**Umjetna inteligencija: Debata**

Napori da se proizvedu misleći strojevi rezultirali su u proteklih 35 godina zanimljivim nizom pokušaja od kojih su neki predstavljali napredak, a neki se pokazali čistim promašajima. Računala su savladala intelektualne zadaće poput šaha i integralnog računa, ali morat će poput jastoga odbaciti svoj oklop kako bi mogli rasti i naučiti nositi se sa stvarnim svijetom. Izvan uskog kruga stručnjaka iz područja umjetne inteligencije drži se da je stvaranje mislećeg stroja nemoguća misija – po svojoj prirodi računala su naprosto nesposobna spoznavati.

U tekstu koji slijedi, John R. Searle s Kalifornijskog sveučilišta u Berkeleyju zastupa tezu da računalni programi nikada neće postati umovi. S druge strane, Paul M. Churchland i Patricia Smith Churchland s Kalifornijskog sveučilišta u San Diegu tvrde da strujni krugovi izrađeni po uzoru na mozak mogu razviti inteligenciju. U pozadini ove debate leži temeljno pitanje – što zapravo znači misliti? Riječ je o izazovnom pitanju koje intrigira ljude (jedina bića za koja se zna da misle) tisućama godina. S pojavom računala (koja za sada ne misle), pitanje „što znači misliti“ pridobiva ponešto drugačije značenje i potiče niz mogućih odgovora. Konačni odgovor, dakako, tek treba utvrditi.

*Je li ljudski um računalni program?*

*- Ne. Program samo barata simbolima, um im pridaje značenje.*

*Autor John R. Searle*

Može li stroj misliti? Može li stroj imati misli na identičan način kao što ih imamo ti i ja? Ako pod strojem mislimo fizički sustav sposoban izvršiti određene radnje (što je gotovo pa samorazumljiva definicija) onda su ljudi biološki strojevi. Budući da ljudi misle, strojevi mogu misliti. Moguće je, ma kako to malo vjerojatno bilo, da ćemo uspjeti napraviti stroj koji misli, složen od različitih materijala, recimo silikonskih čipova i vakumskih cijevi. Taj se zadatak možda pokaže nemogućim, ali za sada mogućnost moramo ostaviti otvorenom.

Posljednjih desetljeća pitanje o kojem raspravljamo razumijevalo se međutim na jedan u cijelosti drugačiji način. Umjesto pitanja Može li stroj misliti? počelo se pitati Može li stroj misliti putem ugradnje računalnog programa? Čini li računalni program sam po sebi mišljenje? Riječ je, međutim, o jednom posve drugačijem pitanju. Prvo se odnosi na fizička, odnosno kauzalna svojstva stvarnih ili mogućih fizičkih sustava eventualno sposobnih misliti, a potonje na apstraktna računska svojstva formalnih računalnih programa koji mogu biti ugrađeni u bilo koju vrstu materije, pod pretpostavkom da ih ta materija može podržati.

Popriličan broj znanstvenika i istraživača koji se bave umjetnom inteligencijom (UI) uvjereno je kako je odgovor na drugo pitanje pozitivan. Uvjereni su da će izradom odgovarajućih programa s preciznim ulaznim i izlaznim informacijama doslovno proizvesti umove. Štoviše, uvjereni su da raspolažu i znanstvenim postupkom utvrđivanja uspjeha ili neuspjeha svog znanstvenog pothvata. Riječ je o Turningovom testu kojega je osmislio otac umjetne inteligencije Alan M. Turning. Onako kako ga se trenutno razumije, Turningov test svodi se na sljedeće: ukoliko kompjutor izvršava radnje na način da stručna osoba ne može razlikovati radnju koju je izveo kompjutor od radnje koju je izveo čovjek koji ima određene kognitivne sposobnosti (recimo da zna zbrajati ili razumije kineski), onda se može reći da i kompjutor ima iste takve kognitivne sposobnosti. Stoga je njihov cilj izraditi program koji će dovoljno dobro simulirati ljudske kognitivne sposobnost da položi Turningov test. Štoviše, jedan takav program, smatra se u tim krugovima, bio bi doslovno um, nimalo različit od ljudskog, a ne tek puka programska simulacija uma.

Dakako, ne zastupaju svi koji se bave UI krajnje stavove poput ovih. Oni koji pitanju pristupaju opreznije sugeriraju da računalni modeli mogu biti itekako korisni u proučavanju uma na isti način na koji su već korisni u meteorologiji, ekonomiji ili molekularnoj biologiji. Kako bismo razlikovali ta dva pristupa, onaj prvi, krajnji pristup nazvat ću jakom UI, ovaj drugi, oprezniji neka bude slab UI. Važno je uočiti u kojoj je mjeri jaki pristup smion. Zastupnici jake UI smatraju da je mišljenje puka manipulacija formalnim simbolima, što je točno ono što računalo i radi – manipulira formalnim simbolima. Taj stav često sažima uzrečica da je um mozgu što i softver hardveru.

