne pomeni, da ne moremo izdelati, medsebojno primerjati in ovrednotiti teorij, ki bi se ujemale z razpoložljivimi podatki. Kljub pomanjkanju intersubjektivnih raziskav imamo namreč za njihovo ovrednotenje na voljo več kriterijev: preprostost, notranja koherenca, koherenca s teorijami iz drugih domen, zmožnost reproduciranja izkustvenih lastnosti, ki jih poznamo iz prve roke, in celo splošno ujemanje z zahtevami zdravega razuma. Čeprav bo morda marsikaj ostalo nedorečeno tudi po upoštevanju vseh teh omejitev, nam opisane smernice omogočajo vsaj izdelati sprejemljive teoretske kandidate, kar je edini način, da jih kasneje primerno ovrednotimo.

Nereduktivno teorijo zavesti bodo sestavljala številna *psihofizična načela*, načela, ki bodo lastnosti fizičnih procesov povezovala z lastnostmi izkustva. Razumemo jih lahko kot nekakšne strnjene prikaze tega, kako se izkustvo poraja iz fizičnega substrata, saj nam bodo povedala, v kakšnih fizičnih sistemih se poraja, katere fizične lastnosti so relevantne za njegovo porajanje in katere oblike izkustva bodo v nekem sistemu prisotne. Naloga je sicer zahtevna, a nobenega razloga ni, da se je ne bi lotili.

V nadaljevanju bom predstavil svoje lastne kandidate za psihofizična načela, ki bi utegnila soditi v teorijo zavesti. Prvi dve sta *neosnovni* in ponujata sistematičen prikaz povezav med procesiranjem in izkustvom na razmeroma

visoki ravni. Ti načeli sta sicer pomembni pri razvoju in omejevanju teorije zavesti, a nista izraženi na dovolj temeljni ravni, da bi lahko šteli za resnično osnovna zakona. Na zadnjem mestu zato omenjam svojega kandidata za *osnovno načelo*, tj. za morebitni vogelni kamen temeljne teorije zavesti. To načelo je še posebej spekulativno, vendar menim, da je tovrstna spekulacija pri izgradnji zadovoljive teorije zavesti nepogrešljiva. Vsa tri načela bom na tem mestu predstavil le na kratko; veliko natančneje sem jih utemeljil v Chalmers (1996).

1. Načelo strukturne koherence. Gre za načelo koherence med *strukturo zavesti* in *strukturo zavedanja*. Spomnimo se, da sem prej z izrazom »zavedanje« označeval različne funkcionalne pojave, ki jih povezujemo z zavestjo, sedaj pa bom z njim označeval nekoliko bolj specifičen proces, ki je sestavni del kognitivne osnove izkustva. Natančneje, k vsebinam zavedanja bom prišteval tiste informacijske vsebine, ki so dostopne centralnim sistemom in obširno vplivajo na nadzor vedênja. Ali strnjeno: zavedanje pomeni *neposredno dostopnost za globalni nadzor*. Nekoliko poenostavljeno lahko torej rečemo, da je vsebina zavedanja (vsaj za sisteme, ki uporabljajo jezik) istovetna vsebini, ki je neposredno dostopna in potencialno sporočljiva.

Čeprav je zavedanje povsem funkcionalni pojem, je kljub temu intimno povezan z zavestnim izkustvom. Znani primeri nas učijo, da kjerkoli je zavest, tam je tudi zavedanje. Kjerkoli je prisotno zavestno izkustvo, tam so v kognitivnem sistemu prisotne tudi ustrezne informacije, ki so na voljo za nadzor vedênja in verbalno sporočanje. In nasprotno, zdi se, da je vedno, kadar so informacije dostopne za izdelavo poročil in izvedbo globalnega nadzora, navzoče tudi ustrezno zavestno izkustvo. To pomeni, da sta zavest in zavedanje neposredno usklajena.

