

Filipović Nenad

KINESKA SOBA I IDENTITET PROGRAMA

APSTRAKT: Argument poznat kao Kineska soba je slavni Serlov argument koji je pomoću jednostavnog misaonog eksperimenta pokazao razne probleme sa tezom prema kojoj je moguće da postoji veštačka inteligencija koja ispoljava razumevanje kakvo je svojstveno inteligentnim bićima kao što su ljudi. Sam argument je bio dovoljno uticajan da u decenijama posle objavljivanja donese veliki broj odgovora i kritika, kao i nekoliko Serlovih pokušaja poboljšanja argumenta. U ovom tekstu obratiću pažnju na jednu atipičnu kritiku Serlovog argumenta, koju je izneo Sprevak. Sprevak za razliku od uobičajenih kritika prihvata da unutar kineske sobe nema razumevanja, ali tvrdi da kineska soba ne može izvršiti svaki program kao što je to Serl mislio, i da utoliko misaoni eksperiment nije dovoljan za jak zaključak koji Serl izvodi. U tekstu ću analizirati Serlov argument; izneti pregled uobičajenih kritika; analizirati Sprevakovu kritiku, i izneti argumente koji pokazuju da se Sprevak, da bi izvukao zaključke koje želi, mora odreći ili dela svoje kritike ili dodati i jednu od tipičnih kritika uz svoju, čineći time svoju kritiku „zavisnom“ od drugih.

KLJUČNE REČI: kineska soba, program, algoritam, funkcija, Serl, Sprevak.

Uvod

Poznati Serlov [Searle] argument *Kineske sobe* usmeren je protiv teze Veštačke inteligencije (VI); takođe usmeren je protiv intuicije koja stoji iza teze VI i vodi poreklo još od Tjuringovog [Turing] testa¹ koju ukratko možemo opisati na sledeći način: kompetentan razgovor na nekom jeziku je dovoljan da pripišemo razumevanje tog jezika onome ko razgovara. Serl nam misaonim eksperimentom *Kineske sobe* pokazuje da postoji nešto pogrešno i sa tezom VI i sa takvom intuicijom. Utoliko što nam argument ukazuje na to da nešto nije u redu sa tezom i intuicijom koje se ne bismo lako odrekli, u godinama posle iznošenja argumenta, dobili smo veliki broj kritika, koje možemo sistematizovati u nekoliko većih grupa: sistemska kritika, robotska kritika, kritika preko virtuelnog uma, itd. Svi ovakvi tipovi kritike pokušavaju da ukažu na to da razumevanje jezika postoji unutar kineske sobe, bilo to u celom sistemu, u određenom virtuelnom entitetu ili bi postojala kada bi sistem bio robotski. Sprevak [Sprevak] je, 2006. god². ponudio radikalno drugačiju verziju

1 Turing, A., 1950, 'Computing Machinery and Intelligence', *Mind*, 59: str. 433–460.

2 Sprevak, M., 2007, 'Chinese Rooms and Program Portability', *British Journal for the Philoso-*

kritike, prema kojoj prihvatamo da nema razumevanja jezika unutar kineske sobe ni na kakav način, ali poriče da mehanizam koji imamo u kineskoj sobi jeste ekvivalentan svakom mogućem kompjuteru, što je jedna od pretpostavki koju je Serl imao u argumentu. U radu ću prikazati Serlov originalni argument, dati kraći pregled najuticajnijih kritika argumenta koje smo imali do sada, izneti i analizirati Sprevakovu kritiku, i ukazati da njegova kritika mora imati manji domet nego što je on mislio, ili da mora podrazumevati neki tip već postojećih kritika poput sistemske kritike.

Deo 1: Teza Veštačke inteligencija i Tjuringov test

Možemo reći da u najgrubljoj podeli postoje dve teze veštačke inteligencije³: Teza Jake veštačke inteligencije i teza Slabe veštačke inteligencije. Teza Slabe veštačke inteligencije (SVI) predstavlja umerenu tvrdnju o mogućnosti(ma) veštačke inteligencije:

(SVI) Moguće je da postoje mašine (veštački entiteti) koje ispoljavaju kvalitete koji nama izgledaju kao razumevanje, inteligencija, i sl., i takve mašine mogu biti od koristi inteligentnim bićima.⁴

Ovakvom tezom ne tvrdimo da možemo stvoriti inteligentne mašine, već naprosto da je moguće da stvorimo mašine koje ispoljavaju neke osobine inteligentnih bića, a da li su zaista inteligentne nije predmet tvrdnje. (Primetite da pitanje šta su inteligentni entiteti ovom tezom još uvek ne stavljamo u prvi plan.) Teza Jake veštačke inteligencije (JVI) predstavlja daleko jaču tvrdnju:

(JVI) Moguće je da postoje mašine (veštački entiteti) koje imaju razumevanje i inteligenciju poput drugih inteligentnih bića kao ljudi.