Stanovište jake UI nema baš puno pristaša među teoretičarima uma. Razlog tome je jasan –jednostavno ga je odlučno opovrgnuti. Bilo tko to može učiniti samo ako se potrudi. Uzmimo ovako. Zamislite jezik kojega ne razumijete. U mom slučaju to je kineski. Kineska mi slova izgledaju kao škrabotine bez ikakva značenja. Zamislite sada da su me smjestili u sobu s brojnim košarama prepunima kineskih slova (simbola). Zamislite također da imam kod sebe Pravilnik na jeziku kojega razumijem, čija pravila govore o međusobnom slaganju kineskih simbola. Pravila određuju simbole isključivo prema obliku – od mene se ne zahtjeva da razumijem značenje znakova. Recimo, pravilo samo kaže da iz košare jedan uzmem znak koji izgleda tako i tako i stavim ga pored znaka iz košare dva koji izgleda tako i tako.

Zamislite nadalje da izvan sobe u kojoj se nalazim postoje ljudi koji razumiju kineski i koji mi svako malo daju male skupine simbola (kineskih slova), ja ih uzimam, baratam njima prema uputama iz Pravilnika i tako posložene ih vraćam natrag. Pravilnik je, dakako, računalni program, ljudi koji su ga napisali su programeri, a ja sam računalo. Košare prepune simbola su baza podataka, male hrpice koje primam su pitanja, a one koje vraćam su odgovori. Pravilnik je napisan tako dobro da „odgovori“ koje dajem nije moguće razlikovati od odgovora koje bi davala osoba čiji je materinji jezik kineski. Primjerice, mogao bih dobiti hrpicu meni potpuno nerazumljivih simbola koji bi konstituirali pitanje: „Koja ti je omiljena boja?“, a ja bih, nakon što sam konzultirao Pravilnik, odvratio meni jednako tako nerazumljivim nizom simbola koji bi konstituirali odgovor: „Omiljena mi je plava, ali volim i zelenu.“ Zadovoljio sam Turingov test premda nemam blage veze s kineskim. No, ne samo što ne znam kineski, nego ga u okviru opisanog sustava nikada nemam niti šansu naučiti. Jednostavno ne postoji način da naučim značenje simbola. Poput računala, baratam (manipuliram) simbolima ali im ne pripisujem bilo kakvo značenje.

Misaoni eksperiment je dokazao sljedeće – ako čovjek ne razumije kineski samo zato što se služi Pravilnikom (računalnim programom) za slaganje kineskih simbola, onda niti jedno digitalno računalo ne razumije kineski samo zato što se služi istim takvim računalnim programom. Računala samo barataju formalnim simbolima prema naputcima računalnog programa. Što vrijedi za poznavanje kineskog, vrijedi za sve druge spoznajne oblike. Puko baratanje simbolima nije po sebi dovoljno da rezultira poznavanjem, percepcijom, razumijevanjem, mišljenjem i ostalim oblicima spoznaje. Budući da su računala po sebi samo strojevi za manipulaciju znakovima, puko pokretanje računalnog programa ne rezultira spoznajom.

Navedeni argument je jednostavan i u cijelosti pobija tvrdnje pristaša jake UI. Prva njegova premisa poziva se na formalni karakter računalnog programa. Programi se definiraju u pogledu funkcije manipulacije simbolima koju obavljaju, a simboli su čisto formalni, odnosno sintaktički. Napominjem da je upravo ovaj formalni karakter računalnih programa ono što računala čini tako moćnima. Isti program moguće je koristiti na beskonačnom nizu različitih računala, a jedno računalo može pokrenuti beskonačno mnogo raznih računalnih programa. Dozvolite mi da rečeno sažmem u aksiom.

*Aksiom 1: Računalni programi su formalni (sintaktički).* Ova je tvrdnja u toj mjeri temeljna da ju je potrebno dodatno detaljnije objasniti. Računalo procesuira informacije tako da ih prvo šifrira vlastitim simboličkim jezikom, a potom manipulira simbolima pomoću skupa precizno zadanih pravila. Skup tih pravila je računalni program. Primjerice, u Turingovoj ranoj računalnoj teoriji simboli su bili '0' i '1', a pravila programa izdavala bi naredbe poput „Ispiši 0 na traku, pomakni se za jedno mjesto u lijevo i izbriši 1“. Ono što je fantastično kod kompjutora je mogućnost da se bilo koja informacija iskazana u nekom jeziku može šifrirati i da se bilo koji zadatak procesuiranja informacija kojega je moguće riješiti pomoću eksplicitnih pravila ujedno može i programirati.

Nadalje, važno je napomenuti sljedeće dvije činjenice. Kao prvo, simboli i programi su čisto apstraktni pojmovi. Lišeni su bilo kakvih fizičkih svojstava koja bi ih definirala pa ih je moguće implementirati u bilo koji fizički medij. Nule i jedinice, uzeti kao simboli, nemaju bitna fizička svojstva, pa po logici stvari nemaju niti uzročna svojstva koja proizlaze iz fizičkih. Ovo je važno naglasiti iz dva razloga. Prvi je taj što postoji tendencija poistovjećivanja računala s određenom tehnologijom – recimo sa silikonskim čipovima – te se slijedom toga drži da je ovdje riječ o problemu fizičkih svojstava čipova. Drug razlog je taj što postoji i mišljenje kako sintaksa (u simbolima) prepoznaje fizičku pojavu koja ima trenutno nepoznatu nam moć uzrokovanja, na isti onaj način na koji tu moć imaju stvarne fizičke pojave poput elektromagnetskog zračenja ili atoma vodika, koje kao izvjesno fizičke imaju uzročna svojstva. Druga činjenica koju je važno napomenuti jest da se simbolima manipulira ne pozivajući se na bilo kakvo značenje. Simboli programa mogu označavati bilo što – što god programeru ili korisniku padne na pamet. U tom smislu, program ima svoju sintaksu, ne i semantiku.

