ANαliZA

časopis za kritično misel

Letnik 7, številka 3, september 2003 · ISSN 1408-2969

RAZPRAVE

Janez Bregant

Ekstrinzičnost mentalnih lastnosti kot problem mentalne vzročnosti

Pedagoška fakulteta Univerze v Mariboru

Citiranje: Janez Bregant, »Ekstrinzičnost mentalnih lastnosti kot problem mentalne vzročnosti«, *Analiza: časopis za kritično misel* 7, št. 3 (2003): 3–15.

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/deed.sl.

Izdajatelj: Društvo za analitično filozofijo in filozofijo znanosti, Ljubljana.

Elektronska različica je bila leta 2026 rekonstruirana iz izvirnih stavnih datotek in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

Janez Bregant
Pedagoška fakulteta Univerze v Mariboru

EKSTRINZIČNOST MENTALNIH LASTNOSTI KOT PROBLEM MENTALNE VZROČNOSTI

Komputacionalizem je vir težav tako za psihologijo in njeno pojasnjevanje dogodkov kot tudi za filozofsko razpravo o mentalni vzročnosti, saj predpostavlja, da vsebine naših želja, prepričanj, namer in ostalih intencionalnih stanj, ker ima vzročno vlogo le njihova oblika oziroma zgradba, niso vzročno relevantne. Intencionalnim stanjem preti nevarnost, da bo tistemu, kar je njihova glavna značilnost, tj. vsebina, in zaradi katere naj bi sploh imela vzročno vlogo, odvzeta kakršnakoli relevantnost pri povzročanju dogodkov, kar imenujemo problem ekstrinzičnosti mentalnih lastnosti. V članku je poleg argumenta za to predstavljena še Dretskejeva rešitev omenjenega problema, ki temelji na razlikovanju med sprožitvenim in strukturalnim vzrokom, ter moj ugovor, ki ji očita, da vodi to v epifenomenalizem mentalnih lastnosti, zaradi česar ne predstavlja prave alternative.

Ključne besede: komputacionalizem, eksternalizem, internalizem, mentalna vzročnost, ekstrinzičnost mentalnega, epifenomenalizem mentalnih lastnosti.

Uvod

Motivacija za eno vrsto težav, ki naj bi jih imeli pri zagovarjanju vzročne učinkovitosti mentalnih lastnosti, izhaja iz stališča, ki je v zadnjih desetletjih v sodobni psihologiji pridobilo na moči in veljavi, pojasnjuje pa, kako moramo razumeti mentalne procese in psihologijo nasploh.

Vzemimo kot primer računalnik z določenim programom, na primer s programom, ki ugotavlja, koliko plastenk radenske je na zalogi. Predpostavite, da se odpravite v trgovino in kupite pet plastenk radenske. Ko pridete na blagajno, prijazna uslužbenka poišče kode na plastenkah in jih skenira. V tem trenutku opravi računalnik vrsto operacij (ugotavljanje števila kupljenih plastenk, odštevanje števila plastenk, ki ste jih kupili, od števila plastenk, ki so bile na zalogi, spreminjanje podatkov o številu plastenk na zalogi ipd.), kar ga na koncu

pripelje do števila plastenk, ki so še na zalogi. Prodano število plastenk radenske tako za računalnik predstavlja vnos (input), zapisan v nekem *formalnem jeziku*, recimo v nizu ničel in enic, število plastenk radenske, ki so še na zalogi, pa njegov iznos (output), zapisan v istem formalnem jeziku. Ko želi uslužbenec, ki nabavlja plastenke radenske, preveriti, koliko jih je še na zalogi, in ugotoviti, ali je čas za nabavo nove pošiljke, mu računalnik pove, da je število plastenk radenske na zalogi, ker jih je bilo v zadnjih petih dneh prodanih petsto, dvesto petdeset.[1]

Podobno naj bi po omenjenem psihološkem stališču veljalo tudi za mentalne procese. Predpostavite, da vas med igranjem košarke nekdo iz nasprotnega moštva nenamerno udari po glavi. Takoj začutite ostro bolečino, ki se počasi razleze po celem telesu. Kako bi to pojasnili? Udarec, ki ste ga dobili, je bil skladno z opisanim primerom računalniškega ugotavljanja števila plastenk radenske na zalogi vnos, bolečina, ki ste jo začutili, potem ko so se v vas končali vsi "računalniški"/nevrofiziološki procesi, pa iznos. Tako je sestavljanje psihološke teorije kot pisanje programa. Za vsak čutni impulz je določen biološko-fizični proces, skozi katerega moramo, da bi na koncu prišli do določenega mentalnega pojava. Stališče v sodobni psihologiji, ki se pri pojasnjevanju mentalnih procesov sklicuje na računalniške procese in trdi, da moramo mentalne procese razumeti kot računalniške procese, se imenuje *komputacionalizem*.

Zakaj pa je komputacionalizem vir težav tako za psihologijo in njeno pojasnjevanje dogodkov kot tudi za filozofsko razpravo o mentalni vzročnosti? Vrnimo se na naš primer računalniškega ugotavljanja števila plastenk radenske na zalogi. Kar moramo v celi zgodbi poudariti, je, da za računalnik to, kaj niz ničel in enic *pomeni*, sploh ni pomembno. Ali niz ničel in enic pomeni, na primer, hitrost pretoka Drave ali debelino snežne odeje na Pohorju ali matično številko občana ali nič, računalnik ne zanima. Vnos istega niza ničel in enic bo po opravljenih *istih* računalniških operacijah vedno pripeljal do iznosa istega niza ničel in enic, različen pa, normalno, vedno do različnega. Očitno je, da pomen ničel in enic za računalniškega programerja ni nekaj, kar je in mora biti vključeno v računalniški program. Pomen takšnih in podobnih nizov za računalnik ni pomemben, tisto, kar je zanj pomembno, je njihova *oblika* ali *vzorec*. Tako računalnik ni "semantični stroj", ampak "sintaktični stroj", ki deluje zaradi oblike simbolov, ne pa zaradi njihovih pomenov. (*glej* Kim, 1996: 145)