Kako bismo utvrdili da li zaista neka mašina poseduje inteligenciju i razumevanje? Intuicija kojom se vodimo pri određivanju toga jeste slična intuicija koja stoji iza Tjuringovog testa:⁵

(ITT) Ukoliko mašina može da vodi razgovor kao kompetentan govornik nekog jezika, onda mašina razume taj jezik, i *ipso facto*, ima inteligenciju.

Tjuringov test, kao i ITT su osporavani nezavisno od Serlovog argumenta⁶, kritikama koje u ovom tekstu neću analizirati, jer uglavnom ili nemaju uticaja na analizu(e) Serlovog argumenta, ili je uticaj koji imaju marginalan.

phy of Science, 58(4): str. 755–776.

³ Za zgodan početni pregled pogledati: Cole, David, “The Chinese Room Argument”, *The Stanford Encyclopedia of Philosophy* (Verzija 2015), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2015/entries/chinese-room/>>.

⁴ *Ibid.*

⁵ Ovde tvrdim da je intuicija koja stoji iza ovog određivanja slična intuicija koja stoji iza Tjuringovog testa, ništa jače od toga u smislu ispravnosti Tjuringovog testa ili neke njegove verzije. .

⁶ Za pregled kritika vezanih za Tjuringov test videti: Oppy, Graham and Dowe, David, “The Turing Test”, *The Stanford Encyclopedia of Philosophy* (Verzija 2016), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2016/entries/turing-test/>>.

Deo 2: Kineska soba

Serlov argument⁷ predstavlja misaoni eksperiment koji treba da pokaže da na način na koji pravimo računare možemo zadovoljiti uslov ITT, i dobiti mašine koje vode razgovor kao kompetentan govornik nekog jezika, ali da to nije dovoljno da zaključimo da u toj mašini postoji razumevanje jezika. Argument ima sledeću formu:

- P1)** Čovek unutar sobe ne može razumeti kineski jezik.
- P2)** Čovek unutar sobe može da izvrši bilo koji program. Dakle,
- Z)** Nijedan program ne može biti dovoljan za razumevanje.

Opravdanje prve premise je misaoni eksperiment po kom je Serlov argument poznat:

Zamislimo da se osoba koja ne govori niti razume kineski jezik, i razume samo engleski jezik⁸, nalazi unutar sobe koja ima dva otvora A i B. Unutar sobe se nalaze gomila papira na kojima su ispisani kineski znakovi, kao i uputstva na engleskom jeziku, koja su sledećeg tipa: Ako kroz otvor A uđe papir na kom su znakovi ABC, kroz otvor B izbaciti papir na kom se nalaze znakovi XYZ. Papira sa kineskim znakovima i uputstva ima dovoljno mnogo, toliko da za svaki papir koji ulazi kroz otvor A, osoba unutar sobe može da nađe uputstvo za taj papir, i papir koji treba da izbaci nazad kroz otvor B. Posle nekog vremena, papiri počinju da ulaze na otvor A, u osoba unutar sobe preko uputstava izbacuje papire kroz otvor B. To je sve što se dešava unutar sobe.

Izvan sobe, s druge strane, osobe koje ubacuju papire kroz otvor A, znaju kineski jezik. Papiri koje ubacuju predstavljaju pitanja na kineskom jeziku, i oni dobijaju odgovore na kineskom jeziku kroz otvor B. Osobama izvan sobe na osnovu tih odgovora izgleda da ko god da se nalazi unutar sobe kompetentan govornik kineskog jezika i da osoba unutar sobe razume kineski jezik.

Serlova intuicija⁹, s kojom se *prima facie* nije teško složiti, jeste da osoba unutar sobe ne razume kineski jezik. Sve što osoba iz sobe radi jeste prepoznavanje znakova na potpuno sintaksnom nivou, i prati uputstva koja su na istom tom nivou - razumevanja nema. Naravno, videćemo u delu koji se bavi kritikama da ovakva intuicija nije neosporna i da je bilo pokušaja da se argument ospori tvrdnjama da razumevanje (negde) unutar sobe postoji; ali zasad, složimo se sa Serlom da osoba unutar sobe definitivno ne može razumeti kineski jezik. Kako se ovo odnosi na tezu JVI? Za to Serlu služi premisa P2.

Premisa P2) treba da pokaže da način na koji funkcioniše osoba unutar kineske sobe jeste takav da on efektivno može da izvrši bilo koji program bilo kog kompjutera ili mašine. Nužno za Serlov argument je da čovek unutar sobe može da izvrši bilo koji

7 Searle, J., 1980, 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: str. 417–57

8 Bilo koji jezik različit od kineskog, engleski jezik nije ključan za argument. (u tom smislu, naravno, nije ni kineski.)

