
Adam Elga

Prevedel Borut Trpin

Prepričanja o lastnem položaju in Trnuljčičin problem*

Poleg negotovosti o tem, kakšen je svet, smo lahko v negotovosti tudi glede lastnega prostorskega ali časovnega položaja v svetu. Rad bi zastavil problem, ki nastane v odnosu med tema dvema vrstama negotovosti, ga razrešil in iz njega izpeljal dva sklepa.

1. Trnuljčičin problem¹

Raziskovalci bodo uspavali Trnuljčico. Njeno spanje bo trajalo dva dneva, vmes pa jo bodo enkrat ali dvakrat za kratek čas zbudili. Ali jo bodo zbudili enkrat ali dvakrat je odvisno od meta kovanca (grb: enkrat, cifra: dvakrat). Po vsakem bujenju jo bodo ponovno uspavali s posebno tableto, ki povzroči, da se svoje budnosti kasneje ne bo spominjala.²

Ko se Trnuljčica prvič zbudi, do katere stopnje je lahko prepričana, da je padel grb?

Prvi odgovor: 1/2, seveda! Na začetku je bila prepričana, da je kovanec pošten in zato je bila njena izhodiščna stopnja prepričanja, da bo padel grb, 1/2. Ko se zbu-

* Elga, A. (2000). »Self-locating belief and the Sleeping Beauty problem«. *Analysis*, 60(2), str. 143–147.

¹ Tako je problem poimenoval Robert Stalnaker (ki se je s takimi primeri srečal v neobjavljenem delu Arnolda Zuboffa). Problem se pojavi kot Primer št. 5 v Piccione, 1997, kjer sta nakazana dva različna odgovora, vendar pa se avtor vzdrži sodbe, kateri odgovor je pravilen (1997: 12–14). Podoben primer z metodo pravične loterije analizira Aumann, 1997. Če njegovo analizo prilagodimo za Trnuljčičin problem, dobimo enak odgovor, kot ga bom zagovarjal v razdelku 2. Za razliko od njegovega argumenta pa moje izvajanje ni vezano na stave.

² Tableta deluje tako, da prepričanja povrne v stanje, v katerem so bila pred začetkom poskusa. Če se zdi obstoj take tablete neverjeten, naj izpostavim, da je mogoče problem zastaviti tudi brez te tablete – relevantno je le, da uspavana oseba verjame, da poskus poteka tako, kot sem ga opisal.

di, ne dobi nobene nove informacije (že ves čas je vedela, da se bo zbudila). Njena stopnja prepričanja, da je padel grb, bi torej morala ostati $1/2$.

Drugi odgovor: $1/3$, seveda! Predstavljajmo si, da se poskus večkrat ponovi. Na dolgi rok se izkaže, da je $1/3$ bujenj t.i. grb-bujenj – bujenj, ko pade grb. Ob vsakem bujenju bi torej morala imeti stopnjo prepričanja $1/3$, da gre za grb-bujenje, in, posledično, stopnjo prepričanja $1/3$, da je padel grb. To velja tudi v trenutnem primeru, ko poskus poteka samo enkrat.

Zagovarjal bom stališče, da je pravilen odgovor $1/3$.

2.

Predpostavimo, da se prvo bujenje zgodi v ponedeljek, drugo (če sploh) pa v torek. Ko se Trnuljčica zbudi, je torej lahko gotova, da je v eni izmed teh treh »situacij«:

G_1 GRB in je ponedeljek.

C_1 CIFRA in je ponedeljek.

C_2 CIFRA in je torek.

Naj izpostavim, da razlika med tem, ali je Trnuljčica v C_1 ali C_2 , ni razlika v tem, kateri možni svet je dejanski, temveč razlika v njenem časovnem položaju znotraj sveta. (Za bolj tehnično obravnavo lahko prevzamemo podobno ogrodje, kot ga predlaga Lewis, 1983. Po tem predlogu osnovne alternative, ki razdelijo stopnje prepričanj, niso možni svetovi, temveč osredinjeni možni svetovi: možni svetovi, ki so vsak zase opremljeni z označenimi posamezniki in časom. V takem ogrodju bi G_1 , C_1 in C_2 predstavljali množice osredinjenih svetov.)

