

Donald Davidson

## JEDINSTVENA TEORIJA MIŠLJENJA, ZNAČENJA I DJELOVANJA \*

Svaki pokušaj razumijevanja verbalne komunikacije mora je posmatrati u prirodnom okruženju kao dio većeg poduhvata. U početku se čini da to ne može biti teško jer jezik komunikacije ne uključuju ništa više nego samo jezičke transakcije među govornicima i interpretatorima te sklonosti za takve transakcije. Ipak, taj zadatak nam izmiče. Jer činjenica da su jezički fenomeni samo bihevioralni, biološki ili fizički fenomeni opisani u egzotičnom vokabularu značenja, referencije, istine, tvrdjenja [assertion] i tako dalje – sama dominacija činjenica ove vrste nad drugom vrstom činjenica ili opisa jedne vrste fenomena nad drugom vrstom fenomena – ne jamči, niti čak ne obećava, mogućnost konceptualne redukcije.

U tome je naš problem. Čini se da je za razumijevanje potrebna neka vrsta redukcije, no značajna redukcija ostaje nedostižna u slučaju jezika. Društvena scena, kada se prikazuje u terminima koji ne pretpostavljaju ono što treba objasniti, prevelika je i preslobodno konceptualno povezana s onim što je karakteristično za govor da bi se otkrila tajna jezičkog značenja. Kad se za prosvjetljenje o prirodi jezika obratimo privatnim interesima i interesima zajednice koji pokreću jezik, gubimo dodir s pitanjima koja nas zanimaju ako ih ne postavljamo. Potrebna nam je druga strategija, drugi način povezivanja govora s njegovim ljudskim okruženjem. U ovom radu skiciram alternativni pristup fundamentalnom objašnjenju jezika.

Neposredno psihološko okruženje jezičnih sposobnosti i postignuća nalazi se u propozicionalnim stavovima, u jezičnim činjenicama, osim što to razumijevanje prati isprepleteni prikaz središnjih kognitivnih i konativnih stavova.

Previše je tražiti da se ti temeljni intenzionalni pojmovi reduciraju na nešto drugo - nešto više bihevioralno, neurološko ili fiziološko, na primjer. Niti možemo analizirati ni jednu od ovih osnovnih triju, vjerovanja, želje i značenja, u smislu jednog ili dva druga; ili barem tako ja mislim, što sam i dokazivao već negdje drugdje. <sup>1</sup> No, čak i ako bismo mogli ostvariti redukciju u ovom osnovnm triju,

---

\* Tekst "A Unified Theory of Thought, Meaning, and Action" preveden je ovdje iz knjige Donald Davidson: *Problems of Rationality*. Oxford University Press, 2004. pp.151-166

<sup>1</sup> Za razmatranja koja potkrepljuju ove tvrdnje, vidi moje tekstove "Belief and the Basis of Meaning", *Synthese*, 27 (1974): 309–23; "Radical Interpretation", *Dialectica*, 27 (1973): 313-328; "Thought and Talk", in Samuel Guttenplan (ed.), *Mind and Language*, Oxford, 1975, pp. 7-23.

rezultati bi bili nedostatni onome što bi se moglo tražiti samo zato što bi krajnja tačka - interpretacija, recimo, govora – bila previše blizu mjesta gdje smo započeli (sa vjerovanjem i željom, ili namjerom, koja je rođena od vjerovanja i želje). Osnovni opis bilo kojeg od ovih pojmova mora započeti iza ili ispod njih, ili u nekom trenutku jednako udaljenom od svih.

Ako je tako, ne možemo temeljiti analizu jezičkog značenja na nejezičkim svrhama ili namjerama koje pokreću upotrebu jezika. Niti će pomoći da se apelira na eksplicitna ili implicitna pravila ili konvencije, samo zato što se oni moraju razumjeti u smislu intencija i vjerovanja. Naravno, nema sumnje da je važno pokazati kako su povezani značenja i intencije. Takve veze daju strukturu propozicionalnim stavovima i odgovara im u sistematskom tretmanu. Ali prema mom sadašnjem planu, intencija i intencionalno djelovanje neće izravno objasniti značenje. Umjesto toga, značenje, vjerovanje i želja bit će tretirani kao potpuno koordinirani elementi u razumijevanju djelovanja. Ove široke, nejasne tvrdnje dobijaju smisao u ovome što slijedi.

Cilj je teorija za interpretiranje riječi govornika, teorija koja također pruža osnovu za pripisivanje vjerovanja i želja govorniku. Teorija ne bi trebala pretpostavljati da su dostupni bilo kakvi propozicijski stavovi govornika, barem ne u potpuno individualiziranom obliku. Pomoći će razmatranje prve Bayesianove teorije odlučivanja, kako ju je razvio Frank Ramsey,<sup>2</sup> koji se bavi dvama od naših tri osnovna elementa, vjerovanjem i željom. Ta će teorija poslužiti kao model za tu vrstu teorije koju želimo, a također će, naravno, bitno pridonijeti izgledu jedinstvene teorije.

Izbor jednog pravca djelovanja u odnosu na drugi, ili preferencija da se jedno stanje stvari postigne radije nego drugo, općenito je proizvod dvaju razmatranja: vrjednujemo neki pravac djelovanja ili stanje stvari zbog vrijednosti koju smo postavili kao njegove moguće posljedice, kao i to koliko vjerujemo u te posljedice, s obzirom na to da obavljamo djelovanje ili da stanja stvari dolaze do realizacije. Prema tome, pri odabiru između pravaca djelovanja ili stanja stvari odabiremo jednu relativnu vrijednost čije su konsekvence, kada su prožete vjerovatnoćom tih konsekvenci, najveće. Pravci djelovanja su obično kockanja, jer ne znamo sigurno kako će se stvari završiti. Dakle, u mjeri u kojoj smo racionalni, prihvaćamo ono za što smatramo da je najbolja moguća opklada (mi "maksimiziramo očekivanu korist").