To usklajenost pa lahko razvijemo še naprej. Ena od osrednjih značilnosti izkustva je, da ima kompleksno strukturo. Vidno polje ima npr. kompleksno geometrijo. Poleg tega so med posameznimi izkustvi različni odnosi, kot so podob nost-različnost ali razlike v relativni intenzivnosti. To pomeni, da je testirančev: izkustva mogoče vsaj deloma opredeliti s temi strukturnimi lastnostmi (podob nost-različnost, zaznani položaj, relativna intenzivnost, geometrična struktura itd.) in jih razstaviti nanje. Naslednja osrednja značilnost je, da vsaka strukturna lastnost izkustva sovpada z ustrezno lastnostjo informacijsko procesne strukture zavedanja.

Vzemimo npr. barvno zaznavanje. Vsako razliko v barvnem doživljanju spremlja ustrezna razlika v procesiranju. Različne fenomenalne barve, ki jih izkušamo, tvorijo kompleksno tridimenzionalno polje, določeno z razlikami v barvnosti, nasičenosti in intenzivnosti. Lastnosti tega polja nam postanejo razvidne na podlagi analize procesiranja informacij: raziskava vizualnih sistemov pokaže, da poteka prepoznavanje in analiziranje valovnih dolžin svetlobe po treh različnih oseh in da so za kasnejšo obdelavo pomembne ravno te tridimenzionalne informacije. Tridimenzionalna struktura fenomenalnega barvnega prostora se torej neposredno ujema s tridimenzionalno strukturo vidnega zavedanja. To pa je povsem v skladu z našimi pričakovanji. Nenazadnje vsako barvno razlikovanje sovpada z neko sporočljivo informacijo in potemtakem z razliko, reprezentirano

v strukturi procesiranja.

Ali če to povemo v nekoliko bolj strnjeni obliki: geometrična struktura vidnega polja se neposredno odraža v strukturi vidnega procesiranja informacij. Vsak geometričen odnos sovpada z neko informacijo, ki je sporočljiva in jo je zato mogoče kognitivno reprezentirati. Če bi pri nekom poznali zgolj strukturo procesiranja v njegovem vidnem in kognitivnem sistemu, njegovih vidnih izkustev resda ne bi mogli *neposredno* opazovati, bi pa zato lahko rekonstruirali njihove strukturne lastnosti.

V splošnem torej za vsako zavestno izkušeno informacijo velja, da je kognitivno reprezentirana. Natančna struktura vidnega polja se ujema z natančno strukturo vidnega procesiranja. Enako velja tudi za izkustva v drugih modalitetah in celo za nečutna izkustva. Notranje duševne podobe imajo npr. geometrične lastnosti, ki so reprezentirane v procesnih stanjih, pri čustvih pa so prisotne strukturne lastnosti (denimo relativna intenzivnost), ki neposredno sovpadajo s strukturnimi lastnostmi procesiranja: večjo intenzivnost spremlja večji vpliv na nadaljnje procese. V splošnem velja, da so strukturne lastnosti izkustva neposredno reprezentirane v strukturi zavedanja ravno zaradi svoje dostopnosti in sporočljivosti,

Načelo strukturne koherence torej sestoji iz opisanega izomorfizma med strukturama zavesti in zavedanja. Osrednji nauk tega načela je, da čeprav kognitivni procesi dejstev o kognitivnem izkustvu pojmovno ne zajemajo, zavest in kognicija nista nepovezani, temveč vlada med njima intimna koherenca.

Načelo strukturne koherence pa ima svoje omejitve. Čeprav nam omogoča, da na podlagi lastnosti procesiranja informacij rekonstruiramo strukturne lastnosti izkustva, so za slednje značilne tudi druge lastnosti in strukturni opis nekaterih med njimi, denimo intrinzične narave občutja rdeče, ne zajame v celoti.

Dejstvo, da so nam scenariji z obrnjenim barvnim spektrom – scenariji, v katerih sta izkustvi rdeče in zelene obrnjeni, strukturne lastnosti pa ostanejo nespremenjene – razumljivi, priča o tem, da strukturne lastnosti izkustvo omejujejo, a ga ne izčrpajo. Nagibanje k prepričanju, da bi v dveh funkcionalno identičnih sistemih z obrnjenimi izkustvenimi domenami ostale strukturne lastnosti nespremenjene, nazorno prikazuje, kako pomembno vlogo igra načelo strukturne koherence v našem pojmovanju duševnega življenja. Načelo resda ni *logično* nujno – navsezadnje si lahko zamislimo, da procesiranje informacij poteka brez izkustev —, a zato predstavlja močno in dobro znano omejitev psihofizične povezave.

