
David Lewis

Prevedel Borut Trpin

Trnuljčica: odgovor Elgu*

1. Problem

Raziskovalci Eksperimentalnega filozofskega laboratorija so se odločili, da izvedejo sledeč poskus. Najprej bodo Trnuljčici povedali vse, kar vam bom povedal v tem odstavku, poskrbeli pa bodo tudi za to, da jim bo povsem verjela. Nato jo bodo v nedeljo uspavali. V ponedeljek jo bodo za kratek čas zbudili. Najprej ji ne bodo povedali, kateri dan je, kasneje pa ji bodo zaupali, da je ponedeljek. Zatem ji bodo izbrisali spomin na budnost. Morda jo bodo za kratek čas zopet zbudili v torek, odvisno od meta poštenega kovanca: če pade grb, jo bodo zbudili samo v ponedeljek, če cifra pa tudi v torek. V sredo se bo poskus zaključil in Trnuljčica se bo lahko zbudila. Ta tri možna kratka bujenja med poskusom bodo nerazločljiva: v ponedeljek bo imela dostop do enakih podatkov ne glede na izid meta kovanca, in če jo zbudijo v torek, bo zaradi ponedeljkovega izbrisa spomina imela dostop do povsem enakih podatkov kot ob ponedeljkovem bujenju. Naučili pa jo bodo, da bo lahko svoje bujenje v sredo, ko se poskus zaključi, razločila od bujenj v ponedeljek in torek, in seveda tudi od vseh prejšnjih in kasnejših bujenj.

Predpostavimo, da je Trnuljčica vzor verjetnostne racionalnosti in ima vedno natančno takšne stopnje prepričanj (subjektivne verjetnosti), kot bi jih morala imeti. Funkcijo njenih stopenj prepričanj bomo preverili trikrat. Naj bo P njena funkcija stopenj prepričanj, takoj ko se zbudi v ponedeljek. Naj bo P_+ njena funkcija stopenj prepričanj, takoj ko izve, da je ponedeljek, in naj bo P_- njena funkcija stopenj prepričanj, tik preden jo uspavajo v nedeljo, ko že ve, kako bo poskus potekal.

* Lewis, D. (2001). »Sleeping Beauty: Reply to Elga«. *Analysis*, 61(3), str. 171–176.

Kakšno stopnjo prepričanja naj bi Trnuljčica, takoj po bujenju v ponedeljek, pripisala hipotezi GRB, da je padel grb? Kakšno stopnjo prepričanja naj pripiše hipotezi CIFRA, da je padla cifra? Adam Elga (2000) trdi, da je $P(\text{GRB}) = 1/3$, $P(\text{CIFRA}) = 2/3$. S tem se ne strinjam in trdim, da je $P(\text{GRB}) = P(\text{CIFRA}) = 1/2$.

Nisem še opredelil, ali se met kovanca zgodi pred ali po ponedeljkovem bujenju. Elgov argument sicer temelji na tem, da kovanec mečejo nakladno; vendar pa meni, in tu se strinjava, da bi moral biti naš odgovor na vprašanje v obeh primerih enak. Moj argument velja za obe možnosti.

2. Skupno izhodišče

Pri najinem nestrinjanju je še posebej zanimivo, da se skoraj v vsem strinjava (vključno z mnogimi stvarmi, s katerimi se marsikdo ne bi). Naj torej najprej orišem najino skupno izhodišče.

Strinjava se, da obstajata dva tipa možnih pripisovanj stopenj prepričanj. Prvi tip se nanaša na to, kateri možni svet je dejanski; drugi tip pa se nanaša na to, kdo in kdaj nekdo je in v kakšni vrsti sveta živi. Tako kot Quine (1969) bomo drugi tip opredelili kot razrede osredinjenih svetov: možni svetovi z označenimi posamezniki-ob-nekem-času znotraj njih. Imenujmo razred osredinjenih svetov osredinjene možnosti. (Prvi tip, neosredinjene možnosti, bi lahko prikazali kot razrede neosredinjenih možnih svetov; vendar to ni potrebno, saj lahko neosredinjene možnosti zaobsežemo pod osredinjene.) Lahko se zgodi, da se dva osredinjena svetova nahajata znotraj istega neosredinjenega možnega sveta: le njuni označeni posamezniki-ob-nekem-času se razlikujejo. V takih primerih ta dva svetova imenujemo *kolocirana*.

