

Kako čitamo tuđe umove: odnos teorije teorije i simulacijske teorije

ANNA KOCSIS

*Diplomski studij filozofije
Hrvatski studiji Sveučilišta u Zagrebu
Borongajska cesta 83d, 10000 Zagreb
E-mail: anna.kocsis1@gmail.com*

SAŽETAK: Suvremeno istraživanje teorije uma obilježeno je dvama dominantnim pristupima. Teorija teorije (TT) tvrdi kako je čitanje tuđih umova proces zaključivanja s obzirom na prethodno formulisane generalizacije nalik znanstvenim hipotezama. S druge strane, zastupnici simulacijske teorije (ST) zastupaju mišljenje kako se teorija uma oslanja na svojevršno zrcaljenje ili simuliranje tuđih mentalnih stanja. Simuliranjem mentalnih stanja drugih se postavljamo u njihovu poziciju, tj. u njihove „mentalne cipele“. Pretpostavka o međusobnoj isključivosti TT-a i ST-a je u velikoj mjeri usmjeravala empirijska istraživanja, no pokazala se neodrživom. Odbacivanjem pretpostavke o međusobnoj isključivosti TT-a i ST-a, pažnja je usmjerena na odnos teorijskih i simulacijskih procesa pri teoriji uma. Za vrstu procesa koji će se odvijati, od velike su važnosti kontekst ili situacija u kojoj se događa proces te vrsta pripisivanog stanja. Uvriježena je podjela TT-a i ST-a s obzirom na pripisivanje kognitivnih i afektivnih mentalnih stanja. Moguće je tvrditi kako je afektivno mentalno stanje subjekta teorije uma, koje zanemaruju i TT i ST, ključan faktor u određenju vrste procesa koji će se odvijati.

KLJUČNE RIJEČI: Afektivno mentalno stanje, kognitivno mentalno stanje, kontekst, simulacijska teorija (ST), teorija teorije (TT), zrcalni neuroni.

Uvod

Društveni život je u velikoj mjeri ovisan o našoj sposobnosti pripisivanja mentalnih stanja pojedincima koji nas okružuju. Zaključivanje o tuđim mentalnim stanjima omogućuje usklađivanje vlastitog ponašanja te manipuliranje tuđim osjećajima, mišljenjima i ponašanjem. Takvo svakodnevno razumijevanje tuđih umova se naziva *pučka psihologija*. Filozofski interes za pučku psihologiju je motiviran dvama pitanjima: i) Koji je status pučke psihologije s obzirom na neku buduću znanstvenu teoriju o odnosu uma i tijela? te ii) Koji nam procesi omogućuju opisivanje, objašnjavanje i predviđanje ponašanja drugih pojedinaca? Najčešće se alternativna objašnjenja tih procesa navode pod zajedničkim nazivom *teorija uma*.¹

Pitanje statusa pučke psihologije je u središtu rasprave između funkcionalista i eliminativista. Kako navode Nichols i Stich (2003: 7): „Temeljna je tvrdnja [...] kako je pučka psihologija empirijska teorija koja treba objasniti ‘pravilnost između podražaja i odgovora na njih’ koja obilježava ponašanje ljudi“. Funkcionalisti tvrde kako je značenje termina za mentalna stanja (*mental state terms*) analogno onom teorijskih entiteta koji smisao dobivaju svojim kauzalnim ulogama, međusobnim odnosima te odnosima prema opažljivim entitetima. Prema tome, značenje termina za mentalna stanja je određeno funkcionalnim definicijama koje se temelje na pučkoj psihologiji. S druge strane, eliminativisti tvrde kako je pogrešno značenje mentalnih termina temeljiti na pučkoj psihologiji jer je ona netočna teorija koju je potrebno odbaciti. Utjecajan eliminativist Paul Churchland u svojim

¹ Mnogi autori umjesto termina „teorija uma“ odabiru koristiti neutralnije temine kao što su čitanje uma (*mind-reading*), mentaliziranje (*mentalizing*) ili zauzimanje perspektive (*perspective taking*). S obzirom da je glavni predmet spora u području odnos između takozvane teorije teorije (TT) i simulacijske teorije (ST), korištenje navedenih termina izbjegava terminološku pristranost „teorije uma“ prema teorijskom taboru u sporu. No u ovom radu se koristi termin „teorija uma“ jer je uvriježeniji od ostalih inačica te prikladniji u određenim kontekstima.

radovima pučku psihologiju često uspoređuje s flogistonskom teorijom te tvrdi kako teorija o odnosu uma i tijela mora biti temeljena na dostignućima neuroznanosti, a ne na našem svakodnevnom razumijevanju ljudskog ponašanja.

Rasprava između eliminativista i funkcionalista je utjecala na odgovor na pitanje o prirodi procesa koji omogućuju primjenu pučke psihologije. Prema Goldmanu (2012), ono oko čega će se funkcionalist i eliminativist složiti je karakterizacija pučke psihologije kao teorije koja nalikuje znanstvenim teorijama i temelji se na određenim generalizacijama sličnima prirodnim zakonima. Stich i Nichols (2011: 304) ovo stajalište o pučkoj psihologiji opisuju na sljedeći način: pučka psihologija je „baza bogata informacijama (*rich body of information*) o umu“ kojom se koriste mentalni mehanizmi odgovorni za teoriju uma. Pod utjecajem takvog shvaćanja, unutar proučavanja teorije uma razvio se takozvani *teorija teorije* (TT) pristup koji je dugo dominirao područjem. Tek krajem 20. stoljeća u radovima Gordona i Goldmana pojavljuje se *simulacijska teorija* (ST) koja se u znatnoj mjeri oslanja na empirijska otkrića iz kognitivne neuroznanosti. ST tvrdi kako je temeljna pretpostavka funkcionalizma i eliminativizma netočna: pučka psihologija nije teorija te je uopće govor o njezinom postojanju u tom obliku zabluda. Psihologiziranje ili mentalno uživljavanje nije zaključivanje iz prethodnog iskustva, već poistovjećivanje s osobom u pitanju putem procesa simulacije.

U prva se dva poglavlja ovog rada prikazuje teorija teorije (TT) i simulacijska teorija (ST). Treće poglavlje tematizira utjecaj koji je na pitanje prirode procesa psihologiziranja izvršilo otkriće zrcalnih neurona. Potom se iznose argumenti protiv pretpostavljene međusobne isključivosti TT-a i ST-a te se pokazuje kako prihvaćanjem tih argumenata pravi predmet rasprave postaje rasvjetljavanje dosega i međusobnog odnosa tih teorija unutar hibridnog pristupa. Sljedeća dva poglavlja propituju utjecaj konteksta i

vrste pripisivanog stanja na procese teorije uma. Posljednje poglavlje iznosi viđenje o ulozi afektivnog stanja subjekta pri procesima teorije uma.

1. Teorija teorije (TT)

Teorija teorije je najšire prihvaćen i do sredine 1980-ih godina jedini pokušaj razumijevanja mehanizama psihologiziranja. TT tvrdi kako je razumijevanje tuđih mentalnih stanja niz uspješnih generalizacija na temelju prethodno opaženih ponašanja i iskustava. Goldman (2005b: 80) opisuje TT na sljedeći način:

subjekt pripisivanja mentalnih stanja je kao znanstvenik koji pristupa tuđim umovima – i svom vlastitom – na jednak način kao i bilo kojem sustavu. Formira vjerovanja, ili možda zaključke o vjerojatnosti, o trenutnom stanju sustava objekta. Iz svoje baze znanja „izvlači“ neke psihološke generalizacije i koristi ih kako bi zaključio na buduća ili prijašnja stanja sustava.

Prema TT-u, pučka psihologija je skup *ceteris paribus* generalizacija kojima se koristimo pri zaključivanju o sadržaju tuđih umova. Formiranje tih generalizacija nalikuje formiranju znanstvenih hipoteza. Dakle, teorija uma počiva na tvrdnji kako pojedinci posjeduju bazu informacija o ljudskim umovima. Hutto i Ratcliffe (2007: 7) opisuju znanje koje je potrebno za primjenu pučke psihologije na sljedeći način: „Predviđanje kako će se neko biće ponašati (ili obrnuto, objasniti njegovo ponašanje) zahtijeva reprezentaciju njegovih stanja uma unutar koje će se neki propozicijski stavovi odnositi prema drugima na prikladno strukturiran način.“

S obzirom na porijeklo sposobnosti psihologiziranja, dominiraju dva suprotstavljena gledišta: teorija stečenog modula, poznata i pod nazivom „teorija o djetetu-znanstveniku“ (*child-scientist theory*), i teorija naslijeđenog modula.² Utjecajan argument u korist teorije stečenog modula dolazi

² Nazivi „teorija stečenog modula“ i „teorija naslijeđenog modula“ preuzeti su iz Gallagher (2007).

