ANαliZA

časopis za kritično misel

Letnik 8, številka 3, oktober 2004 · ISSN 1408-2969

DUH IN TELO

Paul M. Churchland, Patricia S. Churchland

Interteoretska redukcija: področni vodnik za nevroznanstvenike

prevedel Sebastjan Vörös

Citiranje: Paul M. Churchland in Patricia S. Churchland, »Interteoretska redukcija: področni vodnik za nevroznanstvenike«, *Analiza: časopis za kritično misel* 8, št. 3 (2004): 60–72.

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 izvirnih stavnih datotek in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

Paul M. Churchland in Patricia S. Churchland
prevedel Sebastjan Vörös

INTERTEORETSKA REDUKCIJA: PODROČNI VODNIK ZA NEVROZNANSTVENIKE*

"Redukcionizem" je znotraj akademskih krogov sporen termin. Nekaterim se zdi, da označuje pravilen pristop k vsem znanstvenim področjem, saj stremi k interteoretski enotnosti in pravi sistematičnosti opisa pojavov. Šlo naj bi torej za pristop, ki ga velja aktivno izvajati in vneto zagovarjati.

Drugim pa se zdi, da označuje napačno naravnan pristop, ki je ozkogleden in slep za bogastvo pojavov. Je neumen primer "ničkotstva"****, ki je slepo za emergentno kompleksnost in organizacije višjega reda. Gre za pristop, ki se ga velja ogibati.

Slednji ugovor največkrat najdemo pri različnih družbenih znanostih, kot so, na primer, antropologija, sociologija in psihologija. Prvo stališče pa je največkrat prisotno znotraj naravoslovnih znanosti, kot so, recimo, fizika, kemija in molekularna biologija. Povsem pričakovano je torej dejstvo, da je problem redukcionizma najbolj pereč tam, kjer se ta dva intelektualna tokova srečata, torej v moderni nevroznanosti.

Sporno vprašanje je, če lahko upravičeno pričakujemo in stremimo k redukciji vseh psiholoških pojavov na nevrobiološke in nevroračunske [nevro*computational]* pojave. Velik in še vedno ugleden del akademske skupnosti ostaja pri zavračanju te možnosti. Njihovo zavračanje je načelno. Nekateri opozarjajo na obstoj tega, čemur filozofi pravijo takšnosti [qualia], tj. raznovrstne subjektivne kakovosti, ki se nam kažejo v naših čutih: pomislimo na bolečino, vonj rože, občutek rdečosti itd. Za te takšnosti velja, da presegajo vsakršno zmožnost materialistične obrazlage ali redukcije (Jackson, 1982; Nagel, 1974). Drugi opozarjajo na semantično vsebino oziroma *intencionalnost* naših misli, ki naj bi ravno tako bila izven dosega redukcije (Popper in Eccles, 1978; Searle, 1980, 1990). Spet nekateri trdijo, da lahko najbolj pomembne aspekte človeškega obnašanja razložimo zgolj s pomočjo *emergentnih lastnosti* in njim lastnih zakonitosti, tj. s pomočjo lastnosti, ki ireducibilno vključujejo socialni nivo, kot so, recimo, vdanost kaki moralni ideji, dojemanje nekega političnega dejstva ali pre-poznava osebne izdaje (Taylor, 1970, 1987). Spet drugi vidijo

* Prevedeno po Paul M. Churchland in Patricia S. Churchland "Intertheoretic Reduction: A Neuroscientist's Field Guide," V *Philosophy and the Neurosciences* (ur. Bechtel, W., Mandik, P., Mundalc, J. Stufflebcam. R.), 2001, Oxford: Blackwell, str. 419-430. – op. prev.

** Skovanko iz "nothing but" – tj. "nič (drugega) kot" – "nothing-butery" prevajam s slovensko skovanko "ničkotstvo". – op. prev.

navzkrižje s pomembno in globoko utrjeno idejo o *človeški svobodi* (Popper in Eccles, 1978). Končno, nekateri materialisti izpostavljajo problem *večvrstne realizacije*. Opozarjajo namreč na dozdevno dejstvo, da lahko zavestno inteligenco vzdržujejo fizični sistemi, ki so drugačni od človeku lastne biokemije, npr. sistem tranzistorjev, ravno tako kot lahko nacionalno finančno ekonomijo vzdržujejo žetoni, ki so drugačni od trenutno uporabljanih kovancev in bankovcev. Vendar nihče med nami ne verjame, da lahko makroekonomijo reduciramo na kemijo kovin in papirja. Kako lahko torej predpostavljamo, da je mogoče psihologijo zreducirati na nevrobiologijo človeka? (Fodor, 1975)

Namen tega članka je trojen. Prvič, skušala bova predstaviti uporaben pregled splošne narave interteoretske redukcije, pri čemer se bova opirala na številne primere iz zgodovine znanosti. Tukaj je bistvenega pomena, da razširimo svoja obzorja, kajti namesto, da nepretrgoma zremo v problem, ki nam trenutno buri duhove (možnost redukcije psiholoških fenomenov na nevralne fenomene), si raje oglejmo primere, kjer se je prah spora že polegel, saj nas taki primeri lahko marsikaj naučijo. Drugič, pri teh primerih bova dognala, kaj so dejanske vrednosti, ki jo ti primeri kažejo, in kaj slabosti, ki se jih velja ogibati. Na koncu bova skušala uporabiti te zgodovinske lekcije na našem primeru – kognitivna nevroznanost – skušala z njihovo pomočjo spodbiti prej omenjene ugovore in pomisleke.