In zakaj je to problematično? Predpostavimo, da je naša notranja predstava, imenujmo jo *k*, predstava enega dejstva zunanjega sveta, recimo, na mizi so beli tulipani, imenujmo ga *L*. Dejstvo zunanjega sveta, tj., da so na mizi beli tulipani, ki smo ga imenovali *L*, je *vsebina* ali pomen (tisto, na kar se mentalno stanje nanaša), tj. semantika, *k*-ja.[2] Toda ker komputacionalizem predpostavlja, da so notranje predstave pogosto zapisane v sistemu, ki je podoben našemu jeziku, mora *k* imeti tudi svoj formalni ustroj. Tukaj pa se začnejo težave. Če je naša

[1] Za izvirni primer *glej* Kim, 1996: 144.
[2] Mentalna stanja z vsebino, ki je deloma določena z zunanjimi dejstvi, tj. dejstvi sveta, se imenujejo *intencionalna stanja*.

ugotovitev, da je za računalnik pomembna oblika niza, ne pa njegov pomen, pravilna, je tisto, kar določa potek računalniškega procesa, katerega produkt je *k*, njegova oblika. Dejstvo, da so na mizi beli tulipani in ne oranžne gerbere, nima na to, kar notranja predstava *k* povzroča, nobenega vpliva. Tisto, kar je v zvezi s *k* vzročno relevantno, je njegova formalna zgradba (oblika in ne vsebina) in tako njegov formalni ustroj povzroča druge notranje predstave organizma, na primer, bele tulipane je treba dati v vodo.[3] Tako vsebine naših želja, prepričanj, namer in ostalih intencionalnih stanj, ker ima vzročno vlogo le njihova oblika, za psihologijo *niso* vzročno relevantne.

To pa ne predstavlja težave samo za psihologijo, ampak tudi za filozofsko razpravo o mentalni vzročnosti, saj potemtakem vzročna neučinkovitost intencionalnih stanj (enega dela mentalnih stanj, tj. mentalnih stanj z vsebino) izhaja iz domala vseh fizikalističnih stališč o mentalnem. V prvem delu članka bom opisal eksternalizem in internalizem, prevladujoči teoriji o vsebini mentalnih stanj, v drugem pa argument zoper vzročno učinkovitost intencionalnih pojavov, ki temelji na eliminativnosti široke vsebine. V tretjem delu članka bom predstavil Dretskejev odgovor na omenjeni argument, v četrtem pa svoj ugovor.

I. Eksternalizem in internalizem

Vsebina mentalnih stanj je tesno povezana s tem, kaj se dogaja v svetu, tj. s tem, kaj se dogaja zunaj fizičnega subjekta. Glede na to, kaj se dogaja v nas (glede na živčno stanje, v katerem se nahajamo), ko vidimo pred seboj kozarec vode, se ljudje mogoče ne razlikujemo,[4] gotovo pa se razlikujemo glede tega, kaj se dogaja zunaj nas (glede vsebine naših predstav). Zdi se, da so naši opisi dogodkov zunaj nas ali njihova vsebina odvisni (tudi) od socialnega in zgodovinskega okolja, v katerem se nahajamo. Najslavnejši primer ali miselni poskus v sodobni filozofiji duha, ki zagovarja idejo, da tisto, kar želimo in v kar smo prepričani, ni odvisno samo od tega, kar se dogaja v naših možganih, ampak tudi od tega, kar se dogaja v zunanjem svetu, je v sedemdesetih letih prejšnjega stoletja razvil Putnam.

Predpostavite, da v enem delu vesolja obstaja planet, ki se od Zemlje razlikuje le v enem pogledu, tj. vodo, kot jo poznamo mi, nadomešča kemična substanca XYZ, drugače pa je popoln posnetek Zemlje. XYZ ima vse vizualne značilnosti naše vode – nima ne vonja ne okusa, je prozorna, gasi žejo in požare, zmrzne pri 0 stopinjah Celzija, topi sladkor in sol ipd. Tako na Zemlji dvojčici reke, jezera in morja napolnjuje XYZ in ne H_2O kot na Zemlji, njeni prebivalci pijejo in gasijo požare z XYZ in ne s H_2O kot mi, ledene kocke, na katerih strežejo viski, so zmrznjena XYZ in ne H_2O kot pri nas itd. Vsak izmed nas ima na Zemlji dvojčici svojo kopijo, ki je natančen fizični (molekulo po molekulo) posnetek njega

[3] Za izviren primer *glej* Kim, 1996: 145.

[4] Znanstvene raziskave so resda pokazale, da se živčna stanja ljudi, ko, na primer, zaznavajo rdečo barvo, razlikujejo, vendar tukaj zaradi enostavnosti primera predpostavimo, da so vsi ljudje, ko vidijo pred seboj kozarec vode, v istem živčnem stanju.

samega,[5] vsi njeni prebivalci pa govorijo slovenščino, ki se od naše razlikuje samo v tem, da se izraz voda nanaša na XYZ. Tako v mislih vašega dvojnika nastopa "tvoda",[6] v vaših pa voda ne glede na to, da sta popolna fizična duplikata, in ko ste vi prepričani, da je voda mokra, je vaš dvojnik prepričan, da je "tvoda" mokra. Vidimo, da razlike v vsebini misli in prepričanj med nami in našimi kopijami niso odvisne niti od razlik v fizičnih stanjih niti od razlik v mentalnih stanjih, ampak od razlik v našem odnosu in odnosu naših kopij do okolja. Zato imenujemo lastnosti, ki tvorijo vsebino mentalnih stanj, zunanje ali *ekstrinzične*, vsebino *široka*, stališče, ki to zagovarja, pa *eksternalizem*. Putnamov miselni poskus dokazuje, da vsebina želja, prepričanj, namer in ostalih intencionalnih stanj ni odvisna od naših nevrofizioloških stanj, ampak od socialnega in zgodovinskega okolja, v katerem živimo.

