

## **Ksenija Puškarić**

Seton Hall University, 400 S Orange Ave, South Orange, NJ, U.S.A.  
ksenija.puskaric@gmail.com

# **Rosenthalova teorija svijesti i zombi argumenti**

### **Sažetak**

*Narav svijesti te njezin subjektivni i kvalitativni aspekt predstavljaju teško pitanje u okviru naturalističkih teorija koje nastoje objasniti ovaj fenomen po uzoru na znanstveni reduccionistički model. U ovom radu bavim se Rosenthalovom teorijom svijesti koja je u biti reprezentacionalistička – tj. svjesnost se sastoji u obliku misli višeg reda o mentalnom stanju. Fokus je rasprave na pitanju može li ova teorija ponuditi odgovor na problem eksplanatornog jaza, te sukladno tome branim tezu da Rosenthalov indirektno-reduktivni pristup u konačnici nije uspješan. U tu svrhu formuliram dvije verzije zombi argumenata kojima predlažem da: a) možemo zamisliti biće s identičnim sustavom mentalnih reprezentacija kao našim, ali bez svijesti, te da b) možemo zamisliti da postoji biće s mentalnim reprezentacijama višeg reda kao našim, koje nije nikada imalo kontakt sa stvarnošću, čija bi totalna suma svjesnih iskustava bila konfabulirana. Obje verzije zombi argumenata ukazuju na epistemičke probleme u Rosenthalovoj teoriji.*

### **Ključne riječi**

David M. Rosenthal, svijest, mentalna reprezentacija, naturalizam, zombi argument, eksplanatorni jaz, filozofija uma, intencionalnost

## **Uvod: problem svijesti**

Poseban problem u filozofiji uma predstavlja pitanje kako naše svjesno i subjektivno iskustvo može biti reducirano na nekakvo fizičko ili funkcionalno stanje. Kako, primjerice, doživljaj zelene boje ili okus čokolade može biti nekakvo neuralno stanje? Čak i da možemo otvoreno vidjeti ljudski mozak, u njemu ne bi pronašli niti »nešto« zeleno, niti čokoladno. Drugim riječima, nije jasno kako bi ikakvo objašnjenje moglo povezati, s jedne strane, subjektivni doživljaj, recimo, osjeta boli i, s druge strane, aktivaciju određenih neurona ili funkcija. Naime, teško bi bilo shvatiti kakav bi prirodni zakon mogao povezati subjektivnu i objektivnu domenu – tj. *fizičku* i *mentalnu*. Problematičnost objašnjenja mentalnih svojstava u terminima ne-mentalnog referira na tzv. eksplanatorni jaz<sup>1</sup> (Levine 1983), jer čini se da nam uvijek nešto izmiče kada pokušamo objasniti svijest naturalistički. Preciznije, ono što predstavlja osobito »težak problem svijesti« (Chalmers 1996) jest tzv. fenomenalna svijest koju Ned Block definira na sljedeći način:

»Fenomenalno svjesna stanja jesu iskustvena stanja, tj. stanje je pojavno svjesno ako ima iskustvena svojstva. Ukupnost iskustvenih svojstava mentalnog stanja čine ono kako izgleda biti u stanju. Svojstva pojavno svjesnog stanja uključuju iskustvena svojstva senzacija osjeta i percepcija, a također bih uključio i misli, želje, i emocije.« (Block 1995: 230)

1

Ovo je »široki termin« koji obuhvaća problem formuliran na različite načine. Za najznačajniji doprinos vidi: McGinn 1991; Chalmers 1996.

Fenomenalna svijest odnosi se na ono »kako izgleda biti u mentalnom stanju« (Nagel 1974), npr. ono kako izgleda biti u stanju glavobolje, razmišljati o šetnji New Yorkom ili ono kakav je doživljaj okusa vanilije itd. Jedno je pitanje kako shvatiti taj subjektivni aspekt iskustva kao dio fizikalne stvarnosti u kojoj vladaju zakonitost i egzaktnost, a drugo je pitanje koje bi svaka teorija svijesti trebala objasniti što određuje ono *kako* mi iskustvo izgleda (za više o ovoj problematici na hrvatskom jeziku vidi: Berić 1994). U ovom se kontekstu često koristi termin *qualia*, koji referira upravo na kvalitativni aspekt iskustva. Što primjerice određuje različitost osjeta crvene i plave boje? Je li ona određena različitim moždanim procesima, funkcijama, sadržajem itd.?

Rosenthalova teorija HOT-a (engl. *higher order thought*; hrv. »misao višeg reda«) nastoji objasniti svijest u terminima nesvjesnog, te u konačnici nesvjesna mentalna stanja objasniti naturalistički. Cilj je ovakvoga indirektnog, reduktivnog pristupa eliminirati jaz u objašnjenju te ponuditi odgovore na pitanje: što određuje subjektivnost i kvalitativnost svjesnog doživljaja?

Da bi se cjelovito shvatila problematika i značaj Rosenthalova objašnjenja svijesti, cilj je prvog dijela ovog rada razmotriti osnovne značajke reprezentacionističkih teorija kojoj HOT-teorija svijesti pripada, ideju reduktivnog objašnjenja, te skicirati Rosenthalovu metodu indirektno naturalističke redukcije svijesti na ne-mentalno. Drugi dio posvećen je detaljnijem pojašnjenju Rosenthalove teorije te njezine eksplanatorne moći, i to s ciljem da se u posljednjem dijelu razmotre problemi i evaluira uspješnost i plauzibilnost potonje teorije.

Centralna teza ovog rada svodi se na formulaciju HOT-zombi misaonog eksperimenta kojim dovodim u pitanje ideju da je Rosenthalova teorija uspjela zatvoriti eksplanatorni jaz. Tvrdim da postoji epistemološki problem koji nas ostavlja s pitanjem prednosti njegove teorije nad jednostavnijim naturalističkim pristupima, npr. fizikalizmom.

## 1. Naturalistička redukcija i reprezentacionalizam

### 1.1. Reduktivno objašnjenje i naturalizam

Ljudi su u prošlosti mislili da je nemoguće doći do Mjeseca te nisu niti mogli zamisliti, recimo, transplantaciju organa. Život se nekoć smatrao utjelovljenjem *elan vital*, životne sile. Tada naprosto nije bilo jasno kako to da je jedna materija utjelovljuje život, a druga ne. Nama se danas čini nemogućim da će znanost ikada objasniti svijest. Ipak, znanost je uspješna, a znanstvenici su optimisti. »Duh« iz tradicionalne filozofije nastoji se demistificirati i zamijeniti – laicima nerazumljivim – neurološkim pojmovima. Na taj način, svijest više ne bi bila nešto mistično. Sve što postoji može u krajnjoj liniji biti objašnjeno fizikalnim entitetima i prirodnim zakonima. Sva tijela sastavljena su od atoma i veza među njima. Sve prirodne pojave mogu se objasniti fizikalnim zakonima. Zašto bi svijest predstavljala izuzetak? Ovakav stav zastupaju redukcionisti naturalisti, pristaše znanstvenog objašnjenja svijesti.

U znanosti su uobičajene identifikacije tipa voda je H<sub>2</sub>O, toplina je molekularno gibanje itd. Tako bi se i svijest *identificirala* s ili *reducirala* na neurološko stanje. Pogledajmo kako izgleda redukcija vode na H<sub>2</sub>O i koje su njezine implikacije. Svojstva vode kao što su bezbojnost, tekućost itd. – spoznajemo osjetilima, dok molekularnu strukturu spoznajemo različitim mjerenjima i eksperimentima. Ipak, to nije razlog da se tvrdi da voda nije isto što i

H<sub>2</sub>O. Voda, prozirna pitka tekućina, nije ništa više od određenog spoja atoma vodik i kisika.

Posljedica je redukcije jednog entiteta na drugi redukcija objašnjenja. Svojstva vode, dakle, tekućost, bezbojnost itd., ujedno su i objašnjena svojstvima atoma i kemijskih veza u molekuli H<sub>2</sub>O. Zašto voda ključa na 100 °C? Zato što tada pucaju vodikove veze. Time smo reducirali vodu na H<sub>2</sub>O, ali i ujedno dali objašnjenja svojstva vode.  $X$  je reduciran na  $Y$  znači da je  $X = Y$ . Voda je reducirana na H<sub>2</sub>O znači da voda jest H<sub>2</sub>O. Ovo je tipičan primjer jedne znanstvene identifikacije.

Isto tako, mentalna stanja mogu se reducirati na ili su identična s fizičkim procesima u mozgu.<sup>2</sup> Redukcionistička teorija koja zastupa ovakav stav zove se *tipski fizikalizam* ili *teorija identiteta* (Place 1956, Smart 1959). Fenomen svijesti, dakle, predmet je proučavanja neurologije. Cilj je za svaki tip mentalnog stanja pronaći odgovarajuće moždano stanje, a to je na kraju krajeva i empirijsko pitanje.

S obzirom na to da je ovakva teorija nastala u 50-im i 60-im godinama prošloga stoljeća, da je imala znatnih problema s višestrukim realizacijama jednog te istog tipa mentalnih stanja – te s nizom drugih argumenata kao što su zombi argument – s vremenom se javila ideja da se reduktivnom objašnjenju pristupi indirektno, tj. da se na taj način izbjegniju slabosti različitih verzija fizikalizma i funkcionalizma (više: Sesardić 1984) koji reduciraju mentalna stanja na tipične kauzalne uloge.