1. Interteoretska redukcija: nekaj prototipskih primerov

Ker so posamični zgledi najboljši učitelji, predlagava, da si jih na tem mestu nekaj tudi ogledamo. Eden najzgodnejših primerov obsežne interteoretske redukcije je bila redukcija Keplerjevih treh zakonov o gibanju astronomskih teles na Newtonovo mehaniko. Keplerjeva teorija je bila omejena zgolj na gibanje planetov znotraj našega osončja, medtem ko se je pri Newtonovi teoriji zdelo, da gre za pravilen opis gibanja v splošnem. Zato je bil nedvomno sijajen uspeh, ko je Newtonu uspelo pokazati, da se lahko, ob predpostavki, da je masa katerega koli planeta zanemarljivo majhna v primerjavi z maso sonca, vse tri Keplerjeve zakone izpelje iz njegove lastne teorije.

Keplerjevi trije zakoni za gibanje planetov so:

1. Vsi planeti se gibljejo po elipsah okoli Sonca, pri čemer se le-to nahaja v gorišču elipse.
2. Planet pokrije v enakih časih enake ploščine.
3. Za vsak planet velja, da je kvadrat planetovega obhodnega časa sorazmeren kubu planetove velike polosi.

Newtonovi trije zakoni so:

1. Inercijsko gibanje je konstantno in premočrtno.
2. Pospešek = sila/masa.
3. Medsebojna učinka dveh teles sta si vedno enaka in nasprotno usmerjena.

Izkazalo se je torej, da gre pri Keplerjevi razlagi zgolj za poseben primer oz. posebno aplikacijo bolj splošne Newtonove razlage. Prav tako se je izkazalo, da gre pri gibanju astronomskih teles zgolj za poseben primer splošnih inercijskih gibanj masivnih teles, ki jim gibanje usmerjajo sile. Božanski oz. nadnaravni značaj nebes je bil tako izgubljen za vekomaj. Sublunarna in superlunarna domena sta bili združeni v eno samo domeno, znotraj katere je en sam zbir zakonov urejal obnašanje vseh objektov.

Newtonova mehanika je prav tako zaslužna tudi za drugi veliki primer interteoretske redukcije, do katere pa je prišlo šele v 19. stoletju. Porodilo se je namreč vprašanje, ali lahko taista mehanika, glede na to, da ji je doslej uspelo razložiti gibanje tako na astronomski ravni kot na ravni človeške velikosti, ponudi razlago tudi za gibanje na mikroskopskem nivoju.

Poskusi izgradnje tovrstne razlage so ustvarili še eno združitev, katere rezultati so vključevali tudi nepričakovan dodatek, ki se tiče teorije toplote. Če predpostavimo, da se katera koli plinasta snov sestoji iz velikega števila mrgolečih submikroskopskih delcev, ki znotraj dane posode poskakujejo sem ter tja v skladu z Newtonovimi tremi zakoni, potem lahko izpeljemo zakon za pritisk, ki ga bodo ti delci kolektivno povzročali na stenah posode. Ta "kinetični" zakon ima sledečo obliko:

$PV = 2N/3 \cdot mv^2/2$

Ta zakon je imel enako obliko kot takrat že znani "zakon idealnega plina", namreč

$PV = nRT$

(Pri obeh velja, da P = pritisk, V = prostornina.) Čeprav se zakona razlikujeta v zapisu, pa tako "2N/3" kot "nR" označujeta količino plina v posodi (N – število delcev plina v posodi; n – število molov plina). Edina preostala razlika je ta, da ima "kinetični" zakon izraz za *kinetično energijo povprečnega delca* ($mv^2/2$) na mestu, kjer ima "zakon idealnega plina" izraz za *temperaturo* (T). Mar gre pri fenomenu, ki ga poznamo pod imenom "temperatura", pravzaprav za povprečno kinetično energijo na molekularnem nivoju? Ta presenetljiva konvergenca načel je pripravila Bernoullija, Joula, Kelvina in Boltzmanna, da so na to vprašanje odgovorili pritrdilno. Dodatne raziskave so pokazale, da ima povprečna molekulama kinetična energija *vse* vzročne lastnosti, ki jih je klasična teorija pripisovala temperaturi. Skratka, izkazalo se je, da temperatura ni nič drugega kot povprečna molekularna kinetična energija. Newtonska mehanika je tako slavila še eno redukcijsko zmago. Gibanje na vseh treh ravneh je bilo opisano znotraj ene same teorije, dobro znana pojavna lastnost, *temperatura*, pa je bila razložena in predstavljena na nov ter nekoliko nenavaden način.

Ni nepomembno poudariti, kako je pri tovrstni redukciji šlo za enačenje dobro poznane *fenomenalne* lastnosti vsakdanjih predmetov z zelo nevsakdanjo mikrofizično lastnostjo. (Z izrazom "fenomenalno" mislimo tisto lastnost, ki jo lahko zanesljivo izkustveno prepoznamo, vendar pa s sklicevanjem na preprostejše izkustveno prepoznavne elemente ne moremo razložiti, na kakšen

