

## Andrej Ule

Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva 2, p.p. 580, SI-1001 Ljubljana  
andrej.ule@guest.arnes.si

# Kako uloviti pticu misli?

## Nekoliko wittgensteinovskih komentara uz računarsko modeliranje uma

### **Sažetak**

*U ovome eseju autor analizira Wittgensteinovu kritiku nekoliko pretpostavki koje su ključne za velik dio spoznajne znanosti. Ove uključuju koncepte komputacijskih procesa u mozgu (neuralni sistem), mozak kao mašinu, psihofizički paralelizam, misleću mašinu, kao i konfuziju pravila koja slijede ponašanje u skladu s tim pravilom. Po autorovu mišljenju, teoretičari spoznajne znanosti još uvijek nisu ozbiljno razmotrili Wittgensteinovu kritiku tako da oni, pomalo iznenađujuće, često brkaju pitanje »kako ovo djeluje?« s onime »što ovo čini?«. Ali njihova »najpogubnija« pogreška jest njihovo brkanje internih komputacijskih (ili paralelnih) procesa što se odvijaju u mozgu (koji možda uzrokuju mentalna stanja) sa socijalno-baziranim, svakodnevnim kriterijem prepoznavanja i klasifikacije sadržaja mentalnih stanja, te znanja o sadržaju mentalnih stanja.*

### **Ključne riječi**

Ludwig Wittgenstein, spoznajna znanost, komputacijski procesi, modeliranje uma, mentalna stanja

Naslov je ovog eseja metaforičan. Vjerujem da je jednako teško promatrati ljudsku misao dok radi, kao i uloviti (vlastitim rukama) pticu koja leti. Imamo prigodu uloviti pticu samo kad se dogodi da nam se dostatno približi, možda kad časak zastane da bi sjela na granu pokraj nas, ili kad sleti pravo ispred nas da bi ključnula nešto hrane. U tom času ona, međutim, nije »ptica koja leti«, nego ponajviše ptica koja se sprema iznova poletjeti. A ipak, čak i u takvim trenucima, šanse da uhvatimo pticu vrlo su male, što je činjenica vjerojatno dobro poznata svakome tko je već pokušao uloviti pticu. Ljudska je misao veoma nalik ptici. Većinu vremena misli »prolijeću« kroz naše umove, pri čemu mislim da se javljaju u našoj svijesti na trenutak, a onda odmah nestanu, tj. bivaju zamijenjene novim mislima. Istinski pokušaj razumijevanja onoga što se događa unutar naših umova dok mislimo neku misao mijenja tu misao u nešto drugo (drugu misao). Uglavnom na isti način, ne možemo razumjeti kako je moguće da svojom mišlju možemo »dotaknuti« stvari koje su posve različite od misli, koje su možda fizički nedostupne ili čak nepostojeće. No, čim je promatramo na ovaj način, ta nam misao »izmiče«, i na kraju zurimo u prazno ne mogavši shvatiti što se događa. Katkad nam se može učiniti da se misao zaustavila, takoreći da nam stoji jasno pred očima, te da ćemo je moći dohvatiti i razumjeti, i što znači »imati misao u nečijem umu« ili »misliti o ovom i onom«. Ovaj fenomen možemo iskusiti kad, recimo, razbijamo glavu nad nekim nadasve »zakučas-tim« problemom, a neka posebna misao jedna je od pretpostavki koje tvore taj problem. No ipak, čim se koncentriramo na tu posebnu misao i odmak-

nemo od problema što ga ona konstituira, misao se često odmah izgubi, tj. izmakne našem mentalnom horizontu. To je jedan od razloga zašto je teško analizirati pretpostavke o problemu – za nas, one imaju smisla samo ukoliko ih iskusujemo kao dio problema, no ovaj smisao se gubi čim ih počnemo analizirati kao takve.

Sličnu poteškoću prikazuje Wittgenstein u svojim *Filozofskim istraživanjima* (= FI; Wittgenstein, 1976).

»Čudne li stvari, misli – ona nas, međutim, ne pogađa kao nešto čudno dok razmišljam. Misao nas ne pogađa kao nešto tajanstveno dok razmišljam, nego jedino kad kažemo, kao da je to retrospektivno: 'Kako je to bilo moguće?' Kako je misli bilo moguće razmatrati upravo *sami* objekt? Osjećamo da smo pomoću njega stvarnost uhvatili u svoju mrežu.« (FI, par. 428)

Čini se da su misli izuzetno eluzivni »mentalni fenomeni«. Znamo što su one sve dok nas to nitko ne puta, no jednom kad nas zapita to više ne znamo, da parafraziram Augustinovu duhovitu opasku o vremenu. Mnogi drugi »mentalni fenomeni«, primjerice razumijevanje, nakana, želja, osjećaj itd., proizvode slična iskustva.

Prema Wittgensteinovim riječima, to nije ona vrsta pitanja koju obično postavljaju prirodne znanosti, nego

»... nešto što znademo dok nas nitko ne pita, ali više ne znamo ako se pretpostavi da to trebamo obrazložiti, to je nešto na što se trebamo podsjećati. (Nadalje, to je očito nešto, na što se čovjeku zbog nekog razloga teško podsjećati.)« (FI, par. 89)

Tijekom svoje »kasne faze«, Wittgenstein je isticao kako je pogrešno tražiti uzroke i razloge za neshvatljivost i neodredivost mentalnih fenomena u njihovoj tajnovitoj prirodi, primjerice u strogo privatnim »unutarnjim« odnosima koji su navodno dostupni samo čistoj introspekciji, i o kojima se može govoriti samo u prvom licu jednine, a da nijedna druga osoba ne može znati za te procese osim one koja je introspektivna. Prema tom razumijevanju, ljudi mogu doprijeti do hipotetičkih zaključaka o mentalnim stanjima drugih osoba samo prema njihovu vidljivom ponašanju, tako da područje mentalnog po svoj prilici spada u posve različit red stvari i događaja negoli je to područje onoga fizičkog.

Wittgenstein je žestoko odbacio tu ideju tvrdeći da je besmislena. On je odbacio čak i gledište biheviorista, koje su neshvatljivost i neodredivost mentalnih stanja dovele do zaključka da su intelektualnost, mišljenje, svijest i slično, samo pojave bez temelja u stvarnosti, što znači da je fizičko ponašanje ljudi ono jedino što nam preostaje. On govori o gramatičkoj fikciji biheviorizma (FI, par. 308) i kaže:

»Kako se javlja filozofski problem glede mentalnih procesa i stanja te biheviorizma? – Prvi je korak onaj koji posve izmiče opažanju. Govorimo o procesima i stanjima, a njihovu prirodu ostavljamo bez odgovora. Jednom ćemo možda znati više o njima – to mi mislimo. Upravo nas to, međutim, veže uz poseban način gledanja na stvari. Mi, naime, nemamo određen pojam o tome što znači naučiti se bolje poznavati proces... – Tako se sad analogija koja nas je trebala dovesti do boljeg poznavanja naših misli raspada u komade. Moramo tako zaniijekati još neshvaćeni proces u još neistraženome mediju. Sada to izgleda kao da smo porekli mentalne procese. A naravno da ih ne želimo poreći.« (FI, par. 309)

Wittgenstein ističe da je pogrešno govoriti o misaonom procesu kao o »netjelesnom«, zato što se misao ne može odvojiti od govora i ponašanja. Govor tako izražava našu zbunjenost koja, s jedne strane, proizlazi iz pokušaja da na primitivan način odredimo značenje riječi 'misli', a s druge, iz gramatičkog razlikovanja između, primjerice, gramatike riječi 'misli' i one 'jesti'

(FI, par. 339). On ipak dodaje da se razlika u značenjima čini vrlo malom (uglavnom na isti način kao što je mala razlika između toga kad kažemo da su prsti stvarni, a brojevi nestvarni).