9 Searle, J, *op cit.*

program - ako bi on mogao da izvrši samo *neki* program, onda bi zaključak njegovog argumenta bio daleko slabiji¹⁰ - pokazao bi da konkretan program nije dovoljan za razumevanje i inteligenciju, ali to ništa ne bi značilo za programe uopšteno, i utoliko zaključak ne bi napao tezu jake veštačke inteligencije. Zašto Serl misli da čovek unutar sobe može da izvrši svaki program? Njegova ideja jeste da je čovek unutar kineske sobe univerzalna Tjuringova mašina:

“Počinjemo sa dva rezultata iz matematičke logike, Čurč-Tjuringovom tezom, i Tjuringovom teoremom. Za naše svrhe, Čurč-Tjuringova teza tvrdi da za bilo koji algoritam postoji neka Tjuringova mašina koja može implementirati takav algoritam. Tjuringova teza govori da postoji univerzalna Tjuringova mašina koja može simulirati bilo koju Tjuringovu mašinu. Sada, ako postavimo ove dve (TEZE) zajedno, imamo rezultat da univerzalna Tjuringova mašina može implementirati bilo koji algoritam.”¹¹

Serl misli da je čovek koji je unutar sobe jedna univerzalna Tjuringova mašina, da univerzalna Tjuringova mašina može da simulira bilo koju Tjuringovu mašinu, a da za bilo koji algoritam (program) postoji neka Tjuringova mašina koja može da ga implementira. Zbog toga, prema Serlu, čovek unutar sobe može da implementira (izvrši) bilo koji algoritam (program).

Deo 3: Uobičajene kritike *Kineske sobe*

a) Sistemska kritika

Sistemska kritika je najzastupljeniji tip kritike, koji je razmatrao kao mogući odgovor i sam Serl u iznošenju argumenta. Blok p

[Block]¹², Denet [Dennet]¹³, Fjodor [Fodor]¹⁴, Kopeland [Copeland]¹⁵ su među najznačajnijim zastupnicima ovakvog tipa kritike. Glavna ideja je jednostavna: osoba unutar sobe je samo deo, centralna jedinica procesuiranja (CJP), većeg sistema. Osoba nema razumevanje, ali to je zbog toga što razumevanje nije odlika nijednog dela sistema, već sistema kao celine. Sistem u pitanju jeste cela kineska soba sa ogromnom bazom podataka, in-

10 *Ibid.*

11 *Ibid.* Ukoliko bi čovek unutar sobe mogao da izvrši samo *neke* programe, onda bi Serlov zaključak važio samo za veštačke entitete koje izvršavaju te programe.

12 Block, N., 1978, 'Troubles with Functionalism', u C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press.

13 Dennett, D., 1978, 'Toward a Cognitive Theory of Consciousness', u *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press.

14 Fodor, J., 1991, 'Yin and Yang in the Chinese Room', u D. Rosenthal (ed.), *The Nature of Mind*, New York: Oxford University Press.

15 Copeland, J., 2002, 'The Chinese Room from a Logical Point of View', in Preston and Bishop (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, 2002, str 104–

122.

strukcijama, mogućim stanjima i sl. Takav sistem razume kineski jezik, iako CJP ne razume. Možemo posmatrati analogiju sa intencionalnim stanjima kod ljudi. Intencionalna stanja, prema zastupnicima ovakvog tipa odgovora, jesu nešto što pripisujemo celom sistemu osobe (celokupnoj fizičkoj osobi), ne samo onome što odgovara centralnoj jedinici procesuiranja kod osobe (mozak ili deo mozga). To što u kineskoj sobi CJP predstavlja svesna osoba nema uticaja na ishod argumenta; ta osoba je samo deo većeg i komplikovanijeg sistema.

b) Kritika virtuelnog uma

Ova kritika je unekoliko vezana za sistemsku kritiku. Čalmers [Chalmers]¹⁶ i Blok su, barem u nekom momentu, zastupali ovakav oblik kritike. Isto kao i sistemska kritika, u ovoj kritici priznajemo da osoba unutar sobe ne razume kineski jezik samo tako što prati uputstva koja nalazi u sobi; takođe kao u sistenskoj kritici, zastupnici ove kritike tvrde da to nije dovoljno da zaključimo da razumevanja nema. Naime, za njih je bitno da razumevanje nastaje u funkcionisanju programa, iako nijedan deo programa ne razume. Kako razumevanje nastaje? Stavaranjem virtuelnog uma koji ispoljava razumevanje. Ovaj virtuelni um je zaseban entitet u smislu da nije identičan sa CPJ, niti sa sistemom kao celinom. Zastupnici ovakvog tipa odgovora na kinesku sobu tvrde da je teza Veštačke inteligencije shvaćena prejako - ono što se tom tezom zaista tvrdi jeste da razumevanje nastaje pri funkcionisanju sistema, ne da sistem ili neki njegov deo razumeju.