Karakteristika takve teorije je da je ono za što je osmišljena da objasni - osnovne preferencije ili izbore među opcijama - relativno otvoreno za posmatranje,

---

<sup>2</sup> F. P. Ramsey, "Truth and Probability", u: *The Foundations of Mathematics*, New York, 1950.

dok se mehanizam objašnjenja, koji uključuje stupanj vjerovanja i glavne vrijednosti, ne može posmatrati. Stoga se javlja pitanje kako reći kada osoba ima određeni stupanj vjerovanja u neku propoziciju ili koja je relativne jačine njenih preferencija. Očigledni problem je da ono što je poznato (obična, ili prosta preferencija) je rezultanta dvije nepoznanice, stupnja vjerovanja i relativna jačina sklonosti (preferencije). Ako su poznate glavne preferencije neke osobe za ishode, onda bi njegovi izbori pravaca djelovanja otkrili njegov stupanj vjerovanja; i ako je njegov stupanj vjerovanja bio poznat, njegovi bi izbori otkrili komparativne vrijednosti koje on stavlja na ishode. Ali kako obje nepoznanice mogu biti određene jednostavnim izborima ili preferencijama? Ramsey je riješio taj problem pokazujući kako pronaći propoziciju koja se smatra vjerovatnom kao i njena negacija na temelju jednostavnih izbora. Ova pojedinačna propozicija može se koristiti za konstruiranje nepreglednog niza izbora opklada koje daju mjeru vrijednosti za sve moguće opcije i eventualnosti. Tada je rutinsko utvrđivati stupnjeve vjerovanja u sve propozicije.

Ramsey je uspio preokrenuti trik određivanjem ograničenja na dopuštene obrasce jednostavnih preferencija ili izbora. Ta ograničenja nisu proizvoljna, već su dio zadovoljavajućeg opisa o razlozima sklonosti osobe i izbora ponašanja. Ta ograničenja opisuju zahtjev da akter bude racionalan, ne u njegovim posebnim i krajnjim vrijednostima, nego u obrazcu u kojem se međusobno oblikuju i u kombinaciji s njegovim vjerovanjima. Teorija dakle ima snažan normativni element, ali element koji je bitan ako se pojmovi preferencije, vjerovanja, razuma i intencionalnog djelovanja trebaju primjenjivati.

Forma u kojoj se promatra je središnja za sposobnost teorije da izdvoji iz činjenica koje uzete pojedinačno jesu relativno izravno povezane s onim što se može u samom sebi uočiti ili promatrati da važi za druge - da se iz takvih činjenica izvedu činjenice sofisticirane vrste (stupanj vjerovanja, usporedbe razlika u vrijednosti).

Sa stajališta teorije, sofisticirane činjenice objašnjavaju jednostavne, one koje su vidljivije, dok one koje se mogu promatrati konstituiraju dokaznu bazu za testiranje ili primjenu teorije.

Budući da su ograničenja oštro postavljena, različite se stvari mogu dokazati o toj teoriji. Intuicija da ograničenja definiraju neki aspekt racionalnosti, na primjer, može biti potkrijepljena dokazom da samo neko čija su djela u skladu sa tom teorijom radi najbolje što može pod svojim vlastitim svjetlima: nizozemska knjiga ne može biti protiv njega (vidi: Dutch Book Argument-*prim.prev.*)

Bayesianova teorija odlučivanja ne daje definiciju pojmova vjerovanja i preferencije na temelju neintenzionalnih pojmova. Umjesto toga, koristi jedan intenzionalni pojam, običnu preferenciju između klađenja ili ishoda, kako bi dao sadržaj za još dva pojma: stupanj vjerovanja i usporedbe razlika u vrijednosti. Zato

bi bilo pogrešno misliti da teorija daje redukciju intenzionalnih pojmova na nešto drugo. Ipak, to je važan korak u smjeru reduciranja složenih i relativno teorijskih intenzionalnih pojmova na intenzionalne pojmove koji su u primjeni bliže javno vidljivom ponašanju. Prije svega, teorija pokazuje kako je moguće dati korisni sadržaj za dva osnovna i međusobno povezana propozicijska stava bez pretpostavke da je bilo koji od njih prethodno određen.

Kao teorija za objašnjenje ljudskih djelovanja, Bayesianova teorija odlučivanja takve vrste o kojoj raspravljam otvorena je kritici koja pretpostavlja da možemo prepoznati propozicije na koje se usmjeravaju stavovi poput vjerovanja i želje (ili preferencije). Ali naša sposobnost identificiranja i razlikovanja među propozicijama koje neki akter govorne komunikacije razmatra ne treba odvojiti od naše sposobnosti da razumijemo šta on kaže. Općenito saznamo upravo ono što neko želi, preferira ili vjeruje samo interpretiranjem njegovog govora. To je osobito očito u slučaju teorije odlučivanja, gdje su predmeti želje često složene opklade, a ishodi opisani kao kontingentni u odnosu na određene događaje. Jasno je da teorija koja pokušava iznijeti stavove i vjerovanja koja objašnjavaju preferencije ili izbore mora uključivati teoriju verbalne interpretacije ako ne bi trebalo napraviti osakaćene pretpostavke.

Ono što moramo dodati teoriji odlučivanja ili u nju uklopiti, jest teorija interpretacije za aktera govorne komunikacije, način kazivanja onoga što podrazumijeva pod svojim riječima. Ipak, ovaj dodatak mora se dati u nedostatku detaljnih informacija o vjerovanjima, željama ili namjerama..