Ovo otvara prostor za »treću« mogućnost, tj. za pokušaj konceptualiziranja intelekta, mišljenja, iskustva i sličnih fenomena kao »unutrašnjih« fizičkih stanja i fizičkih procesa koji se odvijaju unutar fiziološke strukture pojedinca, npr. u nečijem mozgu, kao rezultat nekih unutrašnjih mehanizama, primjerice proračuna koji se događaju u mozgu. Prednost ovakvog pristupa jest što čuva »intuiciju« glede unutrašnjih procesa i stanja, a to je pojam koji se tako snažno nameće našoj svijesti, dok istodobno odbacuje ideju da su ti procesi i stanja nedostupni znanstvenom istraživanju. Pojam unutrašnjih fizičkih procesa i mehanizama u ljudskome mozgu, koji uzrokuju mentalne pojave ili dovode do njih, nesumnjivo je vrlo privlačan velikom dijelu kognitivne znanosti. On ipak ima svojih načelnih ograničenja, što ih je istaknuo i Wittgenstein. On se usprotivio ne samo ideji da su ne-materijalni procesi što se događaju u pojedincu nedostupni nego i općenito ideji »unutrašnjeg mehanizma« ili »unutrašnjih procesa«. Prema njemu, razlika »vanjski-unutrašnji proces« jednostavno je premala da bi bila od koristi u objašnjavanju predikamenata koji izviru iz svakidašnjeg diskursa, te, još više, iz filozofskog diskursa o mentalnim pojavama.

Prema Wittgensteinu, od svih mentalnih »pojava«, upravo se razumijevanje i, posebice, namjera odupiru našim pokušajima da ih konceptualiziramo kao unutrašnja stanja ili procese. U svezi s ovim fenomenima, upitnima postaju čak i kategorije 'stanja' i 'procesas'.

»Namjera s kojom netko djeluje ne 'prati' djelovanje ništa više negoli što misao 'prati' govor. Misao i namjera nisu ni 'artikulirani' niti 'neartikulirani'; ne treba ih uspoređivati ni s pojedinim tonom koji odzvanja tijekom djelovanja ili govorenja, a niti s melodijom. 'Govorenje' (bilo glasno ili tiho) i 'mišljenje' nisu pojmovi iste vrste; čak i kad mislimo na njih vrlo su tijesno povezani.« (FI, II, xi, 217)

Evo iste misli, izražene još eksplicitnije:

»Značenje nije proces koji prati riječ. Nijedan proces, naime, ne bi mogao imati posljedice značenja. (Isto tako, mislim da bi se moglo reći: računanje nije pokus, jer nijedan pokus ne bi mogao imati osobite posljedice množenja.)« (FI, II, xi, 218)

Dio u zgradama važan je stoga što kognitivna znanost često pretpostavlja da je računanje mentalni proces ekvivalentan nekom fizičkom procesu u procesoru, ili biofizičkom nervnom procesu. Takav bi proces zaista više nalikovao određenoj vrsti »mikro-pokusa« negoli računanju, a tu je ideju Wittgenstein žestoko odbacio.

U svojim *Primjedbama o filozofiji psihologije* (=PFP; Wittgenstein, 1980) Wittgenstein je čak izričito potvrdio da mu se nijedna pretpostavka ne čini prirodnijom od pretpostavke da asocijacija, tj. mišljenje, nije nikakva vrsta mozgovnog procesa, tako da je nemoguće razumijevanje mentalnih na osnovi mozgovnih procesa (PFP, I., par. 903). Kao ilustraciju, on koristi primjer sjemena. Isto sjeme uvijek proizvodi iste biljke, premda ništa u tom sjemenu ne odgovara biljci, tako da se, isključivo na temelju njegovih svojstava ili strukture, ne može zaključiti koju će biljku sjeme proizvesti. Samo bi nam povijest sjemena možda mogla dati odgovor na to koja će biljka iz njega izrasti. Prema Wittgensteinu, na sličan bi način, i skoro bez razloga, potpuno amorfn masa mogla stvoriti organizam. On se tako pita bi li nešto

slično moglo biti slučaj i s našim mislima, govorom ili pisanjem.<sup>1</sup> Kasnije u tekstu, Wittgenstein ide dotle da počinje sumnjati u pretpostavku kako svaki mentalni proces ima »trag« u našem mozgu, koji onda dopušta mogućnost da tu postoje psihološki obrasci koji se ne podudaraju *ni s kakvim* fiziološkim obrascima (PFP, I., par. 905). »Ako to postavlja naglavačke naš pojam uzročnosti«, piše Wittgenstein, »onda je i vrijeme da on bude postavljen naglavačke«. <sup>2</sup> On također dodaje da je predrasuda glede paralelizma između onoga psihološkog i onoga fizičkog posljedica našega primitivnoga tumačenja gramatike. Zato što, ako dopustimo da psihološki fenomeni slijede načelo uzročnosti koje nije fiziološki posredovano, onda pretpostavljamo da *pokraj* tijela postoji neka vrsta duše (PFP, I., par. 906).

Time ne mislim reći da se Wittgenstein odjednom prometnuo u zagovornika dualizma, tj. postavke o spiritualnoj supstanciji koja opstoji odvojeno od tijela, te omogućuje spiritualne fenomene. Zapravo, on je jednostavno htio istaknuti, možda na malo pretjeran način, duboku razliku između ljudskog mišljenja i svih ostalih ljudskih procesa ili stanja, bilo materijalnih ili nematerijalnih. Prema njemu, o mišljenju se ne može govoriti kao o privatnom procesu; ono se može promatrati samo u kontekstu društvenih aktivnosti ljudi. Drugim riječima, mišljenje se događa samo unutar konteksta pridržavanja pravila, unutar različitih lingvističkih igara, te unutar međuigre ljudskih djelovanja. Ti se procesi ne mogu prevesti u procese koji se odvijaju u pojedincima, ili u individualna ponašanja.

Kad kažemo kako je mišljenje unutrašnji proces unutar našega mozga – tj. fiziološki proces unutar tijela koji uzrokuje mišljenje ili jest mišljenje samo – dolazimo u opasnost da to mišljenje odvojimo od intersubjektivne prakse govora, uzajamnog razumijevanja, društvenog funkcioniranja ljudi u različitim situacijama (kontekstima), itd. Takav proces bio bi, na neki način, samodostatan, a to bi ga naposljetku učinilo isto tako »privatnim« kao što su to stanja i procesi pretpostavljene spiritualne supstancije. Sve dok nam ne bude dostupna intersubjektivno provjerljiva (javno pristupačna) i sustavna veza između unutrašnjih stanja i procesa koji se odvijaju unutar mozga, te pojedinačeva djelovanja u intersubjektivnim (društvenim) kontekstima, bit ćemo izloženi Wittgensteinovoj kritici privatnoga govora i privatnih stanja i procesa. Moguće je naravno pretpostaviti kako se takva veza u načelu ne može isključiti, te da će nam možda, jednom u budućnosti, biti moguće da je razvijemo. Zasad se ipak nalazimo pred zbnunjujućim rascjepom između javno dostupne, »izvanjske« prakse ljudskog djelovanja u različitim situacijama (npr. jezičnim igrama) koja određuje značenje naših iskustava, želja, nakana i misli, te »unutrašnje« (fizičke) prirode, tj. strukture tih iskustava, želja i sličnoga.