c) Robotska kritika

Robotskom kritikom priznajemo deo poente Serlu¹⁷ - da razumevanja nema unutar sobe, niti u delu, niti u celini sistema, niti u virtuelnom umu kao zasebnom entitetu. Razlog za to što razumevanja nema, s druge strane, zastupnici ovakve kritike ne nalaze u tome što program ne može biti odgovoran za razumevanje, već što program *u okruženju kao što je kineska soba* ne može biti odgovoran za razumevanje. Kineska soba, osoba unutar nje, ili bilo šta oko kineske sobe, ne može razumeti šta znači reč "lopta". Ali, mi to možemo znati ne samo zato što znamo kako da odgovorimo na pitanja vezana za reč "lopta", već zato što smo učili šta je lopta, igrali smo se sa njom, uzimali je od druge dece, i uopšteno lopta je činila deo našeg života. Ako stavimo program unutar robotskog tela sa sensorima koji simuliraju čula koja mi imamo (video i audio senzori, itd.), pustimo takvog robota da odrasta i uči značenje reči kao što i ljudi to čine, interakcijom u zajednici, kontaktiranjem sa predmetima i drugim bićima, razumevanje u takvim i samo dovoljno sličnim uslovima može nastati. Ono što kineska soba pokazuje nije, dakle, da program ne može biti odgovoran za razumevanje, već da program ne može biti odgovoran za razumevanje bez dodatnih uslova kao što je robotsko¹⁸ telo sa čulima.

16 Chalmers, D., 1992, 'Subsymbolic Computation and the Chinese Room', u J. Dinsmore (ed.), *The Symbolic and Connectionist Paradigms: Closing the Gap*, Hillsdale, NJ: Lawrence Erlbaum.

17 Dennett, D., *op.cit.*

18 Robotsko telo treba smatrati u dovoljno širokom smislu, tako da je robotsko telo svako telo koje ima dovoljno slične receptore i čula kao biološka tela entiteta koji ispoljavaju razumevanje.

Deo 4: Sprevakova kritika

Uobičajene kritike napadaju premisu P1 Serlovog argumenta, ukazujući da razumevanje mora postojati negde unutar sobe. Sprevakova kritika¹⁹, s druge strane, odlazi od te tradicije pokušavajući da ospori premisu P2. Videli smo da je Serlova argumentacija za P2 sledeća:

- A1) Za bilo koji algoritam, postoji neka Tjuringova mašina koja ga može implementirati.
- A2) Univerzalna Tjuringova mašina može simulirati bilo koju Tjuringovu mašinu.
- A3) Osoba unutar kineske sobe je univerzalna Tjuringova mašina.
- Z) Osoba unutar kineske sobe može implementirati bilo koji algoritam.

A1 je tvrdnja koju Serl shvata kao Čurč-Tjuringovu tezu, A2 je ono što Serl zove Tjuringovom teoremom, A3 je tvrdnja da je način na koji osoba unutar sobe funkcioniše ekvivalentan funkcionisanju univerzalne Tjuringove mašine. Dodatne napomene oko terminologije: u iznošenju argumenta sam koristio termine “program” i “izvršiti”, dok sam u predstavljanju argumentacije za P2 koristio termine “algoritam” i “implementirati”. Te parove termina koristim kao sinonimne, jer ono što čini program jeste algoritam prema kom je program sastavljen, dok je implementiranje algoritma ono što bismo nazvali izvršenje programa.

Sprevak kritikuje A1²⁰. On se slaže sa Serlovom karakterizacijom Tjuringove teze i sa time da osoba unutar kineske sobe predstavlja način na koji funkcioniše univerzalna Tjuringova mašina, ali poriče da je A1 istinito. Preciznije, on tvrdi da je Serl pogrešio kada je tvrdnju A1 shvatio kao Čurč-Tjuringovu tezu. Čurč-Tjuringova teza tvrdi da za svaki *input-output* par neke funkcije, postoji neka Tjuringova mašina koja može proizvesti upravo takav *input-output*. Ali, to nije ono što Serl tvrdi, niti ono što je Serlu neophodno za zaključak. Jer univerzalna Tjuringova mašina može da izvrši istu funkciju, da ima iste parove *input-output* kao bilo koji kompjuter, ali to znači da implementira isti algoritam, odnosno, da izvršava isti program. Prema Sprevaku, dakle, kineska soba može reprodukovati *input-output* kompetentnog govornika kineskog jezika, ali to ne znači da implementira program kineskog jezika. Kao što neke mašine ne mogu izvršiti neke programe, ni kineska soba ne može izvršiti neke programe, i prema Sprevaku, govor kineskog jezika na nivou kompetentnog govornika jeste jedan od njih. Od čega zavisi da li neka mašina može implementirati neki program? Prema, Sprevaku, to zavisi od arhitekture sistema. Razmotrimo razlike između paralelnih i serijskih struktura programa: kineska soba je primer programa sa serijskom arhitekturom. Programi koji rade na paralelenoj strukturi doslovno ne mogu da budu izvršeni na mašini sa serijskom strukturom - jer zavise od obavljanja više od jedne operacije u isto vreme, što je nešto što je nemoguće izvršiti u okviru serijske arhitekture. Ako pretpostavimo (što je za Sprevaka razumna pretpostavka) da mozak ima paralelnu strukturu, to znači da je po samoj arhitekturi za čoveka unutar kineske sobe nemoguće da