način pravzaprav prepoznavamo to lastnost.) Od tod sledi, da redukcija ni omejena zgolj na pojmovne sisteme, skrite v višavah teoretične stratosfere. Včasih se izkaže, da gre pri pojmovnem sistemu, ki ga vključimo v neko širšo vizijo, pravzaprav za del našega zdravorazumskega sistema, za del, katerega pojme redno uporabljamo v vzročnih opazovanjih, ki temeljijo na naših vrojenih zaznavnih sistemih. Oglejmo si še nekaj drugih primerov: že pred Newtonom so vedeli, da gre pri *zvoku* pravzaprav za kompresijske valove v atmosferi, pri *višini tona* pa za valovno dolžino; to vedenje je bilo rezultat obsežnejše redukcije zdravorazumskega pojmovanja zvoka in glasbene teorije na mehanično akustiko. Stoletje in pol po Newtonu pa so ugotovili, da so tudi *svetloba* in njene mnogovrstne *barve* enake elektromagnetnim valovom in njihovim različnim valovnim dolžinam; ta ugotovitev je bila rezultat obsežnejše redukcije geometrične optike na elektromagnetno teorijo, kot jo je predstavil Maxwell leta 1864. Tudi *žarečo toploto*, torej še en zdravorazumski pojav, so pričeli pojmovati kot dolgovalovno elektromagnetno valovanje, ko so jo nekoliko kasneje vključili v prej omenjeno teorijo. Očitno je torej, kako dejstvo, da je neka lastnost oz. neko stanje izpostavljeno kateri od naših čutnih zaznav, nikakor ne pomeni, da te lastnosti oz. stanja ne moremo pojmovati drugače, kadar ga vključimo v pojmovni sistem teorije z boljšo opisno močjo.

K temu nadvse pomembnemu dejstvu se bomo vrnili nekoliko kasneje. Za zdaj pa si oglejmo še nekaj primerov interteoretske redukcije. V 20. stoletju je redukcija klasične (valenčne) kemije na atomsko in subatomsko (kvantno) fiziko v 20. stoletju še en nazoren primer pojmovnega poenotenja. Tukaj so struktura atomovih zaporednih elektronskih lupin in lastnosti stabilnih regij pri porazdelitvi elektronov v medatomskih vezeh omogočile sistematično in poučno rekonstrukcijo elektronske strukture mnogih atomskih elementov, klasičnega zakona valenčne porazdelitve in celotne strukture periodnega sistema. Kot se to zelo pogosto pripeti v primeru interteoretske redukcije, je nova teorija lahko razložila prenekatero vprašanje, na katerega stara teorija ni imela ustreznega odgovora, kot, na primer, specifične toplotne kapacitete različnih substanc in interakcijo kemijskih spojin s svetlobo.

Redukcija kemije na fiziko je pomembna tudi zato, ker sam proces še ni zaključen in zelo verjetno tudi nikoli docela ne bo. Eden od razlogov, ki botruje temu dejstvu, se skriva v neskončnem številu možnih kemičnih spojin in njim značilnih kemičnih, mehaničnih, optičnih in termičnih lastnostih. Drugi razlog pa je ta, da je računanje teh raznolikih lastnosti, izhajajoč iz osnovnih principov kvantne mehanike, računsko grozljivo zamudno, tudi če se omejimo zgolj na približne rezultate, k čemur pa smo iz čisto matematičnih razlogov navadno primorani. Potemtakem ni res, da je bilo vso kemijsko znanje uspešno rekonstruirano v pojmih kvantne mehanike. Rekonstruirane so bile le osnove, pa še te zgolj v približkih. Vendar pa nam tukaj naše izkustvo trdno kaže, da je kvantni mehaniki res uspelo uspešno zajeti osnovne elemente kemijske resničnosti. Zato pričakujemo, da bi lahko vsak del kemije približno rekonstruirali s pomočjo kvantnomehanske terminologije, v kolikor bi kaj takega kdaj res bilo potrebno.

Predhodni primeri lepo ponazarjajo, da gre pri interteoretski redukciji v bistvu za odnos med dvema različnima pojmovnima sistemoma, ki opisujeta pojave, ne pa za dve različni domeni pojavov. Končni smisel redukcije je namreč pokazati, da je tisto, za kar smo mislili, da sta dve različni domeni, pravzaprav ena domena, ki pa jo lahko opišemo z dvema (ali večimi) besednjaki.

Brž̌čas najbolj znana redukcija pa je Einsteinova redukcija Newtonovih treh zakonov gibanja na bistveno drugačno mehaniko posebne teorije relativnosti (PTR). PTR je zajela Newtonove zakone na sledeč način: Če (napačno) predpostavimo, da se vsa telesa gibljejo s hitrostjo, ki je bistveno manjša od hitrosti svetlobe, potem PTR vsebuje zbir zakonov za gibanje, ki ga eksperimentalno ne moremo ločiti od Newtonovega starega zbira. Če upoštevamo relativno omejenost človeškega izkustva, tako ni nerazumljivo, da so se zdeli stari Newtonovi zakoni resnični.

Vendar pa so zakoni PTR semantično in logično bistveno drugačni od Newtonovih zakonov, saj svetu pripisujejo znatno drugačne lastnosti, čeprav so v zgoraj omenjenem posebnem primeru eksperimentalno nerazločljivi od Newtonovih zakonov. Tako v vsaki situaciji, kjer Newtonovi zakoni pripišejo telesu neko notranjo lastnost (npr. maso, dolžino, gibalno količino itd.), PTR pripiše *relacijo*, tj. dvomestno lastnost (npr. *x* ima masoglede-na-inercijski-sistem-F itd.), kajti predstavitev vesolja in tega, kar to vesolje vsebuje (enoten 4-D prostor-čas kontinuum s 4-D svetovnimi osmi), je bistveno drugačna kot pri Newtonu.

Tu gre za primer redukcije, kjer posebni primeri in deduktivne posledice nove in bolj splošne teorije niso identični, marveč zgolj podobni stari in bolj ozki teoriji, ki naj bi jo prva reducirala. Rekonstrukcija posebnega primera znotraj nove teorije namreč ustreza stari teoriji v zadostni meri, da lahko pokaže, čemu je stara teorija veljala za ustrezno v okviru določene domene, in kako je možno staro teorijo nadomestiti z novo, ne da bi pri tem izgubili razlagalno in napovedno moč znotraj domene stare teorije; vendar pa nova rekonstrukcija ni docela izomorfna s staro teorijo, kajti izkaže se ne le, da je stara teorija preozka, marveč da je v nekaterih ozirih tudi napačna. Čas in prostor nista različna, kot je predpostavljal Newton, ravno tako kot ne obstajajo neke od inercijskega okvirja neodvisne notranje lastnosti, kot sta npr. masa ali dolžina.