Vsi filozofi pa se s sklepom, da plavzibilnost psiholoških teorij zahteva upoštevanje široke vsebine intencionalnih stanj, ne strinjajo. Poudarjajo, da je s psihološkega vidika, tj. z vidika pojasnjevanja obnašanja, pomembno nekaj drugega. Res je, da ima vaše prepričanje, da voda topi sladkor in sol, preprosto zato, ker se nahajata na različnih planetih (v različnem okolju), drugačno (široko) vsebino kot prepričanje vašega dvojnika, da "voda" topi sladkor in sol. Ker vi mislite na H_2O, vaš dvojnik pa na XYZ, se vajini prepričanji razlikujeta. Verjetno temu ne bi nihče ugovarjal, kljub vsemu pa se zdi, da imate vi in vaš dvojnik nekaj skupnega, nekaj takega, kar je za razlago vajinega ravnanja s stališča psihologije pomembno. Ko vi mislite na vodo, imate pred očmi predstavo o substanci, ki je brez barve, vonja, okusa itd. Kaj pa vaš dvojnik? Tudi vaš dvojnik ima, ko misli na "tvodo", pred očmi predstavo o substanci, ki je brez barve, vonja, okusa itd. Zdi se, da pojem vode, čeprav se nanaša na različni substanci, vedno opisuje isto dejstvo sveta. Tako imata obe misli ali prepričanji o vodi s psihološkega stališča isti pomen.

S psihološkega stališča za Petra in njegovega dvojnika, na primer, ni pomembno to, na katero substanco se njuna prepričanja nanašajo, ali na H_2O ali na XYZ, ampak so zanju pomembna skupna prepričanja, ki jih imata o njej, na primer, nima ne vonja ne okusa, je prozorna, gasi žejo in požare, zmrzne pri 0 stopinjah Celzija, topi sladkor in sol ipd.; rekli bi lahko, da je s psihološkega stališča zanju pomemben *stereotip*, ki ga imata o vodi, skladno s katerim se ne glede na to, na kateri planet ju pošljemo, vedno obnašata na isti način,[7] na primer, če Peter na Zemlji dvojčici opazi požar, ga skuša pogasiti s "tvodo", in če Petrov dvojnik na Zemlji opazi požar, ga poskuša pogasiti z vodo. Zdi se, da razlika med prepričanji o vodi in prepričanji o "tvodi" za pojasnjevanje obnašanja sploh ni relevantna. Relevantno je tisto, kar je vam in vašemu dvojniku skupno, in sicer stereotip o vodi. Tako imata vaše psihološko življenje in psihološko življenje

[5] Dejstvo, da ima vaš dvojček namesto molekul H_2O v svojem telesu molekule XYZ, lahko zanemarimo, saj za argument, ki ga bomo razvili, ni pomembno.

[6] Voda, tj. XYZ, na Zemlji dvojčici (angl. "*twater*").

[7] Na isti način se obnašata tudi, če se spremenijo razmere na njunem domačem planetu, na primer, če na Zemlji čez noč vodo zamenja "tvoda", skuša Peter požar naslednji dan pogasiti s "tvodo", in če na Zemlji dvojčici čez noč "tvodo" zamenja voda, skuša Petrov dvojnik požar naslednji dan pogasiti z vodo.

vašega dvojnika nekaj skupnega, nekaj takega, kar bi prav tako lahko imenovali vsebina.

Vendar takšna vsebina ni odvisna od okolja, ampak od tega, kar se dogaja znotraj vas, tj. od nevrofizioloških lastnosti vaših možganskih procesov. Zato imenujemo lastnosti, ki tvorijo takšno vsebino, notranje ali *intrinzične*, vsebino *ozka*, stališče, ki to zagovarja, pa *internalizem*. Iz povedanega lahko izpeljemo dve trditvi, ki ju zastopajo predstavniki internalizma, in sicer:

(i) ozka vsebina je notranja in ne vsebuje ničesar, kar je zunaj posameznikovega trenutnega prepričanja,

(ii) ozka vsebina je odvisna (ali *supervenira* nad) od trenutnih notranjih fizičnih stanj posameznika. (*glej* Kim, 1996: 203–204)

Internalizem predstavlja za razpravo o mentalni vzročnosti vir težav zato, ker poudarja, da je v zvezi z intencionalnimi stanji vzročno relevantna njihova ozka vsebina, tj. njihova fizična zgradba (nevrofiziološke lastnosti) ali njihov formalni ustroj, in se v tem približuje že omenjenemu komputacionalizmu. Peter in njegov dvojnik ne gasita požarov (se ne obnašata na isti način) zato, ker imata pred seboj H_2O oziroma XYZ, ampak zato, ker imata isti stereotip o vodi, tj. skupek prepričanj, ki so odvisna od nevrofizioloških lastnosti njunih možganskih procesov. Intencionalnim stanjem preti nevarnost, da bo tistemu, kar je njihova *glavna značilnost, tj. široka vsebina, in zaradi katere naj bi sploh imela vzročno vlogo*, odvzeta kakršnakoli relevantnost pri povzročanju dogodkov. Vzročno učinkovita naj bi bila vsebina, ki je odvisna od tega, kar se dogaja znotraj nas, in ne vsebina, ki je odvisna od tega, kar se dogaja zunaj nas. Za naše obnašanje to, na katero substanco se lastnosti kot, na primer, brezbarvnost, prozornost, topljivost itd. nanašajo (vsebina, ki je odvisna od okolja), ni pomembno. To pa je že podobno temu, kako delujejo računalniški sistemi in s tem komputacionalistični teoriji duha.