¹⁹ Sprevak, M. *op.cit.*

²⁰ *Ibid.*

izvrši program koji izvršava mozak. Druga razlika koju Sprevak ističe jeste između atomskih operacija koje program sadrži²¹. Program koji izvršava osoba unutar kineske sobe se sastoji od tri atomske operacije: uporediti, kopirati, lančano povezati. Ako mozak ima više od te tri atomske operacije (što je za Sprevaka takođe razumna pretpostavka), program koji izvršava osoba unutar sobe i mozak nisu isti, niti osoba unutar sobe može izvršiti takav program. Sprevakov argument možemo sumirati na sledeći način:

- S1) Čurč-Tjuringova teza tvrdi da za svaku funkciju postoji Tjuringova mašina koja može proizvesti input-ouput te funkcije.
- S2) Reprokovati funkcije nekog programa ne znači izvršavanje tog programa.
- S3) Univerzalna Tjuringova mašina može reprodukovati svaku funkciju.
- S4) Osoba unutar sobe može reprodukovati svaku funkciju.

Z) Nije dokazano da osoba unutar sobe može izvršiti svaki program.

Sprevak se ne zaustavlja isključivo na ovakvom zaključku. Njegov argument zavisi od opravdanja koje pruža za tvrdnju da osoba unutar sobe ne može da izvrši isti program kao i kompetentan govornik kineskog jezika. Opravdanje koje on pruža jeste navođenje specifične arhitekture programa koji čini razumevanje kineskog jezika, odnosno, specifične arhitekture mozga. Utoliko, on tvrdi da je za razumevanje kineskog jezika potreban program koji ima istu ili dovoljno sličnu arhitekturu kao ljudski mozak, u smislu bioloških i fizičkih karakteristika. Sam Serl je u originalnom argumentu istakao značaj biološke arhitekture mozga kao činjenicu zbog koje teza JVI ne može biti istinita. U tom smislu, Sprevak i Serl²² su došli do dovoljno sličnog zaključka²³. Sprevak ne misli da je to od velikog značaja. Prema njemu, dok je Serlova motivacija za posezanjem za biološkom strukturom neopravdana i predstavlja *deus ex machine*; motivacija koju sada imamo da bismo upotrebili biološku strukturu jeste konkretno opravdanje koje se sastoji od tvrdnji o identitetu programa. Praktično, biologija proizvodi hardver, a arhitektura hardvera određuje koji softver se može izvršiti na njemu. Druga razlika jeste što je Serl, posežući za biologijom, nudio obrazloženje zbog čega program *ne može* biti konstitutivan za razumevanje, dok Sprevak čini upravo suprotno: on tvrdi da program može biti konstitutivan za razumevanje, ali program koji je *dovoljno sličan* programu koji izvršava ljudski mozak. Dakle, Sprevak ne tvrdi potpuno isto što i Serl, iako, poput Serla, ističe važnost biološke i fizičke strukture mozga. Ono što on tvrdi, jeste da program za razumevanje mora biti isti ili dovoljno sličan kao i program koji izvršava mozak. Da li je onda tvrdnja da je program konstitutivan za razumevanje jezika “nepotreban teret”? Serl je tvrdio ne samo da je sistem, odnosno funkcionalnost između fizičkih i bioloških delova odgovoran za razumevanje, već biološka i fizička struktura. Odnosno, za razumevanje nije odgovoran samo algoritam ili program koji naša fizika i biologija izvršavaju, već sama biološka i fizička struktura. Prema Serlu, ne bi trebalo samo replikovati program koji naš

21 *Ibid.*

22 Searle, J, *op.cit*

23 Sprevak, M. *op.cit.*

mozak ima, već bi trebalo replikovati celu strukturu tela za razumevanje. Sprevak ne nalazi ovakav način razmišljanja plauzibilnim. Odnosno, prema Sprevaku, ne postoji ništa specifično u fizičkim i biološkim supstancama i sastavima u našem telu što bi bilo nezamenljivo - ako posmatramo npr. serotonin, nema ničega što bi nam ukazalo na to da nije dovoljno samo ga zameniti nečime što bi igralo ulogu serotonina u programu ili sistemu koji pravimo. Sprevakova poenta na ovom mestu jeste komputaciona i funkcionalistička - treba da zadovoljimo sve funkcije koje delovi mozga imaju, time replicirajući program, ali se pri tome ne moramo služiti identičnim biološkim elementima. Ukoliko insistiramo na identičnim biološkim elementima, ne rešavamo se “nepotrebnog tereta” programa, već bez razloga odbijamo funkcionalizam kao objašnjenje sistema. Ako insistiramo na fizičkom objašnjenju bez uzročnih uloga u funkcionalističkom smislu, onda ne dobijamo kvalitetno objašnjenje funkcionisanja sistema; ako s druge strane insistiramo na fizičkom objašnjenju preko navođenja uzročnih uloga i funkcija, onda nemamo problem sa komputacionom teorijom i ne moramo da insistiramo na potpunoj replici biološkog sistema od bioloških elemenata.²⁴

