THEORIA 1
(2021) : 64 : str. 173-196

https://doi.org/10.2298/THEO2101173S
Originalni naučni rad
Original Scientific Paper

Vanja Subotić

ZDRAVORAZUMSKA PSIHOLOGIJA, ELIMINATIVIZAM I SADAŠNJOST KONEKCIONIZMA

APSTRAKT: *Pre trideset godina, Vilijem Remzi, Stiven Stič i Džozef Geron su u zajedničkom radu izneli argument u prilog sledećeg kondicionala: ako konekcionistički modeli koji implementiraju paralelno distriburirano procesiranje predstavljaju vernu sliku ljudskih kognitivnih procesa, onda je eliminativizam u pogledu propozicijskih stavova ispravna teza. Korolar njihovog argumenta, ukoliko se pokaže kao zdrav, jeste da za zdravorazumsku psihologiju nema mesta u savremenoj kognitivnoj nauci. Ovakvo viđenje konekcionizma – kao hipoteze o kognitivnoj arhitekturi kompatibilne sa eliminativizmom – karakteristično je i za Pola Čerčlanda, radikalnog protivnika zdravorazumske psihologije. Cilj ovog rada je da se ispita u kojoj meri sadašnji metodološki rafinirani konekcionistički modeli, bazirani na neuronskim mrežama dugog kratkoročnog pamćenja, potvrđuju argumente kako Remzija i kolega, tako i Čerčlanda. Argumentovaću u prilog eliminativizma ograničenog uticaja. Naime, tvrdiću da to što konekcionistička kognitivna nauka nema potrebu za zdravorazumskom psihologijom qua teorijom ne povlači za sobom nelegitimnost zdravorazumske psihologije per se u drugim naučnim domenima, ukoliko se zdravorazumska psihologija shvati kao korisna heuristika*.

KLJUČNE REČI: *eliminativizam*, *konekcionizam*, *neuronske mreže*, *propozicijski stavovi*, *zdravorazumska psihologija*.

Uvod

Istoričar Plutarh, opisujući u svojim *Uporednim životopisima* lik i delo Marka Porcija Katona, poznatijeg kao Katon Stariji, navodi sledeću epizodu: pod stare dane, svaki svoj govor u Senatu ratni veteran Katon je završavao rečima *Ceterum censeo Carthaginem esse dellendam* ili „Naposletku, smatram da Kartagina mora biti uništena“ (1835: 393). Čitajući Plutarha, možemo interpretirati ovu Katonovu naviku kao odraz vojničkog *uverenja* i nepokolebljivosti, ili pak kao *želju* za osvetom Kartagini

usled fijaska Rima tokom Drugog punskog rata. Bilo da pripisujemo Katonu verovanja ili želje, u oba slučaja, na osnovu pretpostavke da ga karakterišu određena *mentalna stanja*, možemo *razumeti* ili *predvideti* njegovo ponašanje – recimo, da će nam Plutarh sigurno saopštiti na sledećoj strani da čak i kada je rasprava bila o dozvoljenoj ceni žita, Katon je morao biti konzistentan i završavati govor istim rečima. U filozofiji duha, ovakvo svakodnevno pripisivanje mentalnih stanja – koja se inače izražavaju u formi propozicijskih stavova „Katon veruje da *p*" ili „Katon želi da *p*" –označava se grupnim terminom kao „zdravorazumska psihologija". Ovaj termin se može shvatiti na bar dva načina. S jedne strane, usko gledano, njime se može referirati na *skup ljudskih kognitivnih sposobnosti*, a koje se tiču predviđanja i objašnjenja ponašanja. S druge strane, zdravorazumska psihologija se može posmatrati kao *unutrašnja teorija* o ponašanju drugih ljudi.[1] Mnogi filozofi (Sellars 1956, Lewis 1972) i psiholozi (Premack & Woodruff 1987, Leslie 1987, 1994) su bili skloni tome da zdravorazumsku psihologiju posmatraju kao teoriju koja postulira postojanje verovanja i želja – stanja označena *teorijskim terminima*, čije značenje se specifikuje putem veza sa ostalim teorijskim i opservacionim terminima, i za čije instanciranje je zadužen zaseban *modul* u mozgu.[2]

Kako Stiven Stič (Stephen Stich) primećuje (1983: 2), zdravorazumska psihologija tumači važnu ulogu u istoriji, književnosti i antropologiji: kako bismo drugačije razumeli postupke istorijskih ličnosti, književnih likova, ili pripadnika kultura nesamerljivim našoj ako se ne bismo oslanjali na zdravorazumska objašnjenja njihovog ponašanja? Međutim, eliminativista poput Pola Čerčlanda (Paul Churchland) (1981: 74) smatra da je zdravorazumska psihologija spekulativna, primitivna teorija kojom

1 Ovakva verzija zdravorazumske psihologije se u literaturi naziva još i „teorija-teorija". Važno je primetiti da postoji više suparničkih alternativa ovoj verziji zdravorazumske psihologije – od teorije simulacije (Stich & Nichols 2003, Goldman 2006), interakcionističke teorije (Gallagher 2001), teorije modela osobe (Newen 2015, 2018) i hipoteze narativne prakse (Gallagher & Hutto 2008) – međutim, one neće biti predmet razmatranja ovog rada, iako bi bilo zanimljivo proveriti koje od njih bi mogle da ne budu u koliziji sa konekcionizmom i do koje mere. Takođe, sve pomenute verzije zdravorazumske psihologije su *deskriptivne*, to jest tiču se teoretisanja o unutrašnjim uzrocima ljudskog ponašanja; ali, u poslednje vreme se pojavljuju i filozofi (Andrews 2015, McGeer 2015) koji zastupaju *regulativnu* verziju. Ovom verzijom se insistira da je primarni zadatak zdravorazumske psihologije formiranje normi, kao i šablona prihvatljivog ponašanja, kojim se regulišu svakodnevne društvene interakcije. Ni regulativne verzije neće biti predmet razmatranja u radu, jer za njih *prima facie* nije ni jasno kako bi se povezale sa konekcionističkim modelima u kognitivnoj nauci, budući da su ovi modeli deskriptivni.

2 Naravno, postoje i filozofi koji se ne slažu sa viđenjem da je zdravorazumska psihologija teorija (npr. Clark 1987 izražava umereni skepticizam u odnosu na argumente koje nude Čerland (1981, 1985) i Stič (1983), dok Sharpe (1987) decidirano kritikuje projekat pravljenja teorije od laičkog prisipisivanja mentalnih stanja kao cirkularan). U odeljku **IV** ću se detaljnije osvrnuti na ovakve predloge.

se postuliraju mitovi – nalik tome kako je Homer pripisivao bogovima raspoloženja i emocije na osnovu kojih su se tumačili događaji – te da se od doba antike i Sofokla takva teorija nije menjala niti napredovala. Prema eliminitivističkoj tezi, zdravorazumska psihologija ne bi mogla da preživi redukciju na egzaktnu nauku poput neuronauke, te će stoga naprosto biti odbačena i zamenjena boljom teorijom. Štaviše, postoje opravdani razlozi i za bojazan da verovanja i želje ne bi trebalo da budu predmet istraživanja kognitivne nauke, jer ne izgleda da je uopšte potrebno eksplicirati ih unutar teorije o ljudskoj kogniciji (cf. Stich 1983: poglavlje 11).

Trojica filozofa, Vilijem Remzi (William Ramsey), Stiven Stič i Džozef Geron (Joseph Garon) (1990), sledeći novu konekcionističku paradigmu u kognitivnoj nauci, argumentovali su da ukoliko konekcionistički modeli pružaju adekvatno objašnjenje kognitivnih procesa, onda nema više razloga da se odoleva eliminativizmu – neizbežna budućnost zdravorazumske psihologije je odlaganje na smetlište prevaziđenih spekulativnih teorija gde su već završile teorija flogistona, aristotelovska kosmologija ili alhemija. Rad Remzija i kolega je privukao pažnju filozofske zajednice i provocirao dugogodišnju diskusiju uprkos tome što je ovaj trojac manje radikalan i obazriviji u oduševljavanju konekcionizmom u odnosu na Čerčlanda. Međutim, kako je konekcionistička literatura postajala opterećena tehničkom terminologijom i metodološkim problemima, interesovanje filozofa je početkom dvehiljaditih godina značajno opalo.

Cilj ovog rada je da, sledeći metodološke inovacije na polju konekcionističkog modelovanja, kao i ponovni rast interesovanja filozofa za ovu granu kognitivne nauke, preispita argumente Remzija i kolega (odeljci **I** i **II**), kao i Čerčlandov optimizam (odeljak **III**). Pitanje opstanka zdravorazumske psihologije u dvadeset prvom veku – „Veku mozga", kako ga je nazvao neuronaučnik Stiven Rouz (Steven Rose) – tiče se i statusa naših laičkih uverenja utkanih u humanističke nauke. Stoga, u svetlu recentnih konekcionističkih modela trudiću se da pokažem da *ograničeni* eliminativizam može biti deo metodologije kognitivne nauke, i da kognitivna nauka nema potrebe da se bavi entitetima zdravorazumske psihologije (odeljak **IV**). Naposletku, zaključiću da to što konekcionistička kognitivna nauka može bez propozicijskih stavova, ne znači da treba zabraniti istoričarima, antropolozima ili filozofima da ih koriste zarad teoretisanja o ponašanju ljudi. Naime, kognitivna nauka može da nam pomogne da otkrijemo kognitivne mehanizme u pozadini mentalnih stanja poput verovanja i želja, koji bi trebalo da važe za sve ljude – ma kakve bile njihove idiosinkrazije na kulturnom, demografskom ili društvenom planu i kroz epohe, o čemu ipak svedoče humanističke nauke.

I Struktura argumenta Remzija, Stiča i Gerona

Remzi, Stič i Geron počinju argumentaciju prihvatajući tezu eliminativizma prema kojoj će bolja teorija T_1 *zameniti* teoriju T_0 koja postulira entitete ili procese čije postojanje se dovodi u pitanje. Međutim, postoji suptilna razlika između *ontološki konzervativne* i *ontološki radikalne* zamene teorije (Ramsey et al. 1990: 120). Ukoliko se entiteti i procesi T_0 čuvaju unutar T_1, odnosno T_0 biva redukovana na T_1, radi se o ontološki konzervativnoj zameni teorija. Ukoliko dolazi do potpunog odbacivanja entiteta i procesa T_0, a T_1 čini T_0 empirijski *lažnom* radi se o ontološki radikalnoj zameni teorija. Argument koji nude Remzi, Stič i Geron imaće kao posledicu potvrđivanje vijabilnosti ontološki radikalne zamene teorije – ukoliko je zdrav i valjan. Drugim rečima, ispostaviće se da konekcionizam, kao T_0, čini empirijski lažnom zdravorazumsku psihologiju, kao T_1, jer entiteti poput verovanja i želja koji se postuliraju ne mogu da prežive unutar konekcionističkog okvira.