## 1.2. Reprerzentacionalističke teorije

Zajednička direktna ili indirektna karakteristika reprerzentacionalističkih teorija jest pretpostavka da se jaz u objašnjenju može premostiti u više koraka ili razina u objašnjavaćem lancu. Kratki nacrt reprerzentacionalističke strategije premošćivanja jaza sastoji se od dvije razine:

- (1) ono kako izgleda biti u mentalnom stanju ima strukturu i objašnjeno je intencionalnom strukturom (tj. sadržajem);
- (2) sadržaj se može naturalistički objasniti (npr. tako da se objasni u terminima kauzalnih veza s vanjskim svijetom )

Reprerzentacionalistički teoretičari prvi korak u naturalizaciji svijesti vide u tome da pokažu kako svijest nije jednostavno svojstvo mentalnih stanja, nego da ima strukturu. Ako svijest ima strukturu, onda postoji objašnjenje. O kakvoj je strukturi riječ? Kada usmjerimo pažnju na doživljaj crvene boje, te smo *crvenosti* svjesni kao jednostavnog svojstva iskustva. Naime, *mentalne crvenosti* nismo svjesni kao produkta nekakvog procesa, kao nekakve relacije spram crvenog predmeta i sl., doživljaj *crvenosti* zapravo je *sama crvenost*. Dana je kao takva, jednostavna i bez strukture. Sada usporedimo crvenost s položajem nekog predmeta u prostoru. Primjerice, dok gledamo jabuku na stablu, njezinog položaja svjesni smo kao relacije, tj. svjesni smo položaja jabuke s obzirom na odnos koji ima prema nekom drugom predmetu u prostoru. Jabuka se nalazi npr. 2 metra od površine, 3 metra od promatrača itd.

2

Za svako individualno mentalno stanje postoji individualni moždani proces. Uzmimo npr. da sam u moždanom stanju  $X$  da je moj susjed u stanju  $X$ , te da  $X$  odgovara stanju depresije. Za jedno i drugo neuralno stanje postoji isto

mentalno stanje. Prema tome, slijedi da smo ja i susjed u stanju depresije. Situacija vrijedi i obrnuto, ako sam u depresiji onda sam u moždanom stanju  $\beta$ .

Položaj predmeta uvijek određujemo s obzirom na neki drugi objekt. Dakle, lokacija objekta određena je nekakvom relacijom, tj. ona je relacijsko svojstvo. Slično prostornoj lokaciji, doživljaj crvene boje – po mišljenju reprezentacionalista – također ima strukturu, ali ona je skrivena introspekciji. Naime, mentalnih stanja nismo svjesni kao stanja koja imaju sadržaj, nego naprosto doživljavamo nešto što prepoznajemo kao *crvenost* i određujemo pojmom *crveno*. Ipak, senzorna stanja – kako tvrde reprezentacionalisti – imaju strukturu. To je, dakle, prvi korak u naturalizaciji svijesti.

Sada ćemo detaljnije objasniti što je sadržaj i povezati ga s učenjem reprezentacionalista, čiji su najpoznatiji zastupnici Gilbert Harman, Michael Tye, Fred Dretske, Tim Crane i mnogi drugi (više o reprezentacionalizmu: Crane 2001). Većina se uglavnom slažu oko toga da vjerovanja, želje i nade uvijek jesu o nečemu. Misao je uvijek misao o nečemu, percepcija je uvijek percepcija nečega. Ako postoji odgovor na pitanje: na što je usmjeren um u stanju *m*, onda postoji objekt mentalnog stanja. Kada gledam zalazak sunca, objekt percepcije je sunce.

Mnoga su mentalna stanja usmjerena na objekt (intencionalna), ona jesu *o objektima*. Stoga, reprezentacionalizam tvrdi sljedeće: to da mentalno stanje jest o nečemu implicira da je nešto reprezentirano u umu, i obrnuto, ako je nešto reprezentirano u umu u nekom mentalnom stanju, onda postoji i nešto o čemu je to mentalno stanje. Doživljaj zalaska sunca mentalna je reprezentacija sunca koja odražava stvarni zalazak sunca. Prema tome, sunce je reprezentirano u umu. Kakav je objekt, kako mi npr. sunce izgleda? Okruglo i sjajno, žuto, bliješteće. To *kako* mi sunce izgleda, kako mi je sunce reprezentirano u umu, odgovara sadržaju. Sunce je zahvaćeno pod određenim aspektom, *kao* okruglo, sjajno, crveno i bliješteće. Vidim sunce kao takvo, a ne kao zvijezdu koja isijava fotone. Dakle, sadržaj predstavlja odgovor na pitanje »što je u umu?« ili »kakav je objekt?«. Mišljenje, percipiranje, osjećanje načini su ili modusi kojima um stoji u relaciji prema sadržaju. Tim Crane ovako definira modus:

»Intencionalni modusi jesu relacije u kojima subjekt stoji prema sadržaju intencionalnih stanja.« (Crane 2001: 32)

No kakve veze ima sadržaj s onim kako mi izlazak sunca izgleda? Da je sunce reprezentirano u umu kao četvrtasto i zeleno i doživljaj bi bio odgovarajući. Zalazak mi izgleda tako zato što je sadržaj percepcije »sjajno, žuto, okruglo«. Prema ovoj teoriji *sadržaj mentalnog stanja određuje ono kako izgleda biti u tom stanju*. Osim sadržaja bitan je još i način na koji zahvaćamo sadržaj. U sunce možemo gledati, možemo razmišljati o suncu kao nebeskom tijelu itd. Dakle, sadržaj zajedno s modusom fiksira ili determinira ono kako izgleda *biti u nekom stanju*. Postavimo sada karakteristike reprezentacionalističkih teorija u širi kontekst.

Kao što je već spomenuto, teorija svijesti trebala bi adekvatno objasniti subjektivnost iskustva i njezinu kvalitativnost. Subjektivnost, tj. činjenica da nekako izgleda imati iskustvo, objašnjena je samom naravi sadržaja. Naime, sadržaj je stvar aspekta pod kojim stojimo u relaciji s objektom. Prisjetimo se primjera sa suncem koje može biti doživljeno kao bliješteće ili kao fotonska zvijezda. Kvalitativnost, ono što određuje različitost doživljaja – primjerice, crvene i plave boje, određeno je razlikom u sadržaju perceptivnog iskustva.

Do ove se točke svi zastupnici reprezentacionalističkih teorija međusobno slažu – svjesna mentalna stanja imaju sadržaj. Razilaženja se prvenstveno odnose na pitanje: je li sadržaj ujedno i dovoljan uvjet za svjesno stanje? Odnosno, je li za pojavu svijesti dovoljno imati mentalnu reprezentaciju vanjskog svijeta?

S obzirom na spomenutu dilemu, reprezentacionalističke teorije dijele se u dvije skupine. Jedni tvrde da je za svjesno stanje dovoljno da ono reprezentira vanjski svijet. Drugim riječima, da bi mentalno stanje imalo fenomenalni karakter, dovoljno je da ono nešto reprezentira (tj. da imamo svijest o nečemu). Da bismo imali npr. doživljaj zalaska sunca, dovoljno je da imamo svijest o tome kakvo je sunce, odnosno da imamo mentalno stanje (tj. reprezentaciju prvog reda). Ova teorija naziva se *reprezentacijskom teorijom prvog reda* (ili skraćeno teorija FOR, engl. *first order representationalism*). Njezini najpoznatiji zastupnici su Fred Dretske i Michael Tye.

Druga skupina reprezentacionalista smatra da samo sadržaj prvog reda nije dovoljan za objašnjenje svijesti zato što i nesusjesna stanja imaju sadržaj. Adekvatna bi teorija svijesti ujedno trebala i objasniti u čemu se sastoji razlika između svjesnih i nesusjesnih stanja. Nesusjesna stanja nemaju fenomenalni aspekt. Za razliku od nesusjesnih stanja, svjesna su stanja ona stanja o kojima imamo svijest. Da bi uopće doživljavali zalazak, moramo imati svijest o tome kako nam zalazak izgleda. Ako ne znamo kako nam zalazak izgleda, onda nema toga kako izgleda *vidjeti zalazak*, percepcija je tada nesusjesna. Sadržaj prvog reda imaju i nesusjesna stanja. Međutim, svjesna stanja očituju se u tome da je upravo sadržaj prvog reda reprezentiran u sadržaju drugog/višeg reda. Potrebna je, dakle, reprezentacija mentalnog stanja (tj. prvog reda). Zato se ova skupina teorija naziva *teorijom reprezentacije višeg reda* (HOR, engl. *higher order representationalism*).

S obzirom na to da smo svjesni mentalnih stanja, teoretičari HOR-a nadalje se razilaze na one koji smatraju da je reprezentacija višeg reda u stvari nekakav oblik kvazi percepcije, dok drugi tvrde da je misao bolji kandidat za tu ulogu. Teoriju o iskustvu višeg reda (HOE, engl. *higher order experience*) zastupaju David Malet Armstrong i William Lycan. Oni tvrde da smo svjesni mentalnih stanja pomoću monitoring procesa ili skeniranja mentalnih stanja. Ovaj model objašnjenja naziva se još *kvaziperceptivnim* zato što se skeniranjem formira nekakav oblik perceptivne reprezentacije. Dakle, stanje će biti svjesno tek kada je *skenirano* ili *percipirano*, a time i reprezentirano u sadržaju višeg reda. Doživljaj je zalaska svjestan onda kada je to stanje objekt unutrašnje percepcije.