Ovdje je bitno da to nije samo konceptualni rascjep između supstrata mentalnih stanja – vjerojatno posve objašnjivih pomoću izračunavanja – nego i o rascjepu između toga kako »procesori« rade, misleći pritom na procesore koji tvore fiziološko-fizičku osnovu mentalnih stanja, te toga *što i kako mi jesmo* kad iskustveno doživljavamo ta stanja i izražavamo ih u svojoj društvenoj praksi. To »kako jesmo« bitno je povezano s činjenicom da naše iskustveno doživljavanje i funkcioniranje ima za nas značenje, te da *svoje postojanje osjećamo* upravo zbog tog iskustvenog doživljavanja i funkcioniranja. Razlika između toga dvoga nije samo konceptualna nego i kategorijalna, pa nemamo niti najsitniji trag kako bismo mogli prevladati taj rascjep.

Pravi pojam »rada« procesora uključuje dvojbu koja se odnosi na opis uzročnog procesa, tj. na slijed stanja procesora *spram* pravila po kojima procesor

radi. Taj pojam možemo postići popisivanjem fizikalnih zakona koji upravljaju fizičkim tijekom događaja i inicijalnih stanja procesora u određenim trenucima. A da bismo upoznali spomenuta pravila, mogli bismo, primjerice, predočiti tablicu operacija koja ocrta prijelaze procesora iz jednog u drugo stanje, ovisno o inputu i trenutnom stanju procesora. No ipak, takva bi tablica bila samo idealizirana kreacija, koja prije pokazuje kako bi procesor *trebao* raditi u idealnim uvjetima nego kako on zaista radi. Ona je samo jedan od mogućih opisa pravila rada, vjerojatno usađenih pomoću procesora, a ne opis samoga procesa. Takva pravila možemo također prikazati na tablici koja opisuje operacije što ih procesor može »izvršiti« pod uvjetom da su prisutni odgovarajući uvjeti i odgovarajuća unutarnja stanja. No ipak, kojigod pristup usvojili određujući pravila rada procesora, moramo imati na umu razliku između idealiziranog, *operacionalnog rada* i stvarnoga *tijeka* događaja, tj. stvarnoga slijeda fizičkih stanja procesora. Operacionalni je rad procesora jedna vrsta rada koju mu *pripisujemo* kada ga *tumačimo* kao implementaciju nekoga pravila ili tabularni opis operacije.

Kognitivne znanosti često upućuju na model »Turingova stroja«, za koji se vjeruje da predstavlja općenitu formulu za računarski rad prirodnih ili umjetnih procesora. No ipak, taj je model samo idealizirana struktura rada sustava kojemu se pripisuje računarsko djelovanje, a da nijedan stvarni sustav ne odgovara tome modelu. Razlika leži ne samo u činjenici da turirajući stroj koristi beskonačnu memorijsku traku nego i u tome da je taj »stroj« kadar, nakon nekoliko računarskih koraka, vratiti se točno u isto stanje. Nasuprot tome, stvarni, konačni automati nikad se zaista ne »vraćaju« u isto stanje, nego samo na »istu vrstu« stanja koja pripada u istu ekvivalentnu klasu stanja s obzirom na stanoviti odnos ekvivalencije između stanja. Upitno je stoga je li u ovom slučaju uopće moguće govoriti o računarstvu u smislu »operiranja sa simbolima«.

U svom eseju, koji se pojavio u zbirci eseja o Turingovoj ostavštini, C. Fields izričito je naglasio kako je nužno razlikovati između ponašanja fizičkog sustava u vremenu i (dinamičkog) opisa tog sustava koji koristi algoritmima, a ovaj određuje promjene u takvom sustavu koje se zasnivaju na (diskretnim) promjenama specifičnog parametra (ili parametara) opažljivog (izmjerljivog) u vremenu pojavljivanja. Taj opis jest *naša* metoda opisa koja nam omogućuje da slikovito prikažemo promjene stanja sustava na algoritamske stupnjeve. Da bismo mogli pružiti takav opis, mora nam također biti dostupno neko tumačenje koje slikovito prikazuje stanja sustava u slijedu apstraktnih računarskih stanja, pri čemu se svako  $a_k$  stanje sustava »postize« upotrebom početnih  $k$  stupnjeva algoritma kod posebnog inputa (Fields, 1996: 170). Fields ističe da je ovakva vrsta slikovitog prikazivanja moguća samo ako su sve operacije koje se navode u procesu grafičkog prikazivanja algebarski zamjenjive (tj., gledano iz perspektive kvantne mehanike, mjerenja i promatranja procesa uzajamno su neovisni). Fields upućuje na E. Dietricha, koji je 1989. istaknuo da se takvo tumačenje može konstruirati samo ako se na sustav prethodno, i neformalno, gleda kao da obavlja funkciju koja se može izračunati pomoću algoritma i ako se  $k$  stupnjevi algoritma priznaju kao izračuni nekih prethodno razumljenih nižih funkcija.

Ovdje nam je potreban dostatan broj opaženih stanja sustava kako bismo ih iscrtali na računarska stanja premrežena algoritmom kad ga primijenimo na

1

Slične tvrdnje možemo naći u Wittgensteinovoj *Zettel* (= Z; Wittgenstein, 1970, par. 608).

2

Vidi također Z, par. 610.

dani input. Fields o takvom tumačenju sustava govori kao o »virtualnom stroju« za algoritam. No ipak, ovo tumačenje bilo bi moguće samo ako se ponašanje sustava, počev od početnog stanja (protumačiva kao input za algoritam), uvijek može tumačiti kao prikaz algoritma za dani input.

Naravno da je ovo idealizirana pretpostavka, ili, još bolje, pretpostavljena idealizacija, prije negoli realistički tijek događaja unutar sustava. Sustav stoga – kao virtualni stroj, a ne po sebi – izračunava samo unutar okvira naše interpretacije. Dinamički opis sustava predstavlja tek specifičan odnos koji pruža dostatno točan prikaz slijeda (izmjerivih, opazivih) stanja sustava u uzastopnim vremenskim intervalima. U mnogo slučajeva moguće je taj slijed opisati koristeći skladnu, kontinuiranu funkciju, ali metoda nije bitna – za računarski rad sustava bitna je upravo interpretacija koja njegova uzastopna stanja prikazuje kao rezultat specifičnog algoritma (tj. sustava operacija) primijenjenog na odgovarajuća input stanja sustava što su inputi za virtualni stroj. Razmišljanja su Fieldsa dovela do dvaju važnih zaključaka. Prvi je

»... da bilo koji sustav koji se može tumačiti kao da simulira virtualni stroj za neku funkciju  $f$  jest virtualni stroj za  $f$ . Slično tome, bilo koji sustav koji se može tumačiti kao da simulira univerzalni kompjutor jest univerzalni kompjutor. Ovo obilježje izračunavanja – da *simuliranje izračunavanja jest izračunavanje* – jest ono što ga odvaja od neprotumačenog dinamičnog procesa poput nestalnoga tijeka.« (Fields, 1996: 171)

#### Drugi je

»... da se pojedini fizički sustav često može tumačiti kao različiti virtualni strojevi na temelju različitog niza mjerenja, tumačeći vrijednosti različitih skupina varijabli kao pokazatelj stanja sustava.« (Fields, 1996: 171)

Fieldsovu prvu postavku – da savršena simulacija izračunavanja jest izračunavanje – trebalo bi također čitati i obrnuto, naime, da je svako izračunavanje svoja vlastita »savršena simulacija«, ili takvo tumačenje rada sustava koje ga prikazuje kao savršeno izračunavanje (tj. prikazuje svako stanje sustava kao rezultat izračuna nekih inputa i nekih stanja sustava). Zaključno, ne postoji »izračunavanje po sebi«, tj. postoji bilo koji tijek događaja bez odgovarajućeg (računskog) tumačenja. Ovu misao možemo povezati s Wittgensteinovom zanimljivom kritikom (u *Traktatu*) jednadžbe operacije i predikacije, ili operacije i činjenica (događaja) (Wittgenstein, 1976). Wittgenstein zaista tvrdi da su operacije sastavni dio našega simbolizma, tj. način prikazivanja odnosa između oblika stanja stvari, prije nego što bi bile sama stanja stvari, činjenice ili događaji u svijetu.