Naj bo P funkcija stopenj prepričanj, ki naj bi jo Trnuljčica imela po prvem bujenju. Po prvem bujenju je gotova o sledečem: v situaciji G_1 je, če in samo če je kovanec padel na grb. Potemtakem je izračun $P(G_1)$ zadosten za razrešitev Trnuljčičinega problema. Najprej bom pokazal, da $P(C_1) = P(C_2)$, in potem, da je $P(C_1) = P(G_1)$.

Če bi (po prvem bujenju) izvedela, da je padla cifra, bi to zadoščalo za to, da spozna, da je bodisi v C_1 ali C_2 . 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. Vendar pa bi morala biti njena stopnja prepričanja, da je v C_1 , potem ko izve, da je padla cifra, enaka kot njena pogojna stopnja prepričanja $P(C_1 \mid C_1 \text{ ali } C_2)$ in enako za C_2 . Od tod sledi $P(C_1 \mid C_1 \text{ ali } C_2) = P(C_2 \mid C_1 \text{ ali } C_2)$, in torej $P(C_1) = P(C_2)$.

Raziskovalci mečejo pošten kovanec, da določijo, ali bodo Trnuljčico zbudili enkrat ali dvakrat. To storijo tako, da bodisi

- (1) *najprej* mečejo kovanec in potem Trnuljčico zbudijo bodisi enkrat ali dvakrat glede na izid; ali
- (2) najprej Trnuljčico zbudijo enkrat *in* potem mečejo kovanec, da ugotovijo, ali jo bodo zbudili še enkrat.

Njena stopnja prepričanja (ko se zbudi), da je padel grb, bi morala biti enaka ne glede na to, ali raziskovalci uporabijo metodo (1) ali (2). Brez izgube splošnosti, torej predpostavimo, da uporabijo metodo (2).

Zdaj: če (ko se zbudi) Trnuljčica izve, da je ponedeljek, je to enakovredno spoznanju, da je bodisi v G_I ali C_I . Njena stopnja prepričanja, da je v G_I , predstavlja njeno stopnjo prepričanja, da bo pošten kovanec, ki ga bodo raziskovalci kmalu vrgli, padel na grb. Nerelevantno je, da bodo Trnuljčico zbudili tudi naslednji dan, če in samo če kovanec pade na cifro – v tem primeru bi morala biti njena stopnja prepričanja, da je padel grb $1/2$. Ampak njeno prepričanje, da je padel grb (po tem, ko izve, da je ponedeljek) bi moralo biti enako kot njeno pogojno prepričanje $P(G_I | G_I \text{ ali } C_I)$. Torej $P(G_I | G_I \text{ ali } C_I) = 1/2$, in torej $P(G_I) = P(C_I)$.

Če te rezultate združimo, ugotovimo, da $P(G_I) = P(C_I) = P(C_2)$. Ker se te tri stopnje prepričanj seštejejo v 1, je $P(G_I) = 1/3$

3.

Naj bo G propozicija, da je kovanec padel na grb. Preden Trnuljčico uspavajo, je njena stopnja prepričanja v G $1/2$. Ravno kar pa sem zagovarjal, da bi se njena stopnja prepričanja, ko jo zbudijo v ponedeljek, morala spremeniti v $1/3$. Ta sprememba prepričanja je nenavadna. Do spremembe ne pride, ker bi Trnuljčica prejela nove informacije – že prej je vedela, da se bo zbudila v ponedeljek.³ (Lahko celo predpostavimo, da je Trnuljčica že na začetku poskusa vedela, natančno, kakšna čutna izkustva bo imela, ko se bo zbudila v ponedeljek.) Ta sprememba prepričanj tudi ni posledica neke kognitivne okvare v vmesnem času – spomnimo; tableto za pozabo prejme šele, ko je prvič že povsem zbudjena. Kaj torej upravičuje to spremembo stopnje prepričanja?

Odgovor je, da je Trnuljčica prešla iz situacije, v kateri je bil njen lasten časovni položaj nerelevanten za resnico G , v situacijo, kjer upošteva svoj časovni položaj

³ Ko trdim, da akterka pridobi nove informacije (kot bom trdil tudi v nadaljevanju), trdim, da pridobi podatke, ki izključujejo možne svetove, ki jih prejšnji podatki niso izključevali. Z drugimi besedami: akterka pridobi nove informacije, ko izve resnico propozicije, ki jo lahko izrazimo z večnim stavkom (Quine, 1960: 191) v dovolj bogatem jeziku.

kot relevantnega za resnico G .⁴ Predpostavimo, na primer, da Trnuljčica na začetku poskusa ni prepričana ali je ura 1:01 ali 1:02. V tem trenutku svojega časovnega položaja ne smatra kot relevantnega za resnico G : njena stopnja prepričanja v G ob pogoju, da je ura 1:01, je $1/2$. Prav tako je tudi njena stopnja prepričanja v G ob pogoju, da je ura 1:02, $1/2$.