Sljedeći je aksiom samo podsjetnik na očiglednu činjenicu da misli, percepcije, razumijevanje i slični fenomeni imaju mentalni sadržaj. Upravo temeljem svog sadržaja odnose se na objekte i stanja stvari u svijetu. Ako mentalni sadržaj uključuje jezik, pored sintakse u sadržaj će biti uključena i semantika - jezično razumijevanje zahtijevat će barem nekakav semantički okvir. Ako, primjerice, razmišljam o zadnjim predsjedničkim izborima, kroz glavu će mi prolaziti određene riječi. One će se, međutim, odnositi na predsjedničke izbore samo zato što im ja pripisujem određeno značenje u skladu s poznavanjem engleskog jezika. U tom smislu, za mene te riječi nisu nimalo nalik kineskim simbolima. Tu činjenicu također ću sažeti u obliku aksioma.

*Aksiom 2: Ljudski umovi imaju mentalne sadržaje (semantički su).* Dozvolite mi da sada dodam činjenicu koju je misaoni eksperiment s kineskom sobom dokazao. Ako raspolažemo simbolima samima po sebi i njihovom sintaksom, to još nije dostatno da bismo dobili njihovu semantiku. Puko slaganje simbola (po sintaktičkim pravilima) ne garantira razumijevanje njihova značenja. Ovu ću činjenicu sažeti u treći aksiom.

*Aksiom 3: Sintaksa, sama po sebi ne čini niti je dostatna za semantiku*. Na jednoj razini, rečeno proizlazi iz samih definicija pojmova sintaksa i semantika. Može se, dakako, te pojmove i drugačije definirati. Radi se, međutim, o tome da postoji jasna razlika između formalnih elemenata, koji nemaju intrizično značenje ili sadržaj, i pojava koje posjeduju intrizičan sadržaj. Iz ove tri premise slijedi prva konkluzija.

*Konkluzija 1: Računalni programi nisu bitne sastavnice umova niti su dostatni tvoriti ih*. Konkluzija dokazuje da je stanovište jake UI pogrešno. Ovdje je važno uočiti što se gornjim zaključkom dokazalo, a što ne. Kao prvo, nisam pokušao dokazati da „računalo ne može misliti“. Sve što se može računalno simulirati može se smatrati računalom. Budući da se naš mozak može donekle računalno oponašati, slijedi da su mozgovi, koji nedvojbeno misle, neka vrsta računala. Međutim, iz činjenice da neki sustav može biti oponašan slaganjem (manipuliranjem) simbola i da misli, ne slijedi da je mišljenje ekvivalentno formalnom (sintaktičkom) slaganju simbola. Kao drugo, nisam pokušao prikazati kako samo biološki utemeljeni sustavi poput našeg mozga mogu misliti. Za sada, mozak je jedini misleći sustav za kojeg sa sigurnošću možemo reći da misli, no moguće je da u svemiru naiđemo na druge sustave sposobne proizvesti svjesne misli, štoviše, nije nemoguće da u budućnosti čak i stvorimo umjetne misleće sustave. Te izazove prepuštam drugima. Treće, teza koju zastupaju pristaše jake UI nije da, koliko za sada znamo, računala uz odgovarajuće programe možda misle, odnosno da možda imaju još neotkrivena psihička svojstva. Njihova je teza puno jača, naime da računala zasigurno misle, jer mišljenje je ni više ni manje nego točno ono što računala rade. Četvrto, pokušao sam pobiti upravo tako definiranu tezu pristaša jake UI, dokazujući da program sam po sebi nije bitna sastavnica mišljenja budući da je program zapravo priručnik za formalno slaganje simbola. Iz drugog izvora znamo da manipulacijom čistim simbolima nikada nećemo proizvesti značenje. Argument 'kineske sobe' upravo to dokazuje.

