

Tadej Todorović

Filozofska fakulteta Univerze v Mariboru

Večvrstna realizacija in Turingov test: preslikava

Članek v prvem delu predstavi večvrstno realizacijo in Turingov test. Drugi del je namenjen dokazu, da je način sklepanja pri Turingovem testu in večvrstni realizaciji identičen. V tretjem delu predstavi enega izmed ugovorov Turingovemu testu in njegovo rešitev s pomočjo t.i. *oblikovnega* kriterija. V četrtem delu nato to rešitev preslikamo na večvrstno realizacijo, kar nas pripelje do sklepa, ki je popolnoma nasproten sklepu večvrstne realizacije. To nas vodi do ugotovitve, da je edini razlog za sprejem večvrstne realizacije njena metafizična zamisljivost, kar pa, kot pokažemo, ni dovolj. V petem delu predstavimo možno strategijo zagovarjanja večvrstne realizacije in težave, ki se lahko pri tem pojavijo.

Ključne besede: večvrstna realizacija, mentalna stanja, Turingov test, ugovor iz črne škatle, takšnosti

0 Uvod

V članku bomo v prvem delu najprej predstavili večvrstno realizacijo, kot si jo je zamislil Hilary Putnam v letih 1960–1970, in pa Turingov test, kot ga je predstavil Alan Turing. Poskušali bomo ugotoviti, na podlagi česa sklepamo v primeru večvrstne realizacije in na podlagi česa v primeru Turingovega testa. Ugotovili bomo, da tako Turingov test kot tudi večvrstna realizacija gradita svoji osnovni premisi na isti predpostavki, ki jo bomo imenovali behavioristična predpostavka. V tretjem delu članka bomo izpostavili enega izmed ugovorov Turingovemu testu, ki to predpostavko postavi pred velik vprašaj – ugovor iz črne škatle. S pomočjo *oblikovnega* kriterija bomo nato nadgradili Turingov test. To nadgradnjo bomo zatem preslikali na tezo večvrstne realizacije. Ugotovili bomo, da z dodano nadgradnjo pridemo do sklepa, ki je nasproten tezi večvrstne realizacije, in da je edini razlog za sprejetje večvrstne realizacije njena metafizična zamisljivost. V zadnjem delu članka bomo poskusili ponuditi strategijo zagovarjanja večvrstne realizacije in opozorili na težave, ki se lahko pri tem pojavijo. Poskusili bomo tudi prikazati jasen primer večvrstne realizacije v relaciji s človeškimi možgani, ki

ne bo zgolj analogija, njegov princip bo predstavljal morda premočen, vendar zadosten kriterij za oblikovanje večvrstnih realizacij v problemu duha in telesa.

I Večvrstna realizacija in Turingov test

Večvrstna realizacija je teza, ki jo je Hilary Putnam uvedel v filozofijo duha v seriji člankov v letih 1960–1970.¹ Trdi, da so mentalna stanja (npr. bolečina) lahko realizirana preko različnih fizičnih sistemov. (gl. Bickle, 2016: 1) Če uporabimo primer bolečine: tako mi (ljudje) kot druga živa bitja (hobotnice, psi, mačke) doživljamo subjektivno izkustvo bolečine. Vendar se način nastajanja bolečine v ljudeh, mačkah, psih in hobotnicah (zaradi različne zgradbe možganov) razlikuje, kar vodi v to, da lahko različni fizični sistemi (fizični sistem možganov ljudi, psov, mačk in hobotnic) realizirajo isto mentalno stanje – tj. bolečino. Iz tega sledi, da mentalnih stanj ne moremo enačiti s fizičnimi stanji, ki mentalna stanja realizirajo. Večvrstna realizacija je tako eden izmed glavnih argumentov proti redukcionizmu kot teoriji duha in telesa,² ki trdi, da je mentalno stanje identično fizičnemu stanju, ki ga realizira.

V deduktivni obliki lahko argument iz večvrstne realizacije zapišemo v naslednji obliki:

- 1) Vse mentalne pojave lahko večvrstno realiziramo z različnimi fizičnimi sistemi. (teza večvrstne realizacije)
- 2) Če lahko mentalne pojave večvrstno realiziramo z različnimi fizičnimi sistemi, potem le-ti ne morejo biti identični s specifičnim fizičnim sistemom.
- 3) ($\therefore$) Noben mentalni pojav ni identičen specifičnemu fizičnemu sistemu. (Bickle, 2016: 1)

Pod drobnogled bomo vzeli prvo premiso argumenta – tezo večvrstne realizacije, »resničnost katere so kritiki³ v zadnjih letih poskušali ovreči« (Bickle, 2016, 2.4). Preko analogije med večvrstno realizacijo in Turingovim testom bomo poskušali tezo večvrstne realizacije ovreči ali vsaj postaviti pod vprašaj. Da lahko to naredimo, si najprej oglejmo, kaj je Turingov test.