Načelo strukturne koherence je zelo uporabno, saj nam omogoča posredno razlago izkustva s fizikalnimi procesi. Strukturo barvnega prostora lahko npr. posredno pojasnimo z dejstvi o živčnem procesiranju vidnih informacij. In ker z dejstvi o živčnem procesiranju zajamemo in pojasnimo strukturo zavedanja, lahko z njimi – če vzamemo načelo koherence za samoumevno —, pojasnimo tudi strukturo izkustva. Empirične raziskave nam bodo morda celo dale boljši vpogled v strukturo zavedanja pri netopirju in tako posredno osvetlile Naglovo

begajoče vprašanje, kako je biti netopir. To načelo je prisotno v dobršnem delu obstoječih razlag zavesti (npr. Clark, 1992 in Hardin, 1992 o barvah; Akins, 1993 o netopirjih), čeprav se avtorji nanj pogosto nanašajo le posredno. Ker je tako zelo običajno, ga skorajda vsi jemljejo za samoumevno in predstavlja enega od poglavitnih gradnikov v kognitivnih razlagah zavesti.

Koherenca med zavestjo in zavedanjem je prisotna tudi v nevroznanstvenih prizadevanjih za izoliranje *substrata* (ali *živčnega korelata)* zavesti. Predlagane so bile različne hipoteze. Crick in Koch (1990) npr. domnevata, da bi živčni korelat zavesti lahko bila 40-Hz nihanja, medtem ko Libet (1993) stavi na v času razširjeno živčno dejavnost. Če sprejmemo načelo koherence, je najbolj *neposredni* fizični korelat zavesti zavedanje: proces, pri katerem informacije postanejo neposredno razpoložljive za globalni nadzor. Na posamezne hipoteze lahko torej gledamo kot na predloge o tem, kako do zavedanja pride. Crick in Koch denimo menita, da 40-Hz nihanja omogočajo integracijo informacij v delovnem spominu in jih s tem naredijo razpoložljive za kasnejše procese. In tudi za Libetovo v času razširjeno živčno dejavnost je naravno predpostaviti, da je pomembna le zato, ker naj bi omogočala globalno razpoložljivost. Enako velja za druge predlagane korelate, kot so Baarsov »globalni delovni prostor« (1998), Farahine »visokokakovostne reprezentacije« (1994) in Shalliceovi »izbirni vhodni podatki (inputi) v akcijskih sistemih« (1972). Pri vseh teh predlogih gre za hipoteze o *mehanizmih zavedanja*: mehanizmih, katerih funkcija je, da informacijo naredijo neposredno razpoložljivo za globalni nadzor.

Ker vlada med zavestjo in zavedanjem koherenca, bo mehanizem zavedanja obenem tudi korelat zavestnega izkustva. Vprašanje, *kateri* so tisti mehanizmi, ki v možganih usmerjajo globalno razpoložljivost, je empirično – morda jih je več. Toda če sprejmemo načelo koherence, imamo dober razlog za prepričanje, da procesi, ki *pojasnjujejo* zavedanje, obenem tvorijo tudi *osnovo* zavesti.

2. Načelo organizacijske nespremenljivosti. Načelo pravi, da bosta imela sistema z identično *funkcionalno organizacijo* tudi kvalitativno identična izkustva. Če bi vzročno shemo živčne organizacije podvojili v silikonskem substratu, npr. tako, da bi s silikonskimi čipi nadomestili nevrone in vzorce njihovih interakcij, bi dobili identična izkustva. Ali drugače, pri izkustvu ni toliko pomembna specifična fizična zgradba sistema kakor abstraktna shema vzročnih interakcij med njegovimi sestavnimi deli. Načelo organizacijske nespremenljivosti je seveda kontroverzno. Nekateri avtorji (npr. Searle, 1980) so mnenja, da je zavest odvisna od specifičnega biološkega substrata in da zato silikonski izomorf človeka ne bi bil zavesten. Sam pa sem prepričan, da je mogoče to načelo izdatno podkrepiti z analizami miselnih poskusov.