Ko se Trnuljčica med poskusom zbudi, so z vsemi njenimi podatki skladne tri osredinjene epistemske možnosti:

G_1 : GRB in je ponedeljek.

C_1 : CIFRA in je ponedeljek.

C_2 : CIFRA in je torek.

Elga zapiše, da »ker je zanjo situacija, ko je v C_1 , subjektivno povsem enaka kot, ko je v C_2 , in ker so povsem enake propozicije resnične ne glede na to, ali je v C_1 ali C_2 , celo zelo omejen princip indiferentnosti vodi do sklepa, da mora imeti enako stopnjo prepričanja v obe možnosti« (Elga, 2000: 144). S »propozicijo« meni neosredinjeno možnost. Razlog, da so enake propozicije resnične ne glede na to, ali je Trnuljčica v C_1 ali C_2 , je v tem, da so osredinjeni svetovi, ki so člani C_1 , kolocirani z ustrežajočimi člani C_2 .

Sprejemam tudi Elgov »zelo omejen princip indiferentnosti«.¹ Strinjava se torej tudi v tem:

$$(1) \quad P(C_1)=P(C_2).$$

Ko Trnuljčica med svojo ponedeljkovo budnostjo izve, da je ponedeljek, bi se morala njena funkcija stopenj prepričanj spremeniti iz P v P_+ . Z Elgom se strinja, da do te spremembe pride ob pogojevanju z njenimi novimi podatki, ki jih lahko izrazimo kot $ne-C_2$ ali kot $(G_1 \vee C_1)$. Torej

$$(2) \quad \begin{aligned} P_+(GRB) &= P(GRB \mid G_1 \vee C_1), \\ P_+(CIFRA) &= P(CIFRA \mid G_1 \vee C_1). \end{aligned}$$

Iz celote njenih podatkov na začetku ponedeljkove budnosti sledi, da je GRB resničen, če je G_1 v tem trenutku resničen; in podobno za CIFRA in $(C_1 \vee C_2)$. Njeno pogojno stopnjo prepričanj lahko enostavno izrazimo kot razmerje brezpogojnih stopenj prepričanj. Strinjava se torej, da

$$(3) \quad \begin{aligned} P(GRB \mid G_1 \vee C_1) &= P(G_1 \mid G_1 \vee C_1) = P(G_1)/[P(G_1) + P(C_1)], \\ P(CIFRA \mid G_1 \vee C_1) &= P([C_1 \vee C_2] \mid G_1 \vee C_1) = P(C_1)/[P(G_1) + P(C_1)], \end{aligned}$$

in dalje, da

$$(4) \quad \begin{aligned} P(GRB) &= P(G_1), \\ P(CIFRA) &= P(C_1 \vee C_2) = P(C_1) + P(C_2). \end{aligned}$$

Strinjava se tudi glede

- (5) $P_+(GRB)=1/2=P_+(CIFRA)$ in
- (6) Trnuljčica ne dobi nobenih dodatnih neosredinjenih podatkov, ki so relevantni za GRB proti CIFRA v času od njene funkcije stopenj prepričanj P do funkcije stopenj prepričanj P . Edini podatki, ki jih pridobi, so osredinjeni podatki,² da je trenutno bodisi v ponedeljkovem bujenju ali pa torkovem bujenju, t.j. $(G_1 \vee C_1 \vee C_2)$.

¹ Če ne bi upoštevali kolokacije ustrežajočih članov dveh epistemskih možnosti, bi dobili manj omejen princip indiferentnosti, iz katerega bi sledilo, da $P(G_1) = P(C_1) = P(C_2)$. Na tak način bi po bližnjici prišli do Elgovega sklepa – bistveno preuranjeno in Elga se je tej poti bistro izognil. Sledijo nenavadne posledice: na primer, da za enakovrednost $P(G_1)$ in $P(C_1)$ poštenost kovanca sploh ni relevantna, četudi bi Trnuljčica izvedela, da je kovanec obtežen tako, da je verjetnost padca grba 999 proti 1!