iz područja razvojne psihologije u vidu testa lažnog uvjerenja (*false belief test*). Tim su testom Wimmer i Perner 1983. godine eksperimentalno pokazali kako djeca u razdoblju između treće i četvrte godine života prolaze bitnu kognitivnu promjenu koja im omogućava pripisivanje lažnih vjerovanja drugim pojedincima.³ Test lažnog uvjerenja se sastoji od predstavljanja djeci sljedeće situacije. U prostoriji se nalazi Sally i ima čokoladu u ruci. Sally stavlja čokoladu na stol i odlazi iz prostorije. Nakon nje u prostoriju ulazi Ann koja sprema čokoladu u hladnjak i nakon toga odlazi iz prostorije. Potom se vraća Sally. Tu se situacija prekida i pred djecu se postavlja problem: Sally želi svoju čokoladu; gdje će je potražiti? Djeca mlađa od (otprilike) tri godine odgovaraju kako će je potražiti u hladnjaku, dok djeca starija od (otprilike) tri godine odgovaraju kako će je potražiti na stolu. Teorija stečenog modula tu promjenu objašnjava tvrdnjom kako djeca te dobi uspjevaju postaviti i uspješno testirati hipotezu o postojanju drugih, od njih nezavisnih umova. Nasuprot tome, teoretičar naslijeđenog modula tvrdit će kako uočena pravilnost u razvoju kognitivnih sposobnosti koje omogućuju psihologiziranje, neovisno o kulturnoj sredini, inteligenciji, izloženosti različitim podražajima itd., navodi na zaključak kako se ne radi o procesima kod kojih glavnu ulogu ima učenje putem iskustva. Tvrdi se kako je riječ o urođenim procesima ili urođenom modulu zaduženom za razumijevanje tuđeg ponašanja. Spor oko porijekla sposobnosti psihologiziranja uglavnom se tiče razvojne psihologije i objašnjenja procesa učenja. U kontekstu ovog rada, moguće je zastupati bilo koje od navedenih gledišta.

³ Test lažnog uvjerenja su kao metodu predložili Daniel Dennett i Gilbert Harman u svojim komentarima na istraživanje Davida Premacka i Guya Woodruffa iz 1978. godine. To je istraživanje bilo prvo koje se bavilo mogućnošću postojanja teorije uma kod primata.

2. Simulacijska teorija (ST)

Do promjene smjera rasprave dolazi uslijed pojave simulacijske teorije (ST), koju prvi iznosi Gordon u članku „Folk Psychology and Simulation“ iz 1986. godine. U osnovi simulacijske teorije je tvrdnja kako se pojedinci koriste procesom *mentalne simulacije* koji uopće ne zahtijeva ili zahtijeva vrlo malo informacija i znanja o tome kako funkcioniraju ljudski umovi. Tijek simulacijskog procesa odvija se u tri koraka koje Goldman (2005: 80-81), najznačajniji suvremeni predstavnik ST-a, objašnjava na sljedeći način:

Prvo, subjekt generira zamišljena mentalna stanja (*pretend states*) u vlastitom umu koja trebaju odgovarati mentalnim stanjima objekta. Drugim riječima, subjekt se pokušava postaviti u „mentalne cipele“ objekta. [...] Potom subjekt „ubacuje“ to mentalno stanje u odgovarajuće mehanizme vlastite psihologije [...] i dopušta tim mehanizmima da operiraju nad zamišljenim stanjima i generiraju novo mentalno stanje ili više takvih stanja. [...] Treće, subjekt pripisuje izlazno stanje objektu kao stanje kroz koje će taj objekt proći ili je već prošao.

Dakle, proces mentalne simulacije nije proces zaključivanja temeljen na znanju pučke psihologije ili na znanju o načinu na koji funkcionira ljudski mehanizam za donošenje odluka, već, kako naglašava Goldman (2006: 20), proces zrcaljenja (*mirroring*) ili oponašanja tuđih umova. Kako su subjekti i sami donositelji odluka te posjeduju kognitivne mehanizme potrebne za izvršavanje tog procesa, oni koriste vlastite procese na zamišljenim mentalnim stanjima sličnima početnoj poziciji objekta. Ovakav pristup počiva na sposobnosti subjekta teorije uma da izvede određene operacije. Prema Mitchellu (2009: 1311), riječ je o sljedećim operacijama:

1. subjekt mora zaključiti postoji li između njega i objekta dovoljna sličnost koja bi potkrijepila pretpostavku da bi se objekt u određenoj situaciji ponašao kao subjekt;
2. subjekt mora biti sposoban koristiti svoje simulacije na način koji mu omogućuje da točno „proizvede“ određeno mentalno stanje;
3. subjekt mora uspješno potisnuti „proizvedeno“ mentalno stanje i pripisati ga objektu.

Premda bi na ovakav opis procesa mentalne simulacije pristali svi zastupnici ST-a, među alternativnim karakterizacijama simulacijskih procesa ipak postoje bitne razlike. Primjerice, Gallagher (2007: 64) razlikuje tri verzije simulacijske teorije: *eksplicitna simulacija* (eST), *radikalna simulacija* (rST) i *implicitna simulacija* (iST). Temeljna razlika među navedenim verzijama procesa odnosi se na razinu svjesnosti subjekta koja je potrebna za njihovo provođenje. Budući da je teško dati točan opis načina na koji se svjesnost koristi unutar ST-a, u nastavku koristim *intencionalnost* kao najbližu odrednicu za ulogu svijesti u simulaciji.⁴

Zastupnici eksplicitne simulacije (primjerice Goldman) i radikalne simulacije (primjerice Gordon) slažu se kako je riječ o intencionalnim procesima pri kojima subjekt pokušava unutar sebe proizvesti stanja identična ili što sličnija stanjima objekta. Razilaze se u pitanju na koji način subjekti to čine: Goldman naglašava manipulaciju vlastitim propozicijskim stavovima dok Gordon tvrdi kako se radi o izravnijem zauzimanju tuđe perspektive. Gallagher (2007: 67-68) razliku između tih dviju verzija ST prikazuje sljedećim primjerom: pretpostavimo da ugledam čovjeka koji skače na cesti ispred moje kuće. Na koji mu način pripisujem mentalna stanja? Prema Goldmanu, upotrebljavam svoju imaginaciju i u svoje mentalne procese „ubacujem“ pretpostavljena ili moguća mentalna stanja te osobe te prihvaćam najplauzibilnije objašnjenje ponašanja. S druge strane, Gordon vjeruje kako se ne radi o manipulaciji propozicijama, već o tome da u svoje mentalne procese „ubacujem“, primjerice, ona perceptivna stanja za koja imam razloga vjerovati da ih čovjek doživljava. Gordonovo objašnjenje postavlja izravniju i snažniju vezu između pozicije subjekta i objekta, a pripisivanje intencionalnosti subjektu smatra nužnim tek pri određenju cilja postavljanja u poziciju objekta, ne i za vrijeme samog trajanja procesa u vlastitoj psiholo-

⁴ Pod intencionalnošću se podrazumijeva usmjerenost subjekta teorije uma na proizvodnju točno određenog mentalnog stanja, onog koje posjeduje objekt pripisivanja.

giji. Drugim riječima, kod Goldmana je riječ o procesu koji je u većoj mjeri sličan TT-u utoliko što zahtijeva usredotočenost subjekta i određen napor u postizanju cilja. Kod Gordona se radi o procesu koji je većim dijelom automatiziran: nakon „ubacivanja“ početnih perceptivnih stanja, proces se odvija automatski, neovisno o stanju subjekta.

Oba određenja procesa mentalne simulacije suočena su s mnogim poteškoćama. Kako naglašava Gallagher (2007), glavni je problem u tome što ni jedna od dviju interpretacija nije fenomenološki uvjerljiva, tj. nije u skladu s našim stvarnim iskustvom.⁵ Čini se kako u našem svakodnevnom zaključivanju na tuđa mentalna stanja ne zauzimamo intencionalnu poziciju kakvu zahtijevaju Gordon, a posebno Goldman. Osim toga, nijedna interpretacija ne rješava problem prve točke ili početnog koraka simulacijskog procesa: naime, ukoliko se radi o „ubacivanju“ vjerojatnih mentalnih stanja u psihološke procese subjekta, ostaje upitno na temelju čega ih subjekt smatra najvjerojatnijima, ili, u slučaju Gordonove interpretacije, kako subjekt odlučuje koje zamišljene perceptivne slike „ubaciti“ u proces. Čini se kako je upravo ta sposobnost donošenja odluke o simuliranim stanjima ono što se pokušava objasniti postuliranjem simulacijskog procesa. Zbog toga je eksplanatorno isprazno jednostavno pretpostaviti isti taj fenomen kao početni korak navedenog procesa.

Jedan mogući način obrane od ovih prigovora je zastupanje inačice hibridne teorije između ST-a i TT-a, a drugi zastupanje implicitne simulacijske teorije (iST). Primjerice, moguće je tvrditi kako su generalizacije koje zastupa TT ono što subjektu omogućuje da neka od zamišljenih mentalnih stanja objekta smatra vjerojatnijima od drugih. Motivacija za zastupanje hibridnih teorija dolazi iz više izvora o kojima će biti riječi u nastavku rada. Teorije implicitne simulacije tvrde kako su za pripisivanje mentalnih stanja zaduženi sub-personalni, neurološki procesi za čije se provođenje ne zahtijeva intencionalnost subjekta. Prednost ove vrste ST-a je u tome što je ona

⁵ Zato Gallagher (2007) taj prigovor naziva argumentom „iz fenomenologije“.

u stanju ponuditi objašnjenje za točnost koju svakodnevno susrećemo kod procesa psihologiziranja.⁶ Implicitne su varijante simulacijske teorije dobile znatnu potporu neuroznanstvenim otkrićima o čemu će detaljnije biti riječi u sljedećem poglavlju.