Zelo na kratko: privzemimo (za namene *reductio ad absurdum)*, da je načelo napačno in da torej obstajata dva funkcionalno izomorfna sistema z različnimi izkustvi. Morda je zavesten le eden od njiju ali pa sta zavestna oba, a so njuna izkustva različna. Za lažjo ponazoritev si predstavljajmo, da je prvi sistem sestavljen iz nevronov, drugi pa iz silikona, in da prvi izkuša rdečo, kadar drugi izkuša modro. Ker imata sistema identično organizacijo, si lahko predstavljamo,

da enega postopoma preoblikujemo v drugega, denimo tako, da nevrone v prvem zaporedoma nadomeščamo s silikonskimi čipi z identično lokalno funkcijo. Na ta način dobimo cel spekter vmesnih primerov, od katerih ima vsak identično organizacijo, a nekoliko drugačno fizično sestavo in potemtakem nekoliko drugačna izkustva. Na spektru najdemo tudi sistema *A* in *B*, v katerih smo zamenjali manj kakor desetino prvotne sestave, a se njuna izkustva vendarle razlikujejo. Razen tega, da smo majhen živčni krogotok v A zamenjali s silikonskim krogotokom v *B*, sta sistema fizično identična.

Ključni korak v miselnem poskusu je, da vzročno izomorfni silikonski krogotok namestimo tik ob relevantnem živčnem krogotoku *v A* in ju nato povežemo s stikalom. Kaj se zgodi, ko pritisnemo na stikalo? Po naši hipotezi se bo spremenilo izkustvo sistema iz npr. – za lažjo ponazoritev – rdeče v modro. To je namreč posledica dejstva, da je sistem po spremembi različica *B*, medtem ko je pred njo bil le *A*.

Toda glede na izvorne predpostavke ni nobene možnosti, da bi sistem te spremembe *opazil!* Ker ostane njegova vzročna organizacija nespremenjena, ostanejo nespremenjena tudi vsa njegova funkcionalna stanja in vedênjske dispozicije. S stališča sistema se ni zgodilo nič nenavadnega. Za misel: »Hm, pravkar se je zgodilo nekaj nenavadnega!« preprosto ni prostora. V splošnem namreč velja, da se mora struktura takšne misli odražati v procesiranju, toda videli smo, da struktura procesiranja ostane nespremenjena. Takšna misel bi morala biti popolnoma nepovezana s preostalimi deli sistema in ne bi smela imeti popolnoma nobenega učinka na kasnejše procesiranje. (Če bi na kasnejše procesiranje vplivala, bi se sistema funkcionalno razlikovala, kar pa je v nasprotju z izvorno hipotezo). Na gumb lahko pritisnemo celo večkrat in tako pred »notranjim očesom« sistema pričaramo zaporedno izmenjavo izkustev rdeče in modre. Po naši hipotezi sistem teh »plešočih kvalij« ne bo nikoli opazil.

Po mojem mnenju gre tukaj za *reductio* izvorne predpostavke. Eno od osrednjih dejstev našega doživljanja – dejstvo, ki ga dobro poznamo iz lastnih izkušenj – je, da vselej, kadar smo pozorni in pride do izrazitih sprememb v doživljanju, te spremembe zaznamo: če to ne bi bilo tako, bi pristali pri skeptični možnosti, da nam izkustva vseskozi poplesujejo pred očmi. Ta hipoteza pa ima enak status kakor hipoteza, da je svet nastal pred petimi minutami: četudi logično koherentna, ni plavzibilna. In glede na zelo plavzibilno predpostavko, da spremembe v izkustvu sovpadajo s spremembami v procesiranju, pridemo tako do zaključka, da je izvorna hipoteza nemogoča in da so izkustva dveh funkcionalno izomorfnih sistemov identična. Če se izrazimo še nekoliko bolj tehnično: filozofski hipotezi »odsotnih« in »obrnjenih kvalij« sta mogoči z logičnega, ne pa tudi z empiričnega in nomološkega vidika.