Imam potrebu ovdje naglasiti ove činjenice dijelom i zato što mi se čini da Churchlandi (vidi članak „Bi li stroj mogao misliti“?“; autori Paul M. Churchland i Patricia Smith Churchland, strana 32] nisu baš razumjeli problematiku. Oni smatraju da zagovornici jake UI tvrde kako je u konačnici moguće da računala misle, a da ja odbijam tu mogućnost na temelju zdravog razuma. Međutim, niti zagovornici jake UI tvrde ono što Churchlandi navode, niti se moj argument protiv jake UI temelji na zdravom razumu. Kasnije ću se još osvrnuti na njihove primjedbe. Do tada, na ovom mjestu želim istaknuti, suprotno onome što Churchlandi sugeriraju, da argument kineske sobe također pobija bilo koju tvrdnju u korist jake UI, a koja se poziva na nove paralelne tehnologije inspirirane i modelirane po uzoru na neuronske mreže. Za razliku od tradicionalnog Neumannovog računala, koje funkcionira serijski, korak po korak, sustavi koji koriste paralelnu tehnologiju imaju brojne računalne elemente koji funkcioniraju paralelno i uzajamno djeluju prema pravilima inspiriranima neurobiologijom. Premda su rezultati za sada skromni, ovi modeli potiču korisna pitanja o tome na koji način složeni, paralelni mrežni sustavi poput onih u mozgu mogu zapravo imati ulogu u stvaranju inteligentnih oblika ponašanja.

Paralelna obrada podataka, nalik na onu što se događa u mozgu, svojstvo je međutim koje je irelevantno s obzirom na čisto računalni vid procesa. Bilo koja funkcija koja se može izračunati na stroju koji obrađuje paralelno, može se izračunati i na serijskom stroju. Štoviše, budući da su paralelni strojevi još uvijek rijetki, paralelni se programi obično pokreću na tradicionalnim serijskim strojevima. Budući da je riječ o drugačijem, ali još uvijek samo o načinu obrade podataka, paralelnost je jednako tako podložna argumentu kineske sobe. Štoviše, paralelni sustav obrade podataka, čak i zasebno promatran, podložan je primjedbi sličnoj onoj koja je iznesena u izvornom argumentu kineske sobe. Zamislite sada da je umjesto kineske sobe riječ o kineskoj teretani i da se u njoj nalazi mnoštvo muškaraca koji govore samo engleskim jezikom. Oni bi obavljali iste radnje koje bi obavljali čvorovi i sinapse u paralelnom sustavu kakvoga opisuju Churchlandi, a ishod bi bio isti kao da jedan čovjek manipulira simbolima služeći se Priručnikom. Nitko u teretani ne govori niti riječ kineskog, a ne postoji niti način kako bi sustav kao cjelina naučio značenja kineskih riječi. Ipak, uz prikladnu prilagodbu, sustav bi mogao početi izbacivati odgovore na pitanja na kineskom.

Postoje, kao što sam već ranije naveo, zanimljiva svojstva paralelnih mreža koja im omogućuju da imitiraju procese u mozgu preciznije nego što su to u stanju tradicionalni serijski sustavi. Međutim, prednosti koje paralelni ustroj ima za argumente u korist slabe UI poprilično su irelevantne za odnos između argumenta kineskog zida i jake UI. Churchlandi pokazuju nerazumijevanje kada tvrde da bi dovoljno velika kineska teretana mogla posjedovati mentalne karakteristike više razine naprosto na temelju svoje veličine, odnosno na temelju veličine i složenosti samog sustava. Otprilike onako kako mozak u cjelini ima karakteristike koje nemaju pojedini neuroni. To, dakako, nije isključeno, ali nije ni na koji način povezano s računanjem. Što se računanja tiče, serijski i paralelni sustavi su istovrijedni – svaki izračun koji je moguć u paralelnom sustavu, jednako je moguć u serijskom. Pod pretpostavkom da je čovjek u kineskoj sobi računarski istovrijedan u oba sustava, onda slijedi da ako on ne razumije kineski samo zato što računa, onda kineski ne razumiju niti brojni ljudi u teretani koji se bave isključivo računanjem. Churchlandi su u pravu kada tvrde da je izvorni argument kineske sobe konstruiran imajući u vidu tradicionalnu UI, ali su u krivu kada drže da je UI bazirana na paralelnim sklopovima imuna na argument kineske sobe. Argument je primjenjiv na svaki računalni sustav. Nemoguće je dobiti semantički učitan misaoni sadržaj pukim računanjem, neovisno o tome računa li se serijski ili paralelno. Stoga argument kineske sobe pobija tezu jake UI neovisno u kojoj se formi ona pojavi.

Mnogi impresionirani ovim argumentom su svejedno zbunjeni razlikama između ljudi i računala. Ako su ljudi barem trivijalno gledajući računala i ako ljudi raspolažu semantikom, zašto onda semantiku ne bismo mogli podariti i drugim računalima? Zašto ne bismo programirali naše kućno računalo ili laptop tako da i ono ima misli i osjećaje? Ili, zašto ne bi neka nova računalna tehnologija premostila razliku između forme i sadržaja, između sintakse i semantike? Koje su to zapravo razlike između životinjskih mozgova i računala, zbog kojih argument kineske sobe pogađa računala, dok su mozgovi na njega potpuno imuni? Najočiglednija razlika je u tome što su procesi koji definiraju nešto kao računalo, takozvani računalni procesi, nešto posve neovisno o stroju u koji se implementiraju i koji ih pokreće. U teoriji, računalo bi se moglo napraviti od odbačenih limenki pive, međusobno povezanih žicama i napajanih vjetrenjačama.