Turingov test je preizkus, ki si ga je leta 1950⁴ zamislil Alan Turing. Namenjen je temu, da presodi, ali stroj misli, to pa naredi s pomočjo igre oponašanja. V igri nastopajo trije akterji – dva človeka in stroj. Stroj in prvi človek sta zaprta vsak v svojo sobo. Obe sobi sta komunikacijsko povezani z nadzorno sobo (danes lahko

¹ Recimo v člankih »Psychological Predicates« (Putnam, 1967) in »The Nature of Mental States« (Putnam, 1975).

² Za več o redukcionizmu glej Bregant, 2004: 61.

³ Zangwill, 2002; Shapiro, 2000; Couch, 2004.

⁴ Gl. Turing, 1950.

to zelo elegantno izvedemo s pomočjo lokalne internetne povezave), v kateri je drugi človek. Le-ta nato preko te povezave z obema komunicira – ne vidi v notranjost sobe in ne sliši stroja in prvega človeka. Tako torej drugi človek ne ve, v kateri sobi je stroj in v kateri je prvi človek. Drugi človek (imenujmo ga spraševalec) nato zastavlja kakršnakoli vprašanja obeh akterjema v sobah (enostavna, površinska, zapletena, globoka) z namenom, da bi ugotovil, v kateri sobi se nahaja stroj in v kateri človek. Naloga stroja je, da pripelje drugega človeka (spraševalca) do zmotne identifikacije (torej, da spraševalec identificira stroj kot človeka, in obratno). To bo storil najlažje, če bo spraševalcu podajal odgovore, ki so nerazločljivi od odgovorov človeka – stroj bo torej v odgovorih oponašal človeka. Če se spraševalec v 50% primerov zmoti ali če se ne more odločiti, v kateri sobi je stroj in v kateri človek, potem moramo priznati, da je stroj zavesten – da torej misli (gl. Bregant, 2010: 64–65).

V prvem delu smo predstavili večvrstno realizacijo in Turingov test, v drugem delu pa si bomo ogledali, na kakšen način sklepamo, ko govorimo o Turingovem testu in večvrstni realizaciji.

II Primerjava načina sklepanja pri Turingovem testu in pri večvrstni realizaciji

Ne bo nas zanimal torej Turingov test sam,⁵ temveč samo predpostavka, na kateri Turingov test sloni. Imenovali jo bomo behavioristična predpostavka, ki bo trdila, da: če se stroj obnaša enako kot človek oziroma če med obnašanjem človeka in stroja ni razlike, potem moramo zavoljo konsistence priznati, da stroj misli. Ali kot to povzame Ayer: »Če vem, da se objekt v vseh pogledih obnaša tako, kot se po definiciji obnaša zavestno bitje, potem vem, da ima v resnici zavest« (Ayer, 1978: 178).

Zapisano v prečiščeni obliki: *če se stroj obnaša enako (nerazločljivo) kot zavestno bitje, potem ima stroj zavest*. Zdi pa se, da je ta predpostavka zelo podobna predpostavki teze večvrstne realizacije, ki pravi, da lahko vse mentalne pojave večvrstno realiziramo z različnimi fizičnimi sistemi. Zakaj? Da namreč lahko smiselno trdimo, da tako kot človek, tudi pes in mačka doživljata bolečino, moramo sprejeti behavioristično predpostavko, in sicer *da če se nek organizem obnaša na enak (nerazločljiv) način, kot se obnašamo mi v stanju bolečine, potem ta organizem prav tako doživlja bolečino*. Imenujmo prvo predpostavko »predpostavka Turingovega testa« in drugo »predpostavka teze večvrstne realizacije« in obe zapišimo:

Predpostavka Turingovega testa: če se stroj obnaša enako (nerazločljivo) kot človek v stanju zavesti, potem ima stroj zavest.

⁵ Za več o samem testu in ugovorih nanj glej Bregant, 2010.

Predpostavka teze večvrstne realizacije: če se nek organizem obnaša enako (nerazločljivo) kot človek v stanju bolečine, potem je organizem v stanju bolečine.

Vidimo lahko, da obe predpostavki sledita istemu pravilu. Če se namreč neka stvar obnaša enako (nerazločljivo) kot človek v stanju x, potem je ta stvar v stanju x. Logično sta torej enaki, še več, rečemo lahko, da je to v bistvu ena in ista predpostavka – imenujmo jo *behavioristična predpostavka*, le stanje x je različno (v prvem primeru je stanje x bolečina, v drugem pa zavedanje/mišljenje).

Predpostavka Turingovega testa in predpostavka večvrstne realizacije bi tako, če želimo biti konsistentni, morali biti podvrženi enako strogi kritiki v filozofiji, saj smo ugotovili, da sta v bistvu ena in ista predpostavka. Pa vendar ni tako. Resničnost Turingovega testa je bila v 20. stoletju večkrat postavljena pod vprašaj, že v času nastajanja samega testa⁶ kot tudi kasneje⁷ – kot enega najbolj znanih lahko omenimo ugovor ameriškega filozofa Johna Searla, *argument kitajske sobe*.⁸ Nasprotno pa v resničnost večvrstne realizacije do leta 2000 ni dvomil skoraj nihče. Kot ugotavlja Shapiro: »Čeprav se je nekaj filozofov vprašalo, ali je večvrstna realizacija res v nasprotju z redukcionizmom, /.../ še, kolikor vem, nihče ni resno podvomil v resničnost same teze večvrstne realizacije« (Shapiro, 2000: 635–636).