Nasprotno pa (ob predpostavki, da prepričanja posodablja racionalno), ko Trnuljčico zbudijo v ponedeljek, svoj trenutni časovni položaj smatra kot relevantnega za resnico G : njena stopnja prepričanja v G ob pogoju, da je ponedeljek, je $1/2$, njena stopnja prepričanja v G ob pogoju, da je torek, pa 0. V ponedeljek se njena brezpogojna stopnja prepričanja v G razlikuje od $1/2$, saj predstavlja uteženo povprečje teh dveh pogojnih stopenj prepričanj – se pravi, uteženo povprečje $1/2$ in 0.

Nobeno presenečenje ni, da se to, ali akterka upošteva svoj časovni položaj kot relevantnega za resnico propozicije, skozi čas spreminja. Kar je bolj presenetljivo – in to je prvi sklep –, je, da se tovrstna sprememba lahko zgodi v obdobju, ko akterka niti ne prejme nobene nove informacije niti ne utрпи kognitivne okvare.

Na začetku poskusa je imela stopnjo prepričanja $1/2$ v G . Hkrati pa je bila gotova, da bo ob ponedeljkovem bujenju njena stopnja prepričanja v H padla na $1/3$ – čeprav je bila gotova, da vmes ne bo pridobila nobene nove informacije ali pa utrpela neke kognitivne okvare. Primer Trnuljčice tako predstavlja novo vrsto protiprimera Bas van Fraassenovemu »principu odsevanja« (1984: 244, 1995: 19), celo zelo dodelani varianti, iz katere sledi naslednje:

Vsaka akterka, ki je gotova, da bo imela jutri stopnjo prepričanja x v propozicijo R (in vmes ne bo prejela novih informacij niti ne bo utrpela kognitivne okvare), bi morala že sedaj imeti stopnjo prepričanja x v R .⁵

David Lewis se je nekoč vprašal, »kaj bi se zgodilo s teorijo odločanja, če bi [nandomestili prostor možnih svetov s prostorom osredinjenih možnih svetov]?« in odgovoril: »Ne prav dosti« (Lewis, 1983: 149). Drugi sklep, ki sledi iz Trnuljčičinega problema, pa je, da se nekaj zgodi: poraja se novo vprašanje, kako naj bi racionalni akterji skozi čas posodabljali svoja prepričanja.⁶

⁴ Ko trdim, da akterka upošteva svoj časovni položaj kot relevanten za resnico določene propozicije, trdim, da obstaja tak čas t , da so akterkina prepričanja združljiva s položajem v t , in da se njena stopnja prepričanja v propozicijo ob pogoju, da je v položaju t , razlikuje od njenih brezpogojnih stopenj prepričanj v to propozicijo.

⁵ Za ugotovitev, da je odgovor $1/3$ v nasprotju s principom odsevanja, se zahvaljujem Nedu Hallu.

⁶ Pri nastajanju tega prispevka so mi bili v pomoč Jamie Dreier, Gary Gates, Ned Hall, Vann McGee, Robert Stalnaker, Roger White, Sarah Wright, udeleženci konference na Brown University leta 1999 (tam sem predstavil zgodnejšo verzijo tega prispevka) in anonimni recenzent. Za pomoč se jim zahvaljujem.

Literatura

Aumann, R. J., Hart, S. in Perryi, M. (1997). »The forgetful passenger«. *Games and Economic Behavior*, 20, str. 117–120.

Lewis, D. (1983). »Attitudes *de dicto* and *de se*«. V Lewis, D., *Philosophical Papers, Volume I*, New York: Oxford University Press, str. 133–159.

Piccione, M. in Rubenstein, A. (1997). »On the interpretation of decision problems with imperfect recall«. *Games and Economic Behavior*, 20, str. 3–24.

Quine, W.V. (1960). *Word and Object*. Cambridge, Mass.: The MIT Press.

van Fraassen, B. C. (1984). »Belief and the will«. *Journal of Philosophy*, 81, str. 235–256.

van Fraassen, B. C. (1995). »Belief and the problem of Ulysses and the sirens«. *Philosophical Studies*, 77, str. 7–37.