Na koncu omenimo še nekaj primerov, kjer nova in bolj splošna teorija v ničemer ne prevzame in podpre prikaza realnosti, kot jo slika stara teorija. Takšen primer bi bil neposredna zamenjava (brez redukcije) stare flogistonske teorije z novo kisikovo teorijo izgorevanja, kot jo je predstavil Lavoisier. Stara teorija je trdila, da pri izgorevanju nekega telesa pride do *izgube* duhu podobne substance, flogistona, katerega funkcija pred izgorevanjem je priskrbeti snovi njeno specifično kakovost (npr. kovinskost, lesenost); flogiston naj bi se vezal na žgano apno kot osnovo, ki ostane prisotna tudi po izgorevanju (torej po izgubi flogistona). Flogiston je bil “duh”, ki naj bi kovinam dajal njihovo ustrezno obliko. Po sprejetju Lavoisierove trditve, da med izgorevanjem pride do absorbcije neke popolnoma materialne substance, kisika, je bil flogiston

preprosto odstranjen iz našega opisa sveta.

Drugi primeri teoretičnih entitet, ki so jih eliminirali iz resne znanosti, so kaloričen fluid, rotirajoče kristalne sfere ptolemejske astronomije, štirje temperamenti srednjeveške medicine, vitalni duh predmoderne biologije in svetlobo oddajajoč eter predeinsteinske mehanike. V vseh teh primerih nove teorije niso imele zadostnih virov, da bi rekonstruirale strukturo starejših teorij ali zakone, ki naj bi veljali znotraj teh teorij; vendar so se nove teorije izkazale za bistveno boljše od starih, zato so jih v končni fazi tudi izpodrinile.

Če si na tem mestu ogledamo pot, ki smo jo prehodili, ugotovimo, da imamo na eni strani spektra pare teorij, kjer je stara teorija gladko reducirana na novo, pri čemer je ontologija stare teorije (tj. zbir stvari in lastnosti, ki jih le-ta postulira) preživela, čeprav je bila na novo opisana v novem in bistveno bolj poglobljenem besednjaku. Tukaj navadno srečamo trditve o medteoretični identiteti, kot npr. "Toplota je enaka povprečni molekularni kinetični energiji" in "Svetloba je identična elektromagnetnemu valovanju". V sredini spektra najdemo pare teorij, kjer v viziji nove teorije najdemo zgolj šibak odsev stare ontologije, ki "preživi" le v znatno izpremenjeni obliki. In končno, na drugem koncu spektra naletimo na pare teorij, kjer bistveno bolj uporabna ontologija in znatno bolj uspešni zakoni nove teorije v celoti izpodrinejo staro teorijo in njeno stare ontologijo.

Preden zaključimo ta kratek pregled, bi si veljalo ogledati še primere kjer starejša teorija ni bila niti vključena v novejšo in bojda bolj splošno teorijo niti ni bila iz nje eliminirana. V takih primerih se starejša teorija uspešno upira poskusom prevzema in se kaže kot poseben primer trenutno veljavne splošne teorije. Maxwellova elektromagnetna teorija (od tod naprej bomo jo imenovali EM teorija) je tipičen primer te vrste. Od leta 1864 do 1905 je večina znanstvenikov pričakovala, da bo EM teorija gotovo našla primerno redukcijo v mehaničnih lastnostih etra, tj. elastične substance, v kateri naj bi se EM valovi širili. Sproženih je bilo kar nekaj pomembnih poskusov rekonstrukcije EM fenomenov znotraj mehanske terminologije, ki pa niso nikoli bili izvedeni do konca. Na veliko presenečenje raziskovalcev se je izkazalo, da je obstoj absolutnega medija, v katerem naj bi se širili EM valovi, očitno nekonsistenten z lastnostmi časa in prostora, kot ga je opisala Einsteinova posebna teorija relativnosti iz leta 1905. EM teorija se je tako izkazala za temeljno teorijo in ne zgolj za poseben primer mehanike. Poskus vključevanja je klavrno propadel.

Drug primer je povezan s teorijo obnašanja zvezd, ki jo je zagovarjala klasična astronomija ob koncu devetnajstega stoletja. V tistem času so mnogi verjeli, da bosta mehanika ali fizika lahko v zadostni meri opisali vzorce oddajanja žarkov zvezde. Vendar je sčasoma postalo očitno, da so možne zaloge kemične ali mehanske energije, ki jih vsaka zvezda potrebuje za vzdrževanje enormnih izbruhov termalne in svetlobne energije, omejene na zgolj nekaj deset milijonov let. Ta omejeni obseg je bil v nasprotju z geološkimi dokazi, ki so zgodovino merili v *milijardah* let. Vendar pa se je Lord Kelvin odločil, da bo tvegal, in je geološkim ugotovitvam navkljub določil starost zvezd na nekaj deset milijonov let. Spor se je razrešil šele po odkritju velikanskih količin energije znotraj

atomskega jedra. Zvezdna astronomija je tako dejansko bila reducirana, vendar ne na kemijo ali mehaniko, marveč na kvantno fiziko. Tako je spodletel še en poskus redukcije, le-temu pa je sledil tak, ki je uspel.