Zato bomo sedaj poskusili najti ugovor Turingovemu testu, ki bi bil po analogiji lahko tudi ugovor večvrstni realizaciji.

III Ugovor iz črne škatle in nadgradnja Turingovega testa

Eden izmed ugovorov Turingovemu testu se imenuje ugovor iz črne škatle. Turingov test temelji samo na obnašanju stroja, zanemari pa, kaj se dogaja v samem stroju – torej njegovo zgradbo. Ko sočloveku pripisujemo mentalna stanja (zavest), vemo, kako delujejo možgani, kateri elektrokemični procesi so odgovorni za katero mentalno stanje ipd. Nasprotno nam je delovanje stroja neznano, zato stroju zaradi obnašanja samega ne moremo kar pripisati mišljenja/zavesti (Bregant, 2010: 67).

Oglejmo si podrobneje, kaj želi ta ugovor sporočiti. Prva interpretacija je naslednja: zaradi tega, ker ne poznamo zgradbe stroja, mu samo zaradi enakega obnašanja ne moremo pripisati zavesti. To se zdi absurdno, saj bi mi tudi danes, čeprav ne poznamo vseh elektrokemičnih procesov v možganih, pripisali zavest sočloveku, saj ne želimo pasti v solipsizem. Z drugimi besedami, velika večina ljudi ne ve, kako možgani delujejo, pa vendar soljudem pripisujejo zavest – torej poznavanje delovanja stroja ne more biti pogoj za pripisovanje zavesti. Bolj se zdi verjetna druga interpretacija: zaradi tega, ker ne vemo, ali je zgradba stroja enaka kot

⁶ Gl. Bregant, 2012.

⁷ Za vse sodobne ugovore Turingovemu testu gl. Oppy in Dowe, 2016.

⁸ Za več o argumentu kitajske sobe gl. Searle, 1981.

zgradba naših možganov, stroju ne pripišemo zavesti samo zaradi obnašanja. Torej ljudje v nasprotju s tem, kar je predpostavil Turing, ne pripišemo zavesti sočloveku samo zaradi obnašanja, temveč tudi zaradi podobne zgradbe možganov. Šele ta dva pogoja v konjunkciji tvorita zadosten pogoj, da pripišemo sočloveku zavest. Ali, zapisano v logični obliki: če se x obnaša kot jaz (zavestno bitje) in ima podobno strukturo možganov kot jaz, potem sledi, da je tudi x zavestno bitje.

Zdi se, da je treba ugovor iz črne škatle interpretirati v drugem smislu. To podkrepi tudi protiguovor temu ugovoru (torej poskus reševanja Turingovega testa), ki ga je postavil Copeland: »Oplemeniten Turingov test bi torej moral vsebovati dva elementa: prvič, izvirno posnemovalno igro, s pomočjo katere bomo ugotovili, ali je izpolnjen *behavioristični kriterij*⁹ (tj. da so odgovori¹⁰ računalnika nerazločljivi od odgovorov človeka), in drugič, morali bi pregledati sam računalnik,¹¹ kar bo razjasnilo, ali je izpolnjen tako imenovan *oblikovni kriterij*« (Copeland, 1993: 50–51). Seveda je vprašanje, kaj ta oblikovni kriterij je. Copeland predlaga dve možnosti – oglejmo si obe in njune probleme. Prva verzija *oblikovnega* kriterija pravi, da mora stroj »delati stvari, ki jih dela, na način, ki je približno enak načinu delovanja človeških možganov« (Copeland, 1993: 51). Z drugimi besedami – fizično (elektrokemično) delovanje stroja mora biti zelo podobno fizičnemu delovanju možganov. Ugovor, na katerega takšna verzija *oblikovnega* kriterija nalleti, je, kot pravi Copeland sam, da »... je premočan. Izgleda, da je neupravičeno antropocentrično zahtevati, da računalnik deluje na podoben način kot človek« (Copeland, 1993: 51). Pa je to res neupravičeno? »Izviren Turingov test je brez sramu antropocentričen, saj zahteva, da računalnik zmore proizvesti obnašanje, nerazločljivo od človeškega« (Copeland, 1993: 51).