Za kraj, Sprevak razmatra još jedan moguć Serlov odgovor. Naime, Serl je, pri odgovaranju na kriticizam konekcionista izbacio navodno poboljšan argument kineske teretane. Takav argument možemo opisati ako zamislimo umesto jedne osobe iz kineske sobe punu prostoriju osoba, koje uporedo rade na komplikovanijim uputstvima sa više atomskih operacija i sl., tako dajući odgovore osobama van sobe. Takvu “teretanu” možemo zamisliti dovoljno komplikovanu i punu ljudi tako da može izvršiti svaki program/algoritam, ne samo funkciju poput kineske sobe. Sa ovime se i Sprevak slaže, ali navodi da je problem što kineska teretana pada pod prigovorima Kopelanda²⁵ i ostalih konekcionista/zastupnika sistemskog odgovora. Konekcionista ne tvrde da jedan neuron razume, već da je razumevanje odlika celokupnog sistema. Ova verzija argumenta tako potpada pod naročito efektnu verziju sistemskog odgovora kineskoj sobi. Sprevak navodi i razlog zbog kog je ovu verziju argumenta tako lako odbaciti: lako je zamisliti da smo osoba u kineskoj sobi, i intuiciju da nema razumevanja kineskog jezika unutar sobe je samim time lakše prihvatiti; s druge strane, jako je teško zamisliti kako je biti kineska teretana (sa više ljudi i više operacija), zbog čega nemamo početnu intuiciju da razumevanja kineskog jezika nema unutar kineske teretane.

Deo 5: Nesklad dva dela Sprevakove kritike

Sprevakovu kritiku možemo podeliti na dva aspekta: negativnu i pozitivnu stranu kritike. Negativna strana kritike obuhvata deo kojim se tvrdi da kineska soba ne može izvršiti svaki program, što je Serlu neophodno za zaključak protiv Jake VI. S druge strane, pozitivna strana kritike su Sprevakovi zaključci o tome *kakav* program je odgovoran za razumevanje, i zahtev za repliciranjem programa mozga koji Sprevak na kraju svog teksta iznosi. Dok Sprevak misli da drugi deo kritike na izvestan način proizilazi iz prvog²⁶,

²⁴ *Ibid.*

²⁵ Copeland, *op.cit.*

²⁶ Zbog toga on napominje da repliciranje mozga u njegovom slučaju nije *deus ex machina* za raz-

mislim da je to pozicija koju ne može zastupati; ili, u po Sprevakova najboljem slučaju, to je pozicija koju ne može zastupati ukoliko se ne obaveže *i* na sistemsku verziju odgovora na argument kineske sobe, što umnogome umanjuje značaj njegove kritike.

Sprevak u pozitivnom delu kritike tvrdi da program odgovoran za razumevanje mora replikuje program koji ljudski mozak ima, u smislu da ima istu ili dovoljno sličnu arhitekturu koja je odgovorna za taj program i iste ili dovoljno slične atomske operacije. S druge strane, u odgovoru na moguće poboljšanje Serlovog argumenta u vidu kineske teretane, on se poziva na konekcionističku/sistemsku kritiku²⁷, napominjući da ovakva verzija argumenta potpada pod naročito jaku formu sistemske kritike²⁸, jer za razliku od standardne kineske sobe, nemamo toliko jaku intuiciju da nema razumevanja unutar prostorije tj. u okviru celog sistema. Ipak, pozivanje na sistemski odgovor u ovoj situaciji, ima neugodnu posledicu: sistemski odgovorom tvrdimo da sistem razume, a ne deo sistema. Takvim odgovorom ne tvrdimo ništa o potrebi za replikacijom konkretnog programa za razumevanje. Zamislimo “dvokrevetnu” kinesku sobu: umesto jedne osobe, imamo dve osobe koje izvode isti broj operacija kao i jedna, samo u ovom slučaju paralelno umesto serijski. Neplauzibilno bi bilo reći da te dve osobe repliciraju program prema kom radi naš mozak, jer su atomske operacije iste kao i u originalnoj kineskoj sobi, a broj paralelnih operacija je daleko manji nego pri programu koji izvodi mozak, tako da, prema Sprevakom, takva soba ne bi mogla da implementira programe napravljene za komplikovanije arhitekture. Sistemski odgovor bi bio jako jasan: iako nijedna osoba unutar sobe ne može da razume kineski jezik, to nije dovoljno da zaključimo da sistem kao celina ne može razumeti. (To bi bio sistemski odgovor i za originalnu sobu, na šta ćemo obratiti pažnju za koji momenat.) Sprevak bi, s jedne strane, ako sledimo pozitivan deo njegove kritike, morao da odgovori da takva soba ne može implementirati program za razumevanje, pa samim time, ne može razumeti. S druge strane, ako sledimo njegov odgovor na kinesku teretanu, nije jasno kako bi mogao tvrditi da soba ne može razumeti. Ako je za kinesku teretanu odgovor bio sistemski, onda bi i ovde trebalo da bude sistemski tip odgovora, prema kom soba može razumeti. Ako je za kinesku teretanu poenta odgovora bila u različitoj intuiciji oko toga da li se razumevanje negde nalazi unutar prostorije (u kineskoj sobi imamo jasnu predstavu kako bi izgledalo da smo u sobi, dok u teretani nemamo), onda nije jasno kako se intuicija čuva unutar dvokrevetne kineske sobe (ili osmokrevetne npr.). Nema nikakve posebne intuicije koja se gubi kada u sobu ubacimo tačan broj ljudi koji bi mogao da replikuje konkretan program mozga - ako ta intuicija postoji sa nekoliko osoba u sobi, verovatno da postoji i sa nekoliko više. Zbog toga ni pozivanje na intuiciju ne može biti Sprevakov odgovor na problem dvokrevetne kineske sobe.