3. Simulacijska teorija (ST) i zrcalni neuroni

Simulacijska teorija je ključnu empirijsku potporu dobila otkrićem tzv. *zrcalnih neurona* (*mirror neurons*) u području F5 prefrontalnog korteksa makaki majmuna, a potom i potvrdom njihovog postojanja kod ljudi. Pronalaskom zrcalnih neurona, mogućnost simulacije ili oponašanja opaženih stanja prestaje biti isključivo teorijska pretpostavka. Zrcalni neuroni su skupina neurona specifičnih po tome što imaju i senzorne i motoričke karakteristike. Rizzolatti et al.⁷ (1996: 134) navode kako zrcalne neurone od ostalih neurona razlikuje činjenica da se aktiviraju kada majmuni opažaju smisle- nu radnju eksperimentatora, ali i kada izvršavaju tu istu radnju. Smisle- ne radnje uključuju stavljanje ili uzimanje predmeta sa stola, posezanje za hranom, precizno manipuliranje predmetima, držanje predmeta, uzimanje hrane od drugog pojedinca i sl. Autori također navode kako oko 60% zrcal- nih neurona reagira isključivo na jednu specifičnu radnju. Klasični uzorak ponašanja zrcalnih neurona koji uočavaju Rizzolatti et al. (1996: 134) sastoji se od aktivacije neurona kada eksperimentator posegne za hranom, prestan- ka aktivnosti dok se hrana primiče majmunu te ponovne aktivacije kada majmun posegne za hranom. Premda su neuroni koji se aktiviraju putem vrlo specifičnih tipova vizualnog podražaja (npr. nečijeg lica, boja, fotogra- fije itd.) poznati od ranije, Rizzolatti et al. (1996: 136) naglašavaju kako su zrcalni neuroni iz područja F5 jedini poznati neuroni koji se aktiviraju i pri izvršavanju motoričke radnje istovjetne onoj percipiranoj. Osim toga, što je

⁶ Točnost procesa psihologiziranja je prešutna pretpostavka teorije uma. O točnosti tog procesa bit će riječi u četvrtom poglavlju.

⁷ Riječ je o istom istraživačkom timu koji je 1992. godine objavio otkriće zrcalnih neurona.

neuron selektivniji za određenu radnju, to je podudarnost između percipirane i izvedene radnje veća.

U posljednjih je 20 godina uložen znatan napor u razotkrivanje funkcionalne uloge zrcalnih neurona. Pretpostavka koju prihvaćaju njihovi pronalazitelji je kako zrcalni neuroni igraju bitnu ulogu u razumijevanju smislenih motoričkih radnji. Pritom Rizzolatti et al. (1996: 137) naglašavaju: „Ovim terminom (razumijevanje motoričkih događaja) označavamo sposobnost pojedinca da prepozna prisutnost drugog pojedinca koji vrši određenu radnju, sposobnost da razlikuje opaženu radnju od drugih radnji te da koristi ove informacije kako bi prilagodio vlastito ponašanje.“ Dakle, riječ je o prepoznavanju prisutnosti drugog pojedinca temeljenom na njegovim motoričkim radnjama. No neki neuroznanstvenici tvrde kako uloga zrcalnih neurona nije tek prepoznavanje pojedinaca kao djelatnika, već i tumačenje njihovog ponašanja s obzirom na mentalna stanja koja posjeduju. Primjerice Jacob i Jannerod (2005: 21) tvrde kako je „socijalnu kogniciju moguće direktno izvesti iz motoričke kognicije“. Prema tome shvaćanju, zrcalni neuroni predstavljaju zajednički temelj širokog spektra složenih kognitivnih strategija koje stoje u pozadini pučke psihologije. Gallese (2003: 523) o ulozi zrcalnih neurona navodi: „simulacije djelovanja se koriste ne samo kako bi se proizvela i kontrolirala ciljno orijentirana ponašanja, već i kako bi se osiguralo smisleno tumačenje ciljeva i svrha tuđih djelovanja.“ Blakemore i Decenty (2001: 565) pak tvrde:

Simulacijom akcija drugih pojedinaca i njihovim mapiranjem u pohranjene reprezentacije naših vlastitih motoričkih naredbi i njihovih posljedica [...] moguće je procijeniti ona unutrašnja stanja opaženog pojedinca koja se ne mogu očitati direktno iz njegovih kretnji. Ovaj simulacijski sustav također omogućava pristup informacijama na temelju kojih se mogu predviđati buduća djelovanja pojedinaca.

Prema tome, zrcalni neuroni su prirodan saveznik simulacijske teorije jer se čini kako se pomoću njih mogu objasniti mnogi fenomeni teorije uma na način

koji predviđa ST. U skladu s tim Gallese i Goldman (1998) postavljaju hipotezu kako su zrcalni neuroni rudimentarni dio mehanizma za psihologiziranje. Istovjetnost između mentalnog stanja subjekta i objekta, koja se postiže aktiviranjem zrcalnih neurona, prvi je korak u simulacijskom procesu. Time zrcalni neuroni dobivaju ključnu ulogu u pokušaju empirijskog utvrđivanja točnosti simulacijskog procesa⁸ i predstavljaju rješenje za problem intencionalnosti i „ubacivanja“ prikladnih stanja u simulacijski proces koji muči i teoriju eksplisitne simulacije i teoriju radikalne simulacije. Dakle, zrcalni neuroni su empirijska osnova za tvrdnju kako intencionalnost subjekta nije ključna za proces psihologiziranja, te su važan argument u korist implicitne simulacije (iST).

Ovim područjem znanosti o umu je dugo dominirala pretpostavka o međusobnoj isključivosti teorije teorije i simulacijske teorije. U sljedećem se poglavlju iznose razlozi zbog kojih se ta pretpostavka pokazala neodrživom tako da danas većina filozofa i neuroznanstvenika⁹ zastupa neku vrstu hibridne teorije koja prihvaća određene tvrdnje i TT-a i ST-a.

4. Točnost teorije uma i argument iz pogreške

Točnost procesa psihologiziranja je početna pretpostavka pri proučavanju teorije uma. Sustavne greške su iznimke koje mogu ukazivati na prirodu mehanizama koje preispitujemo.¹⁰ No kao što je već prethodno naglašeno,

⁸ Osim toga, zrcalni neuroni se pokazuju kao elegantno rješenje za treću operaciju koju navodi Mitchell (2009) – potiskivanje ili inhibiranje „proizvedenog“ mentalnog stanja – budući da su zrcalni neuroni mentalne reprezentacije motoričkih radnji koje su rijetko izvedene. Te reprezentacije najčešće ostaju potisnute, no moguće ih je po potrebi „prizvati“ te je zbog te karakteristike bilo prijedloga da se zrcalne neurone interpretira kao reprezentacijsku osnovu u procesu učenja motoričkih radnji (primjerice Jeannerod, 1994).

⁹ Neki od najvažnijih zastupnika hibridne teorije su Goldman 2006; Mitchell 2005, 2009; Stich i Nichols 1995 ; Nichols i Stich, 2003; Keysers i Gazzola 2007.

¹⁰ Sustavne pogreške u teoriji uma poznate su iz različitih područja bihevioralnih znanosti. Primjerice, Pronin et al. (2001) provode istraživanje o tzv. iluziji asimetričnog uvida (*illusion of asymmetric insight*) koja se sastoji u tome da pojedinci pretpostavljaju kako oni imaju veći uvid u tuđa mentalna stanja nego što drugi pojedinci imaju u njihova. Heath (1999) izvještava o tzv. pristranosti o vanjskim poticajima (*extrinsic incentives bias*) koja se sastoji

mehanizmi teorije uma moraju ponuditi i objašnjenje za uočenu točnost procesa. Simulacijska teorija, a posebno iST, nema problem s pitanjem točnosti procesa jer se pozivaju na sub-personalne psihološke i neurološke procese. Točnost se pokazuje problematičnom za zastupnike TT-a jer se postavlja pitanje objašnjenja točnosti u slučajevima u kojima se subjekt suočava sa svim novim i nepoznatim okolnostima. Kako navode Stich i Nichols (2011: 311), takva točnost predstavlja problem za TT jer ukoliko pojedinci nisu bili u mogućnosti skupiti neku vrstu dokazne građe za tu specifičnu situaciju, ne bi smjeli biti točni u svojim predviđanjima. Stich i Nichols (2011: 311) iznose sljedeću pretpostavku:

U onim područjima u kojima smo posebno dobri pri predviđanju ili atribuciji mentalnih stanja u nepoznatim situacijama [...] proces teorije uma vjerojatno nije temeljen na mehanizmima koji zahtijevaju mnogo informacija. Ali u slučajevima u kojima smo loši u predviđanju ili atribuciji mentalnih stanja [...] proces teorije uma vjerojatno nije temeljen na simulacijskim mehanizmima.