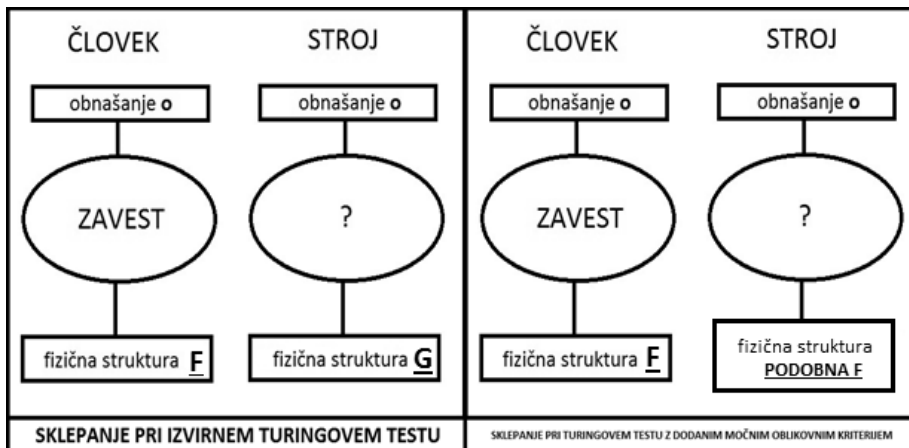
Druga (šibka) verzija *oblikovnega* kriterija pa pravi, da mora biti računalnik modularen. To pomeni, da mora obstajati možnost, da program »zapakiramo« in vključimo v drug program in tako ustvarimo program večje kompleksnosti. Na primer, mora biti možno, da program prenesemo na program, ki nadzira motorične in čutne sisteme robota, rezultat tega pa mora biti robot, ki ima možnost govoriti o tem, kaj dela (Copeland, 1993: 51). Žal ni najbolj jasno, katero verzijo *oblikovnega* kriterija sprejeti. »Mislim, da iz diskusije sledi, da ni nobena od teh verzij očitno dobra ali slaba, niti ni nobena očitno boljša kot druga« (Copeland, 1993: 52). Ne glede na to bomo poskušali podati razlog za sprejetje močnejše verzije *oblikovnega* kriterija.

⁹ Avtor tukaj uporabi besedno zvezo *output criterion*, vendar bomo zaradi jasnosti in že uvedene terminologije uporabljali besedno zvezo *behavioristični kriterij*. Enak prevod uporabi tudi Bregant (2010: 67).

¹⁰ Ponovno avtor uporablja besedno zvezo *conversational output*, kar prevajamo kot odgovori računalnika/človeka – mislim, da s tem obdržimo semantičen pomen besedne zveze.

¹¹ Avtor uporabi besedo »program«, vendar mislim, da avtor misli na pregledovanje delovanja računalnika in programa.

Ko sklepamo samo na podlagi behavioristične predpostavke (če se *x* obnaša kot jaz, ki sem zavestno bitje, potem sledi, da je tudi *x* zavesten), sklepamo na podlagi analogije. Sklepamo namreč tako: jaz (*y*) imam lastnost *z* (zavest) in lastnost *o* (specifično obnašanje). Stroj *x* (ki »zmaga« na Turingovem testu) ima lastnost *o* (specifično obnašanje, ki je nerazločljivo od mojega). Zato ima stroj *x* tudi lastnost *z* (zavest). Ampak sklepanje po analogiji je induktivno, torej ne-nujno, analogija pa je dobra takrat, ko »so podobnosti med primerjanima pojavoma ali dogodkoma precej številčnejše od razlik« (Klampfer, Bregant, Gartner, 2008: 78). V dotičnem primeru imamo samo eno razliko in eno podobnost, kar je za dobro sklepanje po analogiji premalo. Če ponazorimo na primeru: 85% temna čokolada in 99% temna čokolada sta enake barve (obe temno rjavi), pa to še ne pomeni, da če je nekomu dobra 85% temna čokolada, mu bo dobra tudi 99% temna čokolada. Pri sklepanju po analogiji je zato bolj »varno«, če obstaja več skupnih lastnosti, na podlagi katerih sklepamo na neko tretjo lastnost (v primeru Turingovega testa je to zavest). Razlika, ki jo želimo prikazati, je ponazorjena na Shemi 1. Na levi strani vidimo sklepanje pri izvirnem Turingovem testu, kjer sklepamo *zgolj* na podlagi obnašanja – zdi se, da je takšno sklepanje prešibko (ugovor iz črne škatle in argument kitajske sobe). Na desni strani pa sklepamo (z dodanim oblikovnim kriterijem) na podlagi dveh skupnih lastnosti, kar je zagotovo močnejše sklepanje, ki naredi naš argument močnejši.