Međutim, kad se radi o mozgu, premda znanost uglavnom nema pojma kako mozak funkcionira i proizvodi mentalna stanja, ljudi su zapanjeni njegovim anatomskim i fiziološkim specifičnostima. Tamo gdje postoji nekakvo razumijevanje načina na koji procesi u mozgu stvaraju mentalne fenomene poput boli, žeđi, vida ili njuha, jasno je da su u stvaranju mentalnih fenomena uključeni određeni neurobiološki procesi. Žeđ, barem određenu njenu vrstu, uzrokuju signali koje neuroni odašilju u hipotalamusu, a ovi su redom uzrokovani djelovanjem peptidnog hormona angiotensin II. Kauzalni proces ide odozdo prema gore, na način da neuronski procesi niže razine uzrokuju mentalne fenomene više razine. Uistinu, koliko za sada znamo, svaki mentalni događaj, u rasponu od osjeta žeđi do misli o matematičkim teoremima i memorija iz djetinjstva, uzrokovan je određenim neuronima koji šalju signale u okviru određenog neuronskog sklopa.

Zašto bi anatomske i fiziološke specifičnosti mozga bile važne? Napokon, odašiljanje signala može se simulirati na računalima, a ova imaju posve drugačiju fizičku i kemijsku građu od mozga. Anatomske i fiziološke specifičnosti mozga su važne jer mozak, osim što je primjer formalne strukture ili programa također uzrokuje mentalne događaje putem specifičnih neurobioloških procesa. Mozgovi su specifični biološki organi. A njihova specifična biokemijska svojstva omogućuju im da uzrokuju svijest i druge vrste mentalnih fenomena. Računalne simulacije procesa koji se događaju u mozgu nude modele formalnih vidova tih procesa. Međutim, simulacije ne treba brkati s dupliciranjem. Računalni model mentalnih procesa nije ništa stvarniji od računalnog modela bilo koje druge prirodne pojave. Možemo lako zamisliti do zadnje sinapse preciznu računalnu simulaciju djelovanja peptida u hipotalamusu. No, jednako tako možemo zamisliti računalnu simulaciju oksidacije ugljikovodika u motoru automobila ili probavnih procesa u želucu koji probavlja pizzu. Pri tome računalna simulacija zbivanja u mozgu nije ništa stvarnija od računalne simulacije zbivanja u motoru ili želucu. Ne vjerujete li u čuda, teško da ćete pokrenuti automobil pomoću računalne simulacije oksidacije goriva u motoru, niti ćete probaviti pizzu pokretanjem programa koji simulira probavu u vašem želucu. Čini se očiglednim da računalna simulacija spoznaje jednako tako neće proizvesti neurobiološke posljedice koje bi rezultirale spoznajom. Ako je tako, onda su svi mentalni fenomeni uzrokovani isključivo neuropsihološkim procesima u mozgu. Tako dolazimo i do četvrtog aksioma.

*Aksiom 4: Mozak je uzrok umu.*

Vezano za prethodni izvod, smjesta, pomalo banalno, zaključujem sljedeće.

*Konkluzija 2: Bilo koji drugi sistem koji bi mogao polučiti um morao bi imati kauzalnu snagu (barem) jednaku onoj koju ima mozak.*

To je isto kao da kažem da ako električni motor ima pogurati automobil istom brzinom kao i benzinski, električni bi motor morao imati (barem) istu izlaznu snagu kao benzinski. Ova konkluzija ne govori baš ništa o samim mehanizmima. Kako stvari stoje, spoznaja je biološki fenomen: mentalna stanja i procesi uzrokovani su moždanim procesima. Time ne impliciram da samo biološki sustav može misliti, ali doista impliciram da bi bilo koji drugi sustav sposoban misliti, napravljen od silikona, limenki pive ili bilo čega drugoga, morao imati relevantne kauzalne sposobnosti proizvođenja mišljenja ekvivalentne mozgu. Stoga sada mogu izvesti sljedeću konkluziju.

*Konkluzija 3: Bilo koji umjetni predmet koji bi proizveo mentalni fenomen, dakle bilo koji umjetni mozak, morao bi biti sposoban duplicirati posebne kauzalne moći koje ima mozak, a to ne bi mogao napraviti pukim pokretanjem formalnog programa.*

Nadalje, sada mogu izvesti i jednu važnu konkluziju o ljudskom mozgu.

*Konkluzija 4: Način na koji ljudski mozak zbilja proizvodi mentalne fenomene ne može se svesti na puko pokretanje računalnog programa.*

Prvi sam put predstavio usporedbu poznatu kao kineska soba 1980. godine na stranicama časopisa Behavioral and Brain Sciences . Članak se pojavio, kako je i običaj u tom časopisu, zajedno s komentarima kolega znanstvenika, ukupno njih 26. Iskreno, meni se čini da je zaključak koji nudi neprijeporan, ali na moje veliko iznenađenje popratila ga je poplava komentara. Štoviše, na moje još veće iznenađenje, bujica komentara se niti do današnjeg dana ne stišava. Argument kineske sobe očito je mnoge uznemirio. Teza zastupnika teorije jake UI jest da bilo koji sustav, neovisno od čega je složen, limenki pive, silikona ili toaletnog papira, ne samo da bi mogao imati misli i osjećaje, nego da bi ih morao imati, pod uvjetom da primijeni ispravan program, s preciznim ulaznim i izlaznim podacima. Riječ je o duboko anti-biološkom gledištu i čovjek bi pomislio da će ga zastupnici teorije jake UI rado napustiti. Mnogi, naročito oni mlađi, slažu se sa mnom, a začuđuje me i broj i žestina onih koji ju brane. Slijede neke uobičajene primjedbe koje dolaze s njihove strane.