Tri ključne odlike propozicijskih stavova omogućavaju vezu između konekcionizma i eliminativizma, i to su *funkcionalna diskretnost*, *semantička interpretabilnost* i *kauzalna efikasnost* (Ramsey et al. 1990: 121). Tri odlike se sumiraju pod naziv *propozicijska modularnost* po ugledu na Stiča (1983). Objasniću šta svaka od ovih odlika podrazumeva. Funkcionalna diskretnost podrazumeva, prema Remziju i kolegama, da je potpuno smisleno reći kako je subjekat *S* izgubio ili stekao pojedinačno sećanje ili verovanje (1990: 122). Primera radi, pretpostavimo da je *S* zaboravio od koga je pozajmio Plutarhove *Uporedne životopise*, iako nije zaboravio da knjiga jeste pozajmljena, kog dana je pozajmljena i kad treba da je vrati. Naravno, gubitak ovog konkretnog sećanja uslovljava zauzvrat formiranje niza drugih verovanja (recimo, da je *S* neodgovoran, da je odlična ideja zapisivati u notes od koga uzima knjige, itd.), ali i dalje izgleda da gubitak pojedinačnog sećanja može bez problema biti okurentno, trenutno mentalno stanje *S*-a.

Semantička interpretabilnost se tiče semantičkih svojstava propozicijskih stavova. Naime, zdravorazumska je pretpostavka da kada ljudi vide na ekranu da let za Bangkok kasni tri sata, onda će formirati *verovanje da* avion kasni, te nas kao posmatrače ne bi trebalo začuditi što ljudi sa ovog leta sedaju u kafe umesto da pružaju pasoše stjuardesi. Ovakvo verovanje da *p* izražava *zakonolike generalizacije*, to jest regularnosti koje se očitavaju u činjenici da onda kada se ispune određeni uslovi, formirano verovanje će imati nekakve posledice (cf. Ramsey et al. 1990: 121-122). Posledice propozicijskih stavova su u posebnoj vezi sa trećom odlikom: kauzalna efikasnost upućuje na to da su neka semantički interpretabilna mentalna stanja *kauzalno aktivna* dok neka nisu, već su *kauzalno inertna*. Različita verovanja i želje mogu imati istu posledicu u vidu ponašanja na takav-i-takav način, i često je teško utvrditi koje tačno verovanje ili koja tačno želja je iza ponašanja ljudi, ali služeći se zdravorazumskom psihologijom, mi ćemo praviti manje ili više ispravne predikcije i formirati manje ili više tačna objašnjenja ponašanja (cf. Ramsey et al. 1990: 123).

Prihvatanje propozicijske modularnosti je karakteristično za simboličke modele u kognitivnoj nauci, koji se oslanjaju na Fodorovu (1975) *reprezentacijsku teoriju uma* prema kojoj su mentalna stanja u relaciji sa tokenima semantički interpretabilnih mentalnih rečenica. Drugim rečima, ova teorija počiva na ideji da mentalne reprezentacije moraju biti lingvistički strukturirane, ili da postoji *jezik misli*. Fodor vidi propozicijske stavove – to jest, aparat koji zdravorazumska psihologija unosi u kognitivnu nauku – kao ključne reprezentacijske pozite za objašnjenje ljudskih kognitivnih procesa. To znači da je prema njegovom mišljenju zdravorazumska psihologija je empirijski i metodološki valjana teorija, a zakonolike generalizacije kognitivne nauke koje se „izvlače" iz kognitivnih modela moraju uključivati propozicijski, reprezentacijski ili semantički sadržaj, čija istinosna vrednost se evaluira u odnosu na stanje stvari (cf. Fodor 1990: 209).[3]

Simbolički modeli reprezentuju verovanja kao rečenice sa kombinatorijalnom sintaksom i semantikom, to jest kao nizove simbola kojima se operiše pomoću eksplicitnih pravila, i koji mogu biti bilo kad individualno aktivirani nakon što se pohrane u unutrašnju memoriju modela. Argument Remzija i kolega se poziva na modele potpuno različite unutrašnje strukture. Naime, konekcionistički modeli *qua* kognitivni modeli imaju za cilj simulaciju ljudskih kognitivnih procesa putem veštačke neuronske mreže. Ovakve mreže se sastoje od bar tri sloja jedinica: prvog koji procesira ulazne signale, drugog ili „sakrivenog" sloja, gde se skladište informacije kako bi ponovo bile upotrebljene prilikom procesiranja, i trećeg koji pruža izlazne podatke u pogledu uspešnosti obavljanja kognitivnog zadatka. Svaka jedinica ima *aktivacioni prag*, a interakcije između jedinica određeni *stepen jačine*. Kodiranje signala se obavlja na paralelno distribuiran način, što znači da na osnovu procesiranja mreže „emer-

3 Stič (1983: 187-192) interpretira Fodora kao zastupnika i jake i slabe reprezentacijske teorije uma na osnovu pažljive egzegeze. Glavna razlika između ove dve verzije tiče se statusa generalizacija u kognitivnoj nauci: prema slaboj teoriji uma, ove generalizacije nemaju istinosnu vrednost, odnosno ne primenjuju se na mentalna stanja naprema semantičkim svojstvima, već prema čisto formalnim sintaksičkim pravilima. Međutim, obe verzije teorije počivaju na uverenju da je moguće uspešno pomiriti zdravorazumsku psihologiju sa kognitivnom naukom, i obe verzije pate od istovetnih problema prema Stiču. Naime, domeni kognitivnog razvoja, komparativne psihologije, kliničke psihologije i kognitivnih abnormalnosti i protivreče slaboj i jakoj reprezentacionoj teoriji uma, jer takva teorija ne može predstavljati usmeravajuću paradigmu u slučajevima ispitivanja dečije kognicije, životinjske kognicije, ili ljudi sa mentalnim bolestima (cf. Stich 1983: 137-145, 207-208). Razlog za ove poteškoće leži u „protagorskom parohijalizmu" zdravorazumske psihologije (Stich 1983: 7): nemoguće je primetiti regularnosti u kognitivnim procesima „egzotičnih slučajeva" kada zdravorazumska psihologija uzima kao merilo isključivo perspektivu onog koji predviđa ili objašnjava tuđe verovanje. Neću dužiti sa kritikovanjem reprezentacione teorije uma, ali važno je primetiti da ova teorija umire ili preživljava zajedno sa simboličkim modelima: protivnik ovih modela mora biti i protiv reprezentacione teorije uma i *vice versa*.

gira“ *distribuirana reprezentacija* koja se ne može pripisati nijednoj zasebnoj jedinici. Putem različitih algoritama, koji pojačavaju ili smanjuju jačinu interakcija između jedinica na osnovu ulaznog signala, neuronske mreže se „obučavaju“ na određenim skupovima podataka u zavisnosti od specifičnog kognitivnog zadatka koji treba da „obave“, odnosno koji kognitivni proces treba da simuliraju.

Ovi modeli, dakle, operišu na *subsimboličkom nivou*, to jest ne pozivajući se na kombinatorijalnu sintaksu i semantiku kojim bi se strukturirale reprezentacije. Jedan od ubeđenih konekcionista, Pol Smolenski (Paul Smolensky), tvrdio je da su nam opisi oba kognitivna nivoa – i simboličkog i subsimboličkog – korisni radi boljih predikcija i objašnjenja kako ponašanja mreže, tako i ljudskog ponašanja, isto kao što je klasična fizika korisna za opisivanje kretanja tela na makronivou, a kvantna fizika za opisivanje kretanja i interakcije čestica na mikronivou. Iako je, striktno govoreći, „makroteorija pogrešna ukoliko je mikroteorija istinita (...) to ne znači da je eksplanatorno irelevantna“ (Smolensky 1988: 60). Međutim, Stič u komentaru na tekst Smolenskyja insistira da je nužno otići korak dalje ukoliko se prihvati analogija sa klasičnom i kvantnom fizikom: ako je Smolenski u pravu, onda se istiniti i zakonoliki opis kognitivnih procesa nalazi na subsimboličkom nivou, dok su aproksimacije na simboličkom tačne samo u graničnim slučajevima (1988: 53). Zašto onda oklevati i odolevati eliminativizmu?

Remzi, Stič i Geron argument kojim spajaju konekcionizam i eliminativizam iznose u vidu sledećeg kondicionala: ako konekcionistički modeli koji implementiraju paralelno distriburirano procesiranje predstavljaju vernu sliku ljudskih kognitivnih procesa, onda je eliminativizam u pogledu propozicijskih stavova ispravna teza. Korolar njihovog argumenta, ukoliko se pokaže kao zdrav i valjan, jeste da za zdravorazumsku psihologiju nema mesta u savremenoj kognitivnoj nauci, odnosno da treba da dođe ontološki radikalne zamene teorija. Remzi i kolege naglašavaju da ne pružaju dokaz u prilog antecedensa i da ostavljaju otvorenim kako i da li će konekcionistički modeli napredovati. Pogledajmo, najzad, njihov argument, koji ću razložiti po premisama na sledeći način:

(DEF) Propozicijska modularnost = funkcionalna diskretnost *&* semantička interpretabilnost *&* kauzalna efikasnost

(1) Zdravorazumska psihologija *qua* teorija je obavezana na propozicijsku modularnost.

(2) Veštačka neuronska mreža u konekcionističkim modelima ne raspolaže funkcionalno diskretnim, semantički interpretabilnim reprezentacijama koje bi stvorile pojedinačne jedinice, već na osnovu aktivacionih šablona više jedinica „emergira“ distriburirana reprezentacija.

(3) Ako **(2)**, onda se ne može reći da su pojedinačne jedinice unutar mreže kauzalno efikasne, jer ne stvaraju kauzalno aktivne pojedinačne reprezentacije.