Teorija verbalne interpretacije za govornika (ono što ću nazvati "teorijom značenja") mora dati značenje svakom od potencijalno beskonačnih izgovaranja u repertoaru govornika, pa se tako traži rekurzivna teorija koja izvodi značenja izričaja (ili rečenice) iz značenja njegovih dijelova. U tom smislu, naša teorija značenja mora nalikovati teoriji odlučivanja, jer teorija odlučivanja također predviđa izbore između potencijalne beskonačnosti alternativa.

Dispozicije da se izgovori jedna ili druga rečenica očituje se na bezbroj načina. Uočljiv primjer je očitovanje pristanka u činu tvrdnje. (Baš kao što se preferencija manifestira u izboru.) Pretpostavimo da je bilo moguće reći, na ovaj ili onaj način, kada se neka osoba slaže sa rečenicom - kada smatra da je rečenica istinita. Iskrena tvrdnja bi služila tome, ali isto tako mogu i mnogi drugi postupci i stavovi.

Ako bi se moglo smatrati da takvi dokazi podupiru teoriju značenja, došlo bi do značajnog napretka. Jer ipak, kao u slučaju teorije odlučivanja, dokazi bi imali jedan intenzionalni element koji se ne može reducirati (držanje / smatranje istinitim), počeli bismo s jednim stavom koji ne pretpostavlja da možemo otkriti beskonačnu raznolikost propozicionalnih stavova (vjerovanja, želja, intencija i značenja) sa kojima se jedna potpuna teorija nada da će završiti. S ove točke

gledišta, vrsta teorije značenja koju predlažem bolja je od teorija koje ovise o detaljnom poznavanju vjerovanja ili intencija koje prate pojedinačne izjave. Jer takve teorije koriste za dokaze činjenice koje su tako suptilne i teške za identificiranje kao činjenice koje su namijenjene podupiranju. Da bi ponovio moju početnu temu: ne možemo pretpostaviti da možemo napraviti fino granulirane distinkcije između osnovnih propozicionalnih stavova jedne vrste (intencije, na primjer) kako bi konstruirali teoriju drugog niza propozicionalnih stavova (značenja, na primjer) bez pretpostavljanja da su na teža pitanja već dati odgovori.

Volio bih da je teorija značenja slična Ramseyevoj verziji teorije odlučivanja, ali najbolje što znam učiniti izgleda ozbiljno nedostatno za ovaj cilj. Ipak, mislim da se može vidjeti kako je nešto slično moguće, jer znamo u globalu kako zasnovati teoriju značenja na prihvatanju rečenica koje su uzrokovane događajima u svijetu. Takva teorija ne samo da će biti teorija značenja za govornika, već i teorija vjerovanja, jer je rečenica koja se smatra istinitom plus interpretacija jednaka vjerovanju. Budući da znati da neko smatra rečenicu istinitom nije ni znati što znači rečenica, niti koje vjerovanje izražava, takva teorija nije trivijalna. Ali postavljanje stvari na taj način donosi problem: baš kao što je odabir toka djelovanja rezultat vjerovanja i želje, tako je smatranje rečenice istinitom rezultat značenja i vjerovanja. Kako možemo razlikovati uloge svakog od njih u određivanju smatranja istinitim rečenica? Ne možemo se nadati da prvo otkrijemo interpretaciju, a zatim odčitati vjerovanja, jer su vjerovanja povezana s rečenicama koje čine značenja rečenica.

Problem je poput problema u teoriji odlučivanja, ali rješenje nije tako prikladno. Ramsey je smislio genijalan trik za utvrđivanje subjektivne vjerovatnosti jednog događaja tačno bez znanja bilo čega drugog osim jednostavnih preferencija; s tim početkom, dodaci bi mogli biti dodijeljeni općenito na jedinstven način kao što to teorija zahtijeva. U kombiniranoj teoriji značenja i vjerovanja ne postoji takav precizan trik, pa se podaci mogu prilagoditi različitim teorijama interpretacije, pod uvjetom da te teorije budu povezane s odgovarajućim teorijama vjerovanja.<sup>3</sup> Posljedica neodređenosti interpretacije nije prema ovom opisu značajnija ili problematičnija od činjenice da se težina može mjeriti u gramima ili u uncama.

Ako moramo izvesti značenje i vjerovanje iz dokaza o tome što uzrokuje da neko smatra rečenice istinitim, to može biti samo (kao u teoriji odlučivanja) jer pretpostavljamo strukturu. Na strani značenja, plauzibilna struktura je data teorijom istine kakvu je predložio Tarski, ali modificirana na različite načine da se primjenjuje na prirodni jezik. Takva se teorija može, kako sam tvrdio, smatrati

---

<sup>3</sup> Ovo je način izlaganja jedne lekcije koju smo naučili iz Quineove teze o neodređenosti prijevoda.

primjerenom za interpretiranje izjava govornika. (U stvari, to u najboljem slučaju daje samo prvi korak u interpretiranju onoga što govornik misli, navodeći što njegove riječi znače u njihovom doslovnom smislu).