a. Osoba u kineskoj sobi zapravo razumije kineski, premda to ne zna. U konačnici, moguće je nešto razumjeti, a da ne znaš da to razumiješ.

b. Osoba u kineskoj sobi ne razumije kineski, ali postoji jedan (podsvjesni) sustav u njoj koji ga razumije. U konačnici, moguće je imati nesvjesna mentalna stanja i ne postoji razlog da tvoje razumijevanje kineskog ne bi bilo u cijelosti nesvjesno.

c. Osoba u kineskoj sobi ne razumije kineski, ali soba kao cjelina ga razumije. Ti si samo jedan neuron u mozgu i kao što jedan zaseban neuron ne razumije, tako niti ti ne razumiješ. Međutim, kako svaki neuron u mozgu pridonosi razumijevanju, tako i ti pridonosiš razumijevanju. Neuron ne razumije, ali mozak razumije. Ti ne razumiješ, ali sustav razumije.

d. Semantika ionako ne postoji, postoji samo sintaksa. Vrsta je pred-znanstvene iluzije pretpostaviti da u mozgu postoje nekakvi tajanstveni mentalni sadržaji, misaoni procesi ili semantika. U mozgu postoje samo sintaktički procesi manipulacije simbolima kao što je to slučaj u bilo kojem računalu. To i ništa više.

e. Osoba u kineskoj sobi ne pokreće računalni program – samo joj se čini da to radi. Onog trenutka kada svjesni agent prolazi kroz korake računalnog programa, tu se više uopće ne radi o implementaciji programa.

f. Računala bi imala ne samo sintaksu nego i semantiku kad bi se njihovi ulazni i izlazni podaci stavili u odgovarajući kauzalni odnos s ostatkom svijeta. Kada bismo stavili računalo u robota, pričvrstili TV kamere robotu na glavu, instalirali pretvarače koji povezuju informacije koje prikuplja kamera s računalom i dali računalu da upravlja robotovim rukama i nogama, takav bi sustav kao cjelina imao i semantiku.

g. Kada bi program simulirao način rada mozga nekog Kineza, program bi razumio kineski. Preciznije, kada bismo simulirali rad mozga Kineza na razini neurona, tada bi jedan takav sustav razumio kineski jednako kao što bi mozak bilo kojeg Kineza razumio kineski.

I tako dalje.

Navedeni argumenti dijele jednu zajedničku karakteristiku – neformalno su nevaljani jer izravno ne pobijaju sami argument kineske sobe na koji bi se trebali odnositi. Argument kineske sobe počiva na razlici između formalne manipulacije simbolima, čime se bavi računalo i mentalnih sadržaja koje na biološkim osnovama proizvodi mozak. Navedenu sam razliku sažeo, nadam se ne na način koji bi zbunjivao čitatelje, na razliku između sintakse i semantike. Neću se ponavljati odgovarajući na sve gore iznesene prigovore protiv argumenta kineske sobe. Prigovorom „simulator mozga“ (the Brain Simulator, pod g.), koji je inače poprilično popularn, već sam se pozabavio ranije u tekstu. Međutim, pomoći će u razjašnjenju nastalih nedoumica ako pobliže navedem slabosti najraširenijeg prigovora – onog pod c., kojeg nazivam „prigovorom sustava“ (the Systems reply).

Kako to tvrde pristaše navedenog prigovora, naravno da osoba u kineskoj sobi ne razumije kineski, ali čitav sustav – osoba u sobi, soba, priručnik, velike košare prepune simbola – sustav kao cjelina razumije kineski. Kada sam prvi put čuo ovaj prigovor, postavio sam jednom od njegovih zagovornika sljedeće pitanje:“ Hoćeš reći da soba razumije kineski?“ Odgovorio je potvrdno. Hrabar odgovor, no osim što je krajnje neuvjerljiv, ne funkcionira niti na čisto logičkoj razini. Izvorni argument dokazivao je da pomicanje i slaganje simbola samo po sebi ni na koji način ne pruža uvid u njihovo značenje. Ta tvrdnja jednako vrijedi za sobu u cjelini, kao i za osobu u sobi. Navedeno možemo lako dokazati tako da malo proširimo misaoni eksperiment. Zamislimo da sam uspio memorirati sadržaj košara sa simbolima i kompletan priručnik te da sve potrebne računalne radnje potrebne za slaganje simbola znam napamet. Zaboravite na sobu, jer zašto ne bih slagao simbole na otvorenom. Sada više nema ničega u sustavu čega nema u meni (ja sam sustav), a kako ja ne razumijem kineski, ne razumije ga niti sustav.