(4) Konekcionistički modeli su u koliziji sa propozicijskom modularnosti. (sledi iz **(DEF)–(3)**)

(5) Ako konekcionistički modeli predstavljaju vernu sliku ljudskih kognitivnih procesa, onda je zdravorazumska psihologija lažna teorija jer je obavezana na propozicijsku modularnost. (sledi iz **(DEF)–(4)**)
(6) Eliminativizam = teza prema kojoj će bolja, nova naučna teorija u nekom domenu zameniti staru teoriju koju nova čini empirijski lažnom za taj domen.
(7) Ako konekcionistički modeli predstavljaju vernu sliku ljudskih kognitivnih procesa, eliminativizam u pogledu zdravorazumske psihologije je tačna teza. (sledi iz **(5)–(6)**)

II Prigovori Remziju, Stiču i Geronu

Argument koji nude Remzi, Stič i Geron izazvao je dosta polemike. Ovde ću se fokusirati na dve kritike koje dovode u pitanje definiciju u **(DEF)** i premisu **(1)**, budući da smatram da su potencijalno najpogubnije po argument, jer pokušavaju da pokažu kako je konekcionizam kompatibilan sa zdravorazumskom psihologijom, odnosno da ne postoji razlog da se preduzme poslednji korak i prihvati eliminativizam. Na neke kritike koje su dovodile u pitanje premise **(2)** i **(3)**, ne samo što je uspešno odgovoreno (npr. odgovor Stich 1991 na O'Brien 1991), nego su takve kritike počivale na suštinskom nerazumevanju funkcionisanja konekcionističkih modela, na šta možemo blagonaklono gledati imajući u vidu da su u pitanju sami počeci konekcionističkog modelovanja, koje je sa sobom nosilo tehničke detalje, za koje je sigurno trebalo vremena da se usvoje i razumeju unutar filozofske zajednice.

Dalje, premisu **(3)** sam formulisala tako da se izbegne obavezivanje na to da li je nosilac kauzalne efikasnosti unutar neuronske mreže aktivirani šablon *per se* ili dispozicija mreže da stvara takve šablone, a samim tim i kritike koje idu u tom smeru (npr. Botterill 1994). Remzi i kolege imaju u vidu mogući prigovor koji se tiče toga da neuronske mreže mogu imati dispozicionalna verovanja, iako ne i okurentna, ukoliko je nosilac kauzalne efikasnosti dispozicija da se aktiviraju šabloni (1990: 139). Ali, za potrebe njihovog argumenta jedino je važno da *u principu* ne postoji način da se razluče kauzalno aktivna od kauzalno inertnih pojedinačnih stanja u mreži, jer reprezentacije nisu lokalizovane i izolovane, već distriburirane. Međutim, ako je premisa **(1)** lažna, a lažna je u slučaju da je i jedan konjunkt u **(DEF)** lažan, to povlači lažnost ključnog kondicionala u premisi **(5)** od kog zavisi čitav argument.[4]

4 Ovde treba primetiti da strategija pokazivanja da je **(1)** lažno ne mora da podrazumeva i kritiku konekcionizma *qua* hipoteze o ljudskoj kognitivnoj arhitekturi. Neko može potpuno konzistentno da osporava **(1)** i da veruje da je konekcionizam bolja alternativa od simbolizma kada se radi o prirodi ljudske kognicije. Većina filozofa se radije ne bi odrekla aparata propozicijskih stavova, i zaista većina pristalica konekcionističke kognitivne nauke iz filozofskih krugova je posmatrala konekcionizam kao obavezan na *intencionalni realizam*, to jest realizam u pogledu mentalnih repre-

Prva kritika kojom ću se baviti jeste kritika Horgana (Terence Horgan) i Tijensona (John Tienson), izložena u radu iz 1994. godine, kojom se insistira na tome da postoje različiti tipovi funkcionalne diskretnosti, iz čega se dalje izvlači zaključak da određeni tip funkcionalne diskretnosti nije nekompatibilna sa konekcionizmom, te da nema razloga za zastupanje eliminativizma. Prema skiciranom argumentu, Horgan i Tijenson bi pre svega imali problem sa definicijom propozicijske modularnosti u premisi **(DEF)**, a samim tim i sa premisom **(1)**, iz koje posle sledi premisa **(4)**, od koje počinje prelaz ka uvođenju eliminativizma. Naime, u argumentu bi došlo do greške ekvivokacije, jer Horgan i Tijenson koriste termin „funkcionalna diskretnost" na drugačiji način od Remzija i kolege, te bi samim tim premise **(DEF)** i **(1)** bile lažne, a argument ne bi bio zdrav.

Prvi korak Horgana i Tijensona je diferenciranje tri načina na koji ljudi mogu posedovati propozicijski sadržaj: **(i)** okurentno, **(ii)** dispozicionalno, **(iii)** morfološki. Prva dva načina su razmatrali i Remzi i kolege. Međutim, treći način je specifičan: neko poseduje propozicijski sadržaj *P* na morfološki način *M* onda kada je izložen stanjima koja sistematski odgovaraju *P*-u, ali ne usled doživljavanja okurentnog stanja kojim se instancira *P,* već zahvaljujući samoj *strukturi P*-a (cf. Horgan & Tienson 1994: 132). U slučaju konekcionističkog modelovanja, moglo bi se reći da morfološko posedovanje intencionalnog sadržaja ogleda u informacijama „otelotvorenim" u određenim stepenima jačine veza između jedinica. U odnosu na **(i)**, **(ii)**, **(iii)**, Horgan i Tijenson razlikuju pet tipova funkcionalne diskretnosti: **(I)** kauzalno aktivno mentalno stanje S_{ka} je *okurentno*; kauzalno inertno stanje S_{ki} je *dispozicionalno*, **(II)** S_{ka} je *okurentno*; S_{ki} je *okurentno*, **(III)** S_{ka} je *okurentno*, S_{ki} *morfološko*, **(IV)** S_{ka} je *morfološko*, S_{ki} je *okurentno*, **(V)** S_{ka} je *morfološko*, S_{ki} je *morfološko*.

Poenta dvojice filozofa je da je zdravorazumska psihologija obavezana samo na tip **(I)**, dok ostavlja otvorenim empirijsko pitanje da li se ostali tipovi manifestuju unutar propozicijskih stavova (cf. Horgan & Tienson 1994: 134-135). Dalje, konekcionistički modeli su kompatibilni sa tipom **(I)**, a to što mogu da se interpretiraju tako da daju negativan odgovor na empirijsko pitanje koje zdravorazumska psihologija ostavlja otvorenim, a koje se tiče ostalih tipova funkcionalne diskretnosti, ne znači da je konekcionizam *en général* nekompatibilan sa zdravorazumskom psihologijom.[5]

---

zentacija i intencionalnog idioma (to jest, propozicijskih stavova), u istoj meri u kojoj su simbolički modeli klasičan primer ontološke obavezanosti na intencionalni realizam. Intencionalni realizam je u isto vreme i Ahilova peta konekcionizma, jer je strategija ljubitelja simboličkih modela uglavnom uključivala potez upućivanja na to da je tako shvaćen konekcionizam u najboljoj meri biološki plauzibilnija *puka implementacija* simboličke arhitekture, a ne *rivalska* hipoteza o kognitivnoj arhitekture. Stoga čini se da je potrebno ponuditi i dokaz u prilog antecedensa u **(1)**, kojim bi se isključilo da su konekcionistički modeli puke implementacije. Time ću se baviti u odeljku **IV**.

5 Horgan i Tijenson idu toliko daleko da tvrde i sledeće: sve i kad bi zdravorazumska psihologija bila obavezana na ostale tipove funkcionalne diskretnosti, to što takvo nešto ne može da bude

Pogledajmo na osnovu čega Horgan i Tijenson misle da funkcionalna diskretnost tipa **(I)** može biti primećena u konekcionističkim modelima (cf. 1994: 136). Okurentna verovanja zdravorazumske psihologije izgledaju kao da mogu biti predstavljena tokenom aktivacionog šablona veštačke neuronske mreže, jer šabloni imaju kauzalni uticaj na procesiranje slično kao što okurentna verovanja imaju kauzalni uticaj na korpus verovanja koji već imamo, ali koji ulazi u igru tek onda kada je izazvan određenim trenutnim verovanjem. Dispozicionalna verovanja zdravorazumske psihologije najviše podsećaju na dispozicije neuronske mreže da „tokenizuje" aktivacione šablone, koji ukoliko ostanu inertni, jednostavno ne utiču na procesiranje, slično kao što korpus verovanja koji već imamo ne mora uticati na naše ponašanje u svakoj instanci.

Horgan i Tijenson direktno dovode u pitanje tvrdnju Remzija i kolega da je za potrebe njihovog argumenta jedino bitno da *u principu* ne postoji način da se razluče kauzalno aktivna od kauzalno inertnih stanja unutar distriburirane reprezentacije koja emergira iz ponašanja neuronske mreže. Dvojica filozofa, zapravo, preciziraju da, iako su dispozicije sadržane u stepenima jačine na holistički način, to ne znači da su one ikada implicirane u samom procesiranju – naprotiv, niti jedna dispozicija se ne nalazi u procesiranju (1994: 138). Specifične kauzalne uloge zadobijaju aktivacioni šabloni u zavisnosti od zadatka koji neuronska mreža treba da obavi, pri čemu se „tokenizuje" distriburirana reprezentacija *en masse*, i u tom smislu konekcionistički modeli takođe mogu da ispoljavaju paradigmatičnu funkcionalnu diskretnost na kakvu je obavezana zdravorazumska psihologija. „Bilo koji kognitivni proces koji se predstavlja kao neposredan za sistem, i koji se obavlja na osnovu jednog relevantnog inputa u određenom vremenskom trenutku, ispoljava funkcionalnu diskretnost prvog tipa", zaključuju Horgan i Tijenson (1994: 139). Ovo je, dakle, veoma slab vid obavezivanja na funkcionalnu diskretnost.