Druga teorija unutar obitelji HOR-a naziva se *teorijom misli višeg reda* (engl. *higher order thought* – HOT). Kao što sâm naziv teorije kaže, stanje je svjesno onda kada je praćeno mišlju višeg reda. Točka neslaganja unutar teorije HOT-a odnosi se na pitanje: je li potrebna aktualna, stvarno mišljena misao o mentalnom stanju ili je pak dovoljno da mentalno stanje ima dispoziciju da izazove misao višeg reda (Carruthers 2000)? »Aktualist«, kao npr. Rosenthal, inzistira na tome da *nekako izgleda* vidjeti zalazak samo ako pritom stvarno (aktualno) mislimo misao o tome da vidimo zalazak. Misao o tome kako izgleda zalazak sunca čini naše stanje viđenja zalaska svjesnim. Nasuprot »aktualistu«, dispozicionalist – poput Petera Carruthersa – smatra da je za svjesno iskustvo dovoljno da mentalno stanje ima mogućnost izazvati misao višeg reda. U narednom poglavlju posvetit ćemo se Rosenthalovoj HOT-teoriji.

## 2. HOT-teorija svijesti

Po Rosenthalovu shvaćanju, na raspolaganju nam stoje 2 moguće strategije u objašnjenju svijesti:

- (1) svijest objasniti u terminima ne-mentalnog;

(2) svijest objasniti u terminima mentalnih stanja – i to:

(a) mentalnih i svjesnih;

(b) mentalnih i nesvjesnih (ovu opciju prihvaća Rosenthal 1990, 1997).

O prvoj je strategiji već bilo riječi. Radi se o redukciji svijesti na neurološka stanja koja vodi do jaza u objašnjenju. Zbog toga Rosenthal odbacuje prvu opciju. Druga nam strategija ostavlja izbor između dvije opcije. Prvo, svijest se može objasniti u terminima mentalnog, ali ujedno i svjesnog. To je teorija po kojoj su sva mentalna stanja svjesna, a čiji su dominantni zastupnici bili René Descartes i John Locke.<sup>3</sup> U prilog ovoj poziciji svakako idu senzorna stanja jer je teško, primjerice, zamisliti bol, a da nikako *ne izgleda biti* u stanju boli. Ili kako Locke kaže, nema smisla reći da sam cijelo vrijeme gladan, ali da tu glad ne osjećam. Dakle, svijest je neodvojiva od mentalnosti. Svijest je *intrinzično* svojstvo mentalnih stanja. Rosenthal ovakvo shvaćanje mentalnosti smatra neprihvatljivim. Ako se u objašnjenju svjesnog pozivamo na mentalnost, a da su pritom sva mentalna stanja svjesna stanja, onda će objašnjenje biti cirkularno i neinformativno jer se pozivamo na postojanje činjenice koju želimo objasniti. Svijest je pritom primitivna činjenica koja objašnjava, ali koja se kao takva niti može, niti treba objasniti.

Rosenthal smatra da je moguće izbjeći loše strane predloženih strategija i istovremeno zadržati naturalističko objašnjenje svijesti. Rješenje je u tome da se eksplanatorni lanac produži. Svjesna stanja bit će objašnjena *nesvjesnim stanjima*, a ona će dalje biti objašnjena *neuralnim stanjima* mozga ili funkcijama. Pojednostavljeno, slika bi izgledala ovako (strelica stoji za »su objašnjena s«):

*svjesna mentalna stanja* → *nesvjesna mentalna stanja* → *fizička stanja*

Prednost je ovakve eksplanatorne redukcije u više koraka u tome što se jaz lakše premošćuje ili makar ublažava, jer je svijest objašnjena u terminima intencionalnih, nesvjesnih stanja, a ne direktno u terminima neuroloških ili funkcionalnih stanja. Dakle, ako želimo informativno objašnjenje svijesti, onda ujedno trebamo odbaciti tezu da su sva mentalna stanja svjesna. A to u stvari znači da trebamo prihvatiti postojanje nesvjesnih stanja. Ova je teza polazište, a ujedno i motivacija za HOT-teoriju. S obzirom na to da postoje nesvjesna stanja koja imaju sadržaj, jasno je da posjedovanje mentalne reprezentacije neće biti dovoljan uvjet za svijest.

## 2.1. Nesvjesna mentalna stanja

Postojanje nesvjesnih mentalnih stanja posebno je teško prihvatiti za senzorna stanja: bol, percepciju, emocije itd. Ideja o boli *koja ne boli* ili gladi *koja se ne osjeća* toliko je protivna intuiciji da je to neke filozofe, kao npr. Descartesa i Lockeja, navelo na mišljenje da nesvjesna senzorna stanja ne postoje. Rosenthal smatra da je tome razlog taj što, primjerice, za doživljaj boli ne znamo ako *nikako ne izgleda biti* u stanju boli, odnosno ne znamo za bol ako ona nije svjesna (Rosenthal 1997). Međutim, to što ne znamo za nesvjesnu bol ne pokazuje da nesvjesna bol ne postoji.

Drugi razlog zbog kojeg Rosenthal smatra da je teško prihvatiti postojanje nesvjesnih senzornih stanja sastoji se u tome da je teško zamisliti da nesvjesno mentalno stanje ima kvalitetu (Rosenthal 1990, 1997). Razlika između *iskustvene crvene* i *iskustvene zelene* postoji samo ako za nas *nekako izgleda* vidjeti *crveno* i *zeleno*. Naime, mi razlikujemo i klasificiramo senzorna stanja

samo ako su ona svjesna. Međutim, to što ih mi klasificiramo na taj način, ne pokazuje da kvaliteta nesvjesnih senzornih stanja ne može postojati ako nismo svjesni njihove različitosti. To što fiksiramo referenciju pojma *crveno* na svjestan doživljaj crvene ne pokazuje da kvaliteta *crvenosti* nije prisutna kod nesvjesne percepcije.

Osim filozofskih razloga zbog kojih bismo trebali odbaciti stajalište da su sva senzorna stanja svjesna, Rosenthal tvrdi da postoje primjeri iz svakodnevnog života koji idu u prilog suprotnoj intuiciji, dakle, da postoje nesvjesna stanja (Rosenthal 1990, 1997). Nesvjesna stanja predstavljaju najbolje objašnjenje nekih svakodnevnih situacija:

Često se događa da ljudi imaju nekakve emocije, a da ih nisu svjesni dok im netko ne ukaže na to što osjećaju. Zatim, kod manje intenzivnih boli osoba može biti nečim ometena tako da tu bol uopće ne doživljava. Tada je plauzibilno govoriti o konstantnoj boli koja povremeno postaje nesvjesna. To se često događa ranjenim vojnicima koji usmjeravaju pažnju na radnje potrebne za preživljavanje, pa tako potiskuju bol.

Uzmimo slučaj tzv. »odsutnog vozača«. Dok ljudi voze cestom koju dobro poznaju često se »isključe«. Umjesto da obraćaju pozornost na vožnju, vozači odlutaju u razmišljanje o poslovnim problemima, godišnjem odmoru i sl. Iznenada, vozač shvati da je bio odsutan deset minuta i da ne zna što je u tom razdoblju vidio i radio. Vozač ipak nije razbio automobil o stablo, što upućuje na činjenicu da je vidio što se događa na cesti. Postoji senzitivnost u ponašanju s obzirom na promjenu situacije u okolini (npr. izbjegavanje zapreka, motanje volana i kočenje). Najbolje je opisane situacije objašnjenje da je vozač imao perceptivno iskustvo koje nije bilo svjesno.

Uzmimo još i poznati slučaj *blindsight*-a. Pacijenti koji boluju od *blindsight*-a imaju »rupu« u vizualnom polju. Za predmete u tom dijelu vizualnog polja poriču da ih vide. Međutim, iako ne vide, na upite o tim predmetima (kakva je veličina, boja itd.) oni su u stanju točno pogoditi u vrlo velikom broju slučajeva. Naprosto sreća u pogađanju ne može objasniti toliku uspješnost. Pacijenti, čini se vide predmete u tom polju iako toga nisu svjesni. Njihovo je perceptivno iskustvo nesvjesno. Budući da se psiholozi uglavnom slažu o postojanju nesvjesnih stanja, potrebno je reći da je opravdano prihvatiti nesvjesni dio mentalnosti.

Nesvjesna stanja mentalne su reprezentacije vanjskog svijeta. Informacije primljene iz okoline obrađuju se u kognitivnom sustavu i rezultiraju reakcijom: ponašanjem i verbalnom ekspresijom i sl. Primjerice, »odsutni« vozač nije udario u stablo iako nije imao svjesnu percepciju. Vozač je, po Rosenthalovoj terminologiji, imao svijest o cesti iako toga nije bio svjestan. Dakle, on je svjestan svijeta iako nije svjestan svog iskustva o svijetu.

Smatram da ako prihvatimo da nesvjesna mentalnost postoji, onda moramo prihvatiti realnost distinkcije između toga *kako mi svijet izgleda* i *kako mi iskustvo o svijetu izgleda*. Pojam »subjektivnost prema svijetu« odnosi se na različite aspekte pod kojima smo svjesni stvarnosti – npr. šišmišu svijet izgleda drugačije nego vozaču. Da bi imali svijest o svijetu, dovoljno je da vozač i šišmiš imaju mentalne reprezentacije svijeta. I »odsutnom« vozaču

cesta *nekako izgleda* jer postoji subjektivnost vezana za svijet – cesta mu *nekako izgleda*. To objašnjava činjenicu da se nije zabio u stablo. Međutim, vozač nije imao ono *kako subjektu izgleda biti u tom stanju*.