U popratnom članku Churchlandi nude inačicu „prigovora sustava“ zamišljajući zanimljivu analogiju. Pretpostavimo, kažu Churchlandi, da netko tvrdi da svjetlost ne može biti elektromagnetskog porijekla, jer kada mašete magnetnom šipkom u mračnoj sobi, sustav neće proizvesti svjetlo. Nije li, pitaju se Churchlandi, argument kineske sobe upravo to – ne tvrdi li on da ako mašemo kineskim simbolima u mračnoj sobi, samo mahanje simbolima neće baciti svjetlo na razumijevanje kineskog? Znamo, nastavljaju Churchlandi, da su kasnija istraživanja dokazala da je svjetlost vrsta elektromagnetskog zračenja te se pitaju nije li moguće da kasnija istraživanja pokažu i da se semantika u cijelosti sastoji od sintakse. Ne bi li se znanstvena zajednica nadalje trebala pozabaviti i tom mogućnošću?

Poznato je da su argumenti po analogiji slabašni. Prije no što bi jedan takav argument imao smisla, potrebno je sa sigurnošću utvrditi da su slučajevi koji se uspoređuju dovoljno slični da bi bili usporedivi. Ovdje, držim, to nije slučaj. Opis svjetla u terminima elektromagnetskog zračenja je pitanje bazičnog kauzaliteta. Riječ je o uzročno-posljedičnom opisu fizičkih svojstava fizičkog fenomena. Analogija s formalnim simbolima je neuspjela budući da formalni simboli nemaju nikakvih fizičkih svojstava, a time niti moć fizičkog uzrokovanja. Jedinu moć koju simboli po sebi imaju je moć da, kada je stroj upaljen i radi, formalno uzrokuju sljedeći korak programa.

Stoga nema nikakvog razloga čekati rezultate daljnjih istraživanja koja bi otkrila fizička, odnosno kauzalna svojstva nula i jedinica. Njihova jedina relevantna svojstva su apstraktna računalna svojstva, a ta su nam svojstva već poznata. Churchlandi prigovaraju da izbjegavam pobliže objasniti tvrdnju da neinterpretirani formalni simboli nisu isto što i mentalni sadržaji. Uistinu, nisam se baš pretrgao argumentirati navedenu tvrdnju budući da ju držim logički istinitom. Kao što je slučaj kod svih istinitih tvrdnji te vrste, u njenu se istinitost možemo uvjeriti neposredno. Čim zamislimo da je istinita njoj proturječna tvrdnja dolazi do niza nekonzistencija. Uvjerimo se u to. Pretpostavimo da je u kineskoj sobi ipak na djelu neki oblik neopaženog mišljenja na kineskom. Sada se upitajmo što je to što slaganje sintaktičkih elemenata (simbola) pretvara u baš kineski misaoni sadržaj. Uostalom, pretpostavljam da su programeri bili govornici kineskog i da su programirali sustav da procesuira informacije na kineskom.

Sve pet. No, sada zamislite da mi, dok sjedim u kineskoj sobi i prtljam s kineskim simbolima, to prtljanje simbolima bez meni ikakvog značenja dosadi. Tada iz dosade odlučim interpretirati simbole kao poteze u šahu. Koja se semantika dade u tom slučaju iščitati iz sustava? Odaje li sustav semantiku kineskog jezika, semantiku šaha ili obje simultano? Zamislimo nadalje da postoji netko tko sve to promatra kroz prozor i manipulaciju simbolima koju vidi protumači kao predviđanja kretanja na burzi. I tako u beskraj. Nema kraja broju semantičkih interpretacija koje se mogu pripisati simbolima jer, ponovimo to još jednom, simboli su čisto formalni. Nemaju nikakvu intrizičnu semantiku.

Može li se kako analogija koju su postavili Churchlandi spasiti, može li se učiniti koherentnom? Ranije sam napomenuo kako formalni simboli nemaju kauzalna svojstva. S druge strane, naravno da će program uvijek biti implementiran u neki hardver, ovakav ili onakav, a hardver će imati specifičnu moć fizičkih fenomena – moć uzrokovanja. Bilo koje stvarno računalo odavat će svojstva različitih vrste fenomena. Moja računala, primjerice isijavaju toplinu i proizvode razne vrste zvukova. Ima li kakav logički obvezujući razlog zbog kojega uz to ne bi mogli odavati i svijest? Nema. Sa znanstvenog stanovišta, tvrditi da računala ne mogu odavati svijest bilo bi neprihvatljivo. No, argument kineske sobe uopće se ne bavi pobijanjem mogućnosti da računala uzrokuju svijest, niti je riječ o tvrdnji koju bi zagovornici jake UI uopće htjeli braniti. Jer, ako na bilo koji način uzrokuju svijest, onda to čine temeljem fizičkih svojstava koja imaju kao medij u kojem se program implementira. Zagovornici jake UI, upravo suprotno, inzistiraju na tome da su fizička svojstva računala kao medija u kojem se implementira program posve irelevantna za svijest. Bitni su programi, a oni su, kao što znamo, posve formalne prirode. Stoga se analogija između sintakse i elektromagnetizma koju predlažu Churchlandi suočava sa sljedećom dilemom: ili je sintaksa konstruirana čisto formalno na temelju svojih apstraktnih matematičkih svojstava ili nije. Ako je, onda analogija ne drži vodu jer formalno konstruirana sintaksa nema nikakva fizička svojstva, a slijedom toga niti ikakvu fizičku moć uzrokovanja. Ako, s druge strane, sintaksa nije konstruirana čisto formalno, nego treba uzeti u obzir i svojstva fizičkog medija u kojem je implementiran program, onda analogija stoji, ali je potpuno irelevantna za teze jake UI.