Navedena tvrdnja je osnova razumijevanja odnosa između ST-a i TT-a i omogućava izvođenje, a potom i testiranje, pretpostavki o prirodi mehanizama koji leže u osnovi psihologiziranja u pojedinačnim slučajevima. Saxe u vrlo utjecajnom članku „Against simulation: The argument from error“ (2005a) iznosi važan prigovor simulacijskoj teoriji temeljen na sustavnim pogreškama pojedinaca za koje vjeruje kako ih nije moguće objasniti u okviru ST-a. Kao ilustraciju sustavne pogreške navodi rezultate istraživanja koje su proveli Kruger i Gilovich 1999. godine na bračnim parovima.

U istraživanju je pred ispitanike, svakog supružnika individualno, postavljen zadatak da ocijene vlastitu odgovornost u pozitivnim i negativnim događajima u svojoj bračnoj zajednici. Nakon toga je svaki supružnik zamo-
ljen da na istoj skali predvidi odgovore svog bračnog partnera. Ispitanici su

u tome da ljudi pripisuju veću važnost ekstrinzičnim poticajima (kao što je novac) nego onim intrinzičnim kod drugih pojedinaca nego kod sebe samih.

davali jednake ocjene pri procjeni vlastite odgovornost u slučaju pozitivnih i u slučaju negativnih događaja. Predviđali su kako će njihovi partneri sebi pripisivati veću odgovornost za pozitivne događaje, a manju za negativne. Nasuprot njima, neutralni su promatrači predviđali kako će ispitanici u znatno izraženijoj mjeri preuzimati veću odgovornosti za pozitivne događaje, a manju za negativne, nego što je to doista bio slučaj.

S obzirom da ST tvrdi kako koristimo iste mehanizme pri proizvodnji i pripisivanju vjerovanja, sustavne pogreške kao što je pripisivanje prevelike racionalnosti opaženoj osobi nisu problematične jer je prevelika tendencija pripisivanja vlastitih vjerovanja drugim pojedincima predvidljiva pogreška simulacijskog procesa. Drugim riječima, ako sebi pripišem vjerovanje koje smatram racionalnim, predvidljiva je pogreška simulacijskog procesa kako ću imati tendenciju isto vjerovanje pripisati i drugim pojedincima. Saxe ističe kako problem za ST nastupa kod pogrešaka koje ukazuju na asimetričnost između subjekta i objekta. Navedeni primjer pokazuje da postoji neka relevantna razlika između pozicije osobe koja predviđa ponašanje svog surpužnika i pozicije nepristranog promatrača. Saxe (2005b: 362) tvrdi kako simulacijska teorija nije u mogućnosti objasniti ovakav uzorak pogreške jer je on rezultat subjektivih uvjerenja o umu objekta, a ne nečega što je subjekt osobno proživio (tj. simulacije). Dakle, procesi koji omogućuju teoriju uma u većoj su mjeri ovisni o mišljenjima i vjerovanjima subjekta nego što to ST može dopustiti. Pod pretpostavkom ST-a isti *input* bi trebao generirati isti *output* bez obzira na subjekt simulacijskog procesa. Pri sustavnim pogreškama, poput ove koju opisuju Kruger i Gilovich, a pod pretpostavkom istog inputa, uzorak pogreške ne smije ukazivati na specifičnost pozicije određene skupine ljudi u odnosu na drugu. Takav uzorak govori u prilog tome kako se teorija uma tih skupina razlikuje zbog pozadinskih vjerovanja koja posjeduju. Prema tome, takve pogreške su argument u prilog TT-a ili neke vrste hibridne teorije.

Pokazalo se kako točnost teorije uma u određenim situacijama predstavlja poteškoću za TT dok je, s druge strane, njezina netočnost problem za ST. Budući da ni ST (ni u jednoj od prethodno spomenutih inačica) ni TT nisu u mogućnosti ponuditi objašnjenje pogrešivosti ili točnosti mehanizama teorije, postavlja se pitanje mogu li one nezavisno jedna od druge objasniti mehanizme psihologiziranja. Prihvatiti navedenu tvrdnju Sticha i Nicholisa znači odbaciti međusobnu isključivost TT-a i ST-a te pristati na neku formu hibridne teorije koja će se koristiti mehanizmima koje predlažu obje navedene teorije. U tom slučaju cilj istraživanja više nije opovrgnuti TT ili ST, već odrediti međudnos tih dviju teorija, tj. utvrditi u kojoj se mjeri i pod kojim okolnostima pri teoriji uma radi o simulacijskim ili o teorijskim mehanizmima. U sljedeća će dva poglavlja biti ukratko prikazan napredak koji je, s obzirom na kontekst i vrstu mentalnog stanja, ostvaren u tom smislu.

5. Teorija uma i kontekst

Tijekom posljednjih nekoliko godina, neuroznanstvenici su pozornost posvetili utjecaju koji kontekstualne razlike imaju na mehanizme teorije uma. Najšire shvaćeno, kontekstom bi se mogle nazvati sve informacije, subjektu eksterne ili interne, o situaciji koje pojedinac koristi pri psihologiziranju. Pretpostavlja se kako je proučavanjem konteksta pri teoriji uma moguće odrediti „granice“ koje dijele simulacijsku teoriju i teoriju teorije. Opseg kontekstualnih podataka koji utječu na rezultate mjerenja sposobnosti teorije uma govorit će o važnosti simulacijskih ili teorijskih mehanizama u pojedinom slučaju. Cilj empirijskih istraživanja je manipulacijom konteksta utvrditi kojim se informacijama i na koji način pojedinac koristi pri razumijevanju tuđeg uma.

TT i ST imaju različita predviđanja s obzirom na ulogu konteksta. TT predviđa povećanu aktivnost u situacijama u kojima postoji određena „ne-

podudarnost“ – ispitanik će ulagati dodatni mentalni napor u pokušaju razumijevanja tuđeg mentalnog stanja. Zastupnicima ST-a su otvorene dvije strategije tumačenja rezultata. Kako navode Saxe i Wexler (2005: 3), „moguće je da će aktivnost dijelova mozga linearno rasti s povećanjem sličnosti između mentalnih stanja ispitanika i djelatnika, što odražava pojačanu *sna-gu* simulacijskog procesa (koji se temelji na sličnosti). No također je moguće i da će aktivnost dijelova mozga linearno rasti s povećanjem različitosti između mentalnih stanja ispitanika i djelatnika, što odražava pojačani *trud* uložen u simulacijski proces zbog promjene većeg broja parametara.“ Prema tome, indikator odvijanja simulacijskih procesa je prisutnost linearne razlike u aktivnosti određenih dijelova mozga s obzirom na promjene parametara konteksta. U nastavku su prikazani neki od najutjecajnijih eksperimenata koji su se bavili pitanjem prirode mehanizama teorije uma s obzirom na varijacije u kontekstu.

Brass et al. (2007) provode istraživanje u kojem testiraju sljedeću hipotezu: ukoliko zrcalni neuroni omogućuju razumijevanje isključivo motoričkih intencija subjekata, tada promjena konteksta pri istoj motoričkoj radnji ne bi smjela utjecati na aktivnost zrcalnih područja mozga. Drugim riječima, ukoliko putem zrcalnih neurona shvaćam isključivo intenciju osobe da dohvati jabuku s drveta, tada se aktivnost zrcalnih područja mozga ne bi trebala mijenjati bez obzira čini li to osoba u botaničkom vrtu (gdje je zabranjeno branje jabuka) ili u svom voćnjaku. S druge strane, ukoliko zrcalni neuroni barem u određenoj mjeri omogućavaju razumijevanje i socijalnih intencija subjekata, tada bi promjena konteksta trebala izazvati barem minimalne promjene u aktivnosti dijelova mozga sa zrcalnim svojstvima.¹¹

¹¹ Razlika između motoričke i socijalne intencije se najjednostavnije može opisati primjerom: pri sklapanju novog poznanstva, u zapadnoj je kulturi uvriježen običaj rukovanja. Motorička intencija je u tom slučaju motorički plan da se ispruži ruka prema drugoj osobi na odgovarajuću udaljenost i na odgovarajući način. Socijalna intencija je utjecati na ponašanje druge osobe na način da ona prihvati rukovanje te pokazati drugoj osobi dobrotamjernost, ravnopravnost i poštovanje. Primjerice, Jacob i Jeannerod (2005: 23) pod socijalnim intencijama

Dakle, pretpostavljeno je kako ćemo morati uložiti manje napora u razumijevanje intencija osobe koja bere jabuku u svom voćnjaku od intencija osobe koja to čini u botaničkom vrtu. Međutim, rezultati nisu pokazali promjenu aktivnosti onih dijelova mozga u kojima su aktivirani zrcalni neuroni s obzirom na navedene promjene u kontekstu. Na temelju dobivenih podataka, Brass et al. (2007) dolaze do zaključka kako je uloga zrcalnih neurona u razumijevanju motoričkih akcija vrlo ograničene prirode i odnosi se samo na „poznate akcije u tipičnom kontekstu“, pri čemu je „cilj akcije visoko prepoznatljiv i nadovezuje se na odgovarajuće motoričke sheme već prisutne u repertoaru promatrača“ (Brass et al., 2007: 2120).