Na strani vjerovanja, manje je jasno koju strukturu treba pretpostaviti. No ovdje vodeći principi moraju proizilaziti, kao u slučajevima teorije odlučivanja ili teorije istine, iz normativnih razmatranja. Individualiziramo i identificiramo vjerovanja, kao što to činimo sa željama, intencijama i značenjima, na velik broj načina. Ali odnosi između vjerovanja igraju odlučujuću konstitutivnu ulogu; ne možemo prihvatiti velika ili očita odstupanja od racionalnosti bez prijetnje razumljivosti naših atribucija. Ako ćemo shvatiti govor ili djela druge osobe, moramo pretpostaviti da su njihova vjerovanja ugrađena u obrazac koji je u suštinskim pogledima sličan obrascu naših vlastitih vjerovanja. Prvo, onda, nemamo drugog izbora nego projicirati svoju vlastitu logiku na vjerovanja drugoga. U kontekstu sadašnje teorije, to znači da ga uzimamo kao ograničenje mogućih interpretacija rečenica koje se smatraju istinitim da su logički konzistentne jedna s drugom. Drugim riječima, politika je da pretpostavimo da su vjerovanja govornika logički dosljedna (barem do neke tačke).

Međutim, logična dosljednost ne osigurava više od interpretiranja logičnih konstanti, (što god uzimamo da trebaju biti granice logike i popis logičnih konstanti). Daljnja interpretacija zahtjeva pretpostavku daljnjeg dogovora između govornika i interpretatora. Pretpostavka je zasigurno opravdana, a alternativa je da interpretator nalazi da je govornik nerazumljiv. No, teško je precizno odrediti pravila za odlučivanje o tome gdje se sporazumjevanje treba uzeti kao datost. Opća načela relativno je jednostavno izreći: sporazum o zakonima i pravilnostima obično je važniji od sporazuma o slučajevima; više treba favorizirati sporazum o onome što se može otvoreno i javno uočiti nego dogovor o onome što se skriva, zaključuje ili loše opaža; dokazne odnose treba čuvati što se više graniče s tim da su konstitutivni za značenje.

Nije sigurno u kojoj se mjeri ta načela mogu učiniti konačnim - to je problem racionalizacije i kodifikacije naše epistemologije. Ali u jednom ključnom području, onom dokazne potpore, jasno je da je napredak moguć. Očito je da će ispravno interpretiranje riječi govornika znatno ovisiti o tome u kojoj mjeri govornik uzima istinu jedne rečenice kao potporu istine druge. Jer sadržaj rečenica ili predikata koji su više ili manje udaljeniji od onoga što se neposredno posmatra, ovisi o tome što se poduzima kako bi se poboljšala njihova istina ili primjena, dok značenje rečenice više vezano izravno s onim šta se promatra djelomično je određeno teorijskim istinama, njezina istina se povećava.

Ono što je potrebno za adekvatnu teoriju vjerovanja i značenja, nije samo znanje o tome zbog čega govornik drži rečenicu istinitom, već znanje o stupnju

vjerovanja u njezinu istinu. Tada bi bilo moguće otkriti stupnjeve dokazne potpore napominjući kako su promjene u stupnju vjerodostojnosti jedne rečenice bile popraćene promjenama stupnja vjerodostojnosti drugih rečenica.

Stupanj vjerovanja, međutim, sam je udaljen od onoga što općenito agent može opažati ili dijagnosticirati interpretator; kao što smo vidjeli u raspravi o teoriji odlučivanja, stupanj vjerovanja konstrukcija je na elementarnijim stavovima. Zbog toga je Quine, razvijajući svoju teoriju radikalnog prevođenja, ovisio o jednostavnom pristanku i zašto sam, slijedeći Quinea, ovisio o usko povezanom poimanju držanja istinitim. Ovi jednostavni kvalitativni koncepti, upravo sam to tvrdio, nisu primjereni za izgradnju teorije značenja (prijevod, interpretacija).

Teorija značenja kao što je ja vidim i Bayesianova teorija odlučivanja, očigledno su napravljene jedna za drugu. Teorija odlučivanja mora biti oslobođena od pretpostavke neovisno određenog znanja značenja; teorija značenja poziva se na teoriju stupnja vjerovanja kako bi ozbiljno iskoristila odnose dokazne podrške. Ali, navođenje ovih uzajamnih zavisnosti nije dovoljno, jer se niti jedna teorija ne može razviti kao temelj za drugu. Nema načina da jednostavno dodate jednu drugoj, jer kako biste započeli, svaki od njih zahtijeva element izveden iz druge. Ono što se traži je jedinstvena teorija koja daje stupanj vjerovanja, korisnosti na intervalnoj skali i interpretiranje govora bez pretpostavljanja bilo kojeg od njih.

Problem je u velikoj mjeri sličan problemima postavljenim zauzvrat teorijom odlučivanja i teorijom značenja, osim što umjesto dvije postoje tri stvari koje treba apstrahirati iz dokaza koji ne ovise o prepoznavanju onoga što se treba objasniti. Ono što je potrebno za jedinstvenu teoriju je stoga neki jednostavan propozicijski stav koji se može prihvatljivo primijeniti na aktera komunikacije bez detaljnog poznavanja njegovih vjerovanja, preferencija ili verbalnih značenja i iz kojih možemo izvući teoriju stupnjeva vjerovanja, usporedbe razlika u stupnju želje i metod interpretiranja iskaza.

Predlažem da se uzme slijedeći stav kao temeljni: akter komunikacije preferira jednu rečenicu kao istinitu, više nego drugu. Naravno, rečenice moraju biti obdarene sa značenjem za govornika, ali interpretiranje rečenica dio je zadataka interpretatora. Ono što interpretator mora učiniti jest, dakle, informacija o tome koji događaji u svijetu uzrokuju da akter preferira da je jedna, a ne druga rečenica istinita. Jasno, interpretator to može znati bez znanja što znače rečenice, koja su stanja stvari za aktera govora vrijednosti ili što on vjeruje. Ali, preferiranje istinitosti rečenica od strane aktera, jednako je jasno funkcija onoga što akter uzima da rečenice znače, vrijednost koju on postavlja na različita moguća ili stvarna stanja svijeta i vjerovatnost koju on pridaje tim stanjima odvisno od istine relevantnih rečenica. Stoga nije apsurdno misliti da se sva tri stava aktera govora mogu konstruirati na temelju preferencije njegovih preferencija među rečenicama. U isto

vrijeme, postoje važni uvidi u kojima je dokazna baza koju predlažem za jedinstvenu teoriju superiorna u odnosu na dokaznu bazu za teoriju odlučivanja ili teoriju značenja. Teorija odlučivanja obično se ispituje tražeći od subjekta odabir između opcija koje je opisao eksperimentator: u stvari, subjekt odabire između rečenica, a eksperimentator pretpostavlja da zna što te rečenice znače subjektu. Ta se pretpostavka opetovano izaziva eksperimentalnim rezultatima. Naša jedinstvena teorija odbacuje tu pretpostavku uzimanjem podataka koji se biraju među neinterpretiranim rečenicama.