Budući da su tvrdnje koje iznosim poprilično samorazumljive – sintaksa nije isto što i semantika, procesi u mozgu uzrokuju mentalne fenomene – zanimljivo je pitanje kako smo se uopće našli u ovoj zbrci. Kako je netko uopće mogao pomisliti da je računalna simulacija mentalnih procesa isto što i zbiljski mentalni procesi? Uostalom, stvar je kod modela upravo u tome da sadrže samo određena, a izostavljaju ostala svojstva onoga čega su modeli. Nitko uistinu ne očekuje da će se smočiti u bazenu ispunjenom ping-pong lopticama koje predstavljaju modele molekula vode. Zašto bi onda netko pomislio da računalni model misaonog procesa zapravo misli? Dio odgovora na ova pitanja krije se u činjenici da smo naslijedili ostatke biheviorizma prošlih generacija. Turningov test drži nas u stalnom iskušenju da bihevioristički držimo da ono što se ponaša kao da ima neke mentalne procese uistinu te procese i posjeduje. U tome, između ostaloga, leži dio promašene biheviorističke pretpostavke da psihologija, kako bi se smatrala znanošću, mora ograničiti svoje područje istraživanja na vidljive vanjske oblike ponašanja. Paradoksalno, ovi ostaci biheviorističke teorije usko su vezani za još uvijek prisutne ostatke dualističke ontologije. Nitko ne drži da bi računalna simulacija procesa probave bilo što probavila, ali kada je u pitanju spoznaja, ljudi su skloni vjerovati takvim čudima jer ne mogu prihvatiti da je spoznaja biološka činjenica jednako tako kao što je to probava. Um je, smatraju, nešto formalno i apstraktno, nipošto nešto mokro i ljigavo u našim glavama. Polemička literatura o UI uglavnom sadrži radove koji polemiziraju s onim što autori nazivaju dualizmom. Međutim, ti autori ne vide da i sami zrcale dualizam i to u njegovom ekstremnom obliku. Naime, samo ako prihvatite ideju da je um posve neovisan o mozgu ili bilo kojem drugom specifično tjelesnom sustavu, možete se nadati da ćete stvoriti umove dizajniranjem programa.

Povijesno gledano, znanstvenim kretanjima na Zapadu koja su sagledavala čovjeka kao dio redovnog fizičkog, odnosno biološkog poretka, često su se suprotstavljali razni, uvijek iznova prepravljani pokušaji opovrgavanja. Kopernik i Galileo bili su osporavani jer su poricali da je Zemlja centar svemira. Darwin je opovrgavan jer je tvrdio da ljudi potječu od nižih vrsta. Najbolje je stoga na ideju jake UI gledati kao na jedan od posljednjih trzaja protu-znanstvene tradicije. Jaka UI, naime, poriče da um ima bilo kakve fizičke i biološke osnove. Prema zastupnicima jake UI, um je posve neovisan o mozgu. Um je računalni program i kao takav nema nikakve bitne poveznice s bilo kojim specifičnim mu hardverom.

Mnogi koji sumnjaju da bi se umjetnoj inteligenciji mogla pripisati bilo kakva psihološka značajka svejedno misle kako bi računala možda mogla biti u stanju razumjeti kineski i razmišljati o brojevima, ali da ne mogu obavljati one bitno ljudske stvari. Među tim bitno ljudskim stvarima posebno se ističu zaljubljivanje, smisao za humor, sveprisutan osjećaj tjeskobe u postindustrijskom društvu kasne faze kapitalizma i sl. No, oni koji se bave umjetnom inteligencijom s pravom negoduju da se time u sred igre mijenjaju pravila – gol linija se stalno udaljava. Čim umjetna inteligencija uspješno simulira mentalni proces, simulacija gubi psihološki značaj. U ovoj debati obje strane propuštaju uočiti razliku između simulacije (oponašanja) i dupliciranja. Kada je riječ o simulaciji, nije uopće teško programirati računalo da ispiše „Volim te, Suzy.“, „Ha-ha.“ ili „Patim od tjeskobe postindustrijskog društva pod kasnim kapitalizmom.“ (i time oponaša ljubav, humor ili tjeskobu). Važno je još jednom naglasiti da simulacija nije isto što i dupliciranje, a ta činjenica ima podjednaki značaj kod promišljanja aritmetičkih problema kao i kod osjećanja tjeskobe. Ne radi se o tome da računalo uspijeva doći s loptom samo do ruba šesnaesterca, ali ne i do gol linije. Računalo uopće ne igra nogomet.