U prilog tvrdnji o ograničenoj ulozi koju zrcalni neuroni imaju u teoriji uma govore i rezultati istraživanja koje su proveli Saxe i Wexler (2005). Njihovo je istraživanje naglasilo ulogu konteksta u socijalnim situacijama. Ispitanicima su ispričane kratke priče od prosječno 80 riječi. Svaka je priča varirala s obzirom na tri parametra. Prvo, kulturni kontekst protagonista priče je bio *poznat* ili *nepoznat* (ispitaniku poznata pozadina je bila ona iz zapadnjačkog civilizacijskog kruga, a nepoznata je uključivala neku od ostalih kulturnih ili geografskih regija). Drugo, protagonist je imao *normalnu* želju ili *nenormalnu* želju (nenormalna želja je bila ona koja iz perspektive ispitanika krši neku poznatu normu, no u skladu je s *nepoznatim* kulturnim kontekstom protagonista). Treće, protagonistu priče je ta želja bila ispunjena ili neispunjena. Nakon što su čuli priču, ispitanici su zamoljeni da odgovore na pitanje kako će se protagonist osjećati s obzirom na ishod priče – s obzirom na ispunjenje ili neispunjenje svoje želje. Ponuđena su dva odgovora: protagonist će se osjećati *pozitivno* (dobro) i *negativno* (loše). Ispitaniku je najpoznatija situacija bila ona koja uključuje protagonista iz poznatog kulturnog konteksta i normalnu želju. Najneobičajenija situacija je uključivala protagonista iz nepoznatog kulturnog konteksta i nenormalnu želju. Premda

podrazumijevaju djelovanje s ciljem utjecanja na ponašanje ili mentalne reprezentacije objekta.

je za uspješno obavljanje zadatka bilo dovoljno usporediti navedeno mentalno stanje protagonista (želju) s ishodom priče (ispunjenjem/neispunjenjem), tj. nije bilo potrebno uzeti u obzir protagonistov kulturni kontekst, polazna pretpostavka eksperimenta je bila kako će ispitanici pokušati integrirati sve dostupne informacije o protagonistu prije odgovora na pitanje.

Podaci dobiveni metodom fMRI nisu ukazivali na karakteristične linearne razlike niti u jednom dijelu mozga. Osim toga, u skladu s pretpostavkom, kulturni kontekst protagonista je igrao ključnu ulogu u ispitaničkovom sudu o protagonistovom mentalnom stanju. Ispitanici su mentalno stanje protagonista pokušavali dokučiti s obzirom na njegov subjektivni sustav vjerovanja. Pokazalo se kako su ispitanici bili skloni čak i odbaciti mentalno stanje protagonista (njegovu želju) ako se radilo o *nenormalnoj* želji u poznatom kulturnom kontekstu. Interpretirajući dobivene rezultate, Saxe i Wexler (2005: 7) zaključuju:

Ovi rezultati nisu u skladu s jednostavnom interpretacijom simulacijske teorije prema kojoj su drugi umovi u osnovi reprezentirani s obzirom na sličnost s umom promatrača. Odgovor na podražaj [...] nije bio linearno povezan sa sličnošću između umova promatrača i protagonista. Naprotiv, [...] činilo se kako odgovor na podražaj govori u prilog procesa konstruiranja koherentnog modela protagonistova uma, bez referiranja na subjektova vlastita mentalna stanja.

Smatram da je navedeni zaključak problematičan iz sljedećih razloga. Prvo, posebnost slučaja *nenormalne* želje i poznate pozadine ukazuje na sljedeće: budući da simulacijska teorija pretpostavlja kako se pri atribuciji mentalnih stanja subjekt pokušava postaviti u situaciju objekta i proizvesti u sebi identična mentalna stanja, slučaj *nenormalne* želje i poznate pozadine mogli bismo tumačiti i u prilog ST-a. Kako nam je pozadina poznata, lako se postavljamo u poziciju objekta u toj konkretnoj pozadini, a zatim želju otpisujemo jer je nekonzistentna s pozadinom. Ova situacija govori upravo o „snazi“ simulacijskog mehanizma; jednom kada se *dobro* postavimo u poznatu pozadinu, otpisat ćemo želje koje su nekonzistentne s pozadinom jer

naprosto ne možemo simulirati njihovo posjedovanje. Dakle, proces simulacije polazi od pozadine, a ne od navedene želje.¹² Osim toga, plauzibilno je tvrditi da do odbacivanja želje uopće ne bi došlo kada bi se radilo o procesima koje pretpostavlja TT. Svaki pojedinac bi zasigurno mogao iznaći barem neko objašnjenje navedene želje te na temelju njega predvidjeti mentalno stanje objekta. Vjerojatnost tog objašnjenja ne mora biti problem pod pretpostavkom da je ono vjerojatnije od svih drugih objašnjenja koji su subjektu na raspolaganju.¹³

Drugi problem sa Saxeovom i Wexlerovom interpretacijom navedenog eksperimenta je da se nepojavljivanje linearnog uzorka aktivacije dijelova mozga (za koje se pretpostavlja da sadrže zrcalne neurone i sudjeluju u procesima simulacije) može smatrati protuargumentom za ST samo u slučaju da imamo vrlo pojednostavljenu predodžbu funkcioniranja zrcalnih dijelova mozga. Drugim riječima, trebali bismo pretpostaviti da ti dijelovi ne mogu biti ni u kakvom odnosu s dijelovima mozga koji nemaju zrcalne sposobnosti. Ova pretpostavka bi podrazumijevala kako će posljedica promjene u parametrima problema koji se postavlja pred ispitanike biti isključivo promjena aktivacije dijelova mozga sa zrcalnim karakteristikama. S obzirom da se o promjeni parametara konteksta i rezultirajuće promjene u aktivnosti mozga zna vrlo malo, takvu je pretpostavku zasad teško empirijski potkrijepiti.

Zaključak koji se može izvesti iz dosada rečenog jest da kontekst igra važniju ulogu kod teorijskih nego kod simulacijskih procesa. Drugim riječima, u situacijama u kojima pri rekonstrukciji tuđih mentalnih stanja tre-

¹² Ovakvo objašnjenje je posebno u skladu s radikalnom i implicitnom simulacijom koje tvrde kako početna stanja simulacijskog procesa nisu vjerojatna mentalna stanja objekta, već pretpostavljeno vizualno iskustvo objekta (u slučaju rST) ili stanja subjekta koja su „okidač“ za odvijanje simulacijskih procesa (u slučaju iST).

¹³ Doduše stvar nije tako jednostavna. Naime, ako vjerojatnost najvjerojatnijeg objašnjenja u skupu objašnjenja koja su nam na raspolaganju ne prelazi određenu granicu ili određeni stupanj, subjekti neće biti skloni prihvatiti to objašnjenje. Na taj bi način TT mogla objasniti odbacivanje objektove želje navedene u eksperimentalnom scenariju.

ba uzeti u obzir kontekstualne čimbenike, opravdano je za očekivati kako će mehanizmi psihologiziranja u većoj mjeri odgovarati objašnjenjima koje nudi TT. S obzirom da je kontekst u većini slučajeva iznimno važan čimbenik uspješnog psihologiziranja, ovi rezultati ostavljaju relativno mali prostor za ST. U sljedećem poglavlju se ispituje uspješnost TT-a i ST-a s obzirom na vrstu pripisivanog mentalnog stanja – još jednog važnog čimbenika uspješnog psihologiziranja.

6. Teorija uma i vrsta mentalnog stanja

Osim uloge konteksta, za procese teorije uma vrlo se važnim pokazala i vrsta mentalnog stanja koje se pripisuje drugim umovima. Premda je pretpostavka o različitosti mehanizama koji su odgovorni za pripisivanje određenih tipova mentalnih stanja opće mjesto u literaturi, posebno filozofskoj, empirijski se pristup problemu dugo odgađao. Tek su se u posljednjih nekoliko godina počele pojavljivati publikacije s rezultatima istraživanja iz tog područja. Tako je, primjerice, pokazano kako se različite regije u mozgu aktiviraju s obzirom na to percipiramo li osobu koja trpi bol (Singer et al., 2004), ostvaruje novčani zgoditak (Mobbs et al., 2009), osjeća gađenje (Wicker et al., 2003) ili strah (Whalen et al., 2001).