Shema 1

Kot pravi Copeland (za močni *oblikovni* kriterij), se sicer zdi, da je ta kriterij premočan (antropocentričen). Nekdo lahko namreč ugovarja z naslednjim miselnim eksperimentom (imenovali ga bomo miselni eksperiment iz zamisljivosti): lahko si zamislimo bitje, ki bi imelo bistveno drugačno fizično strukturo in delovanje možganov, pa bi vendarle imelo zavest. Vendar tudi če je to zamisljivo, je naloga tistega, ki to trdi, da dokaže, da takšno bitje obstaja. Prav tako z logičnega vidika ne moremo očitati nekomu, ki se s tem ne bi strinjal in bi rekel: »sprej-

mem, da ima to bitje enako obnašanje in drugačno fizično stanje, ampak morda gre ravno zaradi tega, ker ima bitje drugačno fizično stanje od mojega, na mentalni ravni za nek popolnoma drugačen pojav, ki pa si deli z zavestjo skupno lastnost – obnašanje.« »Ne sprejemem, da sta dva sistema, ki sta funkcionalno izomorfna, realizaciji iste vrste samo zaradi tega, ker sta funkcionalno izomorfna« (Shapiro, 2000: 638). Samo dejstvo, da se neko bitje obnaša enako kot mi, torej še ni zadostni razlog za to, da trdimo, da smo oboji tudi v istem mentalnem stanju (tj. da smo zavestni).

Tukaj zagovornika miselnega eksperimenta iz zamisljivosti postavimo v nemogoč položaj, saj se zdi, da več ne obstaja način, na katerega bi lahko dokazal, da ima tudi takšno bitje zavest – samo obnašanje ni več zadosten pogoj za priznavanje zavesti, in ko imamo dva sistema z različnima fizičnima stanjema, ne moremo sploh več govoriti o tem, ali gre za identičen mentalni pojav. Edini način, na katerega lahko ugovarja, je, da še naprej trdi, da si je takšno bitje mogoče zamisliti – v zamišljanju bitja, ki bi imelo zavest in drugačen fizični sistem, ni nič protislovnega. Takšno bitje je torej logično oz. metafizično zamisljivo, ni pa dokazano, da dejansko obstaja v našem svetu (vendar tudi ni dokazano, da ne obstaja). In zaradi metafizične možnosti takšnega bitja mora biti vsak pogoj za pripis zavesti stroju zgolj zadosten, ne pa tudi nujen. Torej, če se stroj obnaša kot človek in če je v podobnem fizičnem stanju (to pomeni, da »možgani« stroja delujejo na podoben način kot naši), potem nujno velja, da ta stroj misli. Ne velja pa, da stroj misli samo takrat, ko izpolni ta dva pogoja (torej obnašanje in *oblikovni* kriterij), saj si je metafizično možno zamisliti bitje, ki bi imelo zavest in bi se obnašalo enako kot človek, pa bi vendar bilo v različnem fizičnem stanju (možgani takšnega bitja bi delovali na popolnoma drugačen način kot pri človeku). V formalni obliki se zato pogoj za pripis zavesti stroju glasi takole: *če se stroj x obnaša kot človek in ima fizično strukturo procesiranja podobno človeškemu možganom, potem ta stroj misli.*

Prišli smo torej do konjunkcije dveh kriterijev (obnašanja in *oblikovnega* kriterija), ki skupaj tvorita zadosten, nesporen pogoj za priznavanje zavesti stroju, hkrati pa smo pustili odprto možnost, da ima lahko neko bitje zavest tudi brez izpolnitve teh dveh kriterijev, pri čemer pa je breme dokazovanja obstoja takšnega bitja na njegovih zagovornikih.

IV Večvrstna realizacija in Turingov test: preslikava

V drugem delu članka smo ugotovili, da večvrstna realizacija in Turingov test temeljita na isti (behavioristični) predpostavki. V tretjem delu pa smo pokazali, da je behavioristična predpostavka za pripisovanje zavesti pri Turingovem testu pomanjkljiva. Da bi to pomanjkljivost odpravili, smo zato obnašanju dodali *oblikovni* kriterij. Ta dva kriterija v konjunkciji tvorita kakovosten zadosten pogoj za priznavanje zavesti. Če je bila behavioristična predpostavka pomanjkljiva pri Turingovem testu, potem je takšno sklepanje pomanjkljivo tudi v primeru večvrstne

realizacije, saj smo ugotovili, da tudi argument iz večvrstne realizacije uporablja isto behavioristično predpostavko. Zato bomo preslikali rešitev, ki smo jo uporabili pri Turingovem testu, na tezo večvrstne realizacije. Z drugimi besedami, pri Turingovem testu smo to pomanjkljivost odpravili tako, da smo dodali *oblikovni* kriterij. Večvrstna realizacija uporablja enako (behavioristično) predpostavko kot Turingov test, zato ima enake pomanjkljivosti. Torej bomo tudi tezi večvrstne realizacije dodali *oblikovni* kriterij.