U slučaju teorije značenja, postoji, naravno, činjenica o kojoj se već ranije govorilo, da zahtijeva stupanj suglasnosti, a to se ne daje izravno. No, čak i ako je jednostavna suglasnost ili smatranje istinitom dovoljna osnova za teoriju značenja, postoji poteškoća da ovi stavovi nemaju očitu ili izravnu povezanost s djelovanjem ili izvorima djelovanja u želji, preferenciji i intenciji. Problem nije toliko da samo djelovanje može biti promatrano kao da se ništa ne može smatrati razlogom za pretpostavku da neko odobri rečenicu ili smatra da je istina, koja ne pretpostavlja puno o akterovim namjerama ili ciljevima ili vrijednostima. Empirijska teorija značenja nas prisiljava na dovođenje želje ili preferencije, kao i vjerovanja, u sliku.

Kako je moguće izgraditi jedinstvenu teoriju iz dokaza o rečenicama koje se preferiraju kao istinite? Ovdje je skica mogućeg postupka.

Već smo vidjeli (ponovno u obliku pregleda) kako doći do teorije značenja i vjerovanja na osnovu znanja o stupnjevima do kojih se rečenice smatraju istinitim. Dakle, ako bismo mogli dobiti stupanj vjerovanja u rečenice pozivanjem na informacije o preferencijama da su rečenice istinite, imali bi uspješnu ujedinjenu teoriju.

Ramseyjeva verzija Bayesove teorije odlučivanja primarno koristi kockanje ili opklade, a to stvara poteškoće za moj projekt. Jer kako možemo reći da akter smatra da rečenica predstavlja kockanje dok ne odmaknemo u procesu interpretiranja njegovog jezika? Kockanje, na kraju, određuje vezu, vjerovatno uzročno-posljedičnu, između pojave određenog događaja (novčić pada na glavu) i određenog ishoda (osvojite opkladu). Čak i pod pretpostavkom da bismo mogli znati kada akter prihvati takvu vezu, izravna primjena teorije također ovisi o uzroku događaja (novčić koji pada na glavu) koji sam po sebi nema nikakvu vrijednost, negativnu ili pozitivnu. Također je potrebno pretpostaviti da vjerovatnost koju akter pripisuje novčiću koji pada na glavu nije kontaminirana mislima o vjerovatnosti osvajanja opklade. U eksperimentalnim testovima teorije odlučivanja pokušava se odabrati situacije u kojima ove pretpostavke imaju šanse biti istinite; ali općenita primjena koju sada imamo na umu ne može biti tako izbirljiva.

Dugujemo Richardu Jeffreyju verziju Bayesove teorije odlučivanja koja ne koristi izravno kockanje, nego obrađuje predmete preferencije, predmete kojima se



pripisuju subjektivne vjerovatnosti i predmete kojima se relativne vrijednosti jednoobrazno pripisuju kao propozicije.<sup>4</sup> Jeffrey je detaljno prikazao kako izvesti subjektivne vjerovatnosti i vrijednosti iz preferencija da su propozicije istinite.

Jeffreyjeva teorija ne određuje vjerovatnosti i korisnosti do istih skupova transformacija kao standardne teorije. Umjesto funkcije korisnosti određene do linearne transformacije, u Jeffreyjevoj teoriji funkcija korisnosti jedinstvena je samo do djelića linearne transformacije; i pripisivanja vjerovatnosti, umjesto da budu jedinstvene nakon što se broj odabere za mjerenje sigurnosti (uvijek Jedan), jedinstvene su samo unutar određene kvantizacije. Ova smanjenja u određivanju konceptualno su i praktično prikladna: između ostalog, dopuštaju donekle istu vrstu neodređenosti u teoriji odlučivanja koju smo očekivali u teoriji lingvističke interpretacije. Baš kao što u teoriji odlučivanja možete koristiti iste podatke koristeći razne funkcije korisnosti izvršavanjem odgovarajućih promjena u funkciji vjerovatnosti, tako možete promijeniti značenja koja pripisujete čovjekovim riječima (unutar granica) pod uvjetom da napravite kompenzacijske promjene u vjerovanjima koja mu pripisujete.

Ostaje jedan očigledan problem. Jeffrey pokazuje kako postići dovoljno rezultata kao što je Ramseyev, zamjenjujući preferencije među propozicijama za preferencije među opkladama. Ali propozicije su značenja ili rečenice s značenjima; ako znamo među kojim propozicijama akter odabire, pretpostavlja se da je naš izvorni problem interpretiranja jezika još jednom riješen. Ono što trebamo je da postignemo Jeffreyjeve rezultate davanjem samo preferencija među neinterpretiranim rečenicama.