2. Nauki za nevroznanost

Po tem pregledu vseh zgoraj navedenih primerov in pa spektra zgledov, ki ga ti primeri definirajo, se lahko vprašamo, česa se lahko iz tega nauči nevroznanstvenik. Eden od naukov je gotovo ta, da gre pri interteoretski redukciji za normalen in razmeroma pogost pojav v zgodovini znanosti. Drugi pa je ta, da je pristna redukcija, če do nje lahko pridemo, nedvomno nekaj pozitivnega. Dobra je iz več razlogov, le-ti pa so znantno močnejši, če jih navedemo v medsebojni povezani.

Prvič, staro teorijo s tem, ko jo prikažemo kot poseben primer znotraj (domnevno resnične) nove teorije, *potrdimo,* tj., dokažemo njeno veljavnost vsaj v grobem orisu ali znotraj neke ustrezno omejene domene. Drugič, ker gre pri rekonstrukciji stare teorije znotraj nove zgolj redko za popolnoma zrcalno podobo, je navadno potrebna *korekcija* določenih delov stare teorije, ti popravki pa kažejo na napredek v našem znanju. Tretjič, redukcija nam omogoča znatno bolj *poglobljeni vpogled* in bistveno bolj *učinkovit nadzor* nad fenomeni v okviru domene stare teorije. Četrtič, s pomočjo redukcije se dokopljemo do *preprostejšega* splošnega pojmovanja narave, saj so očitno raznovrstni pojavi združeni pod isto razlagalno streho. In petič, nova in splošnejša teorija *postane dedič vseh zakonov,* ki so se nakopičili znotraj reducirane stare teorije, saj navsezadnje nudi razlago za iste podatke.

Narobe je, če skušamo nadvse zadovoljivo staro teorijo položiti na Prokrustovo posteljo, tj. izvršiti redukcijo, kadar teorija, ki naj bi nadomestila staro teorijo, ne poseduje zadostnih sredstev, ki bi upravičevala tovrstno nadomestitev. Ali dana sredstva zadoščajo za redukcijo ali ne, pa je kljub človeškim intuitivnim prepričanjem le redko takoj razvidno. Tudi če je redukcija kakšne teorije nemogoča, lahko to priča o napačnosti te teorije in ne o njeni fundamentalni resničnosti. V takem primeru se lahko zgodi, da nova teorija docela izpodrine (eliminira) staro, namesto da bi jo zgolj gladko reducirala. Mogoče bodo človeška pojmovanja, kot so npr. “prepričanja” in “volja”, eliminirana in zamenjana z nekimi docela drugačnimi zgodbami o shranjevanju podatkov in povzročanju obnašanja.

Dejstvo je, da se mnenja o upravičenosti redukcije na področju nevroznanosti in psihologije zelo razlikujejo. Po eni strani se namreč zdi, da je upravičenost redukcije (ali eliminacije) na tem področju bistveno večja kot v vseh zgoraj navedenih zgodovinskih primerih. Kajti, za razliko od svetlobe, toplote ali gibanja nebesnih teles, je za psihološke fenomene v splošnem že znano, da nastanejo kot posledica evolucijskega in ontogenetskega razvoja snovi, natančneje razvoja biološke organizacije. Zato *pričakujemo,* da bomo zmožni razumeti psihološke fenomene s pomočjo biološke osnove.

Po drugi strani pa se nam zdi upravičenost redukcije vprašljiva, kar je lepo izraženo v petih skupinah ugovorov, ki smo jih predstavili na začetku te razprave. Nam lahko zgodovinski primeri, s katerimi smo se seznanili, kakor koli pomagajo pri oblikovanju odgovorov na te ugovore? Predlagava, da si ogledamo vsako skupino posebej.

Prva skupina ugovorov se ukvarja z možnostjo, kako razložiti subjektivni značaj čutnih takšnosti. Vsi negativni argumenti, ki sodijo v to skupino, imajo neko skupno osnovo, namreč to, da si ne moremo predstavljati, kako bi nam lahko kakršen koli opis delovanja nevronov razložil neizrekljive subjektivne fenomene. Očitno je, da ta ugovor postavlja našo zmožnost oziroma nezmožnost, da si kaj predstavljamo, kot merilo za to, da nekaj je ali ni mogoče. Ta zahteva očitno presega to, kar običajno zahtevamo od poskusa. Namreč, kdo pred Jamesom C. Maxwellon si je lahko predstavljal, da bo teorija nabitih energijskih krogel in nihajočih igel pri kompasu zadostovala za razlago fenomena svetlobe? Kdo pred Descartesom, Bernoullijem in Joulom si je mogel zamisliti, da bo lahko mehanika biljardnih krogel razložila na prvi pogled bistveno drugačen fenomen – toploto? Komu bi se zdelo sprejemljivo, da je višina zvoka pravzaprav frekvenca, ali sploh, da sam zvok sestoji iz niza tlačnih valov v atmosferi?

Moramo se spomniti, da je uspešna interteoretska redukcija navadno kompleksen postopek, ki vključuje sistematično rekonstrukcijo vseh ali večine starih pojmov s pomočjo sredstev novega pojmovanja. Rekonstrukcija navadno ni le zelo zapletena, marveč je pogosto tudi silno presenetljiva. Pri tem ne gre za nekaj, kar bi lahko pričakovali, da se bo porodilo v domišljiji kogar koli v obliki retorične zahteve, kot, recimo, vprašanje: "Le *kako* bi lahko bili *Aji* kaj drugega kot *Bji?*".