Međutim, argumentacija Horgana i Tijensona važi samo pod pretpostavkom da kauzalno inertna stanja neuronske mreže *nisu deo procesiranja*. Pa ipak, može se pokazati da ovo nije slučaj. Remzi je u tekstu iz 1992. godine rafinirao filozofske

---

deo naše najbolje teorije o ljudskoj kogniciji ne govori u prilog tome da propozicijski stavovi ne postoje, već da „zdravorazumska psihologija greši tu i tamo u pogledu propozicijskih stavova" (Horgan & Tienson 1994: 131). Ovakva ležernost je u najmanju ruku bizarna: ako zdravorazumska psihologija *qua* teorija nudi pogrešan *explananda* fenomena kojima se bavi, onda je sasvim opravdano tvrditi da tako opisani fenomeni ne treba da figuriraju u kognitivnoj nauci, ili još pre *da ne treba da postoje u okvirima kognitivne nauke*. Primera radi, zdravorazumska fizika *qua* teorija o našem svakodnevnom intuitivnom shvatanju regularnosti u prirodi uključuje i apsolutnu simultanost, međutim apsolutna simultanost je u okviru Einsteinove teorije specijalne relativnosti lažna, to jest uvek mogu postojati posmatrači za koje simultanost neće korespondirati u istom trenutku. Ovde izgleda da slobodno možemo reći da apsolutna simultanost ne treba da postoji u okvirima Einsteinove fizike, uprkos tome što se u svakodnevnom životu oslanjamo na takve konstrukte. Razviću ovu poentu predstavljajući eliminativizam ograničenog uticaja u narednim odeljcima.

implikacije distribuiranih reprezentacija, primetivši da u zavisnosti od stanja aktivnosti mreže, odnosno toga da li je mreža aktivna ili inertna, zavise i njene strukturne odlike. Tako, model koji je predstavljen u koautorskom radu sa Stičom i Geronom, a koji bi služio kao dokaz antecedensa u **(7)**, interpretira se kao da odslikava *strukturni holizam* u pogledu mentalnih reprezentacija.[6] Prema Remziju, kada se mreža posmatra kao inertni sistem, najmanja reprezentaciona struktura je *čitava* mreža, i tada su sve informacije pohranjene u parametrizovanim vezama između *svih* jedinica, odnosno „tragovi najrazličitijih mentalnih stanja se mogu shvatiti kao naslagana jedna na druga u parametrima" (Rumelhart & McClelland 1986, cit. prema Ramsey 1992: 267). Međutim, struktura mreže se menja s aktivacijom sistema, ali prema Remziju upravo bi bilo pogrešno krenuti putem zdravorazumske psihologije i tvrditi da neka od „naslaganih" stanja neće imati kauzalni uticaj na trenutno ponašanje mreže, odnosno da će neka stanja biti „tokenizovana" i uključena u procesiranje, a neka neće.

Šta Remzi zapravo hoće da kaže jeste da ne treba mešati različitost strukture mreže u različitim trenucima aktovnosti sa različitim tipom procesiranja. Različitost strukture uslovljava različit odnos delova prema celini, ali procesiranje signala uvek teče prema algoritmima za učenje na relativno predvidljiv način. Ne postoji način da utvrdimo koja „naslagana" stanja jesu, a koja nisu „tokenizovana" prilikom procesiranja, jer distribucija reprezentacije *emergira* na osnovu *svih* pojedinačnih parametara, ali nije uzrokovana *nekim* od pojedinačnih parametara. Prema tome, ni slaba funkcionalna diskretnost ne može da se pripiše konekcionističkim modelima, naprosto jer je teret dokazivanja na Horganu i Tijensonu da pokažu da *određene* informacije u *pojedinačnim* parametrima bivaju tokenizovane u odnosu na kognitivni zadatak. Šta se dešava u konekcionističkom modelu je da rešenje kognitivnog zadatka sledi iz *kompletnog korpusa* informacija dostupnih neuronskoj mreži.[7]

---

6 Model koji su konstruisali Remzi i kolege sastojao se iz šesnaest jedinica koje kodiraju ulazne signale, četiri skrivene jedinice i jedne jedinice koja kodira izlazni podatak. Ulazni signali su bili iskazi o anatomskim svojstvima pasa, mačaka i riba, a zadatak neuronske mreže je bio da dodeli istinosnu vrednost tim iskazima tako što će u slučaju izlazne jedinice sa stepenom jačine od 0,9 pa naviše obeležiti sa „istina", a sa stepenom jačine ispod 0,1 obeležiti sa „laž". Mreža se obučavala na osnovu algoritma za učenje koji propagira grešku unazad, to jest u slučaju pogrešnog dodeljivanja istinosne vrednosti u poslednjem, trećem sloju, signal o grešci se šalje unazad i shodno tome se naknadno obrađuju informacije u prvom, ulaznom sloju. Cilj modela je bio da prikaže simplifikovan vid pohranjivanja informacija kod ljudi. Metodološke specifičnosti modela ću komentarisati u odeljku **III**, ali budući da nameravam da se osvrnem na daleko ozbiljnije savremene konekcionističke modele u odeljku **IV**, ovaj „model za igranje" (eng. *toy model*) Remzija i kolega nije od suštinskog značaja za evaluiranje njihovog argumenta.

7 Ova linija argumentacije će biti dodatno podržana recentnim konekcionističkim modelima kojima se bavim u odeljku **IV**, jer takvi modeli se obučavaju na osnovu fascinantno velikog broja podataka i sadrže daleko više slojeva, jer im je kompjutaciona moć značajno veća. U takvom moru podataka koji se „ubacuju" u ulazne slojeve jedinica, nakon kojih sledi niz drugih slojeva,

Čini mi se da koren problematičnosti Horganove i Tijensonove kritike leži u tome što nisu raskrstili sa nasleđem simboličke kognitivne nauke, u čijoj osnovi je reprezentacijska teorija uma. Naime, prema reprezentacijskoj teoriji uma, verovanja bi se shvatila kao instancirana sekvenca tokena, koja su prethodno bila pohranjena u vidu rečenica. Instancirana sekvenca tokena verovanja predstavlja okurentne misli, dok su pohranjenje rečenice trajna verovanja kognitivnog sistema i nisu direktno uključena u svesni proces mišljenja. Horgan i Tijenson su ovakvu shemu primenili i na konekcionizam, samo zamenjujući terminologiju: instancirana sekvenca tokena je postala „tokenizovani aktivacioni šablon", a pohranjena verovanja u vidu rečenica su postala „dispozicije u stepenima jačine".

Frensis Igan (Frances Egan) smatra da bi argument Remzija i kolega mogao biti zdrav samo ukoliko bi se ispostavilo da je zdravorazumska psihologija obavezana na to da su propozicijski stavovi realizovani putem diskretnih kompjutacionih struktura, kao i na to da su kauzalno efikasni unutar generalizacija – za šta postoje indicije da nije slučaj (1995: 184). Drugim rečima, Iganova bi se protivila **(DEF)** i **(1)**: konjunkt funkcionalne diskretnosti i kauzalne efikasnosti u definiciji propozicijske modularnosti predstavljaju „slabe karike" **(DEF)**, na osnovu čije lažnosti pada i **(1)**. Njena kritika počiva na strategiji pokazivanja da zdravorazumska psihologija ne postavlja nikakva supstancijalna ograničenja hipotezi o kognitivnoj arhitekturi. Prema tome, i Iganova smatra, poput Horgana i Tijensona, da konekcionizam nije nekompatibilan sa zdravorazumskom psihologijom, tako da nema potrebe za plivanjem u opasnim eliminativističkim vodama.

Iganova predstavlja vlastito minimalističko viđenje zdravorazumske psihologije (1994: 187): to je teorija koja pruža okvir za opisivanje i razumevanje propozicijskih stavova pozivanjem na tipične uzroke i posledice, u odnosu na koje se propozicijski stavovi semantički evaluiraju i kauzalno specifikuju. Ali, fizička ili kompjutacijska realizacija se naizgled uopšte ne moraju pomenuti prilikom teoretisanja o propozicijskim stavovima! Stoga, prema Iganovoj, hipoteza o kognitivnoj arhitekturi je dodatna i nezavisna pretpostavka. U prilog tome, ona navodi niz studija iz razvojne psihologije kako bi argumentovala da „ ne postoji nikakva evidencija kojom bi se podržalo gledište da je razlog za to što mala deca imaju poteškoća sa pripisivanjem verovanja drugima u vezi sa neznanjem u pogledu detalja kognitivne arhitekture (...)" (1995: 189). Šta Iganova ovde zapravo želi da kaže jeste da ne postoji evidencija u razvojnoj psihologiji zahvaljujući kojoj bi se moglo tvrditi da deca imaju verovanja o tome kako (i gde) se kauzalno efikasna stanja realizuju.

---

nema reči o pronalaženju načina da se „tokenizuje" određeni šablon, ili da se govori o funkcionalnoj diskretnosti. Naravno, ne mogu zameriti Horganu i Tijensonu što nisu konsultovali augure da bi saznali kako će izgledati konekcionistički modeli u budućnosti, ali čini mi se da je već devedesetih godina bilo relativno jasno da konekcionizam ne može biti „stišnjen" u kutiju sa funkcionalnom diskretnosti samo kako bi se bar donekle očuvala aparatura propozicijskih stavova, koja se pokazala tako korisnom u drugim filozofskim domenima.

Međutim, ukoliko se prihvati linija rezonovanja Iganove, ispalo bi da svako ko se teorijski obavezuje na neku vrstu kognitivne arhitekture (konekcionisti na konekcionističku, simbolisti na simboličku) ujedno mora i da tvrdi da svaka osoba ima svesno verovanje o tome da ima takvu-i-takvu arhitekturu. Prvo, isključimo trivijalan smisao koji bi ovakva linija rezovanja imala: naravno da kognitivni naučnik ili filozof koji veruje da je, recimo, konekcionizam pravi kandidat za kognitivnu arhitekturu, ujedno veruje da je i njegov vlastiti kognitivni sklop konekcionistički – naprosto njegov *credo* ima takvu posledicu ukoliko je u pitanju racionalna osoba. Šta sledi iz tvrdnji Iganove je da kognitivni naučnik ili filozof, koji veruje da je konekcionizam nekompatibilan sa zdravorazumskom teorijom, *unapred* pretpostavlja posedovanje metaverovanja o određenoj fizičkoj realizaciji kauzalno efikasnih stanja, a onda to „kalemi" na zdravorazumsku psihologiju i tvrdi da je to metaverovanje upravo glavni krivac za nekompatibilnost. Potom, takav naučnik ili filozof prosvećeno prihvata eliminativizam, jer izgleda krajnje *kontraintuitivno* da je potrebno imati verovanje o kognitivnoj arhitekturi kako bismo uopšte bili u mentalnom stanju verovanja, ili umeli da drugima pripišemo to mentalno stanje.