Srećom, ovo nije tako divlji prijedlog za započinjanje kao što zvuči. Jer Jeffreyjeva metoda za pronalaženje subjektivne vjerovatnosti i relativne poželjnosti propozicija ovisi samo o strukturi istinosne funkcije propozicija - o tome kako su propozicije sastavljene od jednostavnih propozicija ponovljenom primjenom konjunkcije, disjunkcije, negacije i ostalih operacija definiranih u terminima od ovih. Ako počnemo s rečenicama umjesto propozicijama, tada će se naš problem riješiti pod uvjetom da se identificiraju konektivi istinosnih funkcija. Jer jednom kad su identificirani konektivi istinosnih funkcija, Jeffrey je pokazao kako podesiti, do željenog stupnja, subjektivne poželjnosti i vjerovatnosti svih rečenica; i ovo je, tvrdio sam, dovoljno da se dobije teorija za interpretiranje rečenica.

Osnovna empirijska datost u metodi koju treba opisati je akterova (slaba) preferencija da jedna rečenica, a ne druga, bude istinita; stoga se može smatrati da su ti podaci iste vrste kao i podaci koji se obično prikupljaju u eksperimentalnom

---

<sup>4</sup> R. C. Jeffrey, *The Logic of Decision*, University of Chicago Press, 2nd edition, 1983.

testu bilo koje Bayesove teorije odlučivanja, pod uvjetom da se interpretacija rečenica među kojima akter bira ne uzima unaprijed kao poznata eksperimentatoru ili interpretatoru.

Jedinstvenost i jednostavnost empirijske ontologije sistema, koja uključuje, kao što to čini samo izgovaranja i rečenice, ključna je za postizanje cilja kombiniranja teorije odlučivanja s interpretacijom. Slijedit ću što je bliže moguće Jeffreya, čija se teorija bavi samo propozicijama, zamjenjujući neinterpretirane rečenice gdje on uzima propozicije. Ovdje je, dakle, analogija Jeffreyjevog *aksioma poželjnosti*, primijenjenog na rečenice umjesto na propozicije:

Ako je  $vjero(s \text{ i } t) = 0$  i  $vjero(s \text{ ili } t) \neq 0$ , onda (D)

$$poželj(s \text{ ili } t) = \frac{vjero(s) poželj(s) + vjero(t) poželj(t)}{vjero(s) + vjero(t)}$$

(Pišem "vjero(s)" za subjektivnu vjerovatnost  $s$ -a i "poželj(s)" za poželjnost ili korisnost  $s$ -a.) Povezujući preferenciju i vjerovanje, ovaj aksiom čini vrstu posla koji se obično vrši opkladama; odnos je međutim drugačiji. Događaji se identificiraju sa rečenicama koje se interpretacijom navode kako se događaj javlja ("Sljedeća karta je mak"). Akcije i rezultati također predstavljaju rečenice ("Igrač se kladi za jedan dolar", "Igrač osvaja pet dolara"). Kockanje ne ulaze izravno, ali element rizika je prisutan, budući da odabir rečenice koja se drži / smatra istinitom obično je rizik za ono što će istovremeno biti zaista istinito. (Pretpostavlja se da se ne može odabrati logično lažna rečenica.) Dakle, vidimo da ako akter odabere istinitu a ne lažnu rečenicu "Igrač se kladi za jedan dolar", on ima priliku da dobije rezultat, za koji može, na primjer, misliti da ovisi o tome hoće li sljedeća karta biti mak ili ne. Tada će poželjnost (za istinitost) rečenice "Igrač se kladi za jedan dolar" biti poželjnost različitih okolnosti u kojima je ta rečenica istinito izvagana na uobičajeni način vjerovatnoćama tih okolnosti. Pretpostavimo da igrač vjeruje da će osvojiti pet dolara ako je sljedeća karta mak i neće osvojiti ništa ako sljedeća karta nije mak; on će tada imati poseban interes u tome da li će istina ili lažnost rečenice "Igrač se kladi za jedan dolar" bit će uparena s istinom ili lažnosti "Sljedeća karta je mak ". Neka ove dvije rečenice budu skraćene sa simbolima "s" i "t". Onda je

$$poželj(s) = \frac{vjero(s \text{ i } t) poželj(s \text{ i } t) + vjero(s \text{ i } \sim t) poželj(s \text{ i } \sim t)}{vjero(s)}$$

Ovo je, naravno, nešto poput Ramseyevog kockanja. Međutim, razlikuje se u tome što ne postoji pretpostavka da su "stanja prirode" za koje se može misliti da određuju ishode, u Ramseyjevom pojmu "moralno neutralna", to jest, nemaju

utjecaja na poželjnost ishoda. Niti postoji pretpostavka da vjerovatnosti ishoda ne ovise samo o vjerovatnostima "stanja prirode" (igrač može vjerovati da ima šansu za osvajanje pet dolara čak i ako sljedeća karta nije mak i da ima šansu da ne dobije pet dolara čak i ako sljedeća karta jest mak ).

"Aksiom poželjnosti" (P) može se upotrijebiti za pokazivanje vjerovatnosti ovisnosti o poželjnosti u Jeffreyjevom sistemu. Uzmite poseban slučaj gdje je  $t = \sim s$ . Onda imamo

$$\text{poželj}(s \text{ ili } \sim s) = \text{poželj}(s) \text{ vjero}(s) + \text{poželj}(\sim s) \text{ vjero}(\sim s) \quad (1)$$

pošto je  $\text{vjero}(s) + \text{vjero}(\sim s) = 1$ , možemo dati rješenje za  $\text{vjero}(s)$ :

$$\text{vjero}(s) = \frac{\text{poželj}(s \text{ ili } \sim s) - \text{poželj}(\sim s)}{\text{poželj}(s) - \text{poželj}(\sim s)} \quad (2)$$