Kalbe et al. (2010) su proveli istraživanje o razlikama u teoriji uma s obzirom na vrstu mentalnog stanja koje se pripisuje. Istraživanje se sastojalo od testiranja teorije uma ispitanika u nekoliko navrata: prije TMS-a,¹⁴ deset minuta nakon izloženosti TMS-u i neposredno nakon izloženosti TMS-u. Test se provodio putem tzv. metode čitanja uma u očima (*reading the mind in the eyes*). Ispitanicima je prikazana skica čovječuljka (nazvanog Yoni)

¹⁴ Metoda TMS (*transcranial magnetic stimulation*) je u primjeni posljednjih desetak godina. Radi se o vrlo slabom elektromagnetskom šoku (u ovom slučaju od 1 Hz) koji je moguće vrlo precizno uputiti u željena područja mozga čije se normalno funkcioniranje nakratko poremeti. Usporedbom podataka dobivenih korištenjem te metode s onima dobivenim standardnim metodama može se preciznije zaključivati o funkciji regije mozga koja je bila izložena šoku.

pogleda usmjerenog u određenom pravcu, nakon čega je Yoniju bilo potrebno pripisati i) kognitivno stanje, ii) afektivno stanje¹⁵ i iii) fizičku karakteristiku. Primjerice, s obzirom na pogled Yonija bilo je potrebno nadopuniti rečenice kao što su *Yoni misli na ...*(Što?), *Yoni misli na voće na koje misli ...* (Tko?), *Yoni voli ...* (Što?), *Yoni voli voće koje voli ...* (Tko?), *Yoni je blizu...* (Čemu?) te *Yoni ima voće koje ima ...* (Tko?). Rezultati istraživanja su pokazali kako se pod normalnim uvjetima afektivne situacije rješavaju nešto brže od kognitivnih. Kalbe et al. (2010: 777) zaključuju kako je plauzibilno pretpostaviti postojanje dvaju mehanizama teorije uma koji su ili potpuno ili u velikoj mjeri nezavisni jedan od drugog. Prvi je zadužen za pripisivanje kognitivnih mentalnih stanja objektu, a drugi za pripisivanje afektivnih mentalnih stanja objektu.

Do istog zaključka dolaze i Sebastian et al. (2012) pri ispitivanju razlika u afektivnom i kognitivnom pripisivanju mentalnih stanja adolescenata i odraslih pojedinaca. Ispitanicima su pokazani stripovi, nakon čega su oni sami trebali odabrati između dvije ponuđene mogućnosti završetka stripa. U svim se pričama radilo o osobi koja promatra ponašanje drugog pojedinca te se od nje pri završetku priče očekuje reakcija na ponašanje tog pojedinca. Rezultati su pokazali kako je učestalost pogreške u slučaju afektivnog pripisivanja očekivano veća kod adolescenata nego kod odraslih pojedinaca. No u slučaju kognitivnog pripisivanja adolescenti i odrasli pojedinci su postigli jednak uspjeh. Na temelju tih rezultata Sebastian et al. (2012) iznose hipotezu kako je mehanizam koji omogućava afektivno pripisivanje složeniji od onog koji omogućava kognitivno pripisivanje. Prema tome, istraživanje navodi na zaključak kako je mehanizam kognitivnog pripisivanja u potpunosti

¹⁵ Kalbe et al. (2010: 776-777) pod kognitivnim stanjima podrazumijevaju propozicijska mentalna stanja, dok pod afektivnim stanjima podrazumijevaju ona stanja čije pripisivanje zahtijeva empatiju. Budući da je takvo shvaćanje termina implicitno i u istraživanju koje provode Sebastian et al. (2012), to je ujedno i način na koji su navedeni termini korišteni u ovoj radu.

razvijen već u adolescenciji. Istovremeno, Sebastian et al. (2012: 63) navode kako se „neuralna baza za sposobnost integriranja afektivnih informacija u odluke donesene na temelju teorije uma nastavlja razvijati između adolescencije i odrasle dobi.“

U kakvom su položaju ST i TT s obzirom na rezultate navedenih istraživanja? Jasno je da rezultati posljednjeg istraživanja (Sebastian et al. 2012) predstavljaju problem za zagovaratelje TT-a prema kojemu je teorija uma posljedica pripisivanja mentalnih stanja vođenog *ceteris paribus* generalizacijama. Te bi generalizacije trebale nalikovati znanstvenim hipotezama koje se usavršavaju iskustvom. Ukoliko se kod pripisivanja kognitivnih stanja radi o procesu kakav postulira TT, nejasno je zašto bi neuralni mehanizam koji leži u osnovi takvog procesa bio u jednakoj mjeri razvijen u adolescenciji i odrasloj dobi. Premda zastupnik TT-a mora tvrditi kako je naša teorija uma nalik cjeloživotnom obrazovanju, čini se kako *ceteris paribus* generalizacije odraslih pojedinaca nisu uspješnije od onih adolescenata.

Ukoliko kognitivno pripisivanje nije dobar kandidat za TT, može li to biti afektivno pripisivanje? Kalbe et al. (2010), Sebastian et al. (2012) i mnogi drugi autori odbacuju tu mogućnost ponajprije zato što je pripisivanje emocija putem prepoznavanja izraza lica područje u kojem je, kako navodi Goldman (2006: 113), ST ostvarila najbolje rezultate zahvaljujući otkriću zrcalnih neurona. Primjerice, poznato je kako su ljudi podložni tzv. emocionalnoj zarazi (*emotional contagion*): već samim uočavanjem emocije na tuđem licu počinjemo osjećati tu istu emociju. Nadalje, ako se kod afektivnog pripisivanja radi o procesu sukladnom TT-u, a kod kognitivnog o procesu sukladnom ST-u, afektivno pripisivanje ne bi smjelo biti brže – pretpostavka je kako su simulacijski procesi u većoj mjeri automatski, a time i brži od procesa teoretiziranja koji se svode na testiranje niza hipoteza od strane racionalnog subjekta. Dakle, plauzibilno je pretpostaviti kako je u oba slučaja riječ o procesima koje bolje objašnjava simulacijska teorija.

Prethodna dva istraživanja ukazuju na razlike koje se sustavno pojavljuju kod mehanizama teorije uma ovisno o tome radi li se o pripisivanju afektivnih ili kognitivnih mentalnih stanja. Proširujući na temelju rezultata iznesenih istraživanja Stichovu i Nicholsovu pretpostavku iz drugog poglavlja možemo ustvrditi sljedeće: prvo, pri pripisivanju kognitivnih mentalnih stanja kontekst igra veću ulogu te time raste mogućnost pogreške; drugo, pri pripisivanju afektivnih mentalnih stanja kontekst igra znatno manju, ponekad i zanemarljivu ulogu te je i mogućnost pogreške stoga znatno manja. Dakle, plauzibilno je tvrditi kako se u prvom slučaju radi o mehanizmima koji su nalik teoretiziranju (kakve pretpostavlja TT), dok se u drugom slučaju radi o mehanizmima nalik simuliranju (kakve pretpostavlja ST). Podjela između simulacijskih i teorijskih procesa s obzirom na vrstu pripisivanog mentalnog stanja je danas uvriježena. U skladu s tim, podjele koje su učestale kod zastupnika hibridnih teorija (primjerice Goldman, 2006) se temelje na razlikovanju procesa koji su nedvojbeno afektivno kontrolirani (kao emocionalna zaraza ili neke vrste imitacije) od onih za koje obično smatramo da su kognitivno kontrolirani (pripisivanje vjerovanja, intencija, pa čak i želja).

7. Teorija uma i afektivno stanje subjekta

Sposobnost uvida u tuđa mentalna stanja je osnova društvenoga života. Bez opisanih su procesa nezamislivi empatija i moral, mnoge društvene institucije i norme. Moralno rasuđivanje ne bi bilo moguće bez uvida u mentalna stanja drugih pojedinaca; bez uvida u to koja će mentalna stanja određena djelovanja izazvati, te u stanja koja su objašnjenje djelovanja. Prema tome, za očekivati je kako će stupanj uspješnosti teorije uma imati veliki utjecaj na moralno mišljenje. Posljednjih godina važna empirijska potvrda za to dolazi iz područja razvojne psihologije. Lane et al. (2010) potvrđuju vezu između razvoja razumijevanja tuđih umova i moralnog zaključivanja kod djece. Navedeno je istraživanje specifično po tome što ispituje posebno važnost kognitivnih i afektivnih procesa pri teoriji uma kod rješavanja moralnih

dilema. Ispitujući kod djece starosti 3.5 i 5.5 godina¹⁶ odnos razvoja teorije uma i moralnog mišljenja longitudinalnom metodom, Lane et al. (2010) dolaze do sljedećeg zaključka: razumijevanje vlastitih i tuđih emocija prvi je korak u razvoju moralnog mišljenja i omogućuje djeci da vlastito ponašanje opravdaju pozivanjem na tuđe fizičke i materijalne potrebe. Dakle, razvojem afektivne teorije uma se smanjuje učestalost pozivanja na hedonističke razloge kod objašnjenja vlastitog ponašanja. Nadalje, razvojem kognitivne teorije uma se povećava učestalost pozivanja na tuđe psihološke potrebe. Opravdanje ponašanja pozivanjem na socijalne razloge kao što su pohvala, nagrada, autoritet pojedinih osoba itd. je u porastu s obzirom na starost djeteta i razvoj inhibitorne kontrole koja djetetu omogućava „zanemarivanje vlastitih inicijalnih reakcija na moralne dileme i odbacivanje osobno nagrađujuće odluke u korist one koja je socijalno prihvatljivija“ (Lane et al. 2010: 884). Prema tome, Lane et al. (2010) sugeriraju kako je afektivno mentalno stanje subjekta teorije uma, tj. procesuiranje emocije, od temeljne važnosti za moralno mišljenje.