Ali, to izgleda kontraintuitivno jer *jeste* kontraintuitivno. Nigde kod Remzija i kolega nismo videli pozivanje na neophodnost metaverovanja u pogledu kognitivne arhitekture, niti bi to trebalo očekivati od bilo koga (osim u trivijalnom smislu), jer bi onda ispalo da čitava kognitivna nauka počiva na jednom *petitio principii*. Štaviše, ovakvo rezonovanje Iganove podseća na već viđene strategije argumentovanja u filozofiji duha: protiv teorije psihofizičkog identiteta se potezao kontraargument da laici nemaju nikavo znanje o finesama neurofizioloških procesa, *ergo* moždani procesi ne mogu biti jednaki mentalnim stanjima, o kojima imamo znanje iz prve ruke. Međutim, takvi napadi su se i pre odbijali (Smart 1962). Na primer, ne bismo rekli da to što laici nemaju medicinski precizno verovanje o tome da srce funkcioniše na takav-i-takav način znači da srce *de facto* ne funkcioniše na takav-i-takav način. Zašto onda zahtevati metaverovanje od laika u pogledu instancirane kognitivne arhitekture, koja mu omogućava da uopšte ima verovanja i metaverovanja, kao uslov za bilo koju teoriju kognicije? Izgleda kao da metodologija kognitivne nauke sugeriše sasvim suprotno od Iganove: zdravorazumska psihologija je dodatna i nezavisna pretpostavka u odnosu na kognitivnu nauku; a teoretisanje o fizičkoj i kompjutacijskoj realizaciji, barem u slučaju konekcionizma, može da funkcioniše i bez kostura zdravorazumske psihologije.

Stoga, čini mi se da ni napad Iganove na Remzija i kolege ne izgleda dovoljno ubedljivo. Njeno osporavanje **(DEF)** i **(1)**, odnosno insistiranje na tome da konekcionizam nije nekompatibilan sa zdravorazumskom psihologijom, počiva na misinterpretaciji metodologije kognitivne nauke. Odgovorivši na dve kritike uperene protiv Remzija i kolega, pokazala sam da je moguće braniti ključni deo argumenta kojim se uspostavlja nekompatibilnost zdravorazumske psihologije sa konekcionizmom, tako da su se stvorili uslovi za dalje razvijanje njihovog argumenta, i proveravanje koliko je u skladu sa savremenim tokovima konekcionističkog modelovanja.

III Još jedan eliminativista ceni vrline konekcionizma: Čerčlandovi argumenti

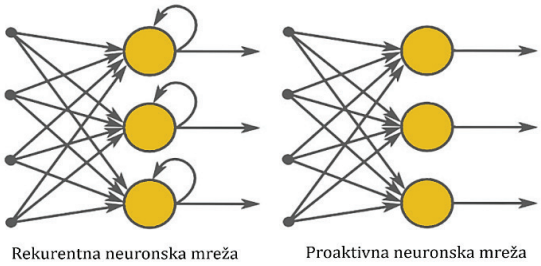
Pre nego što pređem na opisivanje recentnih konekcionističkih modela kako bih podržala antecedens u **(7)**, važno je motivisati eliminativizam uveden u premisi **(6)**.[8] Pogrešno bi bilo misliti da su Remzi i kolege bili na margini filozofske zajednice usled zastupanja eliminativizma, što je utisak koji se može steći na osnovu kritika upućenih njihovom argumentu, budući da je zajednička crta svih kritika izbegavanje eliminativizma po svaku cenu, pa makar i očuvanje aparature propozicijskih stavova u konekcionizmu – za kojom je pitanje da li ima stvarne potrebe. Štaviše, u kasnijim radovima dvojica od inicijalnog tima su odustali od eliminativizma. Remzi (2017) je počeo da dovodi u pitanje osnovnu teorijsku pretpostavku kognitivne nauke da je kognicija suštinski reprezentaciona, a samim tim i intencionalni realizam u filozofiji duha. Stič je već 1994. godine počeo da se premišlja – prve zvanične sumnje u pogledu eliminativizma se mogu naći u Stich & Ravenscroft (1994). Ali, dve godine kasnije, u prvom poglavlju knjige *Deconstructing the Mind*, Stič je priznao da je već prilikom sređivanja koautorskog mansukripta za publikovanje uvideo da je previše naivno pristupao eliminativizmu, a da ga je potom iz „dogmatskog dremeža" trgnula primedba Vilijema Lajkana (William Lycan) da eliminativistički zaključci izneseni u publikaciji slede samo pod pretpostavkom deskriptivne teorije referencije, koju je sam Lajkan kritikovao iz perspektive kauzalne teorije referencije.

Naravno, to što su sami autori odustali od zastupanja eliminativizma ne treba da obeshrabri pokušaje da se vrednost njihovog argumenta sagleda u novom svetlu, ali to jeste ključni razlog zašto se osvrćem i na Čerčlandov optimizam u pogledu potencijalnog konekcionističkog eliminisanja entiteta zdravorazumske psihologije – kako bih pokazala da nisu svi bili spremni da tako lako odustanu od argumenta Remzija i kolega uprkos tome što su oni sami to uradili. Pol Čerčland je počeo da argumentuje u prilog ove teze (i protiv zdravorazumske psihologije) još krajem sedamdesetih, premda se tek njegov tekst iz 1981. godine obično uzima kao manifest eliminativizma. Paralelno sa zastupanjem eliminativizma, Čerčland (1989) piše i niz gotovo egzaltiranih radova o konekcionizmu.

Štaviše, Čerčland je upravo u pojavi konekcionizma video priliku koju je najavljivao zastupajući eliminativizam, to jest da će doći do ontološki radikalne zamene

8 Budući da se specifično bavim vezom između konekcionističke kognitivne nauke i eliminativizma, neću se osvrtati na generalne filozofske prigovore protiv eliminativizma, ili na odbrane od takvih prigovora (za pregled v. Ramsey 2019: odeljak 4), za potrebe uspostavljavnja ove veze važno je samo to da je eliminativizam *odbranjiv*, tako da postoji osnova za načelno prihvatanje konekcionizma. Štaviše, u ne-kompjutacionim teorijama kognicije, koje se oslanjaju na utelovljene, proširene i enaktivističke pristupe, i koje samim tim počivaju na ne-reprezentacijskim pretpostavkama, eliminativizam je prirodna posledica, pre nego li kontroverzna filozofska teza (upor. Ramsey 2019: pododeljak 3.2.3).

teorija o ljudskim kognitivnim procesima ili mentalnim stanjima kako s kognitivna i neuronauka budu razvijale. Prema njegovom mišljenju, zdravorazumska psihologija je degenerativan istraživački program u lakatoševskom smislu jer je priča o zdravorazumskoj psihologiji zapravo priča o „stagnaciji, jalovosti i dekadenciji" (Churchland 1981: 74): tokom vekova, nije bilo moguće proširiti domen za koji važi, tako da su fenomeni poput sanjanja, imaginacije ili mentalnih bolesti i dalje van domašaja zdravorazumske psihologije, kao i slučajevi ne-ljudskih subjekata.[9]



Ilustracija 3.1. Tipovi neuronskih mreža.

S druge strane, Čerčland (1989) vrlinu konekcionizma vidi upravo u tome što kognitivna arhitektura, koja se pretpostavlja konekcionističkim modelima, može biti instancirana i u ljudskom i ne-ljudskom telu. Naime za razliku od simboličkih modela, koji se obavezuju na propozicijsku strukturu kognicije, a samim tim fokusiraju isključivo na organizme koje odlikuje lingvistička kompetencija, konekcionistički modeli su dovoljno apstraktni i opšti da mogu opisivati i kogniciju ne-ljudskih životinja jer su biološki plauzibilni. Pored toga, Čerčland (2007) u primenjivanju *rekurentne neuronske mreže* unutar modela vidi novi način za objašnjavanje svesti, odnosno pravljenje principijelne demarkacije između svesnih i nesvesnih mentalnih stanja.[10]

---

9 Clark (1987) nudi zanimljiv protivargument Čerčlandu kada se radi o teorijskoj stagnaciji zdravorazumske psihologije: teorijski okvir koji se oslanja na propozicijske stavove nije artefakt prošlog vremena, niti produkt eksternh faktora (kulture, životne sredine, ustrojstva društva, itd.), već je intrinsični deo ljudske prirode, to jest *urođen* je. Konekcionizam bi u takvom kontekstu još više bio shvaćen kao blizak eliminativizmu, budući da su konekcionizmu najnaklonjeniji antinativisti: ključni korak u uspostavljanju autonomije konekcionističkih modela u odnosu na simboličke modele sastoji se u pokazivanju da neuronske mreže pre uče na osnovu „iskustva" sa skupovima podataka nego li usled unapred programiranih pravila. Prema tome, konekcionizam može biti interpretiran kao nekompatibilan sa zdravorazumskom psihologijom u daleko jačem smislu: radilo bi se, dakle, o *ontološki kontradiktornim obavezama* dvaju teorija.

10 Laakso & Cottrell (2005) navode i niz drugih prednosti koje, prema Čerčlandovom mišljenju, konekcionizam ima u odnosu na tradicionalnu simboličku kognitivnu nauku, odnosno reprezentacijsku teoriju uma, i te prednosti grupišu prema filozofskim disciplinama u kojima se instanciraju odgovarajuće verzije pomenutih teorija kognicije odnosno uma. Tako, recimo,

Remzi i kolege su imali u vidu konekcionističke modele koji su implementirali *proaktivne neuronske mreže*. Kada se implementira ovakav tip neuronske mreže, kodirane informacije teku od ulaznih ka izlaznim jedinicama posredstvom slojeva u kojima su skrivene jedinice. Međutim, ove neuronske mreže su se pokazale vrlo problematičnim zbog toga što su im sposobnosti učenja i procesiranja informacija u kratkom vremenskom roku krajnje ograničene. Stoga devedesetih godina u igru ulaze *rekurentne neuronske mreže*. Implementacijom jednostavnih rekurentnih mreže, postiže se sledeće: u svakoj iteraciji procesiranja informacija, aktivacioni prag skrivenih jedinica se preslikavaju u tzv. jedinice konteksta, tako da se u sledećoj iteraciji ove jedinice kombinuju sa aktivacionim signalima koji potiču od ulaznih jedinica, čime se iznova aktiviraju skrivene jedinice. Na taj način se stvara *kratkoročna memorija* mreže što zazvrat omogućava vernije simuliranje kognitivnih zadataka koji zahtevaju sekvenciranje i temporalnost.