Moramo u načelu razlikovati između operacija i (prepozicionalnih) funkcija koje prikazuju neke odnose u svijetu činjenica. Operacija prikazuje formalne odnose, tj. prije odnose između oblika negoli između pojedinačnih predstavnika oblika. Wittgenstein to ilustrira upotrebljavajući logičke oblike rečenica.

»Funkcija ne može biti svoj vlastiti argument, dočim operacija može uzeti jedan od svojih vlastitih rezultata kao svoj temelj.« (Tr, 5.251)

»Samo na ovaj način moguć je korak od jednog do drugog termina niza oblika.« (Tr, 5.252)

»Jedna operacija može djelovati protivno učinku druge. Operacije mogu jedna drugu ukinuti.« (Tr, 5.253)

»Operacija, nadalje, može prestati (negacija u slučaju dvostruke negacije, primjerice).« (Tr, 5.254)



U načelu, ne možemo očekivati da funkcije (tj. opisi svojstava predmeta ili odnosa među njima) imaju svojstva operacija. Prema Wittgensteinu, »unutrašnji« ili »formalni« odnosi između stanja stvari, tj. između tvrdnji koje ta stanja opisuju, ne mogu se izraziti razboritim rečenicama, nego samo prikazati pomoću operacije koja »predstavlja tvrdnju kao rezultat operacije, a koja tu tvrdnju izvodi iz drugih tvrdnji (koje su temelj operacije)« (Tr, 5.21). Wittgensteinova tvrdnja da operacije nisu funkcije (predikacije) znači također da operacije nisu niti jedno stanje stvari ili činjenica, nego da su one naše metode prikazivanja odnosa između oblika predstavnika stanja stvari. U tom smislu, računске operacije – ako to zaista jesu operacije što ih predstavljaju sheme Turingova stroja – ne javljaju se u »prirodi« nego unutar simboličkih sustava reprezentacije.

Wittgensteinovo ukazivanje na fundamentalnu razliku između predikacije i operacije dodatno podupire Fieldsovo razlikovanje između izračunavanja i tijeka događaja u nekom fizičkom sustavu. Kako bismo opisali tijek događaja, koristimo rečenicu pomoću koje dovodimo u odnos uzastopna stanja sustava, te ih stavljamo u poseban red, oslanjajući se možda na neku matematičku funkciju koja izražava prirodni zakon pojavljivanja. No ipak, dočim takav odnos nije bitan za sami tijek događaja, neki (simbolički) opis tijeka događaja jest bitan za izračunavanje. Prema Wittgensteinu, izračunavanje podrazumijeva uzastopno (ponekad i uzastopno i paralelno) izvođenje operacija koje možemo prikazati samo koristeći neki simbolički sustav, tj. koristeći prikladne formalne izraze, varijable, tablice operacija ili algoritamske postupke.

U svojim kasnijim radovima, Wittgenstein je obratio pozornost na slična razlikovanja između kauzalnog rada stroja i operacionalnog »rada« koji se javlja unutar predstavljanja stroja kao simbola. Ova razmišljanja osobito su relevantna za raspravu o »mislećim strojevima«, tj. bilo kojem »jakom« programu umjetne inteligencije. Wittgenstein kaže da stroj možemo konceptualizirati idealistički, u smislu apstrahiranja mogućih kvarova, lošeg rada, itd. U takvom slučaju čini se »da stroj ima (posjeduje) takve-i-takve mogućnosti kretanja« (FI, par. 193).<sup>3</sup> Jednom kad upoznamo stroj, zamišljamo da je sve drugo određeno unaprijed, kao npr. pokret koji će on izvesti. Prema Wittgensteinu, u takvom slučaju stroj ili sliku stroja (ili dijagram tijeka) zaista koristimo kao simbol za poseban način funkcioniranja.

To je potpuno različito od predviđanja o stvarnom ponašanju stroja. Tada općenito ne zaboravljamo mogućnost distorzije njegovih dijelova, itd. Stvaran stroj može uvijek učiniti pokret potpuno različit od pokreta što ga je »predvidio« njegov simbolički pandan, zato što je u simboličkom pandanu mogućnost kvara *a priori*, tj. gramatički nemoguća. Wittgenstein otud zaključuje »da je kretanje stroja-kao-simbola predodređeno u smislu različitom od onog u kojem je predodređeno kretanje bilo kojeg danog stvarnog stroja« (FI, par. 193).

No ipak, čak i govor o »idealnom« ili »idealiziranom« stroju, navodno utjelovljenom pomoću stroja-kao-simbola, zavodi čovjeka na pretpostavku da stroj ima (posjeduje) svoje »mogućnosti kretanja«. Wittgenstein zaključuje da

»... ta mogućnost kretanja nije kretanje, ali čini se da ona nije niti puki fizički uvjeti za bilo kakvo kretanje... Jer, dočim je ona iskustveni uvjet za kretanje, netko to može također zamisliti da bude i drukčije.« (FI, par. 193)

3

Vidi također *Remarks on the Foundations of Mathematics* (Wittgenstein, 1972, I., par. 125).

On dalje kaže da se nama čini kako idealna mogućnost kretanja jest neka vrsta

»... sjene samoga kretanja... Mi kažemo: 'iskustvo će pokazati da li ovo povećava mogućnost kretanja', ali ne kažemo: 'iskustvo će pokazati je li ovo mogućnost ovoga kretanja'.« (FI, par. 194)

Govoreći ovako, mi pogrešno tumačimo svoj vlastiti način govora, i to tako što govor koji opisuje operaciju stroja-kao-simbola smatramo opisom stvarnoga rada stroja. Wittgenstein kaže »da smo poput divljaka i primitivnih ljudi, koji slušaju ono što govore civilizirani, ali to krivo tumače, a onda iz toga izvode najčudniji zaključak« (FI, par. 194). Ukratko, stroj-kao-simbol jest poseban izraz pravila rada, tj. prije izraz rada u skladu s pravilima negoli stvarnoga rada. Na sličan način, stroj za spajanje prikazan pomoću odgovarajuće tablice prijelaza stanja prije bi bio simboličkim prikazom rada pomoću pravila negoli stvarnim radom nekoga »logičkoga stroja«, primjerice jednostavnog procesora koji izvodi spajanje.

Očito opet dolazimo do načelne razlike između operacija unutar nekog simboličkog sustava (modela) i stvarnih procesa u svijetu. Bez obzira na to kako savršeno stroj radi (stroj koji izvodi neku logičku operaciju, primjerice), to još ne znači da on zaista izvršava operaciju, zato što se i najsavršeniji stroj može pokvariti; njegov rad može se prekinuti ili se može oštetiti, dočim izvođenje operacije u načelu isključuje takvu »mogućnost«. Prema Wittgensteinu, ovakve su mogućnosti kategorički i gramatički isključene.