(Morda bo kdo oporekal, da silikonski izomorf živčnega sistema ni mogoč iz tehničnih razlogov. To vprašanje ostaja odprto. Načelo nespremenljivosti pravi le, da *če* je tak izomorf mogoč, bo imel enaka zavestna izkustva.)

Čeprav bi lahko o tem povedali še marsikaj, bo za splošni vtis dovolj. Če ponovimo: zgornji miselni poskus izpelje iz dobro znanih dejstev o koherenci

med zavestjo in kognitivnim procesiranjem močan sklep o odnosu med fizično strukturo in izkustvom. Če argument vzdrži, se izkaže, da v množici fizičnih lastnosti k nastanku izkustva prispevajo le *organizacijske* lastnosti. To pa teorijo zavesti močno dodatno omeji.

3. Teorija o dveh vidikih informacije. Prvi dve načeli sta *neosnovni*. Ker vsebujeta kompleksne pojme, kakršna sta »zavedanje« in »organizacija«, se ne gibljeta na pravi ravni, da bi lahko v teoriji zavesti imeli status temeljnih zakonov, jo pa zato močno omejujeta. Naslednji korak je, da najdemo *osnovna* načela – taka, ki se s temi omejitvami skladajo in nudijo razlago zanje.

Moj predlog osnovnega načela temelji na pojmu *informacije*. Pri tem se v glavnem opiram na Shannonovo teorijo (1948), ki pravi, da kjer je informacija, tam so tudi *informacijska stanja*, vpeta v *informacijski prostor*. Osnovna struktura informacijskega prostora sestoji iz *razlikovalnih* odnosov med njegovimi sestavnimi deli, ki določajo, v čem so si ti medsebojno podobni ali različni. Čeprav je informacijski prostor abstrakten predmet, si po Shannonu lahko informacijo zamišljamo kot *fizično utelešeno*, če obstaja nek prostor z diskretnimi fizičnimi stanji, katerih razlike so določene s specifično vzročno potjo. Tako določena fizična stanja namreč že sama po sebi tvorijo informacijski prostor. Če si sposodim Batesonovo (1972) frazo: fizična informacija je *razlika, ki šteje [difference that makes a difference]*.

Načelo dveh vidikov temelji na opazki, da obstaja med nekaterimi fizično utelešenimi in *fenomenalnimi* (ali izkustvenimi) informacijskimi prostori neposredni izomorfizem. Na podlagi opazk, ki smo jih srečali že pri načelu strukturne koherence, lahko zaključimo, da struktura razlik med fenomenalnimi stanji neposredno sovpada s strukturo razlik med fizičnimi procesi, zlasti tistimi, ki igrajo pomembno vzročno vlogo v procesih globalne razpoložljivosti in nadzora. Povedano drugače, fizično procesiranje in zavestno izkustvo sta vpeta v *enak* abstrakten informacijski prostor.

To pa nas pripelje do naravne hipoteze: informacije (ali vsaj nekatere med njimi) imajo dva vidika, in sicer fizičnega in fenomenalnega. Takšna hipoteza ima status osnovnega načela, ki bi utegnilo utemeljiti in pojasniti, zakaj se izkustv poraja iz fizičnega substrata. Do tega pride, ker izkustvo predstavlja en vidik i: formacije, medtem ko je drugi vidik utelešen v fizičnem procesiranju.

Načelo dveh vidikov podpira cela vrsta premislekov, ki pa jih lahko tukaj skiciram le zelo površno. Prvič, če raziščemo fizične spremembe, ki sovpadajo s spremembami v zavestnem izkustvu, ugotovimo, da so te vselej pomembne zaradi vloge, ki jo opravljajo kot *informacijske spremembe* – spremembe v abstraktnem prostoru stanj, določenih z vzročnimi razlikami znotraj specifičnih vzročnih poti. Drugič, veljavnost načela organizacijske nespremenljivosti je odvisna od obstoja neke temeljne *organizacijske* lastnosti, na katero lahko navežemo izkustvo, in informacija je organizacijska lastnost *par excellence*. Tretjič, načelo dveh vidikov vzbuja upanje, da mu bo s strukturo informacijskih prostorov uspelo pojasniti načelo strukturne koherence. Četrtič, analiza kognitivne razlage naših *sodb* in *trditev* o zavestnem izkustvu – sodb, ki so sicer

funkcionalno pojasnljive, a kljub temu tesno povezane s samim izkustvom – namiguje, da osrednje mesto v razlagi pripada informacijskim stanjem, utelešenim v kognitivnem procesiranju. Iz tega sledi, da v informacijski teoriji zavesti med razlago izkustva ter razlago naših izjav in trditev o njem obstaja izrazita koherenca.