Problem za Sprevakom jeste što ima dve opcije povodom kineske teretane: 1) pozvati se na sistemski odgovor; 2) pozvati se samo na intuiciju, drugačiju u slučaju teretane nego u slučaju kineske sobe, bez pozivanja na sistemski odgovor.

liku od Serlovog insistiranja.

27 Sprevak, *op.cit.*

28 *Ibid.*

Ako on zastupa prvu opciju povodom kineske teretane, nije vidljiv razlog kako bi rekao da dvokrevetna kineska soba nije dovoljna za razumevanje kineskog jezika; jer, tipičan sistemski odgovor za dvokrevetnu sobu bi bio kao i za teretanu: sistem kao celina može imati razumevanje kineskog jezika, iako nijedan deo sistema *sam po sebi* ne razume kineski jezik. Štaviše, još lošije po Sprevak, nije jasno zašto bi onda sistemski odgovor bio neodgovarajući za originalnu kinesku sobu. Ako se pozivamo na sistemski odgovor protiv teretane, sistemski odgovor možemo upotrebiti i protiv kineske sobe, dodatni odgovor bi bio izlišan!

Protiv kineske teretane može odgovoriti i na drugi način, pozivajući se samo na intuiciju. U tom slučaju, njegov odgovor ne bi bio izlišan. Ali, ako bi se držao pozitivnog dela svoje kritike, onda bi ga i ovakav odgovor doveo do nezgodnih posledica. Kako smo videli, on u pozitivnom delu kritike tvrdi da je za razumevanje odgovoran program koji funkcioniše u našem mozgu, i da bi veštačka inteligencija morala raditi po istom (ili barem dovoljno sličnom) programu.²⁹ To znači da bi kineska teretana, da bi razumela kineski, morala da ima funkcionalnu arhitekturu istu ili dovoljno sličnu programu našeg mozga. U pitanju bi bila jako specifična kineska teretana, sa slično specifičnom arhitekturom: brojem paralelnih operacija, istim ili dovoljno sličnim atomskim operacijama, istim funkcionalnim odnosima osnovnih delova, itd. Za takvu, i prema pozitivnom delu Sprevakove kritike *samo takvu*, kinesku teretanu bi trebalo reći da razume kineski jezik. Zašto u takvoj teretani postoji razumevanje, a u originalnoj sobi nema? Prema mogućnosti koju razmatramo, jako je intuitivno videti da razumevanja u originalnoj kineskoj sobi nema (staviti se u situaciju osobe unutar sobe, itd.), dok to nije slučaj za *ovakvu* kinesku teretanu (nije jasno kako *bi bilo biti* kineska teretana, itd.). Ali, nije jasno zašto bi tvrdnja o neintuitivnosti važila *samo* za tip teretane koji replicira isti ili dovoljno sličan program koji postoji u našem mozgu. Jer, za teretanu sa duplo više ili duplo manje atomskih operacija, npr. važe iste intuicije: jako je teško staviti se u položaj takve teretane i jako je teško deliti intuiciju da nema razumevanja u teretani kao celini. Čak i za dvokrevetnu kinesku sobu važi nešto slično: dve osobe koje vode paralelene operacije su dovoljno daleko od originalne kineske sobe da intuicija koju imamo o nepostojanju razumevanja jeste značajno manja. Sprevakov glavni problem je što postoji jako širok spektar arhitektura sistema koje mogu implementirati jako širok spektar mogućih programa, iako on tvrdi da insistiranje na istoj ili dovoljno sličnoj arhitekturi i istom ili dovoljno sličnom programu kao onom iz našeg mozga nije *deus ex machina* kao insistiranje na biologiji kod Serla.³⁰ Sprevak se mora odlučiti da odustane ili od pozitivnog dela svoje kritike ili od odgovora na kinesku teretanu i samim time i nezavisnosti svoje kritike. Prva opcija nije sjajna; ali druga je neprihvatljiva, jer bi tom opcijom izgubio “nezavisnost” svoje kritike: morao bi da se poziva na sistemsku kritiku i svoju učini utoliko izlišnom.