² Naj izpostavim: Elga govori o »novih informacijah«. Vendar pa v njegovi terminologiji »osredinjena informacija« sploh ne služi kot »informacija« (prav tam: 145, op. 3). Kljub temu bi jo vseeno lahko imel za »podatek«.

Tukaj se najino skupno izhodišče zaključí, zato se bom posvetil najinima analizama problema, vključno z najinimi različnimi zaključki glede $P(\text{GRB})$.

3. Elgov argument

Elga upošteva zgolj možnost, ko raziskovalci mečejo kovanec po ponedeljkovem bujenju, in iz kasnejšega sklepa na zgodnejše:

$$(E1) \quad P_+(\text{GRB}) = 1/2 \quad (\text{premisa}).$$

$$(E2) \quad \therefore P(G_1)/[P(G_1) + P(C_1)] = 1/2 \quad (E1, 2, 3).$$

$$(E3) \quad \therefore P(G_1) = P(C_1) \quad (E2).$$

$$(E4) \quad \therefore P(G_1) = P(C_1) = P(C_2) = 1/3 \quad (E3, 1).$$

$$(E5) \quad \therefore P(\text{GRB}) = 1/3, P(\text{CIFRA}) = 2/3 \quad (E4, 4).$$

Quod erat demonstrandum; vendar pa opazimo tudi dve nadaljnji posledici.

$$(E6) \quad \therefore P_+(\text{GRB}) = P(\text{GRB}) + 1/6 \quad (E1, E5).$$

$$(E7) \quad \therefore \text{Spremembe stopnje prepričanja iz } P_- \text{ v } P \text{ niso povzročili novi relevantni neosredinjeni podatki; bodisi je sploh niso povzročili relevantni podatki ali pa so bili osredinjeni podatki } (G_1 \vee C_1 \vee C_2) \text{ pomembni za GRB proti CIFRA} \quad (E5, 5, 6).$$

4. Moj argument

Ob upoštevanju obeh možnosti svoj argument izvajam od zgodnejšega k kasnejšemu.

$$(L1) \quad \text{Zgolj novi relevantni podatki, bodisi osredinjeni ali ne, lahko povzročijo spremembo stopnje prepričanja; podatki } (G_1 \vee C_1 \vee C_2) \text{ pa za GRB proti CIFRA niso relevantni} \quad (\text{premisa}).$$

$$(L2) \quad \therefore P(\text{GRB}) = 1/2 = P(\text{CIFRA}) \quad (L1, 5, 6).$$

Quod erat demonstrandum; opazimo pa pet posledic.

$$(L3) \quad \therefore P(G_1) = 1/2 = P(C_1) + P(C_2) \quad (L2, 4).$$

$$(L4) \quad \therefore P(C_1) = 1/4 = P(C_2) \quad (L3, 1).$$

$$(L5) \quad \therefore P(\text{GRB} \mid G_1 \vee C_1) = 2/3, \\ P(\text{CIFRA} \mid G_1 \vee C_1) = 1/3 \quad (L3, L4, 3).$$

$$(L6) \quad \therefore P_+(\text{GRB}) = 2/3, \\ P_+(\text{CIFRA}) = 1/3 \quad (L5, 2).$$

$$(L7) \quad \therefore P_+(\text{GRB}) = P(\text{GRB}) + 1/6 \quad (L2, L6).$$

5. Oris najinega nestrinjanja

Elga zavrača mojo premiso: njegova (E7) je v protislovju z mojo (L1). Jaz zavračam njegovo premiso: moja (L6) je v protislovju z njegovo (E1). Od tod izvira najino razhajanje. Zavreči morava eno izmed premis. Če obdrživa obe (in najino skupno izhodišče), prideva v protislovje. Kaj še lahko povemo?

Elgov razlog za zavrnitev moje premise je, da gre tja, kamor ga vodi argument. Preseneča ga, da lahko pride do »[spremembe stopnje prepričanja] v obdobju, koakterka niti ne prejme nobene nove informacije« – t.j. nobenih novih neosredinjenih podatkov – »niti ne utрпи kognitivne okvare« (prav tam: 146). Trdim, da bi bilo prav tako presenetljivo, in prav tako sumljivo, če bi ugotovili, da je osredinjeni podatek (G_1 v C_1 v C_2) relevanten za GRB proti CIFRA; ali če bi odkrili, da je Trnuljčica utrpela kognitivno okvaro. (Izbris spomina je vsekakor kognitivna okvara; vendar pa se to zgodi kasneje, torej je nerelevantno za spremembo med P. in P.) To mu dopustim; bi pa bilo lepše, če bi bila njegova zavrnitev moje premise neodvisno utemeljena.