Wittgenstein je ipak zaniijekao da bilo koji izraz pravila, tj. pridržavanja pravila, predstavlja krajnje ili izvorno tumačenje ili objašnjenje pridržavanja pravila. On se u tom smislu ne bi bezuvjetno složio s Fieldsovom formulacijom da računarsko tumačenje rada sustava pomoću pravila izračunavanja već predstavlja stvarno pridržavanje toga pravila izračunavanja, tj. da simulacija izračunavanja jest izračunavanje. Zapravo, sve dok je »računarsko tumačenje« samo simbolički izraz precrtavanja fizičkih stanja sustava na algoritamska stanja računarskog sustava, nije moguće govoriti o izračunavanju. Samo izvođenje precrtavanja, naime stvarno praćenje algoritma, jest izračunavanje. Izračunavanje je u tom smislu tek stvarno računarsko tumačenje prije negoli, primjerice, »zamišljeno« formalno tumačenje koje opisuje samo algoritam računa. Tablica precrtavanja ili tablica prijelaza još nisu operacije, nego simbolički izrazi. Povežemo li nazivlje korišteno u *Traktatu* i *Filozofskim istraživanjima*, mogli bismo reći da se stvarno izračunavanje pokazuje samo upotrebom specifičnog algoritma u okolišu poput društvene, javno dostupne prakse (FI, par. 202), tehnike (par. 199) ili običaja (par. 199), u kojemu mi sami moramo sudjelovati logički, dočim da bismo ga *opisali* moramo upotrijebiti neki formalni simbolizam.

No ipak, osjetilno sudjelovanje u praksi nasljedovanja pravila nije tek »adekvatan« odgovor izvršitelja nekih inputa, nego sudjelovanje u životnom obliku bića kojemu nasljedovanje pravila jest osjetilna i bitno neupitna, razumljiva akcija. Ovu neupitnost jednostavno moramo prihvatiti kao temeljnu činjenicu.

»Ono što moramo prihvatiti, ono dano, jesu – tako bi se moglo reći – oblici života.« (FI, II, xi, 226)

Zajednički ljudski oblik života nije nešto fiksirano, nego je to zajednički oblik djelovanja. To je mreža aktivnosti koja određuje naše razumijevanje i djelovanje. Wittgenstein o tome govori u *Zettel* (Z, par. 576).



»Kako možemo opisati ljudski način djelovanja? Pa, samo pokazujući kako se ljudske djelatnosti šire na sve strane. To nije ono što netko upravo sada misli, individualna akcija, nego cjelovito kretanje ljudskih djelatnosti koje tvori pozadinu spram koje opažamo bilo koju pojedinačnu djelatnost i koja određuje naš sud, naše pojmove i reakcije.«

Prema Wittgensteinu, pravila – tj. nasljedovanje pravila, isto kao i operacije i njihovo provođenje – neodvojivi su sastojak ljudskog oblika života, tj. »širenja ljudskih djelatnosti«. Zato ne možemo apstrahirati nasljedovanje pravila i promatrati ga odvojeno od te pozadine – tako, primjerice, pitamo slijedi li neki stroj ili organizam po sebi pravilo.

Budući da mišljenje podrazumijeva sposobnost nasljedovanja pravila, Wittgenstein tvrdi kako nije senzibilno pitati, primjerice, da li stroj misli. Ili, da budem još precizniji, takvo pitanje podsjećalo bi unekoliko na ljudsko biće.

»Može li stroj misliti? Može li osjećati bol? – Zatim, treba li ljudsko tijelo nazvati takvim strojem? Ono je zaista vrlo blizu toga da bude takav stroj.« (FI, par. 359) »Stroj međutim sigurno ne može misliti! Je li to iskustvena tvrdnja? Ne. Mi samo govorimo o ljudskom biću i kako to nekome izgleda da ono misli. Govorimo također o lutkama a nesumnjivo i o duhovima. Pogledajte na riječ 'misliti' kao na alat.« (FI, par. 360)

Wittgenstein, dakle, ne odbacuje potpuno mogućnost da bi stroj mogao misliti, ali samo pod pretpostavkom neke sličnosti s ljudima, makar ona bila samo virtualna. Iz čega bi, međutim, takva sličnost mogla nastati? Naravno, iz naše predodžbe strojnog rada kao dijela ljudskog oblika života i kao logičkog dijela mreže ljudskoga djelovanja, mreže ljudske komunikacije, primjerice, međuljudske suradnje u zajedničkim projektima, igrama, i sl. Sve što možemo reći jest da, kad bismo mogli zamisliti robota koji vjerno oponaša ljudska djelovanja u paradigmatičkim životnim situacijama i ravnopravno sudjeluje u ljudskoj komunikaciji, onda bismo takvog robota mogli zamisliti kao »mislećeg«. To je slično onome kako zamišljamo ponašanje »živih patuljaka« (npr. Pinocchio) ili duhova. No ipak, put od zamišljanja do provedbe je dug, možda i nemoguć.

U posljednjoj rečenici prethodnog citata Wittgenstein daje naslutiti da je to otprilike kao da bi mišljenje bilo oruđe. Može li mišljenje biti oruđe? Možda na način sličan onom na koji su nam ruke »oruđe« za rad. To je nesumnjivo jedan od mogućih pogleda na mišljenje, ali uključuje samo jedan aspekt i ne uzima u obzir cjelinu. Možemo na mišljenje gledati kao da je ono oruđe (stvorenno evolucijom, prirodom, od Boga, itd.), no upravo činjenica da na nj gledamo kao da je oruđe govori da ono po sebi nije oruđe. Na ljude, primjerice, ne možemo gledati kao na »živuće« ili »misleće«, nego jednostavno smatramo da su oni živi i da misle. Na čovjeka bismo mogli gledati »kao da je patuljak«, ali to nije njegova »narav«. Poput toga, na patuljka bismo mogli gledati ne »kao da on jest simulacija čovjeka« nego kao da *biva* simulacijom čovjeka (za nas). Osim toga, na patuljka bih mogao gledati kao da je bio živ, misleće biće, ali to nije njegova »narav«.

Znači li Wittgensteinovo odbacivanje mentalnih pojava, gledanih kao unutarnja stanja unutar pojedinačne nervne i fiziološke strukture, da je on odbacio znanstvenu psihologiju? Znači li to da bismo psihološke pojave trebali tumačiti samo kao društvene pojave, ili barem kao »interpersonalne«, radije nego kao »personalne«? Na Wittgensteina se zaista vrlo često gleda kao na protivnika znanstvene psihologije (vidi npr. Williams, 1999; Rey, 2003), no njegove pronicave primjedbe o filozofiji jednako bi, ili čak i jače, mogle »opravdati« tvrdnju da je on isto tako bio protivnik i filozofije. Mislim da su takvi zaključci preuranjeni. Vjerojatnije je da se Wittgenstein nastojao