Priroda kognitivnih i afektivnih mehanizama pri moralnom mišljenju posebno dolazi do izražaja u tzv. *dvorazinskoj teoriji moralnosti* (dual-process model of morality) koju iznose Geene i Cushman (vidi primjerice Greene et al. 2004, Cushman et al. 2010). U kontekstu rasprave između teorije teorije i simulacijske teorije je zanimljivo na koji način autori opisuju dvije razine o kojima je riječ (Cushman et al. 2010: 49):

Tvrdimo kako kognitivni sustav operira putem „kontroliranih“ psiholoških procesa gdje su svjesno primijenjeni eksplicitni principi, dok su afektivni odgovori generirani putem „automatiziranih“ psiholoških procesa koji nisu dostupni svjesnoj

¹⁶ Prvo ispitivanje je provedeno na djeci starosti 3.5 godina, a drugo na istoj djeci nakon dvije godine. Taj je period života djece posebno važan za razvoj teorije uma. Kako je već prethodno rečeno, u tom periodu nastupa značajna promjena koja djeci omogućuje pripisivanje lažnih vjerovanja drugim pojedincima. Djeca postaju svjesna kako osobe koje ih okružuju imaju vlastita vjerovanja i želje koje određuju ponašanje tih pojedinaca te da ta vjerovanja ne moraju biti niti istinita niti istovjetna njihovim vlastitima. Dakle, djeca uviđaju kako pojedinci koji ih okružuju imaju drugačije interese od njih samih.

refleksiji. Sugeriramo da se distinkcije kognitivno/afektivno te svjesno/intuitivno [...] odnose na istu pozadinsku strukturu moralnog uma.

S obzirom da je u prethodnom poglavlju pokazano kako se kod kognitivne teorije uma radi o procesima teorijskog karaktera (kakve opisuje TT), a kod afektivne teorije uma o simulacijskim procesima, ovaj opis dviju razina gotovo je istovjetan opisu teorije teorije (TT) i simulacijske teorije (ST). Ideja kako je u temeljima moralnog mišljenja afektivno stanje subjekta je osnova tvrdnja dvorazinske teorije moralnosti. Najjednostavnije rečeno, subjekt koji je pod dominantnim utjecajem afektivnog procesuiranja, donosit će deontološke moralne odluke, dok će subjekt pod utjecajem pretežito kognitivnog procesuiranja donositi konzekvencijalističke odluke. Važno je primijetiti osnovnu razliku između ovog viđenja i načina na koji je pripisivanje kognitivnih i afektivnih stanja opisano u prethodnom poglavlju. Rečeno je kako hoće li se pri teoriji uma odvijati simulacijski ili teorijski procesi ovisi o tome pripisuju li se objektu afektivna ili kognitivna stanja. Međutim, ovdje do izražaja dolazi ne vrsta stanja koja se pripisuje objektu već stanje u kojem se nalazi subjekt teorije uma.

Postavlja se pitanje na koji se način TT i ST odnose prema mentalnom stanju subjekta teorije uma. Teorija teorije je objašnjenje uvida u tuđe umove koje se temelji primarno na racionalnosti subjekta i njegovoj sposobnosti učenja iz iskustva. Osim racionalnog subjekta, osnova teorije teorije je i pretpostavka o racionalnosti objekta teorije uma. Dakle, afektivno stanje subjekta nije predmet interesa teorije teorije. Odnos simulacijske teorije prema subjektu nešto je složeniji i ovisi o inačici ST-a koja se zastupa. No niti jedna forma simulacijske teorije ne uzima u obzir afektivno stanje subjekta teorije uma. Teorije eksplicitne i radikalne simulacije često se oslanjaju na element racionalnosti subjekta pri odabiru mogućih mentalnih stanja objekta koji se „ubacuju“ u simulacijski proces. Teorija radikalne simulacije pak shvaća subjektove mentalne sposobnosti doslovno kao simula-

tore stanja objekta. Subjekt je pri simulacijskim procesima isključen (*taken off-line*). Čini se kako je zanemarivanje afektivnih stanja subjekta zajednička karakteristika i ST-a i TT-a. No ukoliko se sjetimo kako je sposobnost teorije uma od ključne važnosti za sposobnost moralnog mišljenja, u slučaju kojeg je empirijski nedvojbeno potvrđena važnost afektivnog stanja subjekta, čini se kako je to ujedno zajednička slabost obje navedene teorije.

Prethodno je opisan eksperiment koji su na bračnim parovima proveli Kruger i Gilovich 1999. godine. Saxe (2005a) citirajući taj eksperiment iznosi argument o važnosti kontekstualnih faktora pri teoriji uma te zaključuje kako asimetričnost procesa teorije uma u slučaju neutralnih promatrača i bračnih partnera ukazuje na važnost pozadinskih vjerovanja subjekata, što je u suprotnosti s tvrdnjama simulacijske teorije. Singer et al. (2004) provode istraživanje o sposobnosti suosjećanja kod romantičnih partnera pri iskusnosti boli. Oba su partnera zajedno izložena električnim šokovima različite jačine. Svaki je pojedinac vizualnim signalima dobivao informacije o jačini šoka kojem je izložen njegov partner te je fMRI metodom bilježena reakcija na vlastitu bol različitog intenziteta, kao i reakcija na pretpostavljenu bol koju osjeća partner. Eksperiment je pokazao kako je reakcija na bol koju je pojedinac osobno iskusio istovjetna reakciji na jednaku razinu boli kada je iskusi partner. Ovaj je rezultat u potpunosti u skladu s pretpostavkama simulacijske teorije. Osim toga, u usporedbi s eksperimentom Kruger i Gilovich, važno je uočiti da je afektivno stanje subjekta – njegov romantični stav prema objektu – upravo ona komponenta konteksta odvijanja teorije uma za koju je plauzibilno pretpostaviti kako objašnjava asimetričnost kod teorije uma dobivenu u slučaju neutralnog promatrača (promatrača koji nema isti afektivni stav prema objektu teorije uma). Dakle, oba navedena eksperimenta ukazuju na važnost afektivnog stanja subjekta pri teoriji uma, što je predmet koji je u potpunosti zanemaren od strane teorije teorije i simulacijske teorije.

Lane et al. (2010) navode razloge zbog kojih je njihov eksperiment, za razliku od gotovo svih prethodno provedenih, uspješno ukazao na odnos između teorije uma i moralnog mišljenja (Lane et al. 2010: 285-286):

Prvi se razlog tiče zadataka iz teorije uma koji su korišteni. Moralne dileme često uključuju emocionalnu komponentu (netko je ozlijeđen ili pati). Dok su drugi koristili zadatke teorije uma koji nisu „prodrjeli“ (*tap in*) u način na koji djeca razumiju emocije, ova studija uključuje mjerenje razumijevanja emocija kod djece kao i neke kompleksne emocionalne izraze ili inhibicije.

Plauzibilno je tvrditi kako je kod mnogih istraživanja relevantnosti kontekstualnih varijabli pri teoriji uma došlo do analogne pogreške u postavkama eksperimenta. Nemoguće je ispitivati varijable na koje je simulacijski proces osjetljiv ukoliko eksperiment nije postavljen na način koji izaziva afektivne reakcije subjekta. U svakom je slučaju moguće tvrditi kako je upravo to razlog zbog kojeg eksperimenti koje su proveli Brass et al. (2007) te Saxe i Wexler (2005) ne pokazuju varijacije kod reakcije dijelova mozga sa zrcalnim neuronima pri promjeni kontekstualnih varijabli. Drugim riječima, ukoliko promjena varijabli konteksta ne uzrokuje promjenu afektivnog stanja subjekta, neće doći niti do promjene simulacijskih procesa. Dakle, sukladno otkrićima iz moralne psihologije, o afektivnom stanju subjekta ovisi hoće li procesi teorije uma biti teorijskog ili simulacijskog karaktera.

Zaključak

Kroz prethodna poglavlja prikazana su i interpretirana neka od najvažnijih istraživanja iz područja teorije uma koja se tiču rasprave između teorije teorije i simulacijske teorije. Pokazalo se kako TT ima problema s objašnjavanjem točnosti procesa teorije uma, dok se ST s problemom susreće pri objašnjavanju njegove netočnosti. Nadalje, mnoga istraživanja pokazuju kako je ST neodrživ s obzirom na relevantnost konteksta za teoriju uma.

Osim toga, ST i TT divergiraju i s obzirom na vrstu mentalnog stanja: čini se kako ST ima važniju ulogu kod pripisivanja afektivnih stanja, a TT kod pripisivanja kognitivnih stanja. Uviđajući brojne probleme koji podjednako muče ST i TT mnogi se filozofi i znanstvenici odlučuju na neku vrstu hibridne teorije.¹⁷ Podjela između simulacijskih i teorijskih procesa s obzirom na vrstu pripisivanog mentalnog stanja je danas uvriježena. Tvrdi se kako su pri pripisivanju afektivnih stanja u pozadini procesi slični simulaciji, dok je kod pripisivanja kognitivnih stanja riječ o nekoj vrsti teoretiziranja. No sustavno primjenjivi kriteriji ili određene karakteristike konteksta na koje su pojedini tipovi procesa osjetljivi još uvijek nisu pronađeni.