II. Eliminativnost široke vsebine

Ali torej pri razvijanju psiholoških teorij obnašanja resnično ne rabimo upoštevati široke vsebine oziroma ali lahko razvijemo plavzibilno psihološko teorijo obnašanja le z upoštevanjem ozke vsebine? Obstajajo vsaj trije razlogi,[8] ki govorijo v prid temu, da je široka vsebina, čeprav sprejmemo idejo, da psihologija formulira vzročne pojasnitve v takšnih, "širokih" pojmih, značilnost, ki jo lahko pogrešamo. Prvi razlog je podobnost v obnašanju med nami in našimi dvojniki, kar smo si že ogledali v primeru z "vodo" in "tvodo", ko različni substanci, na kateri so se nanašala prepričanja, na obnašanje nista imeli nikakršnega vpliva.

Drugič, če je psihološka teorija o obnašanju Zemljanov že razvita in dokazano deluje v vseh razmerah, verjetno takrat, ko boste želeli razviti psihološko teorijo o obnašanju njihovih dvojnikov, ne boste začeli z ničle; verjetno boste zato, ker

[8] Za razloge *glej* Kim, 1996: 205–206.

ni razlik v njihovih stereotipnih prepričanjih, predpostavili, da imajo Zemljani in njihovi dvojniki isto psihologijo, tj., da za ene in druge velja ista psihološka teorija obnašanja. Ali ni potem glede na to smiselno razlike v prepričanjih o vodi in "tvodi" (razlike v substanci, na katero se nanašajo) obravnavati le kot konstante, odvisne od konteksta in prilagojene okolju, v katerem se teorija uporablja, ne pa kot njene sestavne dele? (*glej* Kim, 1996: 205) Če to drži, lahko široko vsebino kot del teoretičnega aparata psiholoških teorij eliminiramo.

Tretji razlog je metafizični in pravi, da mora biti vzrok za premikanje naših teles notranji, fizični, to pomeni, da moramo vzrok za gibanje naših udov, na primer dvigovanje in spuščanje rok, iskati v našem centralnem živčnem sistemu (vzburjenje živčnih vlaken sproži krčenje in raztezanje ustreznih mišic). Vidimo, da to, kar ti živčni dogodki predstavljajo ali pomenijo v zunanjem svetu, za vzročno pojasnitev fizičnega obnašanja sploh ni pomembno, kar se lepo ujema s komputacionalizmom. Kljub temu da se ista živčna dogodka pojavita v dveh različnih okoljih, zaradi česar imata jasno različno široko vsebino, je fizično obnašanje, ki ga povzročita v obeh okoljih, isto.[9] Tako lahko upravičeno trdimo, da so vzroki za obnašanje odvisni (ali supervenirajo nad) od notranjih fizičnih stanj organizma, tj. intrinzičnih živčnih lastnosti, ne pa od ekstrinzičnih relacijskih.

Ker moramo glede na to pričakovati, da bi bile lastnosti, ki bi nastopale v vzročni verigi obnašanja, notranje, široka vsebina psihičnih stanj pa je zunanja, lahko z veliko verjetnostjo predpostavimo, da relacijske mentalne lastnosti sploh niso relevantna za vzročne pojasnitve obnašanja. Poudarjanje pomembnosti široke vsebine za plavzibilnost psiholoških teorij obnašanja tako od nas zahteva odgovor na vprašanje, kako so lahko ekstrinzične mentalne lastnosti pri pojasnjevanju obnašanja vzročno učinkovite. To vprašanje predstavlja problem mentalne vzročnosti, ki ga imenujemo *problem ekstrinzičnosti mentalnih lastnosti*, kot argument ga lahko v standardni obliki zapišemo takole, imenujemo pa *argument zoper vzročno učinkovitost intencionalnih pojavov*:[10]

(1) Lastnosti široke vsebine posameznikovih mentalnih stanj ne supervenirajo nad njegovimi intrinzičnimi lastnostmi (tj. nad tistimi lastnostmi, ki bi jih delil s svojo natančno fizično kopijo).

(2) Vzročna moč posameznikovih mentalnih stanj supervenira nad njegovimi intrinzičnimi lastnostmi (dano vzročno moč ima samo takrat, ko ima isto vzročno moč tudi njegova natančno fizična kopija).

(3) Vzročno moč mentalnega stanja mora v celoti določati niz lastnosti, ki se razlikujejo od lastnosti njegove široke vsebine. (iz 1, 2)

(∴) (4) Lastnosti široke vsebine mentalnega stanja niti delno ne določajo njegove vzročne moči. (iz *3*)

[9] Spet se spomnite na primer z vodo in "tvodo", ko različni substanci, na kateri so se nanašala prepričanja, na obnašanje nista imeli nikakršnega vpliva.

[10] Van Gulick ga imenuje Argument iz odsotnosti individualne supervenience. (*glej* Van Gulick, 1995: 236)

(∴) (5) Lastnosti široke vsebine niso vzročno učinkovite. (iz 4)

∴ (6) Lastnosti široke vsebine niso vzročno relevantne; takšne lastnosti so epifenomenalne. (iz 5) (Van Gulick, 1995: 236–237)

Jasno je, da sklep argumenta za nekoga, ki želi braniti vzročno učinkovitost mentalnih pojavov, katerih posebnost naj bi bila intencionalnost, ni sprejemljiv.

III. Dretskejeva rešitev problema ekstrinzičnosti mentalnih lastnosti

Zanimivo rešitev problema ekstrinzičnosti mentalnih lastnosti je v prvi polovici devetdesetih let prejšnjega stoletja predstavil Dretske. Njegov poskus temelji na ideji, da je skoraj vsak dogodek odvisen od številnih drugih dogodkov, v pravem kontekstu pa je lahko katerikoli izmed njih vzrok zanj, s čimer imajo tudi relacijske mentalne lastnosti vzročno moč, ki si jo tako želimo.