Objašnjenje tzv. tranzitivne svijesti (tj. svijesti o nečemu) u stvari se odnosi na problem kako objasniti mehanizam reprezentiranja, tj. sadržaj, što ga određuje itd. To je »lakši« problem svijesti. »Težak problem svijesti« (Chalmers 1995) sastoji se u tome kako objasniti ono kako mi iskustvo izgleda. Jer doista, ono što odsutni vozač nema upravo je to kako mu njegovo iskustvo ceste izgleda. Da bismo znali kako nam iskustvo izgleda, trebamo imati mentalnu reprezentaciju iskustva, odnosno svijest o iskustvu. Potrebna je, dakle, reprezentacija višeg reda, kako ističe Rosenthal.

## 2.2. *Tranzitivna i intranzitivna svijest*<sup>4</sup>

Uzmimo za primjer »odsutnog vozača« služeći se Rosenthalovom distinkcijom *tranzitivno/intranzitivno* (Rosenthal 1993). Što razlikuje »odsutnog« vozača od »prisutnog« vozača? Prvi ima svijest o cesti, ali nema svijest o tome kako mu *cesta izgleda*. Drugi ima svjestan *doživljaj ceste* – npr. bio bi u stanju reći što je vidio u zadnjih deset minuta vožnje. »Prisutni« vozač ima svjesnu percepciju, dok je »odsutni« nema. Potonji vozač ima svijest o cesti, a da pritom nema svjestan doživljaj. Dakle, svijest o nečemu ne podrazumijeva ujedno i postojanje svjesnog (intranzitivnog) stanja. Kada bi svako stanje koje nas čini tranzitivno svjesnima bilo ujedno i (intranzitivno) svjesno stanje, onda bi to značilo da je vozačeva svijest o cesti ujedno i svjesno stanje. Ukoliko na ovaj način generaliziramo ideju, utoliko bi negirali postojanje nesvjesnih stanja.

Za svijest o vanjskom svijetu dovoljno je imati nesvjesnu mentalnu reprezentaciju vanjskog svijeta. Nesvjesna stanja čine nas svjesnima okoline, tj. tranzitivno svjesnima. Međutim, vidimo da je kod intranzitivno svjesnih stanja subjekt tranzitivno svjestan svog iskustva. Stanje je (intranzitivno) svjesno onda kada smo ga tranzitivno svjesni. Dakle, da pojednostavimo na primjeru:

- (1) »odsutni« vozač: ima tranzitivnu svijest o cesti;
- (2) »prisutni« vozač: ima tranzitivnu svijest o cesti + tranzitivnu svijest o perceptivnom iskustvu ceste.

Dakle, svijest o  $X$  (tj. tranzitivna svijest) ne implicira svijest stanja (tj. intranzitivna svijest). Pogledajmo поближе njihov odnos:

- (1) imati svijest o cesti = biti tranzitivno svjestan ceste;
- (2) imati (intranzitivnu) svjesnu percepciju ceste = biti tranzitivno svjestan viđenja ceste.

Ako  $X$  ima (intranzitivnu) svjesnu percepciju ceste, onda sasvim sigurno stoji da je  $X$  tranzitivno svjestan ceste. Obrnut slučaj *ne* stoji, ako  $X$  ima svijest o cesti, nužno ne slijedi da  $X$  ima (intranzitivnu) svjesnu percepciju ceste. Iz toga da vozač ima tranzitivnu svijest o cesti ne slijedi da ima (intranzitivnu) svjesnu percepciju. Kada bi obrnuto bio slučaj, onda bi objašnjenje svijesti bilo cirkularno jer bi intranzitivno povlačilo tranzitivno, i obrnuto. Prema tome, ako želimo ne-cirkularno, informativno objašnjenje svijesti, onda trebamo odvojiti *tranzitivno* od *intranzitivnog*, a time ujedno i *svjesno* od *nesvjesnog*. Nekada jesmo tranzitivno svjesni svojih mentalnih stanja, ali nekada i nismo. Svjesnost je relacijsko svojstvo mentalnih stanja.

Označimo s  $T\alpha$  tranzitivnu svijest o  $\alpha$ . Svjesno stanje sastoji se u relaciji između:

1. svijesti o cesti koju imamo nesvjesnim gledanjem ceste ( $\alpha$ );

## 2. svijesti o gledanju ceste ( $T(\alpha)$ ).

Kada vozačevo stanje  $\alpha$  stoji u relaciji s  $T(\alpha)$ , onda je  $\alpha$  svjesno stanje. Ako vozač nema svijest o tome da vidi cestu, onda nema  $T(\alpha)$  i  $\alpha$  je naprosto nesvjesno stanje. Na ovaj je način Rosenthal pokazao da svijest ima strukturu, pa je relacijsko svojstvo mentalnih stanja. Pitanja koja još preostaju glase: (1) kakva mora biti relacija između  $\alpha$  i  $T\alpha$  da bi konstituirala svjesnost; te (2) kako smo svjesni mentalnih stanja?

Najprije razmotrimo  $\alpha$  i  $T\alpha$ , stanje prvog reda i tranzitivnu svijest o tom stanju. Možemo zamijetiti da drugi ljudi mogu imati (tranzitivnu) svijest o našim mentalnim stanjima. Primjerice, psihijatar ima svijest o pacijentovoj depresiji. On zaključuje o njegovom mentalnom stanju na osnovu stvari koje mu pacijent ispriča i kako se ponaša. Naravno, psihijatrovu svijest o pacijentovoj depresiji ne čini to stanje svjesnim. Postoji  $\alpha$  (depresija) i  $T\alpha$  (svijest o depresiji), ali relacija nije direktna. Psihijatar *inferencijalno* postaje svjestan, on nema direktan pristup stanju  $\alpha$  kao što pacijent ima. S obzirom na to da drugi ljudi ne mogu svojom tranzitivnom sviješću učiniti naša stanja svjesnima, Rosenthal dodaje sljedeći uvjet: *S-ovo mentalno stanje  $\alpha$  svjesno je ako i samo ako ga je S sâm tranzitivno svjestan.*

Nadalje, pacijent je sâm mogao zaključiti da je depresivan. Na osnovu toga što je pročitao simptome depresije u knjigama i promatrao svoje ponašanje (odsustvo apetita, nezainteresiranost itd.) mogao je postati svjestan toga da je u depresiji. Međutim, iako je S došao do ispravne dijagnoze, ta spoznaja nije učinila depresiju svjesnom. Prema tome, svjesna stanja moraju ispuniti još jedan uvjet: *S-ovo mentalno stanje  $\alpha$  intranzitivno je svjesno ako i samo ako ga je S svjestan ne-inferencijalno i ne-opservacijski.* Sada preostaje još da vidimo koji je mehanizam zaslužan za svijest o mentalnom.

### 2.3. Hipoteza o misli višeg reda (HOT-u)

Predmeta u vanjskom svijetu svjesni smo na različite načine. Primjerice, zeca na cesti možemo biti tranzitivno svjesni tako da ga vidimo ili da sjedimo u fotelji razmišljajući o njemu ili da nam netko kaže da je zec na cesti. Prema tome, načine na koje smo tranzitivno svjesni možemo grupirati u nekoliko skupina: (1) *percepcija* (vid, sluh), *tjelesne senzacije* (bol, glad) i *mišljenje* (sumnja, nada, vjerovanje i sl.).

Koji je od ovih mentalnih procesa zaslužan za svijest o mentalnom? Čini se da ni na koji način ne osjećamo vlastita mentalna stanja kao što osjećamo svoje tijelo ili vidimo stvari. Ne postoji takav organ koji »percipira« mentalna stanja. Iako je tranzitivna svijest moguća i percepcijom, Rosenthal smatra da ne percipiramo mentalna stanja kao što to, primjerice, zastupaju David Malet Armstrong (1968, 1984) i William Lycan (1987), te sukladno tome predlaže hipotezu o misli višeg reda (HOT; više: Rosenthal 2002). Mentalno je stanje svjesno ako je praćeno mišlju višeg reda, i to mišlju u širem smislu: sumnjom, nadom, vjerovanjem itd. Primjerice, vozaču *nekako* izgleda vidjeti zeca na cesti ako i samo ako ima misao o tome da ga vidi. Posjedovanjem misli »vidim zeca«, vozač postaje svjestan svog iskustva, a tek tada govorimo o *svjesnoj percepciji*. Nadalje, HOT – kao što je spomenuto u uvjetima koje tranzitivna svijest mora zadovoljiti – ne smije biti inferencijalna misao, izvedena iz opažanja ili svjedo-

čanstva drugih ljudi. To što pacijent ima teorije o svom stanju, i to što mu psihijatar kaže, ne može utjecati na to da njegovo stanje postane svjesno. Pacijentova misao da je depresivan (zato što je to zaključio iz svog ponašanja) misao je višeg reda koja ga čini svjesnim, ali pošto nema odgovarajuće neposredne relacije s  $\alpha$ , tada  $\alpha$  neće biti svjesno stanje. Da bi stanje bilo svjesno, HOT mora biti neposredna, direktna misao o  $\alpha$ . Dakle, *S-ovo mentalno stanje  $\alpha$  svjesno je ako i samo ako ga je S svjestan na način da ima ne-inferencijalnu misao višeg reda o  $\alpha$ .*

