
Shaun Nichols, Joshua Knobe

Prevedla Mojca Predovnik in Danilo Šuster

Moralna odgovornost in determinizem: Kognitivna znanost o vsakdanjih intuicijah¹

1. Uvod

Razprava med kompatibilisti in inkompatibilisti je gotovo ena izmed najbolj vročih in trdovratnih mrtvih točk v zgodovini zahodne filozofije. Za inkompatibiliste ljudje niso popolnoma moralno odgovorni, če je determinizem resničen oz. če je vsak dogodek neogibna posledica predhodnih pogojev in naravnih zakonov. V nasprotju s tem pa so kompatibilisti prepričani, da, četudi je determinizem resničen, to niti najmanj ne ogroža naše moralne odgovornosti, saj sta determinizem in moralna odgovornost popolnoma združljiva.²

V razpravi med tema dvema stališčema se sklicujejo na vrsto različnih virov, od kvantne mehanike do socialne psihologije in osnovne metafizike. Toda nedavne diskusije se opirajo predvsem na argumente, ki temelje na naših intuicijah o posameznih primerih. Nekateri filozofi trdijo, da imajo ljudje inkompatibilistične intuicije (npr. Kane, 1999: 218; Strawson, 1986: 30; Vargas, 2006); spet drugi so to trditev izzvali in dejali, da se naše običajne intuicije skladajo s kompatibilizmom (Nahmias et al., 2005). Čeprav filozofi konstruirajo vedno bolj sofisticirane

¹ »Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions,« *Nous* 41:4 (2007), 663–685.

² Pravzaprav, kompatibilisti in inkompatibilisti razpravljajo tako o tem, (1) ali je determinizem združljiv z moralno odgovornostjo in (2) ali je determinizem združljiv s *svobodno voljo*. Kot je poudaril Fischer (1999), sta ti dve vprašanji logično neodvisni. Zagovarjamo lahko tezo, da je determinizem združljiv z moralno odgovornostjo, ne pa s svobodno voljo. Tu nas zanima samo prvo od obeh vprašanj – ali je determinizem združljiv z moralno odgovornostjo.

argumente o tem, kaj sledi iz človeških intuicij, pa je bilo do sedaj presenetljivo malo razprave o tem, *zakaj* ljudje sploh imajo takšne intuicije. Relativno malo je torej govora o določenih psiholoških procesih, ki ustvarjajo ali ohranjajo človeške intuicije. Ampak zdi se jasno, da bi lahko vprašanja o izvoru človeških intuicij imela velik vpliv na debate o združljivosti odgovornosti in determinizma. V nekem očitnem smislu je pomembno, da ugotovimo, ali so človeške intuicije posledica nekega splošno zanesljivega procesa ali pa so popačene zaradi procesa, ki nas v splošnem zavaja.

Najin cilj je, da predstaviva in zagovarjava hipotezo o procesih, ki generirajo človeške intuicije glede moralne odgovornosti. Najina hipoteza je, da imajo ljudje inkompatibilistično teorijo moralne odgovornosti, ki pride na dan v določenih primerih, imajo pa tudi psihološke mehanizme, ki nas lahko vodijo do kompatibilističnih presoj v drugih primerih.³ Da bi to hipotezo podprla, predstavlja nove eksperimentalne podatke. Ti podatki kažejo na to, da se lahko odgovori ljudi na vprašanja o moralni odgovornosti zelo razlikujejo – odvisno od tega, kako je vprašanje zastavljeno. Kadar je vprašanje zastavljeno tako, da zahteva bolj abstrakten, teoretičen način razmišljanja, ljudje dajejo povsem inkompatibilistične odgovore. Kadar pa je vprašanje zastavljeno na način, ki sproži čustva, postanejo njihovi odgovori veliko bolj kompatibilistični.

2. Afekt, krivda in pripisovanje odgovornosti

V svojih poskusih razumevanja vsakdanjih pojmov in teorij so se naturalistični filozofi lotili raziskav človeških intuicij o posameznih primerih (npr. Knobe, 2003a, 2003b; Nahmias et al., 2005; Nichols, 2004a; Weinberg et al., 2001; Woolfolk et al., 2006). Osnovni pristop je enostaven. Filozof konstruira hipotetični scenarij, potem pa ljudi vpraša, če je, na primer, oseba v tem scenariju moralno odgovorna. S spreminjanjem detajlov scenarija in opazovanjem, kako to vpliva na intuicije, lahko postopoma pridemo do osnovnih obrisov vsakdanje teorije (angl. folk theory). Metoda je dobra, vendar zahteva pazljivost v izvedbi. Ne moremo preprosto domnevati, da vse pomembne intuicije izvirajo iz iste osnovne vsakdanje teorije. Vedno je namreč mogoče, da se izkaže, da različne intuicije izvirajo iz različnih psiholoških procesov.

Osredotočila se bova predvsem na vlogo *afekta* pri tvorbi intuicij o moralni odgovornosti. Najina hipoteza je, da kadar se ljudje soočijo z zgodbo o osebi, ki se vede nemoralno, to lahko v njih sproži takojšnji čustveni odziv, ki lahko igra ključ-

³ Termin 'teorija' uporabljava v ohlapnem smislu kot oznako za interno reprezentirano množico informacij. In ko trdiva, da imajo običajni ljudje inkompatibilistično teorijo, s tem ne želiva sugerirati, da ima ta teorija privilegiran status glede na psihološke sisteme, ki generirajo kompatibilistične intuicije. Kot bo postalo očitno, se nama zdi, da ostaja odprto vprašanje, ali ima sistem, ki generira inkompatibilistične intuicije, privilegiran status.

no vlogo v njihovih intuicijah o tem, ali je bil dejavnik moralno odgovoren. Še več, ljudje lahko včasih celo označijo tako osebo kot moralno odgovorno, pa čeprav zagovarjajo teorijo odgovornosti, po kateri ta oseba ni odgovorna.