Wheeler (1990) meni, da igra informacija v fiziki univerzuma temeljno vlogo. Po njegovem nauku »biti iz bita« [it from bit] je mogoče fizikalne zakone izraziti v informacijskem besedišču, ki postulira obstoj različnih stanj z različnimi učinki, ne da bi dejansko povedal, kaj ta stanja *so*. Kar namreč šteje, je njihov položaj v informacijskem prostoru. Če je to tako, potem je informacija naravni kandidat za temeljno teorijo zavesti. Po takšnem razumevanju sveta ima informacija resnično temeljno vlogo in dva osnovna vidika, fizičnega in fenomenalnega.

Načelo dveh vidikov je seveda izjemno spekulativno in poddoločeno, zato ostaja veliko ključnih vprašanj brez odgovora. Eno od teh je, ali imajo fenomenalni vidik *vse* informacije. Prva možnost je, da moramo temeljno teorijo dodatno omejiti in pokazati, katera *vrsta* informacij naj bi ga imela. Druga možnost je, da je taka omejitev nepotrebna. Če to drži, potem je izkustvo – glede na to, da so informacije povsod – veliko bolj razširjeno, kakor smo prvotno mislili. Ta zamisel je sprva videti protiintuitivna, a se po skrbnem premisleku izkaže za veliko bolj prepričljivo in elegantno. Kjer imamo opravka s preprostim procesiranjem informacij, so prisotna preprosta izkustva, kjer s kompleksnim procesiranjem informacij, pa kompleksna. Miš, katere struktura procesiranja informacij je preprostejša od človekove, ima posledično tudi preprostejša izkustva; in morda ima termostat s svojo maksimalno preprosto strukturo procesiranja informacij tudi maksimalno preprosta izkustva. Če je izkustvo res temeljna lastnost, bi bilo presenetljivo, če bi se pojavljalo le tu in tam; večina temeljnih lastnosti je razporejenih veliko bolj enakomerno. Nobenega dvoma ni, da gre za odprto vprašanje, toda stališče ni tako nesprejemljivo, kakor se pogosto prikazuje.

Ko je temeljna povezava med informacijo in izkustvom enkrat na mizi, se odpro vrata za daljnosežnejše metafizične spekulacije o naravi sveta. Pogosto npr. slišimo, da fizika svoje temeljne entitete določa le *ekstrinzično*, tj. glede na njihove odnose z drugimi entitetami, ki pa so ravno tako določene ekstrinzično, in tako naprej. Intrinzična narava fizikalnih entitet je dana na stran. Nekateri sicer trdijo, da takšnih intrinzičnih lastnosti ni, toda potem je svet le goli vzročni tok (goli pretok informacij), ki nima nobenih lastnosti, na katere bi se oprla vzročnost. Če dopustimo obstoj intrinzičnih lastnosti, se glede na vse, kar smo povedali doslej, ponuja naravna domneva, da so intrinzične lastnosti fizičnega – lastnosti, na katere se v zadnji instanci opira vzročnost – pravzaprav fenomenalne lastnosti. Z drugimi besedami, fenomenalne lastnosti so notranji vidik informacij. To ponuja tudi odgovor na pomislek o vzročnem pomenu izkustva, pomislek, ki je glede na sprejeto sliko sveta, po kateri je fizična domena vzročno zaprta in ima izkustvena domena v odnosu do fizične le dopolnilno vlogo, povsem legitimen. Informacijski pogled nam pomaga razumeti, da izkustvo, ki je pravzaprav intrinzična narava fizičnega, opravlja pretanjeno vzročno vlogo.