No čini se da ovaj uvjet može biti prelakao zadovoljen. Čovjek može misliti bilo što – primjerice, razmišljati o vješticama znajući da ne postoje, razmišljati da nije depresivan, kada u stvari jest depresivan. Takva misao, naravno, ne može »stvoriti« svjesnu depresiju ako je nema. Što onda razlikuje misao koju čini stanje svjesnim od »lažnog« HOT-a? Postoji razlika u načinu na koji mislimo o stvarima, npr. kada maštamo o jednorogu ili kada mislimo o Eiffelovu tornju. Kada maštamo o jednorogu, onda ne mislimo o njemu kao da on stvarno postoji. S druge strane, kada mislimo o Eiffelovu tornju, onda mislimo na njega s afirmativnom snagom, mislimo da postoji. Isto je i s mentalnim stanjima, moja misao da vidim slova na ekranu misao je koja je afirmativna, dok to ne mogu reći za misao da me boli noga kada me ne boli. Tako će depresija biti svjesna tek kada mislim o tom stanju kao da je stvarno prisutna u umu. Bitan je način na koji HOT referira na mentalno stanje. Misao mora biti asertorički stav, tj. misao koja ima afirmirajuću snagu. Nekako *mi izgleda biti u depresiji* tek kada mislim da sam depresivna *kao* da stvarno jesam. Možemo reći da je afirmativnost način na koji HOT referira na mentalno stanje nižeg reda. Nadalje, HOT je misao da se *ja* (a ne netko drugi) nalazim u *takvom i takvom* mentalnom stanju. Npr. sadržaj HOT-a mogao bi biti nešto poput *ja vidim zalazak sunca*.

S obzirom na to da je sadržaj ovdje izražen u jeziku, nameće se pitanje koje Rosenthal također uzima u obzir: je li jezik nužan za HOT-ove (Rosenthal 1990, 1997)? Sada ćemo se kratko osvrnuti na zahtjeve koje Rosenthalova teorija predstavlja pred svjesna bića.

#### 2.4. Sadržaj HOT-a

Da bi mi zalazak sunca *nekako izgledao*, moram imati neposrednu asertoričku misao sa sadržajem *ja vidim zalazak sunca*. Prema ovakvom određenju, sadržaja HOT zahtijeva sljedeće: (a) jezik; (b) sintaksu, strukturu rečenice; (c) pojam jastva; (d) mentalne pojmove (vidim, želim itd.); te (e) pojmove koji referiraju na vanjski svijest (zalazak, sunce).

Primjerice, može li mačka imati misao *ja vidim zalazak sunca*? Ovako kako izgleda sadržaj HOT-a, očito je da ne može. Ali kako onda životinje ili djeca mogu zadovoljiti ove visoke standarde?

Prvo, Rosenthal smatra da bića mogu misliti i bez jezika (Rosenthal 1990, 1997). Razmišljanje i pojmovi ne zahtijevaju rječnik. Naime, postoje jednostavne rečenice koje se sastoje od npr. jedne riječi. Drugo, pojam jastva kod djece i životinja ne mora biti tako bogat kao kod odraslih ljudi. Dovoljno je da je biće u stanju razlikovati sebe od okoline. Mačka kada misli da vidi muhu, misli o njoj kao o nečemu različitom od sebe. Minimalni koncept jastva sastoji se u mogućnosti razlikovanja *ja* od *ne-ja*.

Što je s mentalnim pojmovima? Kada mačka osjeća bol, ona ne razmišlja putem pojma *boli*. Kakav je onda sadržaj HOT-a? U slučaju bola, bol je objekt misli na koju upućuje HOT sa sadržajem *ovo!* HOT tada demonstrativno referira na

stanje bola, figurativno rečeno »prstom pokazuje« na nj. Dakle, kod bića koja nemaju mentalnih pojmova postoji neki način demonstrativne referencije.

Nadalje, iz Rosenthalovih pretpostavki prividno slijedi da možemo doživjeti samo ono za što imamo pojam. Mogu vidjeti zalazak sunca samo ako imam pojam *sunce* i *zalazak*, što zvuči apsurdno. Primjerice, naši pojmovi o bojama ne mogu obuhvatiti bogatstvo doživljaja zalaska sunca, a ipak doživljavamo mnoštvo boja. Rosenthalova teorija ovdje ima rješenje. Kao što na (nepoznate) predmete u vanjskom svijetu referiramo s *ova stvar*, tako referiramo i na objekte za koje nemamo pojam. Tada, primjerice, za neku nijansu crvene HOT referira kao »ja vidim *ovu* boju«. S druge strane, ne može se zanemariti uloga pojmova za ono kako izgleda *biti u mentalnom stanju*. Prisjetimo se samo patka/zec primjera iz gestalt psihologije. Doživljaj slike ovisi o tome kako konceptualiziramo ono što vidimo. Ako na slici tražimo zeca, vidjet ćemo zeca, a ako tražimo patku, ista će nam slika izgledati drukčije, vidjet ćemo patku. Rosenthalova teorija ovdje lijepo objašnjava ovaj utjecaj pojmova na iskustvo, kao i mnoštvo drugih mentalnih karakteristika na koje ću se fokusirati u sljedećem poglavlju.

### 2.5. Eksplanatorna moć HOT-teorije<sup>5</sup>

Osim našeg perceptivnog aparata koji je »podešen« za percepciju oblika boja itd., postoji i konceptualni aparat koji bogatimo učenjem. Jedan i drugi aparat određuje kako ćemo doživjeti svijet. Ovisno o tome kako konceptualiziramo crtež – tako nam se i doživljaj mijenja, pojam patka ili pojam zeca određuju ono *kako* ćemo vidjeti crtež. Nadalje, uzmimo primjer kušača vina. Isto vino koje je ponuđeno laiku i profesionalnom kušaču, prvome je naprosto kiselo, dok drugome to isto vino ima slojeve okusa koji se razlikuju. Laik nema onoliko pojmova za okus vina koliko ih ima profesionalni kušač. Upravo ta razlika određuje doživljaj okusa vina kod jednog i drugog. Profesionalni kušač doživljava bogatstvo okusa i zato što ima konceptualni aparat koji mu to omogućuje. Rosenthalova teorija može objasniti ulogu pojmova u iskustvu (više: Rosenthalova 2005). *Nekako mi izgleda* vidjeti zalazak sunca onda kada imam HOT sa sadržajem *ja vidim sunce*. To što zalazak sunca doživljavam na taj način ovisi o tome kako mi HOT reprezentira moje iskustvo. Kada bi sadržaj HOT-a bio *ja vidim zelenu kuglu*, onda bi moje iskustvo bilo takvo. Imala bih pogrešnu reprezentaciju vlastite percepcije zato što pojmovi koji figuriraju u sadržaju HOT-a određuju ono kako izgleda *biti u mentalnom stanju*. Upravo sadržaj HOT-a određuje iskustvo, i ujedno predstavlja dovoljan uvjet za svijest. Zbog ove karakteristike HOT-teorija spada u skupinu reprezentacionalističkih teorija. Drugu stvar koju Rosenthalova teorija objašnjava jest introspekcija (više: Rosenthal 1989). Introspektivna svijest nije ništa drugo nego svjesna misao o mentalnom stanju. Primjerice, *mislim da vidim zalazak sunca* svjesna je misao o mentalnom stanju. Prema HOT-teoriji, introspekcijom smo svjesni HOT-a koji prati mentalno stanje, kojih inače nismo svjesni. Postojanje nesvjesnih HOT-ova, kao podvrste nesvjesnih misli, slijedi iz ideje da postoji nesvjesna mentalnost (usp. poglavlje 2.1.). Međutim, HOT-ovi nam se otkrivaju jedino introspekcijom. Naime,  $\alpha$  je svjesno onda kada imamo HOT ( $\alpha$ ). Na isti način, svjestan HOT o mentalnom stanju  $\alpha$  mora zadovoljiti sljedeći uvjet: *HOT ( $\alpha$ ) svjesno je stanje kada imamo svijest o HOT ( $\alpha$ ), odnosno kada imamo HOT (HOT ( $\alpha$ )).*

5

Za više o ovoj temi vidi: Rosenthal 1990.

Kako imamo ovu svijest o HOT ( $\alpha$ )? Tako da imamo HOT o HOT ( $\alpha$ ), ili misao trećeg reda. Dakle, HOT-ova koje su inače neopaženi jer su nesvjesni – postaju svjesne misli tek kada imamo misao trećeg reda, koje HOT ( $\alpha$ ) čini svjesnim. Introspektivna svijest tranzitivna je svijest o HOT ( $\alpha$ ).

HOT-ovi koje čine stanje svjesnim nisu sami nužno svjesne misli jer da jesu – tada bi tranzitivno stanje svijesti ujedno bilo intranzitivno, te bismo tada imali cirkularno objašnjenje. Budući da HOT-ovi nisu svjesne misli, često ne znamo da postoje jer su transparentni prema svojem predmetu  $\alpha$ . Kada bismo kušača vina zapitali kakva je kvaliteta doživljaja vina, on bi rekao da je vino crvene boje, takvog i takvog mirisa itd. Kada obratimo pažnju na to kakvo je iskustvo, naći ćemo samo to kakva je stvar o kojoj imamo iskustvo. To je tzv. *teza o transparentnosti iskustva*.