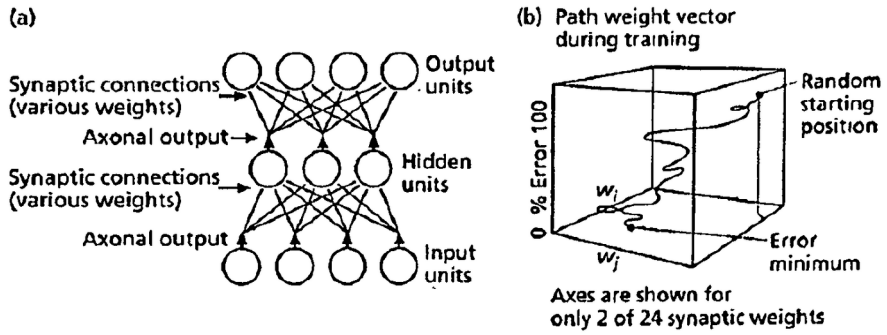
Nadalje, tovrstno retorično vprašanje nas ne more zbegati, če je naša domišljija obveščena o najnovejših teorijah o zaznavnem kodiranju. Ideja, da so vtisi okusa vkodirani kot štiri-dimenzionalni vektorji s koničastimi frekvencami, nam omogoči reprezentacijo prostora vseh človeških okusov, v katerem so vsi sorodni okusi združeni glede na njihove podobnosti, razlike in podobne relacije, kot je npr. vmesnost (Bartoshuk, 1978). Landova retinetalna teorija barvnega vida (Land, 1977) govori o podobni razporeditvi naših barvnih vtisov. Ta teorija vsebuje tudi napoved osnovnih oblik barvne slepote, do katerih pride, kadar je zaradi izgube ene od treh vrst čepkov posameznikov tridimenzionalni barvni prostor reduciran na dve dimenziji.

Pri tem že gre za rekonstrukcijo nekaterih lastnostni pojavov, ki jih preiskujemo, v luči nove teorije. Zdaj je potrebno le še izpeljati rekonstrukcijo do konca, kot je bilo to strjeno v primeru zgodovinskih razlag prej omenjenih objektivnih fenomenalnih lastnosti (toplota, svetloba, višina tona). Nekatere stvari so lahko po svoji naravi res neizrekljivo fenomenalne, saj jih zaznavamo s pomočjo naših osnovnih čutov. Toda zaradi tega niso nič bolj varne pred poučno interteoretsko redukcijo. Zgodovina nas uči ravno o nasprotnem.

Druga skupina ugovorov opozarja na pomen, semantično vsebino oz. intencionalnost naših misli in ostalih mentalnih stanj. Tovrstni antiredukcionistični argumenti so zelo podobni tistim iz prejšnje skupine.

Sklicujejo se predvsem na našo nezmožnost predstavljati si, kako bi lahko pomen bil zgolj posledica interakcije med signali ali procesiranja inertnih simbolov (Searle, 1980, 1990; za zavrnitev Searlovih izpeljevanj glej Churchland in Churchland, 1990. Searle, strogo gledano, nasprotuje le popolnoma računski redukciji; ker pa je ta izrednega pomena za nevroznanost, ga bomo vključili med ostale antiredukcioniste). Pri tovrstnih dokazih gre pravzaprav za argumente iz nevednosti. Vsi imajo sledečo obliko: "Ne morem si *predstavljati*, kako bi deloval nevrokomputacijski proces s smiselnimi reprezentacijami; torej tak proces ni možen." Za kratkoročno spodbijanje tovrstnih izpeljevanj je dovolj, če izpostavimo samo to napako.

Za bolj dolgoročno zavrnitev pa je potrebno nekaj več truda. Potrebno je dejansko predočiti opis, kako možgani reprezentirajo zunanji svet in zakonitosti, ki v njem veljajo. Ravno s tem pa se ukvarjajo moderne teorije o nevronskih mrežah. Te teorije trdijo, da so realnočasovne informacije o svetu kodirane v visokodimenzionalnih aktivacijskih vektorjih, splošne informacije o svetu pa so kodirane v konfiguraciji ozadja mrežnih sinaptičnih uteži. Aktivacijski vektorji so procesirani s konfiguracijami uteži, mimo katerih potujejo, učenje pa ni nič drugega kot prilagoditev posameznikove globalne konfiguracije uteži (glej sliko 22.1) S pomočjo tovrstnih razlag lahko do neke mere že opišemo mnogo različnih stvari, kot so, recimo, prepoznavanje kompleksnih predmetov kljub le pomanjkljivim čutnim podatkom, hiter priklic pomembnih informacij iz obsežnega vsebinsko naslovljivega spomina, zmožnost prepoznavanja majhnih in neubesedljivih razlik in uporaba kompleksne senzomotorične koordinacije (P. M. Churchland, 1989). Še zmeraj vemo znatno premalo, da bi mogli z gotovostjo trditi, da bodo tovrstne hipoteze v zadostni meri razložile reprezentacijske zmožnosti razuma. Ravno tako pa je napačno tudi nasprotno prepričanje, namreč, da bodo take hipoteze nujno neustrezne. Gre za empirično vprašanje in potrebno je počakati na končno sodbo.



Slika 22.1*

Slika 22.1 (a) Preprosta neposredna umetna "nevronska mreža". Vnosi so

* (a) *Synaptic connections (various weights)* – sinaptične povezave (različne uteži); *Axonal outputs* – aksonalni izhodi; *Output units* – izhodne enote; *Hidden units* – skrite enote; *Input units* – vhodne enote. (b) *Path weight vector during training* – vektor poti uteži med treningom; *Random starting position* – naključna začetna točka; *Error minimum* – minimum napake; *Axes are shown for only 2 of 24 synaptic weights* – Prikazani sta osi za le 2 od 24 sinaptičnih uteži. – op. prev.