Izvorna predpostavka večvrstne realizacije se je glasila takole: če se nek organizem obnaša enako (nerazločljivo) kot človek v stanju bolečine, potem je ta organizem v stanju bolečine. V tretjem delu članka smo ugotovili, da obnašanje samo ni dovolj dober kriterij za pripisovanje identičnega mentalnega stanja, zato je potrebno dodati *oblikovni* kriterij. Tako dobimo naslednji zadostni pogoj za pripisovanje bolečine nekemu organizmu: če se nek organizem obnaša enako (nerazločljivo) kot človek v stanju bolečine *IN ima podobno fizično strukturo kot človek (oblikovni kriterij)*, potem je TA organizem v stanju bolečine. Če sedaj to modificirano predpostavko zapišemo v abstraktni obliki, dobimo naslednje: *če se nek sistem obnaša enako (nerazločljivo) kot človek v stanju x IN ima podoben način delovanja »možganov« kot človek, potem je ta sistem v stanju x*. Če to velja za Turingov test, potem mora veljati tudi za večvrstno realizacijo, ker je predpostavka pri argumentu iz večvrstne realizacije identična tisti pri Turingovem testu.

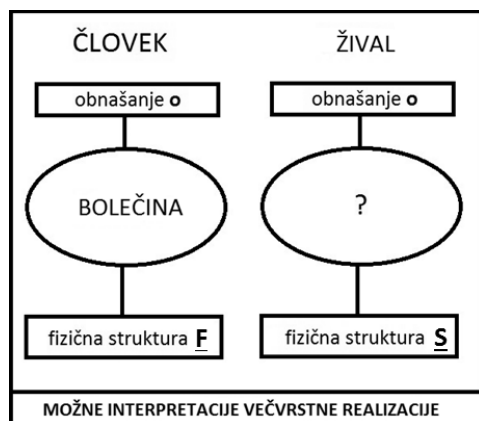
To pa je v popolnem nasprotju s tezo večvrstne realizacije, ki trdi, da lahko vse mentalne pojave večvrstno realiziramo z različnimi fizičnimi sistemi. Zagovornik teze večvrstne realizacije trdi (v logični obliki) dva zadostna pogoja za pripisovanje istega mentalnega stanja (bolečine). Prvič trdi, da: če se nek organizem obnaša enako (nerazločljivo) kot človek v stanju bolečine *in NIMA podobnega sistema delovanja možganov kot človek*, potem je ta organizem v stanju bolečine. Hkrati pa mora zagovornik večvrstne realizacije seveda sprejeti tudi naš izboljšan pogojnik (*če se nek organizem obnaša enako (nerazločljivo) kot človek v stanju bolečine IN ima podoben sistem delovanja možganov kot človek, potem je ta organizem v stanju bolečine*), saj priznava, da tudi soljudje izkušajo bolečino. Zagovornik večvrstne realizacije mora tako popolnoma zanemariti sistem delovanja (fizično stanje) organizma, saj se mora strinjati z obema pogojnikoma in s tem, da če ima organizem lastnost O (obnašanje) in lastnost F (človeku podobna fizična struktura), potem je ta organizem v stanju B (bolečina), in s tem, da če ima organizem lastnost O in nima lastnosti F, potem je ta organizem v stanju B (bolečina). V logičnem zapisu to izgleda takole: $((O \wedge F) \vee (O \wedge \neg F) \rightarrow B)$

Vendar leži razlog za priznavanje $(O \wedge F) \rightarrow B$ v sklepanju na podobnost (analogija), ki je zaradi dveh skupnih lastnosti kakovostno in empirično podkrepljeno. Medtem ko je sklepanje $(O \wedge \neg F) \rightarrow B$, če priznamo, da samo obnašanje ni dovolj za pripisovanje bolečine – kar zaradi ugovora iz črne škatle in ugovora kitajske sobe skorajda moramo –, utemeljeno zgolj z razlogom, da je to metafizično za-

misljivo. »Resnico teze večvrstne realizacije je potrebno oceniti na podlagi empiričnih dokazov« (Shapiro, 2000: 637). Sama metafizična zamisljivost torej ne more biti dovolj za sklepanje na obstoj večvrstne realizacije, saj je teza večvrstne realizacije empirična teza. Če si lahko nekaj metafizično zamislimo, to še ne pomeni, da kaj takšnega v dejanskem svetu tudi obstaja. Kot pravi Shapiro: »Morda pa je samo eno fizično možno elektrokemično stanje, ki lahko realizira bolečino« (Shapiro, 2000: 641). V vsakem primeru je očitno, da je breme dokazovanja – *onus probandi* – kot smo zapisali za Turingov testi, tudi v tem primeru večvrstne realizacije na plečih zagovornikov le-te.

V Kako dokazati večvrstno realizacijo

Kako naj torej zagovornik večvrstne realizacije dokaže svojo tezo, če obnašanje samo ni dovolj dober kriterij, v primeru *oblikovnega* kriterija pa sploh nismo več prepričani, ali res gre za večvrstno realizacijo? Sama metafizična zamisljivost prav tako ne more služiti kot dokaz za obstoj večvrstne realizacije, saj je večvrstna realizacija empirična teza. Prej smo omenili, da je zaradi nezmožnosti vpogleda v izkustvo nekega drugega organizma skoraj nemogoče prepričljivo dokazati večvrstno realizacijo (dvomimo lahko že v enakost izkustva sočloveka), razen seveda preko analogije s samo eno skupno lastnostjo, kar pa se zdi premalo. Kaj torej zagovorniku večvrstne realizacije preostane? Ali kot je zapisal Nagel: »Vedeti hočem, kako je *netopirju* biti netopir. In če si to poskušam predstavljati, me omejujejo možnosti mojega duha, in te možnosti so za nalogo neprimerne. Ne morem je izvršiti niti ob zamišljanju dodatkov k mojemu sedanjemu izkustvu niti ob zamišljanju segmentov, ki mu jih postopoma odvzema, niti ob zamišljanju nekakšne kombinacije dodatkov, odvzemanj in sprememb« (Nagel, 1974/1990: 402).