Transparentnost HOT-ova objašnjava transparentnost iskustva. Iskustvo je transparentno zato što nismo svjesni HOT-ova, a upravo HOT reprezentira *crvenost* kao »iskustvenu« crvenost. Tada je prirodna posljedica da smo svjesni *crvenosti* kao svojstva vina, a ne kao svojstva iskustva. Nadalje, Rosenthal nadalje tvrdi da je svijest toliko problematična za objasniti upravo zbog te transparentnosti HOT-a (Rosenthal 1993). Iz tog nam se razloga čini da je svjestan doživljaj *intrinzično svojstvo* mentalnih stanja, tj. da nema strukturu. Kada nas boli, onda nismo svjesni boli kao relacije tog stanja i misli o boli ( $\alpha$  i  $T(\alpha)$ ). Dakle, svojstva HOT-a objašnjavaju transparentnost iskustva, prividnu intrinzičnost svijesti i njenu problematičnost. U sljedećem poglavlju nastavljam s problemima koje Rosenthalova teorija otvara.

### 3. Problemi s Rosenthalovom teorijom i HOT-zombi argumenti

#### 3.1. Problem pogrešne reprezentacije, placebo i sadržaja HOT-a

Rosenthalova teorija dozvoljava situacije u kojima HOT-ovi mogu pogrešno reprezentirati mentalna stanja:

»HOT model dozvoljava mogućnost da se HOT-ovi događaju bez pratećeg stanja nižeg reda kao nominalnog objekta, ali takav događaj može rezultirati ničim, što se svjesnosti tiče.« (Rosenthal 2005: 237)

Mogućnost HOT-a bez korespondirajućeg stanja prvog reda prigovor je koji je već bio razmatram u različitim formama (npr. Byrne 1997; Levine 2001). Poznati primjer fenomena »lažne« boli dolazi iz zubne ordinacije. Naime, ponekad je moguće kada nam zubar buši zub da osjećamo bol iako smo pod anestezijom. Premda nas ne može boljeti (govorimo o mentalnom stanju prvog reda), mi svjesno osjećamo bol. Možda se radi o tome da nas je strah bušenja zuba, ali taj strah praćen je HOT-om čiji je sadržaj *boli me zub*, ili naše snažno uvjerenje o boli može izazvati samu bol. Poznati su nam učinci placebo i samosugestije.

U normalnim okolnostima strah je praćen HOT-om sa sadržajem *ja sam u strahu*. Naime, nesvjesno mentalno stanje straha daje sadržaj HOT-u. Ali što se onda događa kod zubara? S obzirom na to da je ono *kako izgleda biti* u mentalnom stanju određeno sadržajem HOT-a,<sup>6</sup> tada možemo imati svjestan doživljaj boli iako nas je samo strah (u smislu stanja nižeg reda). Boli kao nesvjesnog stanja prvog reda zapravo nema, već samo HOT o boli. Međutim, po Rosenthalovu mišljenju, to je dovoljan uvjet za svjesni doživljaj boli (Rosenthal 2005).

Mislim da se problem pogrešne reprezentacije svodi na pitanje: što u slučaju »lažne« boli daje sadržaj HOT-u? Budući da bi HOT u »normalnim okolnostima« kod stvarne boli referirao na nesvjesnu bol, pa i od tuda i »crpio« sadržaj, misteriozno je na koji način HOT dobiva sadržaj u slučaju misreprezentacije? Dakle, prigovor se svodi na pitanje: kako misao kao takva može biti zaslužna za ono *kako izgleda biti* u svjesnom stanju? Čak i ako je veza između straha i HOT-a nečim ometena, odnosno u slučaju da postoji i kauzalno objašnjenje, nije jasno kako točno HOT dobiva kvalitativni sadržaj.<sup>7</sup> Sjetimo se da reprezentacionalistička teorija mora moći objasniti fenomenalnu svijest u terminima intencionalne strukture. Ova je situacija misteriozna, te sukladno tome ostavlja prostora za varijantu zombi argumenta (Balog 2000). U sljedećem dijelu pokušat ću predložiti dvije verzije dotičnog argumenta koji otkrivaju zanimljive implikacije za Rosenthalovu teoriju.

### 3.2. *HOT-zombi argumenti*

U povijesti filozofije uma zombi argument odigrao je značajnu ulogu, stoga ću ukratko izložiti osnovnu ideju. Možemo zamisliti nekoliko vrsta zombija, štoviše mnogo je filozofa formuliralo različite verzije ovog argumenta (Levine 2001; Nagel 1974; Stoljar 2001), ali Chalmersova je, možemo slobodno reći, najpopularnija (Chalmers 1996).

Zombi je u filozofskoj terminologiji biće koje se ponaša i izgleda poput čovjeka, ali nema svijest. To su bića koja imaju samo moždana stanja, nemaju nikakva mentalna stanja, a ponašaju se kao da ih imaju. Njihova neuralna stanja na isti su način uzrokovana izvana kao i čovjekova, pri čemu imaju iste posljedice na ponašanje, npr. verbalnu reakciju. Kada bi udarili zombija, on bi rekao *boli me* – iako nema apsolutno nikakvih mentalnih stanja. Ovakav argument koristi se protiv tzv. *tipskog fizikalizma*. Druga varijanta misaonog eksperimenta od nas zahtijeva da zamislimo zombija koji, uz moždana, ima samo nesvjesna mentalna stanja, dok mu se ponašanje ne razlikuje od osobe u svjesnom stanju. Možemo zamisliti da taj zombi više *boli me*, a da za njega *nikako ne izgleda* biti u stanju boli. Dakle, njegovo stanje ima iste kauzalne uloge kao kod normalnih ljudi, ali nema svijest. Ovakva verzija zombi argumenta usmjerena je protiv tzv. *funkcionalizma*.

U oba slučaja, zombijeva i odgovarajuća čovjekova stanja ili imaju iste kauzalne uloge, ili/isto moždana stanja, iako zombi nema svijest. Dakle, izvodi se zaključak da se ono što razlikuje svijest od ne-svijesti ne sastoji niti u kauzalnim vezama, niti u neurološkim stanjima. Time što smo odredili kauzalne relacije tipa mentalnog stanja, nismo iscrpili sve ono što čini mentalno stanje svjesnim. Zaključci zombi argumenta također variraju, ali se slažu u jednom – naime, čini se da funkcionalističkoj i fizikalističkoj analizi izmiče ono najbitnije, a to je fenomenalni karakter iskustva. Joseph Levine tako tvrdi da upravo zombi argument zapečaćuje eksplanatorni jaz kao epistemološki problem (Levine 2001: 10).

Neki fizikalisti (npr. Balog 2012; Loar 1990) tvrdili su da je zombi logički moguć, ali ne i metafizički, pa da stoga ne ugrožava fizikalističku teoriju. S druge strane, Peter Carruthers tvrdio je da to što možemo shvatiti i zamisliti

6

Rosenthalova pozicija može se iščitati indirektno iz: Rosenthal 2005: 238. Riječ je o mjestu gdje autor tvrdi da bi zombi s HOT-om, ali bez nižeg stanja, ipak imao svijest.

7

Balog (2000) iznosi sličan prigovor s drugačijim zaključkom, tvrdeći da je HOT irelevantan za kvalitativni aspekt svijest.

zombije bez kontradikcije ukazuje na mogućnost da oni postoje (Carruthers 2000). Međutim, zamislivost i nezamislivost govore nešto o našim pojmovima stvari, a ne o samim stvarima u vanjskom svijetu. Mogućnost zamišljanja neće nas dovesti do toga kakve stvari jesu.

Što je s Rosenthalovom teorijom, je li HOT-teorija ranjiva istim argumentom koji predstavlja problem za fizikalizam i funkcionalizam, tj. za teorije koje nastoje direktno reducirati svijest na ne-mentalno? Katalin Balog sugerirala je da HOT-teorija nije imuna na takvu vrstu prigovora jer da ipak možemo zamisliti biće sa svjesnim unutarnjim životom koje se ne razlikuje od našeg, iako ono ne bi imalo senzorna stanja (Balog 2000: 218). Iz toga što Rosenthal dozvoljava mogućnost pogrešne reprezentacija – tj. situacija u kojima postoji HOT bez korespondirajućeg nesusjesnog stanja nižeg reda – Balog smatra da je HOT-zombi moguć (Balog 2000). Odgovarajući na njezinu sugestiju o HOT-zombiju, Rosenthal se slaže da bi HOT-zombi, u smislu bića koje nema senzorna stanja nižeg reda, a ima subjektivno nerazlučivu svjesnost *kao da ima*, uistinu bio paradoksalan – zbog toga što biće (na mentalnoj razini višeg reda) identično nama u stvari ne bi imalo nesusjesna senzorna stanja, tj. stanja nižeg reda (Rosenthal 2000: 238). Zombi bi tada imao samo HOT bez odgovarajućeg stanja nižeg reda, kao i u slučaju konfabuliranih stanja, placebo i samosugestije.

Uvjetno se slažem s njegovim stavom, ali pritom tvrdim da nije isključio verziju HOT-zombi argumenta koja se otvara zbog mističnosti HOT-a koji čini da kvalitativni aspekt mentalnog stanja naprosto »oživi«. S obzirom na sve što znamo o HOT-mehanizmima, a očito je da na koncu ne znamo točno kako svijest nastaje, moguće je (u epistemičkom smislu) da se svijest ne dogodi, unatoč normalno funkcionirajućem HOT-sustavu. Dakle, ova zombi hipoteza dovoljna je da se demonstrira jaz u objašnjenju kao epistemološki problem. Stoga, pokušat ću formulirati ovu verziju HOT-zombi argumenta i ukazati na daljnji problem s Rosenthalovom teorijom.