»osloboditi« svih psiholoških, filozofskih i drugih teorija ljudske intelektualnosti, govora, komunikacije, itd., zato što je pokušao izložiti ono što po njemu nije bila nikakva teorija, nego manifestacija »čvrstoga temelja«, tj. nedvosmisleno vladanje pravilima i jezikom. Budući da su i neke druge psihološke i filozofske teorije pokušale izgraditi ovaj »temelj« i objašnjavale ga upotrebljavajući prirodnoznanstvene, psihološke i filozofske pretpostavke i teorije, one su samo dodatno otežale jasnu prezentaciju (FI, par. 109, 122), pa se Wittgenstein tome žestoko suprotstavio. U svojim *Primjedbama o filozofiji i psihologiji*, on je čak nabacio svojevrstan plan analize psiholoških pojmova, i to radije svakodnevnih negoli onih što ih je iznašla znanost za svoju vlastitu upotrebu (PFP, II., par. 63, 148). To znači da dok stvaramo psihološke pojmove ne smijemo prekinuti vezu s time kako ih svakodnevno razumijevamo. Upravo ova crta razlikuje psihologiju od nekih prirodnih znanosti koje tragaju za najvećom mogućom nezavisnošću od svakodnevnog razumijevanja njihovih predmetnih područja. Očuvanje sustavne veze između svakodnevnog razumijevanja psiholoških pojmova i pojava ne znači uklanjanje znanstvene psihologije, ili njezino svodenje na bihevizizam. Kao što kaže McGrinn, Wittgensteinova kritika psiholoških pojmova ne znači da ti pojmovi zahtijevaju bihevioralne kriterije primjene, nego je ona »pokušaj da pokažemo kako ne možemo izvesti ideju onoga što dano psihološko stanje naprosto jest pomoću introspekcije« (McGrinn, 1997: 130). Prema McGrinnu, pouka argumenta protiv privatnog jezika

»... nije ta da naši psihološki pojmovi moraju imati javne kriterije, nego da samo podsjećajući se na gramatiku svojih uobičajenih psiholoških pojmova možemo shvatiti bit, ili narav, dane vrste psihološkoga stanja.« (McGrinn, 1997: 130)

Slazem se s ovim tumačenjem.

Takvo razumijevanje psihologije i psiholoških pojmova nesumnjivo je posve suprotno nakanama velikoga dijela kognitivne znanosti, a posebice nakanama radikalnih predstavnika komputacionalizma. Wittgensteina stoga, po meni, ne možemo opisati kao preteču kognitivne znanosti, metafore kompjutera ili sličnih stvari (premda je bio dobar Turingov prijatelj).

Pogledajmo dva takva tumačenja: esej »Zašto je Wittgenstein trebao biti komputacionalist« G. Reya (2003) i knjigu *Poziv na kognitivnu znanost* J. Leibera (1991). Prema Reyu, ono što odvaja Wittgensteina od kognitivne znanosti jest uglavnom njegovo pretjerano bavljenje problemom prvoga/trećega lica. Po Wittgensteinu, psihološki pojmovi temelje se na razlikama između prvoga i trećega lica (Rey, 2003: 240). Wittgenstein napušta teoriju značenja koja pretpostavlja da svaka riječ ima ujednačen referencijalan odnos prema predmetima. Rey ističe da je Wittgenstein umjesto toga skicirao teoriju koja uključuje ulogu pojma u kontekstu, jezičnim igrama ili obliku života (Rey, 2003: 241). Prema toj teoriji,

»... mentalni predikati poput 'nada se', 'iščekuje', 'žarko ljubi' jesu široko relacionalni, mnogo više negoli to dopušta tradicionalna koncepcija 'unutrašnjih procesa'.« (Rey, 2003: 241)

Rey je ispravno ukazao na činjenicu da je čak i Wittgensteinova uporaba samog pojma 'ponašanja' očito intendirala takvoj širini.

Rey je raspravio tri izvora podrške Wittgensteinovim sugestijama glede relacionalne prirode mentalnih predikata i ljudskoga ponašanja u suvremenoj filozofiji uma: eksternalističke intuicije glede sadržaja, verifikacionističke i »kriteriološke« teorije značenja, te funkcionalističke koncepcije mentalnih stanja. Rey ipak podržava svaku od ovih teorija, osobito funkcionalističke pristupe značenju i mentalnim stanjima s nekim »mentalističkim« pretpo-

stavkama. Prva od njih jest *Umjereni mentalizam (UM)* koji predlaže barem dvije temeljne vrste mentalnih stanja: ona informacionalna, koja predstavljaju svijet; te direkionalna, koja svojega agenta usmjeravaju prema ili od nekog predočenog stanja (Rey, 2003: 245). Druga je *Kauzalno/komputacijsko-reprezentacionalna teorija misli (KRTT)*, koja tvrdi

»... da bismo mozak trebali smatrati kompjutorom koji u stvarnom vremenu izvodi operacije o logički složenim unutrašnjim reprezentacijama, zameci kojih stoje u nekim ko-varijacijskim odnosima bilo s obrascem stimulusa ili s nekim okolnim pojavama.« (Rey, 2003: 249)

Treća pretpostavka jest *nadopuna KRTT nekim unutrašnjim mehanizmom* koji proizvodi iskustva *qualia*, npr. iskustva boja, okusa hrane, osjećaja boli, itd. Rey tvrdi

»... da takva stanja uključuju osobit odnos prema posebno ograničenim predikatima u sustavu unutrašnje reprezentacije stvorenja.« (Rey, 2003: 258)

Rey vjeruje da mentalna stanja pridolaze stanjima agentova mozga i da nam odgovarajući komputacijski mehanizmi mogu pružiti neke dodatne »izvanjske kriterije« mentalnih stanja<sup>4</sup> koje nam ne mogu pružiti neki čisto bihevioralni kriteriji (Rey, 2003: 257, 259). On može sačuvati govor od doslovnih unutrašnjih procesa bez pretpostavke njihove principijelne privatnosti. Ova teorija »treba neku priču s određenim psihološki prihvatljivim detaljima što ih ona pruža« (Rey, 2003: 262), a koje nije bilo u Wittgensteina. Rey nadalje »vjeruje kako on sumnja da su samo filozofi očarani naivnom, introspektivnom slikom uma, ili da je samo jezik njegov izvor« (Rey, 2003: 262). Tu je vjerojatno u pravu. Naivna slika uma ili jezika jest skup jednostavnih analogija ili modela uzetih iz prirode i projiciranih na čovjeka. Iстина je da

»... svakodnevi jezik može trajati i biti bez ikakvih dobrih ili loših analogija; ljude mogu zanijeti loše analogije ako nema neke senzibilnije da ih nadomjesti.« (Rey, 2003: 262)

Moramo ipak biti svjesni sugestivne »moći« ovih analogija kako bismo spriječili njihovu nerefleksivnu upotrebu u znanosti i filozofiji. Rey pritom tvrdi da su takve naivne slike većinom naše nedragovoljne reakcije na stvari koje izgledaju i djeluju poput naših suodredbenica. Ove slike projiciramo u stvari koje se odnose prema toj reakciji u nama samima, i zaista ih nerado projiciramo u stvari koje ne izazivaju tu reakciju (Rey, 2003: 264).

Suprotno Reyu, ja mislim da upravo ideja »unutrašnjih procesa« u našim mozgovima koji proizvode mentalna stanja jest jedna takva snažna analogija koja postulira neke provizorne izvanjske kriterije za mentalne procese i njihovo uzročno objašnjenje, ali nas ona ipak zavodi. Pokušao sam pokazati da se Wittgenstein suprotstavlja upravo pojmu unutrašnjeg procesa, bili ne-materijalnog ili materijalnog. Wittgensteinovu primjedbu da

»... to ne označava proces koji prati riječ. Nijedan proces, naime, ne bi mogao imati posljedice značenja.« (FI, II, xi, 218),

možemo proširiti na intenciranje, promatranje aspekata, računanje itd., te prema tome na mnoge vrste mišljenja. To su Wittgensteinove najvažnije tvrdnje i ne možemo ih mimoći ili »poboljšati« nekom »znanstvenijom« slikom unutrašnjih procesa. Wittgensteinova kritika metafore stroja jasno

4

Rey upućuje na (ne)poznatu Wittgensteinovu primjedbu kako »unutrašnji proces potrebuje izvanjske kriterije« (FI, par. 580).