Shema 2

Tudi če dokažemo, da možgani živali delujejo na popolnoma drugačen način, bo nasprotnik večvrstne realizacije vedno imel možnost, da zanika, da gre za enako subjektivno kvalitativno izkustvo (takšnost). Če dokažemo, da se mehanizmi de-

lovanja možganov živali in ljudi razlikujejo, imamo še vedno na voljo dve interpretaciji (glej Shemo 2). Prvič: gre za večvrstno realizacijo bolečine, saj obstaja (po argumentu iz podobnosti) ena podobnost – enako obnašanje. Drugič: gre za različno izkustvo (žival ne čuti bolečine, temveč nekaj drugega), *ravno zaradi tega*, ker se fizična struktura razlikuje. »Če se realizirajoči vrsti resnično razlikujeta v njihovih vzročno relevantnih lastnostih, potem se zdi, da gre za različni vrsti. Toda če sta različni, potem nista ista vrsta, in tako to ni primer, ko je ena sama vrsta večkrat realizirana« (Shapiro, 2000: 647).

Kako torej rešiti večvrstno realizacijo? Odgovor se morda skriva v plastičnosti možganov. »... možgani so presenetljivo plastični. Preprosto povedano, videti je, da so lahko identična mentalna stanja realizirana v različnih možganskih (fizičnih) stanjih v različnih ljudeh ali pa v različnih možganskih stanjih v isti osebi« (Shapiro, 2000: 642). En način dokazovanja bi torej lahko ležal v nevrofiziologiji. Če bi dokazali, da lahko človek izkusi bolečino, ko je v dveh različnih fizičnih stanjih, ki bi se razlikovali v »vzročno relevantnih lastnostih – tj. v lastnostih, ki pripomorejo k zmožnosti, namenu, cilju in drugim stvarjem, ki opredelijo, da je T¹² takšne vrste« (Shapiro, 2000: 646), potem bi lahko rekli, da je bolečino možno večvrstno realizirati. Problem je seveda določiti vzročno relevantne lastnosti – kdaj se mehanizma na elektrokemični ravni razlikujeta do te mere, da lahko temu rečemo večvrstna realizacija, in pa kako takšno mero postaviti.

Shapiro v svojem članku *Multiple Realizations* poskuša prikazati vzročno relevantne lastnosti na primeru različnih odpiralcev buteljk, z različnimi mehanizmi (glej Shapiro, 2000: 643–644). To Shapiro šteje za jasen primer večvrstne realizacije, še vedno pa nimamo kriterija oziroma vsaj paradigmatičnega primera v samem problemu duha in telesa, ki bi nam omogočal jasno razločevanje in ocenjevanje različnih fizičnih sistemov, in preko katerega bi lahko ugotovili, ali gre za primer večvrstne realizacije ali ne – vprašanje je torej, kaj bi lahko navedli kot primer jasne večvrstne realizacije v samem problemu duha in telesa.

Odgovor, ki ga bomo podali, se nanaša na osnovno predpostavko večvrstne realizacije (če se neka stvar obnaša enako (nerazločljivo) kot človek v stanju x, potem je ta stvar v stanju x) in služi zgolj postavljanju kriterija različnosti (vzročno relevantnih lastnosti), ki bi ga morali vzpostaviti (znotraj človeških možganov), da bi lahko zagotovo rekli, da je nek pojav večvrstno realiziran. Če bi torej v človeških možganih našli dva mehanizma, preko katerih bi lahko subjekt izkusi bolečino, in če bi se ta dva mehanizma razlikovala tako, kot se razlikujeta delovanje človeških možganov in delovanje računalnika, potem bi lahko zagotovo rekli, da je takšna bolečina večvrstno realizirana. Zelo verjetno je, da je ta pogoj premočan (je zadosten pogoj, ni pa nujen) – takšne razlike v delovanju možganov najverjetneje ne bomo našli –, vendar nam predstavi jasen primer razlike v dveh mehanizmih,

¹² T je črka, ki jo avtor uporabi za neko mentalno stanje – recimo bolečino.

kjer kot eden izmed mehanizmov nastopajo človeški možgani, ne pa, kot je v navadi filozofov, analogije z odpirachi, mačkami in mišelvami.