Ukoliko Rosenthalova teorija nije zatvorila eksplanatorni jaz, utoliko će biti smisljeno postaviti koherentno pitanje: zašto mentalno stanje praćeno HOT-om dovodi do svjesnog doživljaja, tj. fenomenalne svijesti? Kako je točno jedna misao kadra to učiniti? Rosenthal uvjerljivo argumentira da pojmovi koji figuriraju u HOT-u doista određuju ili oblikuju svjestan doživljaj (Rosenthal 1993, 2005). Međutim, čini se da Rosenthal nema zadovoljavajući odgovor na pitanje o tome na koji točno način HOT može *nesvjesno* stanje učiniti *svjesnim*, već samo kaže da HOT čini da *qualia* (kvalitativni aspekt iskustva) naprosto »oživi«. <sup>8</sup> S obzirom na to da ne postoji točno objašnjenje kako se taj ponešto magični trenutak događa, možemo zamisliti da se naprosto i ne dogodi. Kao što u nekim slučajevima asertorička misao o boli može u principu samosugestijom izazvati bol, a nekada ne. Ovaj nedostatak u objašnjenju može se iskoristiti za sljedeću formulaciju HOT-zombi argumenta:

- (1) ako je Rosenthalova teorija zatvorila eksplanatorni jaz između fenomenalne svijesti i nesusjesne mentalnosti, onda nije zamislivo da jedno biće posjeduje nesusjesna mentalna stanja prvog reda popraćena adekvatnim HOT-ovima, a da nema fenomenalnu svijest;
- (2) ono što je zamislivo, to je konceptualno i epistemološki moguće. <sup>9</sup>
- (3) moguće je zamisliti zombija, biće koje ima nesusjesna stanja prvog reda i HOT-ove koji na adekvatan način prate stanja nižeg reda, a da biće nema fenomenalnu svijest;
- (4) ako je HOT-zombi zamisliv, onda je i moguć (što slijedi iz 1, 2, 3);

(5) dakle, Rosenthal nije zatvorio eksplanatorni jaz između fenomenalne svijesti i nesvjesnog (što slijedi iz 1, 4).

Nadalje, tvrdim da je moguće na osnovu Balogine sugestije konstruirati još jednu verziju HOT-zombi argumenta koji ukazuje na dodatni problem Rosenthalove teorije (Balog 2000: 218), nazovimo ga »HOT-zombi u bačvi«. Zamislimo biće koje je kroz cijeli život uvijek imalo samo HOT-ove, a nikada nesvjesna senzorna stanja nižeg reda kao objekte HOT-ova. Ta se hipoteza zasniva na sljedećoj Rosenthalovoj ideji:

»... model otvara spekulaciju da bi HOT bez pratnje stanja nižeg reda mogao biti za subjekta introspektivno nerazlučiv od HOT-a s adekvatnim pratećim stanjem nižeg reda kao objektom.« (Rosenthal 2000: 237)

U tom bi slučaju sav iskustveni sadržaj zombija dolazio od misli, a ne od iskustva, a što je po prethodnim Rosenthalovim postavkama teoretski moguće. Zamislimo, dakle, zombija čiji se cijeli život svodi na konfabulirane misli i doživljaje, tj. na jedan veliki produkt uma. Ovaj HOT-zombi bio bi nešto slično žrtvi skeptičkog scenarija »mozga u bačvi«. Po Rosenthalovu mišljenju, njegova bi cjelokupna »lažna«, ali ipak svjesna iskustva bila subjektivno nerazlučiva od stvarnih svjesnih iskustava. Pitanje je, dakle, koja bi uopće bila razlika između svijesti zombija u bačvi i normalnog svjesnog bića? Subjektivno ne bi bilo nikakve razlike, već bi razlika bila u zombijevoj »praznoći« na razini mentalnih stanja nižeg reda. Međutim, treba primijetiti da bi u slučaju svijesti »zombija u bačvi« ipak još nešto nedostajalo, a to je kontakt s realnošću. Dakle, radi se o situaciji da bi zombi samo *površno* imao svijest kao i »normalan« čovjek, u smislu sadržajno identičnih HOT-ova, iako zombijeva svijest ne bi bila ista. Prema tome, ovdje slijedi tipičan zaključak zombi argumenta – svjesnost se ne može reducirati na HOT o mentalnom stanju. Rosenthalovoj teoriji svijesti ipak izmiče nešto esencijalno – a to je da svijest mora biti karakterizirana nečim izvan sebe, kao što tranzitivna definicija (u jednoj užoj, realističnoj interpretaciji) i predviđa.

Zaključno, smatram da je Rosenthal prebrzo isključio probleme koji slijede iz HOT-zombi hipoteza, ali i implikacije koje iz takvog argumenta proizlaze. Nekakav eksplanatorni jaz u njegovoj teoriji još uvijek postoji, tipičan za naturalističke teorije.

## Zaključak

Početna pretpostavka pri formuliranju problema jaza u objašnjenju bila je ideja da je neshvatljivo kako subjektivnost i kvalitativnost senzornih stanja objasniti kao svojstva fizičkih/funkcionalnih stanja. HOT-teorija, s obzirom na to da ne reducira svijest direktno na fizička/funkcionalna stanja, obećavala je da može dati plauzibilno objašnjenje svojstva svijesti u terminima mentalnog, ali zapravo nesvjesnog. Je li HOT-teorija uspjela ispuniti obećanje i premostiti eksplanatorni jaz?

Ova teorija u velikoj mjeri objašnjava određena svojstva svijesti: *subjektivnost* i *kvalitativnost*. Činjenica da mi iskustvo *nekako izgleda*, odnosno da se subjektivnost sastoji u tome da mi HOT reprezentira mentalno stanje. S obzirom na to da je HOT-misao koja ima sadržaj, možemo reći da je sadržaj višeg

reda odgovoran za subjektivnost svjesnih stanja. Bol ima subjektivni aspekt samo ako imam misao da me boli. Svijest o tome da mene boli objašnjava subjektivnost boli, tj. to da bol doživljam kao svoju.

Drugo, kvalitativni aspekt svjesnih stanja (engl. *qualia*) objašnjen je razlikom u sadržaju HOT-a. Ono kako izgleda *crvenost* vina određeno je sadržajem HOT-a (npr. *ja vidim crveno vino*). Da je viđenje vina reprezentirano u sadržaju HOT-a kao *vidim zeleno vino*, onda bi doživljaj vina bio zelenast. Dakle, nesvjesni HOT o mentalnom stanju objašnjava zašto mi moje mentalno stanje uopće *nekako izgleda* – i povrh toga, objašnjava zašto mi *tako izgleda*.

HOT-teorija objašnjava svijest u terminima intencionalnih, nesvjesnih stanja (nižeg i višeg reda) i tako izbjegava problem cirkularnosti koja slijedi iz pretpostavke da su sva mentalna stanja svjesna. Ovdje je učinjen prvi korak u redukciji. Sljedeći korak sastoji se u objašnjenju nesvjesnih stanja u terminima fizičkih.

Nesvjesna stanja imaju sadržaj koji se može objasniti pozivanjem na kauzalnu vezu s vanjskim svijetom. Sadržaj mentalnih stanja može biti objašnjen funkcionalistički i/ili fizikalistički. Drugom se linijom argumentacije može pokušati pokazati da sadržaj figurira u psihološkim zakonitostima. S obzirom na to da su psihološki zakoni prirodni zakoni, sadržaj će po definiciji biti prirodni entitet. Na taj bi se način moglo sačuvati naturalističko objašnjenje.

Produživanje reduktivnog lanca *svjesno* → *nesvjesno* → *neuralna stanja/funkcije* ublažio bi se jaz u objašnjenju. Pravac u kojem treba tražiti rješenje problema svijesti, po Rosenthalovu mišljenju, svodi se na to kako objasniti svjesna stanja u terminima nesvjesnih. Na tu je liniju ukazao distinkcijom *tranzitivno/intranzitivno*, dok se sâm mehanizam svjesnih stanja sastoji u relaciji između dvaju nesvjesnih stanja  $\alpha$  i HOT ( $\alpha$ ). S obzirom na to da svijest ima strukturu, pitanje *kako je svijest realizirana u neuralnim stanjima ili funkcijama* više ne bi trebalo izazivati čuđenje o tome kako je moguće da je svijest na kraju krajeva fizičko stanje.