stavlja načelnu razliku između *operacija* u *simboličkome* modelu rada stroja, te mogućeg stvarnoga (krivog) rada stroja. To nije razlika samo između stroja uzetog idealno, apstrahirajući od mogućeg pogrešnog funkcioniranja, i stvarnoga rada stroja, nego je to kategorijalna razlika između nasljedovanja pravila u simboličkome modelu i slijeda događaja (stanja). Rey ukratko upućuje na problem nasljedovanja pravila i slaže se s »normativističkim« pojmom nasljedovanja pravila, prema kojem

»... pravilo može biti »normativno« ukoliko bi poseban mehanički ostvaren algoritam mogao biti najbolje idealiziranim objašnjenjem aktualne operacije u mozgu.« (Rey, 2003: 253 /bilj. 48/)

Rey se slaže s mišljenjima Howicha, Pietroskog i drugih, da se Kripkeov glasoviti paradoks o nasljedovanju pravila

»... mora baviti izvorima idealizacije (...), i ne normativnošću pravila – ili prisutnošću ‘okruženja koja standarde pretvaraju u standarde’ – nego time je li opravdano konstruirati nervni sustav kao da ostvaruje posebnu mentalnu nadležnost.« (Rey, 2003: 253)

No ipak, ako algoritam može biti najbolje idealizirano objašnjenje aktualne operacije u mozgu, onda *nije* stvarni (aktualni) mozgovni proces onaj koji izvodi operaciju ili slijedi pravilo, nego smo to *mi* koji pomoću »normativnog« algoritma objašnjavamo taj proces. Posve je nejasno govoriti o »mehanički ostvarenom algoritmu u mozgu« zato što to »ostvarenje« nije algoritam. Mozak »radi« algoritamski samo pomoću našeg (aktivnog) tumačenja kao računalnog procesa (vidi raspravu o Fieldsovim postavkama), a ne »per se«. <sup>5</sup>

U *Pozivu na kognitivnu znanost* J. Leiber (1991) tvrdi da Wittgensteina s kognitivnom znanošću povezuje samo njegovo odbacivanje »stare paradigme« psihologije, tj. dualističke teorije utjelovljene duše i načela da svoja mentalna stanja, i to u njihovoj cjelovitosti, poznajemo samo pomoću introspekcije (Leiber, 1991: 46, 70). Ukidanje stare otvara put za »novu paradigmu«, to će reći za komputacijsku teoriju uma (Leiber, 1991: 65). Leiber vjeruje da je Wittgenstein »kognitivni naturalist« (Leiber, 1991: 61, 159). On navodi učenje jezika, prepoznavanje lica, percepciju aspekta i slijeđenje pravila kao neke važne opće istraživačke projekte koji povezuju Wittgensteina s kognitivnom znanošću. Leiber je žestoko kritizirao Wittgensteinovo gledište kao oprečno znanstvenoj psihologiji. Razmotrit ću samo jedno pitanje u Leiberovoj knjizi, naime: tvrdnju da Wittgensteinov prikaz značenja i razumijevanja prirodno vodi do komputacijskoga gledanja na um (Leiber, 1991: 67–68), te da je on anticipirao komputacijski model uma (Leiber, 1991: 109). Leiber upućuje na Wittgensteinovu postavku da značenje izraza jest njegova uloga u jezičnoj igri. Ovu postavku on tumači kao da značenje izraza jest njegovo mjesto u formaliziranom postupku (Leiber, 1991: 77–78). Leiber tu upućuje na Wittgensteinov primjer jezične igre trgovca u *Filozofskim istraživanjima*.

Netko je od trgovca mješovitom robom zatražio pet crvenih jabuka. Trgovac prvo otvara policu s natpisom ‘jabuke’, zatim gleda riječ ‘crven’ na popisu i nasuprot njega nalazi uzorak boje; zatim izgovara niz glavnih brojeva – sve do broja ‘pet’ i za svaki broj uzima jabuku iste boje kao što je uzorak na polici. Upravo na ovaj i slične načine čovjek operira s riječima (FI, par. 1). Ovu jezičnu igru Leiber tumači tako da značenja riječi ‘pet’, ‘crven’ i ‘jabuka’ jesu samo njihove uloge u »uzastopnim fizičkim postupcima« (Leiber, 1991: 66). Prema Leiberu, trgovčevo razumijevanje izraza »pet crvenih jabuka« sastoji se u njegovu vladanju tehnikom, naime, jednostavno u vještini dohvaćanja i grupiranja. Leiber vjeruje da kompjutor može izvesti isti po-

stupak kao i trgovac. Tako on zaključuje s tvrdnjom da je Wittgensteinov prikaz značenja i razumijevanja anticipirao komputacijski model uma.

Vidjeli smo da se Wittgenstein žestoko suprotstavio komputacijskome modelu uma (kao i reprezentacijskoj pretpostavci mišljenja i govora). Leiber prvo mora pokazati da uzastopni postupci uključeni u trgovčevoj jezičnoj igri jesu zapravo algoritamski te ih može izvesti i kompjutor, a zatim, da se i druge jezične igre događaju na isti način, naime izvodeći neke algoritamske postupke što ih može izvesti i kompjutor.

Prvo, teško je uvidjeti kako trgovčev postupak u jezičnoj igri uistinu jest algoritamski, zato što njegove radnje mogu biti daleko složenije nego što u prvi mah izgleda. Moglo bi biti da potpuni opis fizičkoga postupka nadilazi mogućnost bilo kojega algoritma, tako da se oni ne mogu simulirati na kompjutoru. Sigurno je da danu jezičnu igru možemo tumačiti u ideliziranome modelu, no to bi bio samo formalni simbol jezične igre u kojoj se mogu »napraviti« neke algoritamski definirane operacije, ali ne i stvaran proces jezične igre. Sigurno je da naše svakodnevno razumijevanje te jezične igre kao uzastopnog procesa može biti vrlo blisko algoritamskome modelu igre, no to razumijevanje ne mora biti blisko čisto fizičkom procesu. Naše intuitivno »gramatičko analiziranje« trgovčeva ponašanja u jezičnoj igri moglo bi biti veoma različito od stvarnih »elemenata« njegova ponašanja modeliranih pomoću kompjutora, zato što mi obično slijedimo jezični opis igre, a ne odgovarajući fizički opis događaja.

Leiber isto tako nije pružio nikakav dokaz za tvrdnju da sve jezične igre možemo prikazati algoritamski, kao u trgovčevu slučaju. On upućuje na Wittgensteinovu tvrdnju da su jezične igre dio naše prirodne povijesti, te zaključuje da mi posjedujemo prirodnu povijest formaliziranja (Leiber, 1991: 68). U njegovu prikazu, sve jezične igre uključuju formalizirane postupke, ali tu postavku ne podupire neki neovisan dokaz. Isti prigovor koji smo uputili jezičnoj igri trgovca možemo primijeniti na sve jezične igre. Ljudska sposobnost da izvodi uzastopne postupke može biti, zapravo, posljedicom naše prirodne povijesti, no ti postupci, kao i naša stručnost u njihovom provođenju, mogli bi ipak biti presloženi za bilo koji egzaktan formalni opis pomoću algoritma što bi ga moglo izvesti računalo.