Čeprav je pri izdelavi znanstvene teorije to metafizično domnevo najbolje pustiti ob strani, pa je pri nekaterih filozofskih vprašanjih precej relevantna.

8. Zaključek

Teorija, ki sem jo predstavil, je resda spekulativna, a se ponaša s statusom kandidatke. Po mojem mnenju bosta načeli strukturne koherence in organizacijske nespremenljivosti prisotni v vsaki uspešni teoriji zavesti, medtem ko je prihodnost teorije o dveh vidikih informacij manj gotova. Zaenkrat gre bolj za idejo kakor za teorijo, in če naj kdaj postane razlagalno uspešna, jo bo treba natančneje specificirati in jo predelati v izrazitejšo obliko. Toda premisleki o tem, kaj na njej je sprejemljivega in kaj ne, kje je uspešna in kje ne, lahko privedejo le do boljše teorije.

Večina obstoječih teorij zavest bodisi zanika, pojasni nekaj drugega ali jo označi za večno uganko. Upam, da mi je uspelo pokazati, da se problemu lahko približamo, ne da bi ga zato nehali jemati resno. Za korak naprej bodo potrebne nadaljnje raziskave, bolj pretanjene teorije in natančnejše analize. Težek problem je le še en težek problem in nobenega razloga ni, da bi ostal nepojasnjen za vse večne čase.[*]

*[Argumente iz tega članka sem natančneje predstavil v svoji knjigi *The Conscious Mind* (Oxford University Press, 1996). Francisu Cricku, Peggy DesAutels, Matthewu Eltonu, Liani Gabora, Christofu Kochu, Paulu Rhodesu, Greggu Rosenbergu in Sharon Wahl se zahvaljujem za vse komentarje.]

Nadaljnje branje

Problemi zavesti so bili v sodobni literaturi deležni nemajhne pozornosti. Za nekaj pojmovnih razjasnitev glej Block, 1995; Nelkin, 1993 in Tye, 1995. Za prispevke, ki poudarjajo težave pri pojasnjevanju izkustva v fizičnem kontekstu, glej mdr. Hodgson, 1988; Jackson, 1982; Levine, 1983; Lockwood, 1989; McGinn, 1989; Nagel, 1974; Seager, 1991; Searle, 1991; Strawson, 1994 in Velmans, 1991. O reduktivnih pristopih si lahko preberete več v Churchland, 1995; Clark, 1992; Dennett, 1991; Dretske, 1995, Kirk, 1994; Rosenthal, 1996 in Tye, 1995. Čeprav je bilo poskusov izgradnje natančnejših nereduktivnih teorij v literaturi razmeroma malo, najdete v Hodgson, 1988 in Lockwood, 1989 nekaj zamisli, ki se gibljejo v tej smeri. Dve odlični zbirki sodobnih člankov o zavesti sta Block, Flanagan in Güzeldere, 1996 in Metzinger, 1995.

Literatura

Akins, K. (1993). »What is it like to be boring and myopic?« V Dahlbom, B. (ur.), *Dennett and his Critics*, Oxford: Blackwell.

Allport, A. (1988). »What concept of consciousness?« V Marcel, A. in Bisiach, E. (ur.), *Consciousness in Contemporary Science*, Oxford: Oxford University Press.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge:

Cambridge University Press.

Bateson, G. (1972). *Steps to an Ecology of Mind*. San Francisco: Chandler Publishing.

Block, N. (1995). »On a confusion about the function of consciousness«. *Behavioral and Brain Sciences* 18 (2), str. 227–247.

Block, N., Flanagan, O. in Güzeldere, G. (ur.) (1996). *The Nature of Consciousness: Philosophical and Scientific Debates*. Cambridge, MA: MIT Press.

Chalmers, D. J. (1996). *The Conscious Mind*. New York: Oxford University Press.

Churchland, P. M. (1995). *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA: MIT Press.

Clark, A. (1992). *Sensory Qualities*. Oxford: Oxford University Press.

Crick, F. and Koch, C. (1990). »Toward a neurobiological theory of consciousness«. *Seminars in the Neurosciences* 2, str. 263–275.

Crick, F. (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribners.