Kroz detaljnu analizu važnih aspekata Rosenthalove teorije ukazala sam na nekoliko problema i nedostataka u njegovom indirektno reduktivnom pristupu. Smatram da HOT-teorija ostavlja neka otvorena pitanja – primjerice, problem pogrešne reprezentacije, tj. što u tim slučajevima objašnjava sadržaj misli višeg reda? Nadalje, u kojoj mjeri je Rosenthalova teorija uspješna po pitanju premošćivanja eksplanatornog jaza? U ovom tekstu formulirala sam HOT-zombi argumente s ciljem isticanja da Rosenthal nema zadovoljavajući odgovor na pitanje o tome kako adekvatno formirana misao o mentalnom stanju može izazvati svjestan doživljaj? Ideja da se svijest naprosto dogodi (ili nekada čak i ne dogodi iako postoji HOT) – tj. da magično oživi – naprosto ostavlja dojam da se nismo pomaknuli od onog pitanja kako određeno moždano stanje može izazvati subjektivni doživljaj *zelenog*, a koji nije nigdje u mozgu nego u umu, pa se sukladno tome ni ne može kvantitativno opisati. Prvim HOT-zombi argumentom ukazala sam na mogućnost da možemo zamisliti biće koje ima HOT-sustav kao i naš, ali da nema svijest. Time sam uspostavila epistemičku mogućnost HOT-zombija koji ukazuje na eksplanatorni jaz kao epistemički problem. Druga varijanta HOT-zombi argumenta odnosi se na zamislivost HOT-zombija koji bi imao samo konfabulirana svjesna iskustva – dakle, imao bi svijest, ali ne takvu kao našu svijest koja je svojom dubinom iscrpljuje i premašuje naše pojmove jer je »usidrena« u stvarnost. Dakle, spomenuti »HOT zombi u bačvi« argument ukazuje na problem da Rosenthalova teorija ipak ostavlja eksplanatorni (i konceptualni jaz) između svijesti i nesvjesnog jer negdje na koncu priče izostavlja esencijalni aspekt svijesti kao »prozora u svijet«.

S obzirom na ove zaključke HOT-zombi argumenata konačan prigovor glasi: ako Rosenthalova teorija nije uspjela zatvoriti eksplanatorni jaz između fenomenalne svijesti i nesvjesnog, onda se mora postaviti pitanje: na koji je način HOT-teorija bolja od drugih naturalističkih teorija, npr. fizikalizma koji direktno reducira svijest na nešto ne-mentalno, ili pak jednostavnijeg reprezentacionalizma? *Prima facie*, jednostavnost teorije uzima se kao eksplanatorna prednost, pa stoga izgleda da Rosenthalova kompleksna teorija ponešto gubi na plauzibilnosti. Iako HOT-teorija dobro objašnjava određena svojstva svijesti, ipak ostaje dojam da još nismo došli do konačnog i uvjerljivog odgovora na pitanje o prirodi svijesti.

## Literatura

- Armstrong, David Malet (1968): *A Materialist Theory of the Mind*, Routledge, London.
- Armstrong, David Malet; Malcom, Norman (ur., 1984): *Consciousness and Causality. A Debate on the Nature of Mind*, Blackwell, Oxford.
- Balog, Katalin (2000): »Comments on David Rosenthal's 'Consciousness, Content, and Metacognitive Judgments'«, *Consciousness and Cognition* 9 (2000), str. 215–219, doi: <http://dx.doi.org/10.1006/ccog.2000.0438>.
- Balog, Katalin (2012): »In Defense of the Phenomenal Concept Strategy«, *Philosophy and Phenomenological Research* 84 (2012) 1, str. 1–23, doi: <https://doi.org/10.1111/j.1933-1592.2011.00541.x>.
- Berić, Boran (1994): *Kognitivizam i svijest*, Hrvatski Kulturni Dom, Rijeka.
- Block, Ned (1995): »On a Confusion about a Function of Consciousness«, *Biheavioral And Brain Sciences* 18 (1995) 2, str. 227–287, doi: <https://doi.org/10.1017/s0140525x00038188>.
- Block, Ned (1990): »Inverted Earth«, *Philosophical Perspectives* 4 (1990), str. 53–79, doi: <https://doi.org/10.2307/2214187>.
- Byrne, Alex (1997): »Some like it hot: Consciousness and Higher-Order Thoughts«, *Philosophical Studies* 86 (1997) 2, str. 103–129.
- Carruthers, Peter (2000): *Phenomenal Consciousness. A Naturalistic Theory*, Cambridge University Press, Cambridge.
- Chalmers, David J. (1995): »Facing up to the Problem of Consciousness«, *Journal of Consciousness Studies* 3 (1995) 2, str. 200–219.
- Chalmers, David J. (1996): *The Conscious Mind. In Search of a Fundamental Theory*, Oxford University Press, Oxford.
- Crane, Tim (2001): *Elements of Mind*, Oxford University Press, Oxford.
- Horgan, Terence (1984): »Functionalism, Qualia, and the Inverted Spectrum«, *Philosophy and Phenomenological Research* 44 (1984) 4, str. 453–470, doi: <https://doi.org/10.2307/2107613>.
- Levine, Joseph (1983): »Materialism and Qualia: The Explanatory Gap«, *Pacific Philosophical Quarterly* 64 (1983) 4, str. 354–361, doi: <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>.
- Levine, Joseph (2001): *Purple Haze. The Puzzle of Conscious Experience*, Oxford University Press, New York.
- Lycan, William (1987): *Consciousness*, MIT Press, London – Cambridge (MA).
- Loar, Brian (1990): »Phenomenal States«, *Philosophical Perspectives* 4 (1990), str. 81–108, doi: <https://doi.org/10.2307/2214188>.
- Locke, John (1959): *An Essay on Human Understanding*, Dover, New York.
- McGinn, Colin (1991): *The Problem Of Consciousness. Essays towards a Resolution*, Basil Blackwell, Oxford.
- Nagel, Thomas (1974): »What Is It Like to Be a Bat?«, *Philosophical Review* 83 (1974)

- 4, str. 435–450.
- Place, U. T. (1956): »Is Consciousness a Brain Process?«, *British Journal of Psychology* 47 (1956) 1, str. 44–50, doi: <https://doi.org/10.1111/j.2044-8295.1956.tb00560.x>.
- Rosenthal, David M. (1986): »Two Concepts of Consciousness«, *Philosophical Studies* 49 (1989), str. 329–359, doi: <https://doi.org/10.1007/bf00355521>.
- Rosenthal, David M. (1990): »Theory of Consciousness«, *Davidrosenthal.org*. Dostupno na: <https://davidrosenthal.org/DR-A-Theory.pdf> (pristupljeno 29. 5. 2025.).
- Rosenthal, David M. (1997): »A Theory of Consciousness«, u: Ned Block, Owen Flanagan, Güven Güzeldere (ur.), *The Nature Of Consciousness*. Philosophical Debates, MIT Press, Cambridge (MA), str. 729–754.
- Rosenthal, David M. (1990): »The Independence of Consciousness and Sensory Quality«, *Davidrosenthal.org*. Dostupno na: <https://davidrosenthal.org/DR-Independence.pdf> (pristupljeno 29. 5. 2025.).
- Rosenthal, David M. (1989): »Thinking that One Thinks«, *Davidrosenthal.org*. Dostupno na: <https://davidrosenthal.org/DR-TTT.pdf> (pristupljeno 29. 5. 2025.).
- Rosenthal, David M. (1990): »Why Are Verbally Expressed Thoughts Conscious?«, *Davidrosenthal.org*. Dostupno na: <https://davidrosenthal.org/DR-Why.pdf> (pristupljeno 29. 5. 2025.).
- Rosenthal, David M. (1993): »State Consciousness and Transitive Consciousness«, *Consciousness and Cognition* 2 (1993) 4, str. 355–363, doi: <https://doi.org/10.1006/ccog.1993.1029>.
- Rosenthal, David M. (2000): »Metacognition and Higher Order Thoughts«, *Consciousness and Cognition* 9 (2000) 2, str. 231–242, doi: <https://doi.org/10.1006/ccog.2000.0441>.
- Rosenthal, David M. (2002): »The Higher-Order Model of Consciousness«, u: Rita Carter (ur.), *Consciousness*, Weidenfeld & Nicolson, London, str. 45–47.
- Rosenthal, David M. (2005): *Consciousness and Mind*, Oxford University Press, Oxford.
- Sesardić, Neven (1984): *Fizikalizam*, Istraživačko Izdavački Centar SSO Srbije, Beograd.
- Smart, John Jamieson Carswell (1959): »Sensations and Brain Processes«, *Philosophical Review* 68 (1959) 2, str. 141–156.
- Stoljar, Daniel (2001): »The Conceivability Argument and Two Conceptions of the Physical«, *Philosophical Perspectives* 15 (2001), str. 393–413.
- Shoemaker, Sydney (1982): »The Inverted Spectrum«, *Journal of Philosophy* 79 (1982) 7, str. 357–381.

**Ksenija Puskarić**

### **Rosenthal's Theory of Consciousness and the Zombie Arguments**

#### **Abstract**

*The nature of consciousness and its subjective and qualitative character present themselves as a hard problem of consciousness for any naturalistic theory that tries to explain this phenomenon in terms of the scientific reductive model. In this paper, I deal with Rosenthal's theory of consciousness, which is essentially representationalist – i.e., consciousness consists in the form of higher-order thoughts about mental states. The focus of the discussion is on the question of whether this theory can offer an answer to the problem of the explanatory gap, and accordingly, I defend the thesis that Rosenthal's indirect-reductive approach is ultimately unsuccessful. On that note, I formulate two versions of the zombie argument, in which I propose that: a) we can imagine a being with an identical system of mental representations to ours, but without consciousness, and that b) we can imagine a being with higher-order mental representations like ours, who has never had contact with reality, whose total sum of conscious experiences would be confabulated. Both versions of the zombie argument point to epistemic problems in Rosenthal's theory.*

#### **Keywords**

David M. Rosenthal, consciousness, representationalism, naturalism, zombie argument, explanatory gap, philosophy of mind, intentionality