Posljednjih je godina razvoj moralne psihologije ukazao na važnost afektivne komponente podražaja pri moralnom mišljenju. S obzirom da je moralno mišljenje temeljeno na razumijevanju tuđih mentalnih stanja, rezultati moralne psihologije mogu biti indikativni za daljnje istraživanje u području teorije uma. Teorija teorije (TT) i simulacijska teorija (ST) podjednako su neosjetljive na afektivna stanja subjekta teorije uma, što se pokazuje kao njihova velika slabost. Afektivno stanje subjekta je moguće shvatiti kao temeljnu karakteristiku konteksta teorije uma koja, kao i u slučaju moralnog mišljenja, „odlučuje“ u kojem će se smjeru odvijati procesi teorije uma.¹⁸

¹⁷ Primjerice, Gordon (2005) uvodi distinkciju između konstitutivnog i imitativnog zrcaljenja, dok Goldman (2006) govori o visokoj i niskoj simulaciji. Radi se o gruboj podjeli teorije uma na dva područja, ovisno o pretpostavljenoj važnosti simulacijskih i nesimulacijskih procesa. Međutim, obje teorije u svojoj osnovi ostaju simulacijski orijentirane.

¹⁸ Zahvaljujem se doc. dr. sc. Tomislavu Janoviću na vrijednim savjetima i pomoći pri izradi ovog rada.

Bibliografija

Blakemore, S i J. and Decety, J. (2001) „From the perception of action to the understanding of intention“, *Nature Reviews Neuroscience* 2, 561–567.

Brass, M., Schmitt, R. M., Spengler, S. i Gregely, G. (2007) „Investigating action understanding: Inferential processes versus action simulation“, *Current Biology* 17, 2117-2121.

Cushman, F. A., Young, L., i Greene, J. (2010) „Our multi-system moral psychology: Towards a consensus view“, u: J. Doris et al. (ur.) *The Oxford Handbook of Moral Psychology*, New York: Oxford University Press, 47-71.

Gallagher, S., (2007) “Logical and phenomenological arguments against simulation theory,” u: D. Hutto i M. Ratcliffe (ur.) *Folk Psychology Re-assessed*, Berlin: Springer, 62-78.

Gallese, V. (2003) „The Manifold Nature of Interpersonal relations: The Quest for a Common Mechanism“, u: C. Firth, i D. Wolpert, D. (ur.) *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others*, Oxford: Oxford University Press, 159-182.

Gallese, V. i Goldman, A. (1998) „Mirror neurons and the simulation theory of mind reading“, *Trends in Cognitive Sciences* 12, 493-501.

Goldman, A. (2005) „Intention, mind reading, and simultion“, u: S. Hurley, i N. Chater, (ur.) *Perspectives on Imitation: From Neuroscience to Social Science*, Cambridge, Massachusetts: The MIT Press, 78-93.

Goldman, A. (2006) *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*, Oxford: Oxford University Press.

Goldman, A. (2012) „Theory of mind“, u: E. Margolis, R. Samuels i S. Stich, (ur.) *The Oxford Handbook of Philosophy of Cognitive Science*, New York: Oxford University Press, 402-425.

Gordon, R. (1986) "Folk psychology as simulation", *Mind and Language* 1: 158–171.

Gordon, R. (2005) „Intentional agents like myself“, u: S. Hurley, i N. Chater (ur.) *Perspectives on Imitation: From Neuroscience to Social Science*, Cambridge, Massachusetts: The MIT Press, 95-106.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., i Cohen, J. D. (2004) „The neural bases of cognitive conflict and control in moral judgment“, *Neuron* 44 (2), 389–400.

Heath, C. (1999) „On the Social Psychology of Agency Relationships: Lay Theories of Motivation Overemphasize Extrinsic Incentives“, *Organizational Behavior and Human Decision Processes* 78 (1), 25–62.

Hutto, D., Ratcliffe, M. (2007) "Introduction," u D. Hutto i M. Ratcliffe (ur.), *Folk Psychology Re-assessed*, Berlin: Springer, 1-22.

Jacob, P. i Jeannerod, M. (2005) „The motor theory of social cognition: A critique“, *Trends in Cognitive Sciences* 9, 21–25.

Jeannerod, M. (1994) „The representing brain: neural correlates of motor intention and imagery“, *Behavioral and Brain Sciences* 17 (2), 187-245.

Kalbe, E., Schlegel, M., Sack, A.T., Nowak, D. A., Dafotakis, M., Bangard, C., Brand, M., Shamay-Tsoory, S., Onur, O. A. i Kessler, J. (2010) „Dissociating cognitive from affective theory of mind: a TMS study“, *Cortex* 46 (6), 769–80.

Keysers, C. i Gazzola, V. (2007) „Integration simulation and theory of mind: from self to social cognition“, *Trends in Cognitive Sciences* 11 (5), 194-196.

Kruger, J. i Gilovich, T. (1999) „‘Naive cynicism’ in everyday theories of responsibility assessment: On biased assumptions of bias“, *Journal of Personality and Social Psychology* 76(5), 743-753.

Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, Jennifer. i Kerr, D. C. R. (2010) „Theory of mind and emotion understanding predict moral development in early childhood“, *British Journal of Developmental Psychology* 28, 871–889.

Mitchell, J. P. (2005) „The false dichotomy between simulation and theory-theory: the argument’s error“, *Trends in Cognitive Sciences* 9 (8), 363-364.

Mitchell, J. P. (2009) „Inferences about other minds“, *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 364, 1309–1316.

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A.J., Schweizer, S., Frith, C. i Dalgelish, T. (2009) „A key role for similarity in vicarious reward“, *Science* 324, 900.

Nichols, S. i Stich S. (2003) *Mindreading*, Oxford: Oxford University Press.

Pronin, E., Kruger, J. i Ross, L. (2011) „You don’t know me, but I know you: the illusion of asymmetric insight“, *Journal of Personality and Social Psychology* 81 (4), 639-656.

Rizzolatti, G., Fadiga, L., Gallese, V. i Fogassi, L. (1996) „Premotor cortex and the recognition of motor actions“, *Cognitive Brain Research* 3 (2), 131–141.

Saxe, R. (2005a) „Against simulation: The argument from error“, *Trends in Cognitive Sciences* 9 (8), 174-179.

Saxe, R. (2005b) „On ignorance and being wrong: A reply to Gordon“, *Trends in Cognitive Sciences* 9 (8), 362.

Saxe, R. i Wexler, A. (2005) „Making sense of another mind: The role of the right temporo-parietal junction“, *Neuropsychologia* 43 (10), 1391-1399.

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S. J., De Brito, S. A., McCroy, E. J. P. i Viding, E. (2012) „Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults“, *Social Cognitive & Affective Neuroscience* 7 (1), 53-63.

Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J. i Firth, C. D. (2004) „Empathy for pain involves the affective but not sensory components of pain“, *Science* 303, 1157–1162.

Stich, S., Nichols, S. (1995) „Folk Psychology: Simulation or tacit theory?“, u: M. Davies i T. Stone, (ur.) *Folk Psychology: The Theory of Mind Debate (Readings in Mind and Language)*, Cambridge, Massachusetts: Blackwell Publishing, 123-159.

Stich, S., Nichols, S. (2011) „Folk psychology“, u S. Stich (ur.) *Collected Papers (Volume I: Mind and Language 1972-2010)*, New York: Oxford University Press, 300-319.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001) „A functional MRI study of human amygdala responses to facial expressions of fear versus anger“, *Emotion* 1 (1), 70–83.

Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V. i Rizzolatti, G. (2003) „Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust“, *Neuron* 40, 655–664.

How do we Read Other People's Minds: Theory Theory and Simulation Theory

ANNA KOCSIS

ABSTRACT: Contemporary research into the theory of mind is marked by two dominant approaches. Theory-theory (TT) is the position that mindreading is a process of inference by using previously formulated generalisations similar to scientific hypothesis. On the other hand, the proponents of the simulation theory (ST) would claim that the theory of mind is a form of mirroring of other's mental states. By simulating the mental states of others, we are „walking in their shoes“. The assumption of mutual inconsistency between TT and ST has had a strong influence on directing empirical research in the field but has since proven to be implausible. Rejection of the assumption of mutual inconsistency has shifted attention towards their relationship in theory of mind. The type of process involved in theory of mind is strongly dependent on the context, situation and the type of mental state that is being ascribed. An accepted division of labour between TT and ST is based on ascribing cognitive or emotional states. It seems plausible that affective mental states of the subject, which go unnoticed by both TT and ST, are a crucial factor in determining the type of processes involved.

KEYWORDS: Affective mental states, cognitive mental states, context, mirror neurons, simulation theory (ST), theory theory (TT).