Spomnimo se, da obstajajo dobri razlogi, ki govorijo v prid temu, da je široka vsebina, čeprav sprejmemo idejo, da psihologija formulira vzročne pojasnitve v "širokih" pojmih, značilnost, ki jo lahko pogrešamo. Najprepričljivejši je verjetno metafizični, ki pravi, da mora biti vzrok za premikanje naših teles notranji, fizični, tj., da moramo vzrok za premikanje našega telesa iskati v našem osrednjem živčnem sistemu (vzburjenje živčnih vlaken sproži krčenje in raztezanje ustreznih mišic). To, kar ti živčni dogodki *predstavljajo* v zunanjem svetu, za vzročno pojasnitev fizičnega obnašanja ni pomembno. Kljub temu, da se ista živčna dogodka pojavita v dveh različnih okoljih, zaradi česar imata jasno različno široko vsebino, je fizično obnašanje, ki ga povzročita v obeh okoljih, isto. Zato lahko upravičeno trdimo, da so vzroki za obnašanje odvisni od notranjih fizičnih stanj organizma, tj. intrinzičnih živčnih lastnosti, ne pa od ekstrinzičnih relacijskih. Ker lahko glede na to pričakujemo, da so lastnosti, ki nastopajo v vzročni verigi obnašanja, notranje, široka vsebina mentalnih stanj pa je zunanja, moramo zaključiti, da psihične relacijske lastnosti za vzročne pojasnitve obnašanja sploh niso relevantne.

Dretskejev pristop k zavrnitvi tega temelji na razlikovanju med dvema vrstama vzrokov, *sprožitvenimi*[11] in *strukturalnimi*,[12] poskuša pa nas prepričati, da je takšna distinkcija za vzročne razlage obnašanja sistemov pomembna. Vzemimo naslednji primer. Peter želi brati, zunaj pa je že temno. Stopi k steni in pritisne stikalo. Prižge se luč. Vprašanje, na katerega želimo najti odgovor, je, zakaj se je prižgala luč. Odgovor je preprost. Luč se je prižgala, ker je Peter pritisnil stikalo. Pritisk stikala je vzrok, prižig luči pa učinek. Če Peter ne bi pritisnil stikala, se luč ne bi prižgala. V istih pogojih (imenujmo jih *S*) bo pritisku stikala (dogodek *a*) vedno sledilo prižig luči (dogodek *b*). Takšen vzrok Dretske imenuje *sprožitveni vzrok*.

Včasih pa odgovor, da se je luč prižgala zato, ker je Peter pritisnil stikalo, ni dovolj. Vedeti želimo več, in sicer, zakaj se vedno, ko pritisnemo stikalo, prižge

[11] Angl. *triggering cause*.

[12] Angl. *structuring cause*.

luč. Tudi odgovor na to vprašanje je dokaj preprost. Po pritisku stikala se luč vedno prižge zato, ker sta stikalo in luč med seboj na poseben način povezana z električnimi žicami. Električarjev poseben način povezave stikala in luči z električnimi žicami je dogodek (dogodek *o*), ki povzroči, da vsakemu pritisku stikala (dogodek *a*) sledi prižig luči (dogodek *b*). Takšen vzrok Dretske imenuje *strukturalni vzrok*.[13]

Med sprožitvenim in strukturalnim vzrokom pa obstaja nekaj razlik. Ena izmed njih je naslednja. Vrnimo se na naš primer in predpostavimo, da povežem stikalo in luč z žicami tako, da pritisku na stikalo vedno sledi prižig luči. Če pritisnem stikalo danes, se bo prižgala luč, če pritisnem stikalo jutri, se bo prižgala luč itn. Da bi prižgal luč jutri, stikala in luči ne rabim ponovno povezati. Tako bo strukturalni vzrok za prižig luči vedno isti, in sicer medsebojna povezava stikala in luči, za katero sem poskrbel danes. Vzročna zveza, v kateri nastopa strukturalni vzrok, je zato vedno zveza med enim vzrokom (enkratna povezava stikala in luči) in mnogimi učinki (večkratno prižiganje luči). To pa *ne* drži za sprožitveni vzrok. Pritisk stikala, ki povzroči, da se prižge luč danes, ni isti kot tisti, ki bo povzročil, da se bo luč prižgala jutri. Tako sprožitveni vzroki niso nikoli isti. Vzročna zveza, v kateri nastopa sprožitveni vzrok, je zato vedno zveza med enim vzrokom (enkraten pritisk stikala) in enim učinkom (enkratni prižig luči).

Včasih pa se zdi, da bi lahko namesto dveh vzrokov, sprožitvenega in strukturalnega, šlo le za enega. Vrnimo se k Dretskejevemu primeru, opisanem v opombi številka 13 na prejšnji strani, in predpostavimo, da terorist pozabi, v kateri avto je namestil bombo. Čez nekaj dni nujno potrebuje avto in ukrade ravno tega, v katerem je podtaknjena bomba. Ko zavrti ključ, da bi prižgal motor, avto raznese. Ali je terorist res hkrati sprožitveni in strukturalni vzrok za eksplozijo avtomobila? Drži, da je terorist ustvaril pogoje, ki so bili potrebni, da je prišlo do eksplozije, in da je sam, s tem ko je zavrtel ključ, bombo tudi aktiviral. Vendar to pomeni le, da je bila v oba dogodka (in tako tudi vzroka) vpletena ista oseba, in ne, da sta oba vzroka, sprožitveni in strukturalni, ista. Okoliščine, ki jih je terorist ustvaril zato, da bi prišlo do eksplozije, so eno, to, da je zavrtel ključ in bombo aktiviral, pa drugo. Dogodka sta kljub temu, da v obeh nastopa ista oseba, še vedno različna.