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

Dretske, F. I. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Edelman, G. (1989). *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Farah, M. J. (1994). »Visual perception and visual awareness after brain damage: A tutorial overview«. V Umilta, C. in Moscovitch, M. (ur.), *Consciousness and Unconscious Information Processing: Attention and Performance 15*, Cambridge, MA: MIT Press.

Flohr, H. (1992). »Qualia and brain processes«. V Beckermann, A., Flohr, H. in Kim, J. (ur.), *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin: De Gruyter.

Hameroff, S. R. (1994). »Quantum coherence in microtubules: A neural basis for emergent consciousness?«. *Journal of Consciousness Studies* 1 (1), str. 91- 118.

Hardin, C. L. (1992). »Physiology, phenomenology, and Spinoza's true colors«. V Beckermann, A., Flohr, H. in Kim, J. (ur.), *Emergence or Reduction?: Prospects for Nonreductive Physicalism*, Berlin: De Gruyter.

Hill, C. S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.

Hodgson, D. (1988). *The Mind Matters: Consciousness and Choice in a Quantum World*. Oxford: Oxford University Press.

Humphrey, N. (1992). *A History of the Mind*. New York: Simon and Schuster.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Jackson, Frank. (1999). »Epifenomenalne takšnosti«. *Analiza* 2 (3), str. 68-77.

Jackson, F. (1994). »Finding the mind in the natural world«. V Casati, R., Smith, B. in White, S. (ur.), *Philosophy and the Cognitive Sciences*, Vienna: Hölder-Pichler-Tempsky.

Kirk, R. (1994). *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford: Oxford University Press.

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

Levine, J. (1983). »Materialism and qualia: The explanatory gap«. *Pacific Philosophical Quarterly* 64, str. 354-361.

Lewis, D. (1994). »Reduction of mind«. V Guttenplan, S. (ur.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell.

Libet, B. (1993). »The neural time factor in conscious and unconscious events«. V Block, G. R. in Marsh, J. (ur.), *Experimental and Theoretical Studies of Consciousness* (Ciba Foundation Symposium 174), Chichester: John Wiley and Sons.

Loar, B. (1990). »Phenomenal states«. *Philosophical Perspectives* 4, str. 81-108.

Lockwood, M. (1989). *Mind, Brain, and the Quantum*. Oxford: Blackwell.

McGinn, C. (1989). »Can we solve the mind-body problem?«. *Mind* 98, str. 349-366.

Metzinger, T. (1995). *Conscious Experience*. Paderborn: Schöningh.

Nagel, Thomas. (1990). »Kako je biti netopir?«. V Hofstadter, Douglas R. in Dennett, Daniel C. (ur.), *Oko duha: Fantazije in refleksije o jazu in duhu*, Ljubljana: Založba Mladinska knjiga, str. 399–421.

Nelkin, N. (1993). »What is consciousness?«. *Philosophy of Science* 60, str. 419-434.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Rosenthal, D. M. (1996). »A theory of consciousness«. V Block, N., Flanagan, O. in Güzeldere, G. (ur.), *The Nature of Consciousness*, Cambridge, MA: MIT Press.

Seager, W. E. (1991). *Metaphysics of Consciousness*. London: Routledge.

Searle, J. R. (1980). »Minds, brains and programs«. *Behavioral and Brain*

Sciences 3, str. 417-457.

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

Shallice, T. (1972). »Dual functions of consciousness«. *Psychological Review* 79, str. 383-393.

Shannon, C. E. (1948). »A mathematical theory of communication«. *Bell Systems Technical Journal Tl*, str. 379-423.

Strawson, G. (1994). *Mental Reality*. Cambridge, MA: MIT Press.

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

Velmans, M. (1991). »Is human information-processing conscious?«. *Behavioral and Brain Sciences* 14, str. 651-669.

Wheeler, J. A. (1990). »Information, physics, quantum: The search for links«. V Zurek, W. (ur.), *Complexity, Entropy, and the Physics of Information*, Redwood City, CA: Addison-Wesley.

Wilkes, K. V. (1988.), »Yishi, Duh, Um and consciousness«. V Marcel, A. in Bisiach, E. (ur.), *Consciousness in Contemporary Science*, Oxford: Oxford University Press.