Pol Čerčland smatra da, u onom momentu kada pažnju posvetimo aktivnostima reprezentacija namesto reprezentacijskom sadržaju, do izražaja dolaze sledeće ključne sposobni mozga: sposobnost fokusiranja na određeni podskup senzornih stimulusa, pojmovne interpretacije tog podskupa, kratkoročno pamćenje putem kog se interpretacija dovoljno dugo zadržava da bi se pobudila sposobnost obnavljanja reprezentacijskog „narativa" koji stvaramo o svetu iz kog dolaze stimulusi (2007: 12). Upravo ove sposobnosti mozga sugerišu, prema Čerčlandu, da je svest opisiva pomoću instancirane rekurentne neuronske mreže. Drugim rečima, dok proaktivna neuronska mreža može da pruži samo reakciju na *trenutno* pronađeni i fiksirani šablon na osnovu ulaznih signala, rekurentna mreža pamti i po potrebi reaguje u vremenu na *kauzalni* niz *zapamćenih* aktivacionih šablona, čime se pozadinsko znanje uvek može „osvestiti", kao i kontrolisati šta je od tog pozadinskog znanja relevatno za sadašnji trenutak.

Na ovaj način, ponašanje veštačke neuronske mreže se može videti kao *kontinualna funkcija* kako svih ulaznih signala, odnosno perceptivnih stimulusa, tako i trenutnog praga aktivacije (Churchland 2007: 14). Ovo zauzvrat znači da je buduće ponašanje prilično pouzdano predvidljivo u periodima od nekoliko sekundi pa sve do

---

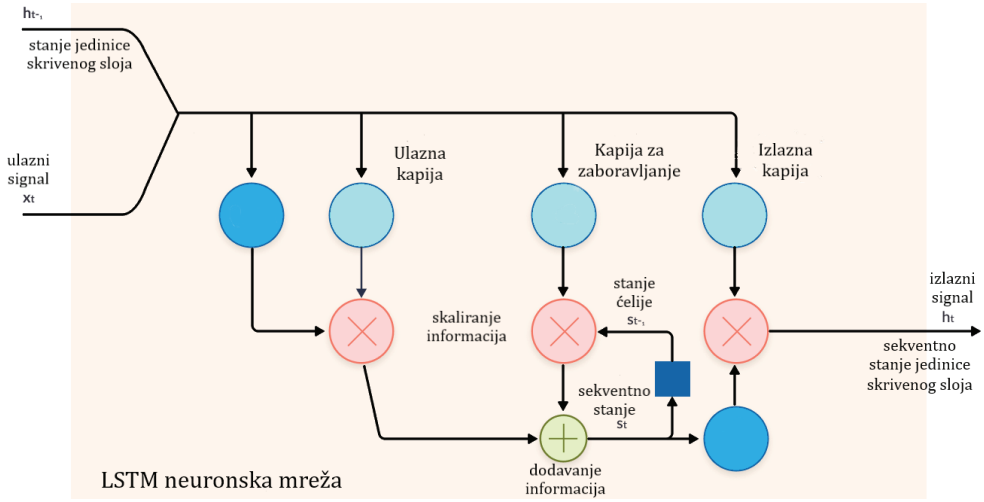
unutar *filozofije nauke*, konekcionizam predstavlja alternativu deduktivizmu, i to zahvaljujući tome što dolaženje do objašnjenja u nauci može da se modeluje putem aktiviranja vektora prototipa namesto praćenja sleda logičkih inferencija. Drugim rečima, unutar konekcionizam parira naučniku koji traži objašnjenje jer on znanje o fenomenima obogaćuje i uvećava putem generalizacija, što je slično kontekstno-senzitivnom procesiranju neuronske mreže koja se kalibriše nakom obučavanja na novom materijalu, a naučni uvid se može razumeti preko brzine paralelnog procesiranja u neuronskoj mreži (Laakso & Cottrell 2005: 120-125). Posebno je interesantno to da je Čerčland pokušavao da primeni konekcionizam i u domenu *etike*. Naime, on je tvrdio da je konekcionizam pogodan čak i za shvatanje odnosno modelovanje i toga kako dolazimo do etičkih i legalnih objašnjenja, jer bi se etičke i socijalne smernice u nekom društvu posmatrale kao protipi koji se formiraju na osnovu situacija i životnog iskustva, a ne kao utvrđene norme (Laakso & Cottrell 2005: 126).

nekoliko dana ili nedelja, kao što je redovno leganje u krevet u jedanaest noću ili obedovanje triput dnevno, ali modelovani kognitivni sistem postaje spontan kada se radi o specifičnim ponašanjima u bilo kom trenutku dalje budućnosti (Churchland 2007: 15). Ovakvo emergiranje spontanosti je za Čerčlanda, u stvari, krajnji dokaz superiornosti konekcionističkih modela – sanjarenje, imaginacija, i sve ostale finese ljudskog ponašanja koja izmiču zdravorazumskoj psihologiji bivaju inkorporirane u konekcionizam zahvaljujući rekurentnim neuronskim mrežama. Stoga, čini se da Čerčlandu njegov eliminativistički *credo* u dvadeset prvom veku deluje kao još prirodniji i neizbežniji usled postepenog razvoja konekcionizma: sve ono u pogledu čega su se mnogi filozofi kolebali da konekcionizam može da objasni bez zdravorazumske psihologije izgleda kao da nije van domašaja rekurentne kognitivne arhitekture. U narednom odeljku ćemo videti da Čerčlandovo poverenje u metodološke novitete konekcionizma ne predstavlja ni lošu opkladu niti naivni optimizam.

IV Da li će zdravorazumska psihologija uspeti da preživi nove neuronske mreže?

Krajem devedesetih godina primećen je ozbiljan nedostatak rekurentnih mreža, koji se naziva u literaturi *problemom gradijenta* (Staudemeyer & Morris 2019: 17-18, Hochreiter & Schmidhuber 1997: 1735-1736), i tiče se sveprisutne pojave da gradijent, čiji intenzitet produkuje najveće promene u polju aktivacionih šablona neuronske mreže, prolazeći kroz slojeve postaje sve manji dok se ne zaokruži na nultoj vrednosti i ne nestane, ili pak dok eksponencijalno poraste. Ovaj problem nastaje kada se akumuliraju greške u procesiranju, i postanu cirkularne usled strukture rekurentne neuronske mreže i primene algoritma propagiranja greške unazad. Stoga, istraživači veštačke inteligencije (Hochreiter & Schmidhuber 1997) konstruišu neuronske mreže dugog kratkoročnog pamćenja (LSTM, od engleskog *long short-term memory*), koje su uspešnije u odnosu na rekurentne neuronske mreže u pogledu rešavanja zadataka prepoznavanja vremenski trajnijih i kompleksnijih šablona, kao i vremenskog sleda naizgled nepovezanih događaja, kao i robusnog pohranjivanja informacija tokom dužih vremenskih perioda nego što je to bilo moguće sa rekurentnim mrežama. Ovakve mreže su našle svoju primenu u kontrolisanju senzorno-motornih reakcija robota, prepoznavanju rukopisa i glasova, prevođenju jezika, predviđanju cene hartija od vrednosti na berzi, i generalno za sve zadatke koji se zasnivaju na *temporalnim sekvencama*, odnosno zahtevaju klasifikaciju i predikciju podataka na kojima se mreža obučava.[11] Struktura LSTM mreže je sledeća (Hochreiter & Schmidhuber 1997: 6-9,

11 LSTM mreže su postale popularne van akademskih krugova 2014. godine, kada su kompanije poput Gugla, Fejsbuka i Majkrosofta počele da ih naširoko koriste u svrhe smanjivanja greške pri prevođenju, prepoznavanje glasa korisnika, ili za obučavanje virtualnih asistenata Siri i Alekse. To je za posledicu imalo intenziviranje akademskog rada na LSTM u poslednje dve godine.



Ilustracija 4.1. Proces obrade informacija u LSTM neuronskoj mreži.
Materijal upodobljen sa internet adrese:
https://adventuresinmachinelearning.com/keras-lstm-tutorial/

Staudemeyer & Morris 2019: 1755): jedinice unutar nje, pored ulaznih i izlaznih logičkih kola, imaju i kolo za zaboravljanje, kao i ćeliju, koja je zadužena za pamćenje informacija, čiji tok je regulisan upravo trima kolima. Ulazno i izlazno kolo za neku ćeliju S_t može ujedno da koristi i ulazne signale raznih drugih ćelija unutar mreže, kako bi se procenili efekti pohranjivanja neke informacije u S_t. Sva kola se aktiviraju sigmoidnom funkcijom.

Ukratko ću predstaviti dva modela koji koriste LSTM u različite svrhe. Važno je napomenuti da je prvi model bliži istraživačima veštačke inteligencije, a drugi neuronaučnicima pre nego kognitivnim naučnicima, i da su metodološke inovacije u konekcionizmu dvadeset prvog veka mahom potekle od ekspertize ljudi iz druge branše. Međutim, oba modela imaju važne kognitivne posledice, i neposredno se tiču i hipoteze o tome koje viđenje kognicije je ono „pravo“. Prvi model o kome će biti reči jeste model *generisanja prirodnog jezika*, što je komponenta sistema koji mogu da održavaju dijalog, kao što su virtualni asistenti Siri i Aleksa. Kembrički informatičari Ven (Tsung-Hsien Wen) i Jang (Steve Young) su nameravali da konstruišu konekcionistički model koji će biti konkurentski modelima baziranim na unapred kodiranim pravilima, budući da su se ovakvi modeli previše rigidnim za obuhvatanje prirodnosti ljudske jezičke aktivnosti. Kombinujući LSTM sa rekurentnim mrežama, ovaj dvojac programira neuronsku mrežu koja generiše izričaje direktno iz samog dijaloga (2019: 1). Naime, efektivnost mreže je evaluirana na osnovu četiri scenarija: dijalog u pogledu pronalaženja restorana ili hotela, i kupovanja laptopa ili televizora. Svaki od dijaloga treba da reprezentuje intenciju koju modelovani sistem prenosi. Ljudski ispitanici su pronađeni putem servisa

Amazon Mechanical Turk, i njihov zadatak je bio da prvo predlože odgovarajuću realizaciju svakog od scenarija u terminima njihovog maternjeg jezika, a onda da evaluiraju ponašanje neuronske mreže, koje se evaluiralo takođe i analizom korpusa (Wen & Young 2020: 11). Specifična uloga LSTM sastojala se u tome da osigura generisanje isključivo onih izričaja kojima se intendira značenje, i to tako što će semantički kontrolisati korpuse na kojima se mreža obučava da ne bi došlo do preterane generalizacije. Rezultati pokazuju da su ispitanici ocenili dijaloge koji emergiraju iz konekcionističke modela baziranog na LSTM arhitekturi kao *informativnije* i *prirodnije* za razliku od rivalskih modela, kao i da više preferiraju te modele jer im deluje da bi *mogli da se uključe* u takav dijalog (Wen & Young 2020: 19).