Preden pa Dretske predstavi svojo sliko mentalne vzročnosti, najprej pokaže, kakšno je razmerje med sprožitvenim in strukturalnim vzrokom ter notranjimi in zunanjimi lastnostmi. Ker se njegov primer za to, da imajo predmeti dve vrsti lastnosti, pogosto navaja, poleg tega pa nam bo pomagal tudi osvežiti utrujen

[13] Sam ponazori oba vzroka z naslednjim primerom. Terorist podtakne v generalov avto bombo, ki tam mirno čaka tako dolgo, dokler general ne sede v avto in zavrti ključa, da bi ga vžgal. Ko se to zgodi, bomba eksplodira (njeno detonacijo je sprožila obrnitev ključa, ki naj bi vžgala avto) in general umre. Kdo ga je ubil? Odgovor je, da terorist. In kako? S podtaknitvijo bombe. Kaj je tukaj sprožitveni vzrok in kaj strukturalni? Generalovo dejanje, tj. vžgig avtomobila, je sprožitveni vzrok, teroristovo dejanje, tj. podtaknitev bombe, pa strutkuralni. In strukturalni vzrok je tisto, kar bi bilo tako v sodnih kot tudi v moralnih preiskavah vzrok za generalovo smrt. (*glej* Dretske, 1995: 122)

spomin, si ga oglejmo.[14] Vzemimo kot primer bankovec. Bankovec ima določeno obliko, recimo, da je pravokoten, in določeno strukturo, zgrajen je iz atomov in molekul. Prav tako ima določeno barvo, recimo, da je siv, in določeno velikost, na primer desetkrat pet centimetrov. To so njegove intrinzične ali notranje lastnosti.

Ima pa tudi določeno vrednost, zaradi katere lahko na primer, z njim kupimo kilogram jabolk, ki je odvisna od njegovih zgodovinskih in socialnih dejstev, tj. od tega, kdo ga je izdal in kje je bil natisnjen. To so njegove ekstrinzične ali zunanje lastnosti. Če se izdajatelj in tiskarna spremenita, se spremeni tudi njegova vrednost. Če ga recimo natisne Peter v svoji garaži in tako spremeni omenjena dejstva, se spremeni tudi status denarja in njegova vrednost (z njim ne moremo več kupiti kilograma jabolk). Tako je pomen ali vsebina bankovca (njegova vrednost) odvisna od njegove relacije do izdajatelja in tiskarne (zgodovinskih in socialnih dejstev). Takšne lastnosti, ki temeljijo na relaciji med predmetom in drugimi stvarmi, imenujemo relacijske lastnosti.

Če smo materialisti in verjamemo v obstoj takšnih stvari, kot so prepričanja in želje, ki imajo svoj pomen ali vsebino, lahko predpostavimo, da je tudi njihov pomen ali vsebina odvisna od njihovih relacij do drugih stvari. Petrovo prepričanje, da je pred njim stol, se verjetno res nahaja v njegovi glavi, vsebina njegovega prepričanja, to, kar verjamem, ali to, na kar se moje prepričanje nanaša (stol), pa skoraj zagotovo ne.[15] Tako je razlikovanje med intrinzičnim in ekstrinzičnim merilo za to, kaj je fizično in kaj mentalno, tipična značilnost prvega je, da gre za notranje pojave, drugega pa, da gre za relacijske. In kako je lahko mentalno pri vzročnem pojasnjevanju obnašanja kljub vsemu učinkovito? Dretske pravi, da so intrinzične lastnosti obnašanja (nevrofiziološke) tisto, kar ga sproži, ekstrinzične (zgodovina in okolje) pa tisto, kar ga omogoči.

Njegovo teorijo o mentalni vzročnosti si oglejmo s pomočjo naslednjega primera. Vzemimo sistem X z različnimi notranjimi stanji, med katerimi je tudi takšno, imenujmo ga F, ki povzroča določeno vrsto obnašanja, imenujmo ga G (ali primerki F-ja povzročajo primerke G-ja). F je notranje stanje, sestavljeno iz intrinzičnih F_i in ekstrinzičnih lastnosti F_e ($F = F_i + F_e$); če je X živi sistem, so lahko F_i električne in kemične, F_e pa družbene in zgodovinske lastnosti. Recimo, da v afriški savani opazimo X, recimo antilopo, ki poskuša zbežati (G) pred drugim živim sistemom Y, recimo levom. Opazujemo dalje in ugotovimo, da vedno, ko se lev približa, antilopa poskuša zbežati. Vprašanji, ki se postavljata, sta dve: (i) zakaj antilopa beži in (ii) zakaj beži pred levom. Pri odgovoru na prvo vprašanje se Dretske opre na sprožitveni vzrok, pri odgovoru na drugo pa na strukturalni.

[14] Za primer *glej* Dretske, 1995: 129.

[15] Podobno ni v knjigi, v kateri se nahajajo besede, pomena teh besed, ampak so tam le črke, ki te besede sestavljajo (pomen istih besed se lahko od bralca od bralca razlikuje, odvisen pa je od kulturnega in družbenega okvira).

Odgovor na (i), zakaj antilopa beži, je, da zato, ker vidi, da se ji približuje lev. Približevanje leva je v antilopi povzročilo dogodek *F*,[16] dogodek *F* pa je povzročil, da je antilopa pred levom začela bežati. Dogodek *F* velja kot vzrok za *G* zato, ker je v obliki elektrokemičnih impulzov v antilopi vzbudil zaznavno predstavo o levu (tj., da je *F* vzrok za *G* zaradi svojih intrinzičnih lastnosti F_i). Če se F_i ne bi pojavil, antilopa približujočega se leva ne bi videla in ne bi začela bežati. Takšno razlago Dretske *imenuje sprožitvena vzročna razlaga*.