Zagovorniki večvrstne realizacije morajo torej na podlagi nevrofizioloških dejstev poiskati različna fizična stanja, ki realizirajo identična mentalna stanja, ugotoviti, kaj neko fizično stanje naredi za bistveno različno od drugega, in postaviti kriterije, na podlagi katerih bomo lahko presojali, ali obstajata dve fizični stanji, ki isto mentalno stanje realizirata na bistveno različen način. Povedati morajo torej, kaj so vzročno relevantne lastnosti, ki naredijo dve fizični stanji različni.

VI Sklep

V članku smo pokazali, da se večvrstna realizacija in Turingov test pri postavljanju lastnih premis naslanjata na isto predpostavko. S pomočjo ugovora iz črne škatle smo prikazali, da je ta (behavioristična) predpostavka pomanjkljiva in je kot takšne ne moremo sprejeti. Zato smo predpostavko v primeru Turingovega testa nadgradili s pomočjo *oblikovnega* kriterija. Nadgradnjo smo nato preslikali na večvrstno realizacijo, kar nas je vodilo v sklep, nasproten sklepu večvrstne realizacije. Ugotovili smo, da je edini razlog za sprejetje večvrstne realizacije njena metafizična zamisljivost, kar pa ni dovolj, saj je teza večvrstne realizacije empirična teza. Zato smo poskušali izdelati načrt, kako bi lahko na kakovosten način dokazali obstoj večvrstne realizacije: s pomočjo nevrofiziologije se poišče dve različni fizični stanji v človeku, obe korelirani z izkustvom istega mentalnega stanja (npr. bolečine). Če bi odkrili dve takšni stanji in prikazali bistvene razlike med njunima delovanjema, bi s tem dokazali tudi večvrstno realizacijo.

Multiple Realizability and Turing's Test: The Transference

The first part of the article introduces the following two concepts: multiple realizability and Turing's test. The aim of the second part is to prove that both Turing's test and multiple realizability use an identical line of reasoning. In the third part we present one of the objections against Turing's test, which shows that the reasoning in part two is fallacious, and show how it can be solved by adding the so-called *design* criterion. This solution is then transferred to the multiple realizability thesis in part four, bringing us to a conclusion opposite to that of multiple realizability. This then leads us to the conclusion that the metaphysical conceivability of multiple realizability is the only reason for accepting it, which, as we show, is insufficient. In the fifth part we present a possible defence strategy of multiple realizability and the problems that may arise.

Keywords: multiple realizability, mental states, Turing's test, black box objection, qualia

Literatura

- Ayer, A. (1978). *Language, Truth and Logic*. Harmondsworth: Penguin.
- Bickle, J. (2016). »Multiple Realizability«. V Zalta, E. N. (ur.), *The Stanford Encyclopedia of Philosophy* (pomlad 2016). Dostopno na: <http://plato.stanford.edu/archives/spr2016/entries/multiple-realizability/>
- Bregant, J. (2004). *Misel kot vzrok: ali so mentalna stanja vzročno učinkovita?* Maribor: Pedagoška fakulteta Maribor.
- Bregant, J. (2010). »Ali lahko stroj misli?«. *Analiza*, 14(4), str. 55–72.
- Bregant, J. (2012). »Turing se je motil.« *Analiza*, 16(1/2), str. 5–31.
- Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell Publishers Ltd.
- Couch, M. (2004). »Discussion: A Defense of Bechtel and Mundale«. *Philosophy of Science*, 71, str. 198–204.
- Klampfer, F., Bregant, J. in Gartner, S. (2008). *Telovadnica za možgane: uvod v kritično mišljenje*. Maribor: Svarog.
- Nagel, T. (1974/1990). »Kako je biti netopir?«. V Hofstadter, D. in Dennett, D. (ur.), *Oko duha*, Ljubljana: Mladinska knjiga, str. 399–410.
- Oppy, G., in Dowe, D. (2016). »The Turing Test«. V Zalta, E. N. (ur.), *The Stanford Encyclopedia of Philosophy* (pomlad 2016). Dostopno na: <http://plato.stanford.edu/archives/spr2016/entries/turing-test/>
- Putnam, H. (1967). »Psychological Predicates«. V Capitan, W. H. in Merrill, D. D. (ur.), *Art, Mind, and Religion*, Pittsburgh: University of Pittsburgh Press, str. 37–48.
- Putnam, H. (1975). »The Nature of Mental States«. V Putnam, H., *Mind, Language and Reality: Philosophical Papers, Vol. 2*, Cambridge: Cambridge University Press, str. 429–440.
- Searle, J. (1981). »Minds, Brains, and Programs«. *Behavioral and Brain Sciences*, 3(3), str. 417–457.
- Shapiro, L. (2000). »Multiple Realizations«. *The Journal of Philosophy*, 97(12), str. 635–654.
- Turing, A. (1950). »Computing Machinery and Intelligence«. *Mind*, 59(236), str. 433–460.
- Zangwill, N. (1992). »Variable Reduction Not Proven«. *Philosophical Quarterly*, 42, str. 214–218.