Ima i drugo zanimljivo pitanje, slično onome što sam ga postavio u poglavlju u kojem sam raspravljao o Reyovoj postavci; pitanje, naime, je li Leiber zaista uspio pokazati da procesi u mozgu koji se zasnivaju na algoritmu mogu objasniti našu sposobnost da slijedimo pravila i da li kompjutorska simulacija tih procesa odgovara ljudskom nasljedovanju pravila. Kompjutor se zaista može »ponašati u skladu s pravilom«, a da ga stvarno ne slijedi, zato što mu nedostaje nadležnosti koja se za to traži. Ova se nadležnost, prema Wittgensteinu, razvija unutar zajednice koja provodi praksu nasljedovanja pravila, pridržava se nekih običaja, osposobljava se u različitim tehnikama nasljedovanja pravila, itd. Leiber bi, ako hoće pokazati da sve jezične igre sadrže formalno opisive algoritamske procese, trebao pokazati kako bi se tu aktivnost moglo prevesti u algoritamske procese koje bismo možda



mogli simulirati na kompjutoru. U protivnom, on bi trebao odbaciti razliku što ju je Wittgenstein povukao između ponašanja »u skladu s pravilom« i temeljitoga poznavanja toga pravila.

D. Proudfoot i B. J. Copeland nedavno su predstavili opsežnu kritiku Leiberove knjige, koja sam i sam slijedio, premda je njihova kritika ekstenzivnija, pa uključuje, primjerice, i kritiku reprezentacijskog razumijevanja značenja, govora i percepcije u komputacijskim modelima kognitivne znanosti. Proudfoot i B. J. Copeland odbacuju Leiberovo uvjerenje da je Wittgenstein ciljao na znanstveno razumijevanje uma koje nadilazi ono svakodnevno razumijevanje, te da je njegovo objašnjenje spoznaje bilo po svojoj nakani kauzalno, tj. proto-znanstveno (Proudfoot, Copeland, 1994). Autori ističu kako je razlika između konceptualnih i kauzalnih objašnjenja slična razlici između »što« i »kako« pitanja što ih koriste mnogi kognitivni znanstvenici (primjerice N. Block). U svojoj kritici funkcionalizma, Block tvrdi da pitanje »Kako procesor radi?« nije *pitanje na koje treba odgovoriti kognitivna znanost*. To pitanje možda spada u drugu disciplinu, teoriju elektroničkoga kruga. Moramo ipak razlikovati to *kako nešto radi* od toga *što ono radi*. Prema Blocku, pitanje *što* primitivni procesor radi jest dio kognitivne znanosti, no pitanje *kako* on radi to nije (Block, 1990: 257). Proudfoot i Copeland tvrde da »primitivna komponenta – »krajnja točka« – u objašnjavanju npr. značenja i razumijevanja jest prirodno ponašanje ljudskoga bića na razini običnog životnog iskustva«. No ipak, razjašnjavanje uloge prirodnog ljudskog ponašanja u stvaranju oblika života i jezičnih igara koje generiraju naš uobičajeni pojam uma – stvar je filozofije. Istraživanje uzročnog mehanizma koje leži u temelju takvoga ponašanja to nije (Proudfoot, Copeland, 1994: 514–515). Oni tvrde da »kognitivna znanost – utoliko ukoliko jest daljnje istraživanje onoga što za Wittgensteina jest primitivna komponenta – nije ništa drugo doli 'znanost realizacije'«. Ova uloga zasigurno ne pristaje uz tekuće pojmove mnogih filozofa uma. Oni tako zaključuju da, »suprotno Leiberovu gledištu, Wittgenstein nije nikakav kognitivni znanstvenik, čak ni onaj pod velikom krinkom«, a nije bio *ni* mentalist niti bihevorist, *ni* predlagatelj Leiberove »stare paradigme« niti kognitivni znanstvenik. Stare i nove paradigme ne predstavljaju istinsku dihotomiju, kao što to mnogo ljudi pretpostavlja.

Ove misli prikladno okončavaju moja razmišljanja o tome kako uloviti pticu misli u mrežu kognitivne znanosti. Po mome mišljenju, teoretičari kognitivne znanosti nisu još ozbiljno razmotrili Wittgensteinovu kritiku »unutrašnjih procesa« (i stanja), pa iznenađujuće često brkaju pitanje »kako to radi?« s onim »što to radi?«. Slično tome, oni miješaju uzročno objašnjenje s operacijama u simboličkome modelu (ili, još općenitije, s nasljedovanjem pravila), isto kao i ponašanje u skladu s pravilom s nasljedovanjem pravila. »Najozbiljnija pogreška«, međutim, jest miješanje unutarnjih komputacijskih (ili paralelnih) procesa u mozgu koji mogu uzrokovati mentalna stanja s vanjskim, društveno zasnovanim svakodnevnim kriterijima prepoznavanja, klasifikacije i poznavanja sadržaja mentalnih stanja.

S engleskoga preveo  
Šimun Selak



## Bibliografija:

Block, Ned (1990), »The computer model of mind«, u: D. N. Osherin – H. Lesnik (ur.), *An Invitation to Cognitive Science, Vol. 3, Thinking*. Cambridge/MA: MIT Press, 247–289.

Dietrich, Eric (1989), »Semantics and the Computational Paradigm in Cognitive Psychology«, *Synthese* 79: 119–141.

Fields, Chris (1996), »Measurement and Computational Description«, u: Peter J. R. Millican i Andy Clark, *Machines and Thought, Vol. 1*, Oxford: Clarendon Press, 165–177.

Leiber, Justin (1991), *An Invitation to Cognitive Science*. Oxford: Basil Blackwell.

McGinn, Marie (1997), *Wittgenstein and the Philosophical Investigations*. London, New York: Routledge.

Proudfoot, Diane; Copeland, B. Jack (2004), »Turing, Wittgenstein and the Sciences of the Mind. A Critical Notice of Justin Leiber 'An Invitation to Cognitive Science'«, *Australasian Journal of Philosophy* 74.

Rey, Georges (2003), »Why Wittgenstein Ought to Have Been a Computationalist?«, *Croatian Journal of Philosophy* 9, 231–264.

Ule, Andrej (1997), *Operationen und Regeln bei Wittgenstein*. Frankfurt/M.: Peter Lang Verlag.

Wittgenstein, Ludwig (1972), *Remarks on the Foundations of Mathematics*. Cambridge/MA: MIT Press.

Wittgenstein, Ludwig (1970), *Zettel (= Z)*, u: Ludwig Wittgenstein, *Schriften* 5. Frankfurt/M.: Suhrkamp.

Wittgenstein, Ludwig (1974), *Tractatus logico-philosophicus (= Tr)*. London: Routledge & Kegan Paul.

Wittgenstein, Ludwig (1976), *Philosophical Investigations (= FI)*. Oxford: Blackwell.

Wittgenstein, Ludwig (1980), *Bemerkungen über Philosophie der Psychologie (= PFP)*. Frankfurt/M.: Suhrkamp.

## Andrej Ule

### How Can One Catch a Thought-Bird?

#### Some Wittgensteinian Comments to Computational Modelling of Mind

*In this essay I analyse Wittgenstein's criticism of several assumptions that are crucial for a large part of cognitive science. These involve the concepts of computational processes in the brain which cause mental states and processes, the algorithmic processing of information in the brain (neural system), the brain as a machine, psychophysical parallelism, the thinking machine, as well as the confusion of rule following with behaviour in accordance with the rule. In my opinion, the theorists of cognitive science have not yet seriously considered Wittgenstein's criticism so they, quite surprisingly, frequently confuse the question "how does it work?" with "what does it do?". But their most "deleterious" mistake is their confusion of the internal computational (or parallel) processes taking place in the brain (which possibly cause mental states) with socially-based, everyday criteria of recognition and classification of, and knowledge about, the content of mental states.*

#### **Key words**

Ludwig Wittgenstein, cognitive science, computational processes, modelling of mind, mental states