Riječima, vjerovatnost propozicije ovisi o poželjnosti te propozicije i njezine negacije. Nadalje, lako je vidjeti da ako je rečenica  $s$  poželjnija od proizvoljne logičke istine (kao što je " $t$ " ili " $\sim t$ "), onda njena negacija (" $\sim s$ ") također ne može biti poželjnija od logičke istine. Pretpostavimo da (s Jeffreyjem) dodjeljujemo broj 0 bilo kojoj logičkoj istini. (To je intuitivno razumno budući da je akter ravnodušan prema istini tautologije.) Tada se (2) može napisati drugačije:

$$\text{vjero}(s) = \frac{1}{1 - (\text{poželj}(s) / \text{poželj}(\sim s))} \quad (3)$$

Odjednom je očigledno da  $\text{poželj}(s)$  i  $\text{poželj}(\sim s)$  ne mogu oboje biti više ili oboje manje poželjni od 0, poželjnosti bilo koje logičke istine, ako  $\text{vjero}(s)$  pada u intervalima od 1 prema 0. Ako (opet slijedeći Jeffreyja) nazovemo jednu opciju dobrom ako se preferira u odnosu na logičku istinu i lošom ako se preferira logična istina u odnosu na nju, tada (3) se pokazuje da je nemoguće za neku opciju (rečenicu) i njezinu negaciju oboje, da budu dobre ili oboje da bude loše.

Uzimajući " $\sim(s \text{ i } \sim s)$ " kao naš uzorak logičke istine, možemo postaviti ovaj princip isključivo u smislu poželjnosti:

$$\begin{aligned} \text{Ako } \text{poželj}(s) > \text{poželj}(\sim(s \text{ i } \sim s)) \text{ onda} & \quad (4) \\ \text{poželj}(\sim(s \text{ i } \sim s)) & \geq \text{poželj}(\sim s), \text{ i} \\ \text{ako } \text{poželj}(\sim(s \text{ i } \sim s)) > \text{poželj}(s) \text{ onda} & \\ \text{poželj}(\sim s) & \geq \text{poželj}(\sim(s \text{ i } \sim s)) \end{aligned}$$

Budući da se i negacija i konjunkcija mogu definirati u smislu Shefferove crte " $|$ " ("ne oboje"), onda se (4) može drugačije napisati:

$$\begin{aligned} \text{Ako } \text{poželj}(s) > \text{poželj}((t|u)|((t|u)|(t|u))) \text{ onda} & \quad (5) \\ \text{poželj}((t|u)|((t|u)|(t|u))) \geq \text{poželj}(s|s), \text{ i} & \\ \text{ako } \text{poželj}((t|u)|((t|u)|(t|u))) > \text{poželj}(s) \text{ onda} & \\ \text{poželj}(s|s) \geq \text{poželj}((t|u)|((t|u)|(t|u))). & \end{aligned}$$

Značaj formulacije (5) za sadašnje svrhe je ovaj. Ako pretpostavimo da je " $|$ " neki proizvoljan operator funkcije istinosne vrijednosti koji tvori rečenice iz parova rečenica, tada vrijedi sljedeće: ako (5) vrijedi za sve rečenice  $s$ ,  $t$  i  $u$ , i za neke  $s$  i  $t$ ,  $\text{poželj}(s|s) = \text{poželj}(t|t)$ , tada " $|$ " mora biti Shefferova crta (mora imati logička svojstva "ne oboje"); nijedna druga interpretacija nije moguća.<sup>5</sup>

Tako podaci koji uključuju samo poželjnost među rečenicama čije značenje nije poznato onome ko ih interpretira dovelo je (s obzirom na ograničenja teorije) da se identificira jedan rečenični konektiv. Budući da su sve logički ekvivalentne rečenice jednake po poželjnosti, moguće je sada interpretirati sve druge istinosno-funkcijske rečenične konektive, budući da se sve mogu definirati primjenom Shefferove crte. Na primjer, ako se utvrdi da za sve rečenice  $s$ ,

$$\text{poželj}(s|s) = \text{poželj}(\sim s),$$

možemo zaključiti da je tilda znak za negaciju.

Sada je moguće izmjeriti poželjnost i subjektivnu vjerovatnost svih rečenica, jer primjena formula poput (2) i (3) zahtijeva identifikaciju samo istinosno-funkcionalnih iskaznih konektiva. Stoga je jasno iz (3) da ako su dvije rečenice jednake po poželjnosti (a preferiraju se za logičku istinu) i njihove negacije su također jednake po poželjnosti, rečenice moraju imati istu vjerovatnost. Isto tako, ako su dvije rečenice jednake po poželjnosti (a preferiraju se za logičku istinu), ali negacija jedne preferira se pred negacijom druge, tada je vjerovatnost prve manja od one druge. Ovo, uzeto zajedno s odgovarajućim aksiomima egzistencije, dovoljno je za uspostavljanje skale vjerovatnosti. Tada je lako odrediti relativnu poželjnost svih rečenica.<sup>6</sup>

U ovom trenutku vjerovatnosti i poželjnosti svih rečenica su u teoriji određene. No, nijedna cjelovita rečenica još nije interpretirana, iako su identificirani istinosno-funkcionalni rečenički konektivi, i tako se mogu prepoznati rečenice

<sup>5</sup> Dužan sam Stigu Kangeru što mi je pokazao zašto raniji pokušaj rješenja ovog problema ne bi uspio. Također je dodao neka potrebna usavršavanja ovom prijedlogu.