Odgovor na (ii), zakaj beži pred levom, pa je, da zato, ker ve, da je lev plenilec in zanjo predstavlja nevarnost, kateri se mora, če hoče ostati živa, za vsako ceno izogniti. Tako je odgovor na to vprašanje tesno povezan z učenjem. V procesu učenja se je pri antilopi vedno, ko se je začel približevati lev, pojavil dogodek *F*, ki je postal vzrok za njeno bežanje. Toda sedaj je dogodek *F* postal vzrok za *G* zato, ker je antilopi povedal, da je lev nevaren sovražnik, ki bi jo rad pojedel in ji hkrati naročil, kaj mora storiti, da bi ostala živa (tj. *F* je vzrok za *G* zaradi svojih ekstrinzičnih lastnosti F_e). Brez pojava dogodka *F* antilopa ne bi vedela, kaj mora storiti, ko zagleda približujočega se leva. Takšno razlago Dretske imenuje *strukturalna vzročna razlaga*.

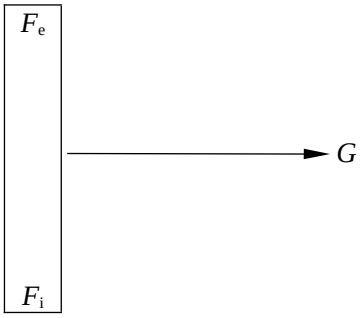
Ko je obnašanje (*G*) posledica takšne vrste učenja, so njegov strukturalni vzrok, ker so vse informacije o levu zunanje, ekstrinzične lastnosti F_e notranjega stanja *F*. Ker se brez informacij o levu antilopa pred njim ne bi nikoli pognala v beg, lahko rečemo, da so te informacije nujni pogoj za *G*. Žal pa niso tudi zadosten. Čeprav antilopa še tako dobro ve, da je lev nevaren, se ne bi spustila v beg, če njeni možgani ne bi dobili informacij v obliki nevrofizioloških impulzov, da se ji lev tudi približuje. Tako so sprožitveni vzrok njenega bežanja, ker so vse informacije v zvezi z videnjem notranje, intrinzična lastnost F_i notranjega stanja *F*. In tudi sprožitveni vzrok moramo obravnavati kot nujni pogoj za antilopin beg.

Kako torej izgleda Dretskejeva slika mentalne vzročnosti? Analogno gornjemu primeru obstajata dva vzroka, ki eden brez drugega ne moreta povzročiti fizičnega dogodka, sprožitveni in strukturalni. Opraviti imamo z vzročno verigo, v kateri F_e in F_i skupaj povzročita obnašanje (*G*), s čimer Dretske zavrne *četrto* premiso argumenta zoper vzročno učinkovitost intencionalnih pojavov. Povedano drugače, ekstrinzična lastnost mentalnega stanja skupaj z njegovo intrinzično lastnostjo povzroči fizični dogodek. Tako je lahko mentalno kljub svoji ekstrinzičnosti vzročno učinkovito, zaradi česar lahko upravičeno pričakujemo njegov nastop v vzročnih verigah obnašanja.

[16] Ali notranje stanje *F*.

IV. Ugovor

Dretskejeva ideja je na kratko ta, da sta sprožitveni in strukturalni vzrok dela iste vzročne verige, pomembno pa je, da lahko obnašanje povzročita samo oba skupaj, tj. zgolj v primeru pojava obeh so izpolnjeni nujni in zadostni pogoji za nastop fizičnega učinka. Sam pravi nalsednje: "/…/ čeprav so intrinzične lastnosti možganov tiste, ki so odgovorne za sprožitev fizičnega obnašanja, pa so njegove ekstrinzične lastnosti tiste, ki so relevantne za to, da lahko do njega sploh pride." Slika, ki nam to nazorno prikazuje, je naslednja (F_e predstavlja strukturalni vzrok, F_i pa sprožitvenega):



(Polna puščica označuje relacijo vzročnosti.)

Dretske tako očitno pravi, da sta oba vzroka, sprožitveni, ki ga predstavljajo intrinzične lastnosti mentalnega stanja, in strukturalni, ki ga predstavljajo njegove ekstrinzične lastnosti, sama zase le *delna* vzroka za nastop določenega pojava. *Popolnoma* zadosten vzrok za pojav določenega dogodka dobimo šele z njuno združitvijo. Podobno kot pravimo, da je avto zletel s ceste zato, ker je bil voznik pijan in ker je bila slaba vidljivost, naj bi trdili tudi, da sem popil vodo zato, ker sem bi žejen in ker so v mojih možganih potekali določeni elektrokemični procesi.

Zakaj pa se takšno stališče izogne kršitvi načela *vzročne zaprtosti fizičnega prostora*, do česar pride takrat, ko mentalni vzrok obravnavamo kot nujno sestavino pravega vzroka za nastop fizičnega dogodka?[17] To bi lahko imelo usodne posledice za Dretskejev fizikalisitični nazor. Toda do tega ne pride, ker so tako intrinzične kot tudi ekstrinzične lastnosti mentalnih dogodkov *fizične*, zaradi česar se Dretske izogne temu, da bi vzročna veriga tekla po področju, ki ni fizično.

Ali njegov odgovor torej res predstavlja uspešno rešitev mentalne vzročnosti? Mislim, da ne. Zdi se namreč, da pripisuje vzročno moč zgolj fizičnim lastnostim. Drži, da so mentalna stanja s široko vsebino po njegovi teoriji vzročno učinkovita, kar je nesporno velik napredek pri raziskovanju vzročne moči mentalnih intencionalnih stanj. Težava nastopi, ko se spomnimo, da je vprašanje mentalne vzročnosti, tj., ali lahko mentalno povzroča fizično, že od

[17] Ta princip pravi, da ima fizični učinek vedno tudi fizični vzrok in se običajno jemlje kot pogoj za to, da se lahko imenujemo fizikalisti.

kritike Davidsonovega anomalnega monizma naprej formulirano v pojmih lastnostih. To pomeni, da govorimo o vzročni učinkvitosti mentalnega zgolj takrat, ko mentalne lastnosti povzročajo fizične dogodke. Ali Dretskejeva rešitev izpolni ta pogoj?