Drugi model se, pak, tiče klasifikacije mentalnih stanja preko obučavanja neuronske mreže na skupovima podataka iz elektroencefalograma (EEG) a pomoću tehnike dubokog učenja.[12] Duboko učenje je statistička tehnika „izvlačenja" šablona od strane neuronske mreže na osnovu značajnog broja podataka i značajnog broja slojeva jedinica. Specifičnost LSTM u ovom slučaju jeste da u kombinaciji sa dubokim učenjem, ove mreže sa 97,5% preciznosti *predviđaju* u kom mentalnom stanju se nalaze ljudski ispitanici, o kojima su informacije dobijene zahvaljujući elektroencefalografijama (Dutto & Nandy 2020: 181-183).

Oba modela imaju kognitivne posledice: u slučaju prvog, slobodno možemo reći da sugeriše da razumevanje intendiranog značenja reči sagovornika nije urođeno, već takav kognitivni proces, prema ovom modelu, ima važan razvojni aspekat pre nego nativistički; dok drugi model sugeriše da naša mentalna stanja mogu biti dostupna i transparentna i trećem licu, uprkos tradicionalnom internalističkom insistiranju na privilegovanom pristupu prvog lica. Međutim, još simptomatičnije jeste to da je funkcionisanje oba modela objašnjeno u terminima „predviđanja", ili „prirodnosti", uz gotovo otvorenu antropomorfizaciju prvog modela od strane ispitanika. Pa ipak, nigde unutar opisa modela i strukture LSTM nije bilo ni mesta ni potrebe za pomen entiteta zdravorazumske psihologije.

Sada je, najzad, moguće i uvesti poziciju *eliminativizma ograničenog uticaja*, prema kojoj zdravorazumska psihologija nije kompatibilna sa konekcionizmom, i konekcionizam *per se* nema potrebu za enitetima zdravorazumske psihologije, o čemu svedoče recentni modeli, koji inkorporiraju LSTM. Međutim, to ne znači da zdravorazumska psihologija ne može biti korisna verbalna heuristika, koju svakodnevno koriste i laici i eksperti: kako Watson (2019) primećuje, postoji izražena tendencija da se funkcionisanje algoritama neuronskih mreža opisuje preko antropomorfnih termina, što za njega ima loše posledice, jer skreće pažnju sa etičkih izazova koje postaju prominentne sa napretkom u istraživanjima veštačke inteligencije. Ovakav način go-

12 EEG meri u mikrovoltima moždanu električnu aktivnost ispitanika koji na sebi imaju elektrode. Rezultat testiranja je *elektroencefalografija*, ili dijagram koji odražava moždane talase.

vora je u osnovi ontološki neutralan, odnosno ne kaže ništa o tome da li postoje entiteti zdravorazumske psihologije (slično predmetno neutralnoj analizi odavno zastupanoj u Smart 1962), stoga se otvara mogućnost da se i u humanističkim naukama govori o verovanjima i željama, bez da se postuliraju entiteti.

Kako bih približila ovu ideju filozofiji duha, poslužiću se uvidima Robert Šarpi (Robert Sharpe) i Endi Klark (Andy Clark), koji su – u pokušajima da sačuvaju zdravorazumsku psihologiju od napada iz pravaca koji su im bili bliski i prihvatljivi – pružili odlična sredstva za argumentovanje protiv takvog čuvanja zdravorazumske psihologije po cenu bilo čega. Šarpi (1987: 392) primećuje da metafore i bogat mentalistički vokabular sugerišu da postoji korpus verovanja koje određene kulture imaju o ponašanju njihovih članova; i taj korpus verovanja zauzvrat može biti informisan teorijama iz psihologije, sociologije, ili ekonomije na takav način da se stvori teorijski okvir oko mentalističkog vokabulara. Međutim, zapravo se radi o tome da su se neki tehnički termini samo ustalili u običnom govoru (poput, recimo, Frojdovo nesvesno). Očigledno je da je takav korpus verovanja određene kulture informisan disciplinama psihologije, sociologije ili ekonomije jer je predmet istraživanja pomenutih disciplina, i odatle potiču zahtevi, kao što je Šarpijev, da s jedne strane nauka ipak morak biti u skladu sa zdravim razumom, a sdruge strane da nisu svi delovi zdravorazumske psihologije teorijski poziti. Klark (1987: poglavlje 10), sa svoje strane, tvrdi da je istinitost zdravorazumske psihologije nezavisna od rezultata borbe između konekcionizma i simboličke kognitivne nauke, to jest da je zdravorazumska psihologija rezervisana za „potpunu drugačiju arenu".

Sklona sam da se složim sa Šarpijem da nisu svi delovi zdravorazumske psihologije poziti – dobar deo zdravorazumske psihologije se može posmatrati kao verbalna heuristika koja nam olakšava da razumemo subjekte koji nismo mi, i koji su nam udaljeni bilo vremenski, bilo prostorno, bilo biološki. I u tom smislu je i Endi Klark u pravu kada naglašava da zdravorazumska psihologija pripada drugoj areni u odnosu na kognitivnu nauku. Ali, onaj deo zdravorazumske psihologije koji jeste teorijski može biti eliminisan iz konekcionizma. Teško je uvideti kako se istinitost zdravorazumske psihologije može očuvati unutar ontološki kontradiktorne hipoteze o ljudskoj kogniciji, jer, podsetimo se, autonomija konekcionizma je bazirana na razvojnom i empirističkom aspektu – nasuprot nativističkih tendencija kako simboličke kognitivne nauke, tako i reprezentacijske teorije uma. Prema tome, teorijski pristupi eksplanandumu izgledaju kao dijametralno suprotni i nekompatibilni. Pritom, ne treba mešati eksplanatornu vrlinu jednostavnosti sa zahtevom da eksplanandum mora biti zdravorazumski: ne samo što jedno iz drugog ne sledi, nego nikome ne bi palo na pamet da takav zahtev postavi pred fizičara koji se bavi domenima koji izmiču prostodušnosti naivne, zdravorazumske fizike, kao što su teorija struna ili standardni model elementarnih čestica. Zašto bi se onda takav zahtev smatrao legitimnim u kognitivnoj nauci ili neuronauci?

Zaključak

Osamdesetih godina je izgledalo da će kognitivna nauka uspeti da izgradi most između suprostavljenih „dvaju kultura“ humanistike i empirijske nauke, koje su se udaljile usled sve veće specijalizacije polja bavljenja naučnikā. Razlog za takva očekivanja je predstavljala integracija zdravorazumske psihologije u teorijski okvir simboličke kognitivne nauke, koji je metodološki usmeravao modelovanje kognitivnih procesa.

Katonovo završavanje svakog govora sa sada naširoko poznatim *Ceterum censeo* moglo je da se razume ne samo iz istoriografske ili etnološke perspektive, već i na osnovu kognitivnih mehanizama koji leže iza njegovih verovanja i želja, kao i naših verovanja i želja. Kognitivni kontinuum između nas i Katona ispoljavao bi se, prema tome, u kompjutaciono-reprezentacijskom karakteru naših kognitivnih procesa: kombinatorijalna sintaksa i semantika, koje uređuju naša mentalna stanja, i čine ih dostupnim kako nama tako i kognitivnim naučnicima čija su primarna briga, održavaju specifičnost ljudske vrste ma koliko da su istorijske epohe i divergentna društva različiti.

Pa ipak, s pojavom konekcionizma, počinju da dolaze do izražaja usamljeni glasovi filozofa poput Čerčlanda, za koje je uporni opstanak zdravorazumske psihologije bio jednak stagnaciji i neadekvatnosti teorije spremne za smetlište odbačenih spekulacija. Novi talas konekcionističke kognitivne nauke podstakao je Remzija, Stiča i Gerona, kao i Čerčlanda, da tvrde da zdravorazumska psihologija može i treba biti eliminisana jednom za svagda iz kognitivne nauke: izolovani propozicijski stavovi se ne mogu locirati u neuronskim mrežama distriburianih reprezentacija, pa samim tim ni identifikovati njihova kauzalna uloga. Međutim, mnogi filozofi i kognitivni naučnici – od Džerija Fodora, preko Horgana i Tijensona, pa sve do Frensis Igan – nisu bili spremni da odu bez borbe, bilo da su im stavovi prema simboličkoj kognitivnoj nauci bili ekstremno pozitivni bilo oprezno skeptični, jer „ako bi zdravorazumska psihologija propala, to bi predstavljalo do sada neviđenu i najveću intelektualnu katastrofu u istoriji naše vrste; ako smo zaista do te mere pogrešili u pogledu prirode našeg uma, onda nam je to najveća greška u pogledu svega do sad“ (Fodor 1987: xii).

Analizirajući recentne konekcionističke modele, pokazivala sam kako konekcionistička kognitivna nauka ima niz metodoloških kečeva u rukavu: od rekurentnih mreža, koje su se pojavile nedugo nakon rada Remzija i kolega (u kom su se oslanjali na proaktivne neuronske mreže), pa sve do mreža dugog kratkoročnog pamćenja ili LSTM, koje se obučavaju na izuzetno velikim i raznovrsnim skupovima podataka, i poseduju izuzetne kompjutacione moći. Kada se posmatra upravo ovaj metodološki razvoj konekcionizma, koji nije bio vođen teorijskim okvirom poput zdravorazumske psihologije, već pre potrebama i znanjima poteklim iz istraživanja veštačke inteligencije i neuronauke, eliminativizam nimalo ne izgleda kao kontroverzna ili tabu pozicija. Moj cilj je bio da branim eliminativizam ograničenog uticaja, to jest moja pozicija u ovom radu je bila da naprosto u konekcionizmu ne postoje dubinski razlozi za

očuvanjem entiteta koji se postuliraju zdravorazumskom psihologijom *qua* teorijom.