kodirani kot vzorec ali vektor aktivacijskih vrednosti na vhodnih enotah. Vzorec se prenese na skrite enote, a se spremeni, ko gre preko vmesnih sinaptičnih povezav z različnimi utežmi. Vsaka skrita enota nato sešteje prejete obtežene vnose in na podlagi rezultata privzame ustrezno aktivacijsko vrednost. To privede do novega vzorca ali vektorja aktivacijskih vrednosti na skritih enotah. Zgodba se ponovi na zadnjem sloju izhodnih enot, ki prevzamejo tretji vzorec aktivacij. Mreža tako skrbi za transformacije med vektorji. Za kakšno vrsto transformacije gre, pa je odvisno od specifične konfiguracije sinaptičnih uteži. S pomočjo sigmoidnih izhodnih funkcij na prvih dveh ravneh lahko mreže izračunajo približno vrednost vsakršne izračunljive transformacije. (b) Shematski prikaz abstraktnega "prostora uteži" za preprosto mrežo. Učenje v taki mreži poteka kot zaporedno prilagajanje njene konfiguracije uteži, tako da bi postopno zmanjšali njene napake pri izvedbi. Ta postopek se nadaljuje, dokler mreža končno ne izvede transformacije med vhodnimi in izhodnimi podatki, ki je bila implicitno prisotna v mnogih primerih, na katerih smo to mrežo urili (povzeto po P. M. Churchland, 1998).

Tretja skupina ugovorov poudarja, da pri tem, kar mi razumemo kot človeška zavest, ne gre le za neko intrinzično lastnost določenega bitja, marveč za bogato matrico posameznikovih odnosov do soljudi, običajev in institucij njegove kulture. Ker je redukcionistični pristop k zavesti in obnašanju omejen na mikroskopske dejavnosti znotraj posameznikovih možganov, mu ne bo nikoli uspelo zajeti več kot le droben delež tega, kar je razlagalno pomembnega.

Na tovrstne ugovore najbolje odgovoriva tako, da jih sprejmeva. Človeško obnašanje je nedvomno funkcija prej omenjenih faktorjev. Kultura, v kateri nek posameznik prebiva, bo imela velik vpliv na razvoj posameznikove osebnosti. To pa pomeni, da mora vsaka ustrezna nevrokomputacijska razlaga človeške duševnosti vsebovati tudi opis, kako možgani reprezentirajo ne le grobe lastnosti fizičnega sveta, marveč tudi značaj kognitivnih bitij, s katerimi je posameznik v interakciji, ter značilnosti družbenega, moralnega in političnega življenja, v katerem ta posameznik živi. Možgani družbenih živali se navsezadnje naučijo igrati vlogo interaktivnega elementa znotraj skupnosti možganov, in ta vloga je izredno pomembna kognitivna prednost. Moramo spoznati, kako to počnejo.

To je nedvomno izredno velik izziv, ki pa mu nevroznanstveniki zaenkrat ne posvečajo preveč pozornosti, kaj šele da bi se ga resneje lotili. Vendar to ni nič presenetljivega. Če je izredno težavno opisati že samo to, kako navadna muha določi svoj položaj v prostoru, gre pri opisu sramu ljubljene osebe, značaja kakega politika ali poznavanja skrivnih načrtov našega sovražnika za toliko težjo nalogo. Vemo pa, da umetne nevronske mreže, ki jih izurimo s pomočjo primerov, lahko prepoznavajo neverjetno subtilne vzorce in podobnosti v naravi ter nanje tudi reagirajo. Če to velja za fizične vzorce, čemu ne bi isto veljalo tudi za družbene vzorce? Tu ne gre za kakršno koli principielno težavo, marveč zgolj za izredno velik izziv.

Morebiti se zdi nerealno pričakovati, da nam bo kdaj uspelo podati izčrpen splošni opis nevralne in vedenjske poti nekega posameznika. Povsem možno je,

da zaradi izredne kompleksnosti živčnega sistema ne bomo mogli nikoli priti dlje od zgolj uporabnih približkov za želeni idealni opis. Vzemimo primer kemije in njeno razmerje do kvantne fizike. Tudi tukaj matematika kompleksnih dinamičnih sistemov narekuje določene omejitve glede na to, s kolikšno mero natančnosti lahko rekonstruiramo kemična dejstva s pomočjo fizikalnih principov. To pomeni, da naša redukcija nikoli ne bo popolna, vendar pa smo vseeno upravičeno prepričani, da so kemični pojavi le refleksija kvantnih fizičnih pojavov na makro ravni. Podobno kot za kemične pojave velja tudi za psihološke.

To nas pa pripelje še do četrtega sklopa ugovorov, ki opozarjajo na nevarnost, ki naj bi jo redukcija predstavljala človeški svobodi. Tukaj bova kratka. Ali obstaja kakršna koli svoboda, ki se razlikuje od relativne avtonomnosti, značilne za vse kompleksne dinamične sisteme, ki so delno izolirani od preostalega sveta, je docela empirično vprašanje. Namesto da skušamo pokazati konsistentnost nevroznanosti s tem ali onim vnaprej ustvarjenim pojmovanjem svobode, dovolimo raje znanstvenim raziskavam, da nas *poučijo*, na kakšen način in v kolikšni meri so človeška bitja res "svobodna". To bo nedvomno povzročilo spremembe ali celo popolno zavračanje nekaterih trenutno veljavnih pojmovanj svobode. Toda to bo bistveno bolje, kot pa da postavimo trenutno zmedo za standard, s katerim se bodo bodoče teorije primorane uskladiti.

Peta in zadnja skupina ugovorov trdi, da za psihologijo velja ireducibilni abstraktni status, kar utemeljujejo s tem, da lahko zelo različni fizični sistemi enako dobro realizirajo abstraktno organizacijo, ki tvori kognitivno ekonomijo nekega sistema. Le kako bi lahko reducirali psihološke fenomene na nevrobiologijo, ko pa lahko povsem drugačni fizični substrati služijo enakemu namenu?