<sup>6</sup> Za detalje vidi Jeffrey, op. cit.

logički istinite ili lažne po logici rečenice. Ostaje skicirati metode koje bi mogle dovesti do potpune interpretacije svih rečenica, tj. konstrukcije teorije istine za jezik aktera komunikacije. Pristup je onaj o kojem sam raspravljao u nizu članaka, a inspiriran je radom W. V. Quinea na radikalnom prevodjenu.<sup>7</sup>

Da bi akter komunikacije imao određenu subjektivnu vjerovatnoću za rečenicu, on mora da drži tu rečenicu istinitom ili lažnom ili da ima određeni stepen verovanja u njenu istinu. Budući da akter komunikacije pripisuje značenje rečenici, njegov stupanj povjerenja u istinitost rečenice također je njegov stupanj vjere u istinu koju izriče rečenica. Vjerovanja, kao što se očituju u stavovima prema rečenicama, su ključ za značenje. Već smo primijetili da su logički ekvivalentne rečenice ujednačene u poželjnosti. To samo po sebi nije izravna pomoć u dobivanju značenja logički ekvivalentnih rečenica, iako to, kao što smo vidjeli, pomaže u interpretaciji istinosno-funkcionalnih konektiva. Obrasci rečenica koje se drže istinitim ili lažnim također će dovesti do otkrivanja egzistencijalnih i univerzalnih kvantifikatora, a time i struktura koje uzimaju u obzir implikacije i logičke istine u kvantifikacijskoj logici. Otkriti strukture koje računaju na implikacije i logičke istine u kvantifikacijskoj logici jest otkriti *logički oblik* općenito, tj. naučiti kako se rečenice sastoje od predikata, singularnih pojmova, kvantifikatora, varijabli i slično. Simbol identiteta trebalo bi lako locirati, s obzirom na njegovu ulogu u promoviranju implikacija zasnovanih na supstituciji; neko ko drži istinitim rečenice oblika " $a = b$ " i " $Fa$ " također će smatrati istinitom rečenicu u formi " $Fb$ ", bez obzira šta predikat " $F$ " znači (osim ako je intenzionalan - jedan od problema sa kojim se ovdje nosim).

Daljnji koraci u interpretaciji zahtijevaju određenu razradu empirijske osnove teorije; bit će nužno svjedočiti ne samo sklonostima aktera komunikacije ka rečenicama, nego i događajima i objektima u svijetu koji uzrokuju njegove sklonosti (a time i njegova vjerovanja). Tako će biti vidljive okolnosti pod kojima je kod aktera komunikacije prouzročeno da dodjeljuju visoku ili nisku vjerovatnost rečenicama poput "Kiša pada", "To je konj" ili "Moja noga je ozlijeđena", koja daje najočitiiji dokaz za interpretaciju ovih rečenica i predikata u njima. Interpretator, primijetivši da akter komunikacije redovno dodjeljuje visok ili nizak stupanj vjerovanja rečenici "Kafa je gotova" kad je kafa gotova ili nije, pokušat će (makar probno čekati povezane rezultate) s teorijom istine koja kaže da je izgovor rečenice aktera komunikacije "Kafa je gotova" istinit ako i samo ako je kafa gotova.

---

<sup>7</sup> Vidi poglavlje 2 of W. V. Quine, *Word and Object*, The Technology Press and John Wiley, 1960, i članke 9–12 in Davidson, *Inquiries into Truth and Interpretation*, Oxford University Press, 1984.

Prilično očigledno, interpretacija zajedničkih predikata uveliko ovisi o indeksnim elementima u govoru, kao što su demonstrativi i vremena, budući da su to oni koji omogućuju predikatima i singularnim pojmovima da budu povezani s objektima i događajima u svijetu. Kako bi se smjestili indeksni elementi, teorije istine kakve je Tarski predložio moraju se mijenjati; priroda tih izmjena raspravljana je drugdje.<sup>8</sup>

Interpretacija predikata manje izravno povezanih s neupadljivim promatranjem u velikoj mjeri ovisi o uvjetnim vjerovatnostima koje pokazuju ono što akter komunikacije smatra kao dokaz za primjenu njegovih više teorijskih predikata. Očekuje se da će takvi dokazi pomoći u objašnjenju pogrešaka u primjeni opservacijskih predikata pod manje idealnim uvjetima.

Pristupu problemima značenja, vjerovanja i želje, koji sam iznio, nije, siguran sam da je to jasno, namjera da baci izravno svjetlo na to kako se u stvarnom životu međusobno razumijemo, a još manje kako svladavamo prve pojmove i naš prvi jezik. Angažirao sam se u pojmovnoj vježbi koja je usmjerena na otkrivanje zavisnosti između naših osnovnih propozicionih mentalnih stanja na razini dovoljnoj da izbjegnemo pretpostavku da ih možemo dohvatiti - ili ih razumljivo pripisati drugima - jedan po jedan. Moj način obavljanja ove vježbe bio je pokazati kako je u načelu moguće doći do svih odjednom.

Ono što čini zadaću uopće izvedivom jeste struktura normativnog karaktera misli, želje, govora i djelovanja nametnuta ispravnim pripisivanjem stavova drugima, a time i interpretacije njihovog govora i objašnjenja njihovih djelovanja. Ono što sam rekao o normama koje upravljaju našim teorijama o intenzionalnom pripisivanju, je na ključnim tačkama sirovo, nejasno ili previše kruto. Način da se poboljša naše razumijevanje takvog razumijevanja jest poboljšati naše shvaćanje standarda racionalnosti koji su implicirani u svim interpretacijama misljenja i djelovanja.

Preveo s engleskog: Nijaz Ibrulj

---

<sup>8</sup> Vidi Alfred Tarski, “*The Concept of Truth in Formalized Languages*”, u: *Logic, Semantics, Metamathematics*, Oxford University Press, 1956. O primjeni općega pristupa Tarskog prirodnom jeziku raspravlja se u Davidson, *Inquiries into Truth and Interpretation*.