Trdim, da je moja zavrnitev Elgove premise neodvisno utemeljena. Od kje jo je Elga sploh dobil? V času P_+ , v Elgovem primeru kovanec mečejo za tem, Trnuljčica ve, da bo prišlo do meta poštenega kovanca. Imamo tudi znan princip, da morajo biti stopnje prepričanj o prihodnjih slučajnih dogodkih enake znanim slučajem (glej Mellor, 1971; Lewis, 1980). Prav iz tega principa sva dobila skupno (5), ki ureja tako primere, ko kovanec mečejo pred ponedeljkovim bujenjem in po njem. Po tem principu bi sledilo tudi, da $P_+(GRB) = \text{možnost}^*(GRB) = 1/2$ in da $P_+(CIFRA) = \text{možnost}(CIFRA) = 1/2$, iz česar sledi Elgova premisa (E1).

Naj omenim, da ta princip vključuje predpogoj, ki smo ga izpolnili, ko smo pridobili (5), v Elgovem izvajanju na premiso (E1) pa ni izpolnjen. Zamislimo si, da obstaja prerok, za katerega vemo, da zelo uspešno napoveduje prihodnost, zaradi česar zelo resno obravnavamo hipotezo, da po nekakšni vzvratni vzročni verigi prejema vesti iz prihodnosti. Prerok nam le redko neposredno pove, kaj se bo zgodilo, nam pa pogosto svetuje, kakšno stopnjo prepričanja naj imamo glede nekega izida. Občasno so njegovi nasveti v nasprotju z znanimi možnostmi (ang. *chances*), po katerih bi sicer uravnavali svoje stopnje prepričanj. Kaj naj torej storimo? Če je prerok dovolj pogosto pravilno napovedal prihodnost, potem trdim, da lahko znane možnosti zanemarimo.

Ko torej Trnuljčica med ponedeljkovo budnostjo izve, da je ponedeljek, ali enakovredno ne- C_2 , pridobi podatke – osredinjene podatke – glede prihodnosti: da trenutno ni v njej. To so novi podatki: preden je izvedela, da je ponedeljek, jih ni imela. Naj poudarim, podatkov ne dobi od preroka ali preko vzvratne vzročnosti, hkrati pa tudi ne z usklajevanjem svoje stopnje prepričanja z znanimi možnostmi.

* Op. p.: V izvorniku *chance*, slučaj oz. slučajna možnost.

Vest je relevantna za GRB, ker njeno stopnjo prepričanja dvigne za $1/6$; glej mojo (L7). Elga se strinja; glej njegovo (E6). Predpogoj je torej pomemben in zato ne moremo izpeljati, da $P_+(GRB) = \text{možnost}(GRB)$ in $P_+(CIFRA) = \text{možnost}(CIFRA)$. Priznam, da je to nova in presenetljiva posledica predpogoja, zato se Elgu zahvaljujem, da je problem izpostavil. Kljub tej posledici smatram predpogoj kot prepričljiv, neodvisno od tega, da bi rad šel tja, kamor me vodi moj argument.³

Literatura

Elga, A. (2000). »Self-locating belief and the Sleeping Beauty problem«. *Analysis*, 60(2), str. 143–47.

Lewis, D. (1980). »A subjectivist's guide to objective chance«. V Jeffrey, R. C. (ur.), *Studies in Inductive Logic and Probability*, Vol. 2, Berkeley: University of California Press, str. 263–293. Ponatisnjeno v Lewis, D., *Philosophical Papers*, Vol. 2., Oxford: Oxford University Press.

Mellor, D. H. (1971). *The Matter of Chance*. Cambridge: Cambridge University Press.

Quine, W. V. (1969). »Propositional objects«. V Quine, W. V., *Ontological Relativity & Other Essays*, New York: Columbia University Press.

³ Za pripombe se zahvaljujem Adamu Elgu in uredniku *Analysis*.