S premiso tega sklopa ugovorov se bomo verjetno strinjali v glavnem vsi. Toda iz nje ne sledi antiredukcionistični sklep. To je razvidno, če si pobližje ogledamo enega od primerov iz naše zgodovine znanosti. Temperatura je, kot smo že omenili, enaka povprečni molekularni kinetični energiji. Vendar pa, strogo gledano, to velja zgolj za pline, kjer se lahko molekule prosto gibljejo v balistični maniri. V trdnem agregatnem stanju, kjer delci oscilirajo sem ter tja, pa energija konstantno niha med kinetično in potencialno energijo. V plazmi, segreti do visokih temperatur, sploh ne moremo govoriti o molekulah, saj je vse razstavljeno na subatomske delce. Tukaj je temperatura kompleksna mešanica različnih energij. V vakuumu, kjer sploh ni mase, pa temperatura sestoji iz distribucije valovnih dolžin – "krivulja črnega telesa" – EM valov, ki potujejo skozenj.

Ti primeri lepo kažejo, kako je lahko redukcija vezana na specifično domeno: v primeru plina velja temperatura za eno stvar; v primeru trdne snovi za nekaj drugega; v primeru plazme spet za nekaj tretjega; v primeru vakuuma za nekaj četrtega itd. (Pri vseh gre za "temperaturo", saj so medsebojno povezane in so podvržene enakim zakonom ravnotežja.) Vendar pa zaradi tega ne bomo trdili, da je klasična termodinamika avtonomna, ireducibilna znanost, ki je za vedno izvzeta iz zgodbe spodaj ležeče mikrofizične domene. Ravno nasprotno, zgoraj

navedeno nas uči, da obstaja več kot le en način, na katerega se energije manifestira na mikrofizični ravni.

Podobno je z vizualnim doživetjem – le-to je lahko ena stvar pri sesalcih, rahlo drugačna stvar pri hobotnici in bistveno drugačna stvar pri potencialnem androidu, zgrajenem iz kovine in polprevodnikov. Vendar gre pri vseh za vizualna doživetja, saj so deležni nekih skupnih abstraktnih lastnosti na višji opisni ravni. Tega, da je nevrobiologija zmožna razlož vse psihološke pojave pri ljudeh, tako ne ogroža možnost, da lahko neka *druga* teorija, recimo elektronika polprevodnikov, ponudi razlago za psih loške pojave pri *robotih.* Taki redukciji si ne bi nasprotovali, marveč se medsebojno dopolnjevali,

Drugod sva navedla bolj natančne in izčrpne opise, kako sodobna dejavnost v nevroznanosti pojasnjuje dejstva v psihologiji in kognitivni teoriji (P. S. Churchland, 1988; P. M. Churchland, 1989). Tukaj bova zaključila z dvema svarilnima opazkama. Prvič, čeprav sva bila v tem članku izredno optimistična glede možnosti redukcije psihologije na nevroznanost, se zavedava, da bo dejanska izvedba tovrstne redukcije izredno dolgotrajna in silno zahtevna naloga. V pričujočem članku nama je šlo predvsem za to, da ovrževa zagovornike nemožnosti in da umestiva redukcijske teženje nevroznanosti v ustrezen zgodovinski kontekst.

Drugič, ne smemo domnevati, da bo psihologija kot znanost v postopku redukcije izginila ali da bo njena vloga omejena na pasivno tarčo nevralnih razlag. Nasprotno, kemija ni izginila po kvantno-mehanski razlagi njenih osnov; ravno tako ni izginila biologija, kljub kemijskim eksplikacijam njenih osnov. Še več, tako kemija kot biologija sta bistveno pripomogli k oblikovanju in razvoju spodaj ležečih znanosti. Nedvomno taka prihodnost čaka tudi psihologijo in nevroznanost. Na tej ravni kompleksnosti se interteoretska redukcija ne zdi kot nenaden prevzem ene discipline s strani druge, marveč je bolj podobna dolgemu in počasi zorečemu zakonu.

Literatura

Bartoshuk, L. M. (1978). “Gustatory system”. V Masterton, R. B. (ur.), *Handbook of Behavioral Neurobiology*, 1. zv., *Sensory Integration.* New York: Plenum Press, str. 503-567.

Churchland, P. M. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: MIT Press.

Churchland, P. M. in Churchland, P. S. (1990). “Could a machine think?”. *Scientific American,* 262, str. 32-37.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Understanding of the Mind/Brain.* Cambridge, MA: MIT Press.

Fodor, J. A. (1975). *The Language of Thought.* New York: Crowell.

Jackson, F. (1982). “Epiphenomenal qualia”. *Philosophical Quaterly,* 32, str.

127-136.

Land, E. (1997). “The retinex theory of color vision”. *Scientific American, 237,*

Nagel, T. (1974). “What is it like to be a bat?”. *Philosophical Review,* 83, str. 435-450.

Popper, K. and Eccles, J. (1978). *The Self and the Brain.* New York: Springer.

Searle, J. (1980). “Minds, brains, and programs”. *Behavioral and Brain Sciences,* 3, str. 417-457.

Searle, J. (1990). “Is the brain’s mind a computer program?”. *Scientific American,* 262, str. 26-31.

Taylor, C. (1970). “Mind-body identity: A side issue?”. V Boršt, C. V. (ur.), *The Mind/Brain Identity Theory*, Toronto: Macmillian, str. 231-241.

Taylor, C. (1987). “Overcoming epistemology”. V Baynes, K., Uohman, J. in McCarthy, T. (ur.), *After Philosophy: End or Transformation?,* Cambridge, MA: MIT Press, str. 464-488.