Nivo pronalaženja i opisivanja kognitivnih mehanizama kakav nam nudi konekcionizam takav je da može obuhvatiti ne samo ljudske kognitivne procese, već i životinja, nasuprot tradicionalnoj simboličkoj kognitivnoj nauci. Prema tome, ne samo što je u pitanju inkluzivnija teorija o kogniciji, već upravo jer nema pozivanja na teorijski okvir verovanja i želja čoveka možemo posmatrati kao koekstenzivnog ostalim organizmima u pogledu kognitivnih mehanizama koji rukovode ponašanje. Specifičnost ljudske kognicije bi mogla da se ogleda ne u sposobnosti reprezentovanja sve kompleksnijeg sadržaja koji je lingvistički strukturiran, već u brojnim *kompleksnijim* neuronskim mrežama koje uče na osnovu *specifičnog iskustva* s okolinom, a čiju kompleksnost i diverzitet i dalje ne razumemo dovoljno. I upravo ovde se može naći prostor za humanistiku: dok je humanistika više zainteresovana za ono što čoveka *razlikuje* od ostalih organizama, to jest za njegovo vlastito iskustvo ili perspektivu, kao i za njegovo oblikovanje okoline čiji je deo, kognitivna nauka će pre biti fokusirana na traženje kognitivnih regularnosti između vrsta.

Zdravorazumska psihologija može biti verbalno korisna heuristika u humanističkim disciplinama. Recimo, na sličan način na koji nam antropomorfni način govora o tehnikama istraživača veštačke inteligencije pomaže da laički shvatimo način funkcionisanja novih tehnologija, iako nije ni precizan ni tačan (Watson 2019), tako nam može biti lakše da razumemo i ljude različitih kultura ili u dalekoj prošlosti koristeći se aparaturom propozicijskih stavova. Ponekada je takva heuristika jedini relevantan medijum prenošenja informacije: Katonova želja za uništenjem Kartagine je relevantna informacija o Katonu u okvirima političke i vojne situacije tadašnjeg Rima, i bilo bi krajnje neprikladno očekivati od nekog savremenog istoričara da nam namesto toga ispriča kako je u Katonov mozak skup kompleksnih neuronskih mreža na osnovu čijeg „ponašanja“ emergira njegovo ponašanje u Senatu. Ali, u sličnoj meri bi onda trebalo smatrati krajnje neprikladnim i da nam savremeni konekcionista opisuje svoje modele preko vokabulara zdravorazumske psihologije. [13]

Vanja Subotić
Institut za filozofiju,
Čika Ljubina 18-20, Beograd
vanja.subotic@f.bg.ac.rs

13 Ovaj rad je nastao u okviru projekta Dinamički sistemi u prirodi i društvu: filozofski i empirijski aspekti, ev. br. 179041, koji se odvija uz podršku Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije. Zahvaljujem se Dušku Preleviću na komentarisanju ranije verzije rada, kao i na dragocenom razgovoru o ovoj temi.

Reference

Andrews, K. 2015. "The Folk Psychological Spiral: Explanation, Regulation, and Language", *Southern Journal of Philosophy* 53, 50-67.
Botterill, G. 1994. "Beliefs, Functionally Discrete States, and Connectionist Networks: A Comment on Ramsey, Stich, and Garon", *British Journal for the Philosophy of Science* 45(3), 899-906.
Churchland, P. M. 1981. "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 78, 67-90.
_____. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States", *Journal of Philosophy* 82, 8-28.
_____. 1989. "On the Nature of Theories: A Neurocomputational Perspective", u: *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. (Bradford Books/MIT Press), 153-194.
_____. 2007. "Catching Consciousness in a Recurrent Net", u: P. M. Churchland, *Neurophilosophy at Work* (Cambridge University Press), 1-18.
Clark, A. 1987. "From Folk Psychology to Naive Psychology", *Cognitive Science* 11, 139-154.
_____. 1993. *Associative Engines: Connectionism, concepts, and representational change*. (Cambridge, MA: Bradford Books/MIT Press).
Dutta, S. & Nandy, A. 2020. „An Extensive Analysis on Deep Neural Architecture for Classification of Subject-Independent Cognitive States", u: *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 180–184.
Egan, F. 1995. "Folk Psychology and Cognitive Architecture", *Philosophy of Science* 62(2), 179-196.
Fodor, J. 1975. *Language of Thought*. (Cambridge: Harvard University Press).
_____. 1987. *Psychosemantics*. (Cambridge, MA: Bradford Books/MIT Press).
_____. 1990. "Semantics, Wisconsin style", u: D. J. Cole et al. (ur.), *Philosophy, Mind, and Cognitive Inquiry*. (Kluwer Academic Publishers), 209-228.
Gallagher, S. 2001. "The Practice of Mind: Theory, Simulation, or Primary Interaction?" *Journal of Consciousness Studies* 8, 83-108.
____. & Hutto, D. 2008. "Understanding Others through Primary Interaction and Narrative Practice", u: J. Zlatev, T. Racine, C. Sinha, E. Itkonen (ur.) *The Shared Mind*. (Amsterdam: John Benjamins), 17-38.
Goldman, A. I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. (Oxford: Oxford University Press).
Hochreiter, S. & Schmidhuber, J. 1997. "Long Short-Term Memory", *Neural Computation* 9 (8), 1735-1780.
Horgan, T. & Tienson, J. 1995. "Connectionism and the Commitments of Folk Psychology", *Philosophical Perspectives* 9, 127-152.
Laakso, A. & Cottrell, G. W. 2005. "Churchland on Connectionism", u: B. L. Keeley (ur.), *Paul Churchland*. (Cambridge University Press), 113-154.
Leslie, A. M. 1987. "Pretense and Representation: The Origins of Theory of Mind", *Psychological Review* 94, 412-426.

____. 1994. "Pretending and Believing: Issues in the Theory of ToMM", *Cognition* 50, 211-38.
Lewis, D. 1972. "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50, 249-58.
McGeer, V. 2015. "Mind-making Practices: The Social Infrastructure of Self-knowing Agency and Responsibility", *Philosophical Explorations* 18(2), 259-281.
Newen, A. 2015. "Understanding Others – The Person Model Theory", u: T. Metzinger & J. M. Windt (ur.), *Open MIND.* (The MIT Press), 1-28.
____. 2018. "The Person Model Theory and the Question of Situatedness of Social Understanding", u: A. Newen, L. De Bruin & S. Gallagher (ur.), *The Oxford Handbook of 4e Cognition*. (Oxford: Oxford University Press), 469-493.
O'Brien, G. 1991. "Is connectionism commonsense?" *Philosophical Psychology* 4(2), 165-178.
Plutarch. 1835. *Lives of Illustrious Men*, Vol. I (prev. na engl. sa starogrč. J. Langhorn & W. Langhorn). (London: Henry G. Bohn).
Premack. D. & Woodruff, G. 1978. "Does the Chimpanzee Have a Theory of Mind?" *Behavioral and Brain Sciences*, 4, 515-526.
Ramsey, W., Stich, S., & Garon, J. 1990. "Connectionism, Eliminativism, and the Future of Folk Psychology", *Philosophical Perspectives* 4, 499-533. Preštampano u: D. J. Cole i dr. (ur.), *Philosophy, Mind, and Cognitive Inquiry*. (Kluwer Academic Publishers), 117 -144.
Ramsey, W. 1992. "Connectionism and the Philosophy of Mental Representation", u: S. Davies (ur.), Connectionism: Theory and Practice. (Oxford University Press), 247-277.
____. 2017. "Must Cognition be Representational?" Synthese 194, 4197-4214.
____. 2019. "Eliminative Materialism", u: E. Zalta (ur.), Stanford Enyclopedia of Philosophy, https://plato.stanford.edu/archives/spr2019/entries/materialism-eliminative/
Sellars, W. 1956. "Empiricism and the Philosophy of Mind", Minnesota Studies in Philosophy of Science 1, 253-329.
Sharpe, R. A. 1987. "The Very Idea of Folk Psychology", Inquiry 30, 381-393.
Smart, J. J. C. 1962. "Sensations and Brain Processes", u: V. C. Chappell (ur.), The Philosophy of Mind. (Engelwood Cliffs: Prentice–Hall, Inc).
Smolensky, P. 1988. "On the Proper Treatment of Connectionism" Behavioral & Brain Sciences 11, 1-74.
Staudemeyer, R. C. & Morris, E. M. 2019. "Understanding LSTM – A Tutorial into Long Short Term Recurrent Neural Networks", arXiv:1909.09586v1
Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. (Cambridge, MA: Bradford Books/MIT Press).
____. 1988. "From Connectionism to Eliminativism", *Behavioral & Brain Sciences* 11, 53-54.
____. 1991. "Causal holism and commonsense psychology: a reply to O'Brien", *Philosophical Psychology* 4(2), 179-181.
____. & Ravenscroft, I. 1994. "What *is* Folk Psychology?" *Cognition* 50, 447-468. Preštampano u *Deconstructing the Mind*, 115-136.
____. 1996. "Deconstructing the Mind", u: *Deconstructing the Mind.* (Oxford: Oxford University Press), 1-96.
____. & Nichols, S. 2003. *Mindreading*. (Oxford: Oxford University Press).

Watson, D. 2019. "The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence", *Minds & Machines* 29, 417-440.
Wen, T. & Young, S. 2019. „Recurrent Neural Network Language Generation for Spoken Dialogue Systems", *Computer Speech & Language* 63, 1-22.

Vanja Subotić

Folk Psychology, Eliminativism, And The Present State of Connectionism
(*Summary*)

Three decades ago, William Ramsey, Steven Stich & Joseph Garon put forward an argument in favor of the following conditional: if connectionist models that implement parallelly distributed processing represent faithfully human cognitive processing, eliminativism about propositional attitudes is true. The corollary of their argument (if it proves to be sound) is that there is no place for folk psychology in contemporary cognitive science. This understanding of connectionism as a hypothesis about cognitive architecture compatible with eliminativism is also endorsed by Paul Churchland, a radical opponent of folk psychology and a prominent supporter of eliminative materialism. I aim to examine whether current connectionist models based on long-short term memory (LSTM) neural networks can back up these arguments in favor of eliminativism. Nonetheless, I will rather put my faith in the eliminativism of the limited domain. This position amount to the following claim: even though that connectionist cognitive science has no need whatsoever for folk psychology *qua* theory, this does not entail illegitimacy of folk psychology *per se* in other scientific domains, most notably in humanities, but only if one sees folk psychology as mere heuristics.

KEYWORDS: connectionism, eliminativism, folk psychology, neural networks, propositional attitudes.