Vrnimo se na primer z antilopo in levom in predpostavimo, da je *F* antilopino mentalno stanje, F_i in F_e pa njene intrinzične oziroma ekstrinzične lastnosti. Po Dretskejevi teoriji požene antilopo v beg skupno delovanje sprožitvenega vzroka, tj. intrinzičnih lastnosti mentalnega stanja *F*, in strukturalnega vzroka, tj. njegovih ekstrinzičnih lastnosti. Drži, da je mentalno stanje tisto, ki je vzrok za antilopin beg, vendar je do njega prišlo zaradi *F*-jevih fizičnih, ne pa mentalnih lastnosti, saj so, kot smo omenili malo prej, tako intrinzične kot tudi ekstrinzične lastnosti fizične. Kar resnično požene antilopo v beg, so njene fizične lastnosti.

Posledica tega je, da skladno z Dretskejevim odgovorom mentalne lastnosti nimajo nobene vzročne vloge ali, kot se v sodobni razpravi o mentalni vzročnosti reče, da so *epifenomeni*. Tako se njegovo stališče, ker obravnava mentalne lastnosti kot vzročno nerelevantne oziroma vzročno neučinkovite, spogleduje z epifenomenalizmom. Sicer ne gre za epifenomenalizem v klasičnem smislu, saj dopušča vzročno učinkovitost mentalnih dogodkov, ampak za *epifenomenalizem glede mentalnih lastnosti*[18] (tudi Davidson je zašel sem s svojim anomalnim monizmom).

Zakaj pa je takšen epifenomenalizem problematično? Zato, ker se ne bi mreža vzročnih relacij našega sveta, če bi mentalne lastnosti vseh njegovih dogodkov spremenili v fizične, tj. iz vseh dogodkov našega sveta bi mentalne lastnosti povsem izginile, spremenila niti za malo. Še vedno bi hodili na zabave, ko bi želeli kramljati s prijatelji, jemali aspirine, ko bi se želeli znebiti bolečin, pili, ko bi želeli pogasiti žejo, in jemali dežnike, ko ne bi želeli, da bi dež pokvaril našo novo frizuro. Vzrokom sledijo učinki zaradi njihovih lastnosti (fizičnih, mentalnih), vendar so za nastop učinkov odgovorne le tiste lastnosti, zaradi katerih dejansko pride do učinka. Ene lastnosti vzroka so za nastop učinka relevantne, druge pa ne: če npr. žoga prileti v okno in se to razbije, je za njegovo razbitje relevantna teža žoge, ne pa njena barva.[19] (*glej* Šuster, 2001: 25) Če želimo ohraniti mentalno vzročnost, moramo tako pokazati, da vpliva nek dogodek na spremembe v fizičnem svetu zato, ker ima določeno *mentalno lastnost*. Vzročna učinkovitost mentalnih lastnosti je torej tisto, kar moramo dokazati in zagovarjati.

Zdi se torej, da Dretskejev poskus rešitve vzročne učinkovitosti mentalnih intencionalnih stanj s široko vsebino, ker ne predpostavlja vzročne moči mentalnih lastnosti, ne obrodi sadov.

[18] Angl. mental property epiphenomenalism.

[19] Govorimo o vzročno relevantnih lastnostih ali vzročno relevantnih značilnostih, npr. količina eksploziva (in ne oblika, barva …) je vzročno relevantna za velikost kraterja, ovinek na cesti je vzročno relevanten za to, da je prišlo do nesreče, teža kamna (in ne starost, zgradba …) je vzročno relevantna za nezgodo ipd. Tako moramo priznati pravi Kim, da obstajajo med dogodki vzročne relacije zaradi lastnosti, ki jih imajo. (*glej* Šuster, 2001: 25)

EXTRINSIC MENTAL PROPERTIES AS A PROBLEM OF MENTAL CAUSATION

Computationalism is a source of problems for psychology and its explanations of behaviour as well as for philosophical discussions about mental causation since it supposes that contents of our beliefs, desires and other intentional states, because it is their structure or shape that has causal powers, are not causally relevant. Intentional states are at risk that their content, which is their main characteristic and because of which they play causal roles at explaining our behaviour, will lose a causal relevance of any sort, which is called the problem of extrinsicness of the mental. The article also introduces Dretske's solution of this problem based on a distinction between two kinds of causes, triggering and structuring, and my objection to it saying that such a move leads to mental property epiphenomenalism and therefore does not represent a true alternative.

Keywords: computationalism, externalism, internalism, mental causation, extrinsicness of the mental, mental property epiphenomenalism.

Literatura

Corbi, J. E. in Prades, J. L. (2000). *Minds, Causes and Mechanisms*. Oxford: Blackwell

Dretske, F. (1995). "Mental Events as Structuring Causes of Behaviour". V Heil, J. in Mele, A. (ur.), *Mental Causation*, Oxford: Clarendon.

Heil, J. in Mele, A. (ur.) (1995). *Mental Causation*, Oxford: Clarendon.

Horgan, T. (1997). "Kim on Mental Causation and Causal Exclusion". *Philosophical Perspectives*, 11, str. 165–184.

Jackson, F. (1996). "Mental Causation: The State of the Art". *Mind*, 105, str. 377–413.

Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. (1996). *Philosophy of Mind*. Boulder: Westview.

Kim, J. (2001): *Mind in a Physical World*. Cambridge: MIT Press.

Van Gulick, R. (1995). "Who's in Charge Here? And Who's Doing All the Work?". V Heil, J. in Mele, A. (ur.), *Mental Causation*, Oxford: Clarendon.