Zaključak

Odustajanje od pozitivnog dela kritike, iako ne sjajan, nije poguban ishod po Sprevaka. Negativnim delom kritike, on je ukazao na bitnu razliku između identiteta programa

29 *Ibid.*

30 *Ibid.*

i identiteta funkcije, koja izgleda kao da je prošla nezapaženo u Serlovoj formulaciji argumenta. Jasno je zašto različiti sistemi nekada ne mogu implementirati *iste* programe; s druge strane, skok na neophodnost identiteta programa je nešto što odatle ne sledi. Intuitivno, nije jasno zašto bi sledilo. Kao što možemo zamisliti bića sa bitno drugačijom strukturom mozga, odnosno, arhitekturom za program koji izvršavaju, tako možemo zamisliti da različiti programi mogu biti odgovorni za razumevanje. Bitno pitanje koje ostaje za razmatranje u tom slučaju jeste šta je odgovorno za razumevanje. Identična funkcija (par input-output) je premali zahtev, jer barem to ostaje poenta originalne sobe. Identičan program izgleda kao neopravdano jak zahtev, barem prema opravdanjima koje Sprevak nudi. Balansiranje između ta dva puta je preostala opcija, ali tačan kriterijum zahteva dodatno istraživanje koje zasad nije jasno u kom smeru bi teklo.

Nenad Filipović

Literatura:

- Chalmers, D., 1992, 'Subsymbolic Computation and the Chinese Room', u J. Dinsmore (ed.), *The Symbolic and Connectionist Paradigms: Closing the Gap*, Hillsdale, NJ: Lawrence Erlbaum.
- Churchland, P., 1985, 'Reductionism, Qualia, and the Direct Introspection of Brain States', *The Journal of Philosophy*, LXXXII: str. 8–28.
- Churchland, P. and Churchland, P., 1990, 'Could a machine think?', *Scientific American*, 262(1): str. 32–37.
- Cole, David, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.),
URL = <<http://plato.stanford.edu/archives/win2015/entries/chinese-room/>>.
- Copeland, J., 2002, 'The Chinese Room from a Logical Point of View', in Preston and Bishop (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, 2002, str 104–122.
- Dennett, D., 1978, 'Toward a Cognitive Theory of Consciousness', u *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press.
- Fodor, J., 1987, *Psychosemantics*, Cambridge, MA: MIT Press.
- Fodor, J., 1991, 'Yin and Yang in the Chinese Room', u D. Rosenthal (ed.), *The Nature of Mind*, New York: Oxford University Press.
- Oppy, Graham and Dowe, David, "The Turing Test", *The Stanford Encyclopedia of Philosophy* (Verzija 2016), Edward N. Zalta (ed.),
URL = <<http://plato.stanford.edu/archives/spr2016/entries/turing-test/>>.
- Penrose, R., 2002, 'Consciousness, Computation, and the Chinese Room' in Preston and Bishop (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence* 2002, str. 226–249.

- Searle, J., 1980, 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: str. 417–57
- Searle, J., 1984, *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.
- Searle, J., 1990a, 'Is the Brain's Mind a Computer Program?', *Scientific American*, 262(1): str. 26–31.
- Searle, J., 1990b, 'Presidential Address', *Proceedings and Addresses of the American Philosophical Association*, 64: str. 21–37.
- Sloman, A. and Croucher, M., 1980, 'How to turn an information processor into an understanding', *Brain and Behavioral Sciences*, 3: str. 447–8.
- Sprevak, M., 2007, 'Chinese Rooms and Program Portability', *British Journal for the Philosophy of Science*, 58(4): str. 755–776.
- Turing, A., 1948, 'Intelligent Machinery: A Report', London: National Physical Laboratory.
- Turing, A., 1950, 'Computing Machinery and Intelligence', *Mind*, 59: str. 433–460.
- Weiss, T., 1990, 'Closing the Chinese Room', *Ratio*, 3: str. 165–81.

Chinese Room and Program Identity (Summary)

The Chinese room argument is famous argument introduced by John Searle, in which Searle presented various problems with the claim that it is possible for the artificial intelligence to have understanding of a language in a way in which intelligent beings such as humans have that capacity. The argument was influential enough to, in decades following it, spark numerous responses and critiques, along with a few alleged improvements to it from Searle. In this article, I will analyze one atypical critique of Searle's argument, made by Mark Sprevak. Sprevak, unlike the other critics of the argument, agrees with Searle that understanding does not exist in Chinese room in any way, but he claims that Chinese room cannot execute every possible program, like Searle claims. Because of that, Searle cannot conclude the strong conclusion he wants from The Chinese room argument. In this article, I will analyze Searle's argument, I will give a brief overview of typical responses to it, and I will analyze Sprevak's response. In the last section, I will present argument that shows that Sprevak, if he wants to keep his conclusions, must either give up one part of his response, or accept one of the typical responses to Searle's argument, thus making his own response dependent on the response from others.

Keywords: Chinese room, program, algorithm, function, Searle, Sprevak.