Poglejmo si, na primer, Watsonovo (1987) zanimivo diskusijo o zločinih Roberta Harrisa. Watson navaja dolge citate iz časopisnega članka o tem, kako je Harris kruto moril nedolžne ljudi, ne da bi pri tem pokazal vsaj kanček obžalovanja za to, kar je storil. Potem pa enako podrobno tudi opisuje, kakšne zlorabe je Harris doživljal v času odrasčanja. Po branju teh nazornih detajlov bi bilo za bralca skoraj nemogoče, da bi se odzval tako, da bi mirno izpeljal vse posledice, ki bi sledile po njeni ali njegovi teoriji o moralni odgovornosti. Vsak normalen bralec bo imel bogat nabor odzivov, ne le abstraktno teoretiziranje, temveč tudi čustva groze in gnusa. Bralčeve intuicije o takšnem primeru se lahko zaradi čustev zamajejo tako, da pristane pri sklepih, ki nasprotujejo njegovim bolj abstraktnim, teoretskim prepričanjem o naravi moralne odgovornosti.

No, lahko bi pomislili, da je zelo malo verjetno, da bi te vrste učinki vplivali na naše reakcije na običajne filozofske primere. Večina filozofskih primerov je povsem hipotetičnih in zelo skopo opisanih (velikokrat samo z nekaj stavki). Na prvi pogled se morda zdi, da ni verjetno, da čustveni odzivi kakorkoli vplivajo na človeške intuicije v takih primerih. Ampak naraščajoča zbirka eksperimentalnih dokazil nakazuje, da je to zdravorazumsko mnenje napačno. Dokazila kažejo, da imajo čustva pomembno vlogo celo pri človeških intuicijah o šibko opisanih in povsem hipotetičnih primerih (Blair, 1995; Greene et al., 2001; Nichols, 2002; Haidt et al., 1993).

Morda se zdi nenavadno, da imajo čustva tolikšno vlogo in kar nekaj različnih modelov o vlogi čustev pri vrednostnem razmišljanju je bilo predlaganih. Nekatere od njih bomo podrobneje obravnavali v razdelkih 5, 6 in 7. Medtem pa bi želela opozoriti na dejavnik, za katerega se zdi, da vpliva na čustvene odzive ljudi. Nedavna študija Smarta in Loewensteina (2005) kaže, da kadar je prestopnik za osebe bolj »določen« (angl. determinate), osebe izkusijo bolj negativna čustva in so zato do njega kaznovalno strožji. V študiji poskusne osebe igrajo igro, v kateri lahko med seboj bodisi sodelujejo bodisi odklonijo sodelovanje. Vsaki osebi je dodeljena identifikacijska številka, nobena oseba pa ne pozna številke drugih. Eksperimentator položi številke tistih, ki so odklonili, v posebno ovojnico. Tisti, ki med seboj 'sodelujejo', lahko kasneje odločajo o tem, ali naj 'pobeglega' kaznujejo ali ne. 'Sodelujoči' je obveščen, da bo izbral številko iz kuverte in s tem določil, kateri 'ubežnik' bo kaznovan (ali ne). Manipulacija je bila neverjetno prefinjena. Pri *nedoločenem* pogoju se osebe odločijo za mero kaznovanja, *preden* izvlečejo številko; pri *določenem* pogoju pa se odločijo za mero kazni *po tem*, ko izvlečejo številko. Kljub tej majhni razliki sta Smart in Loewenstein opazila precejšen učinek – osebe so pri določenem pogoju dodelile hujše kazni, kakor pri nedoločenem pogoju. Osebe so v poskusu tudi vsaka zase izpolnjevale vprašalnik na temo, ko-

liko jeze, krivde in sočutja so občutile. Osebe pri določenem pogoju so občutile več jeze in krivde, kakor osebe pri nedoločenem pogoju. Na koncu sta Smart in Loewenstein z mediacijsko statistično analizo ugotovila, da določenost vpliva na dodeljeno kazen tako, da izzove močnejša čustva.

Kot bomo videli, so v vseh preteklih študijah o človeških intuicijah glede moralne odgovornosti nastopali določeni dejavniki, torej so bile načrtovane na način, ki bi težil k temu, da izzove čustvene odzive. Najina študija je priložnost, da vidimo, kako se človeške intuicije spreminjajo, če je spodbuda oblikovana na način, ki čim manj sproža čustvene odzive.

3. Intuicije o svobodni volji in odgovornosti

Inkompatibilistični filozofi tradicionalno trdijo, da običajni ljudje verjamejo, da so človeške odločitve neodvisne od determinističnih zakonov ter da je determinizem nezdržljiv z moralno odgovornostjo (npr. Kane, 1999; Strawson, 1986). Te trditve niso temeljile na sistematičnih empiričnih raziskavah, ampak bolj na anekdotah in neformalnih opazovanjih. Tako piše Kane, "Po moji izkušnji je za večino običajnih ljudi naravno izhodišče inkompatibilizem" (1999: 217). (Kot bo razvidno kasneje, meniva, da Kane tukaj dejansko zajame nekaj globljega glede naših intuicij.) V zadnjih letih so si filozofi prizadevali tovrstne trditve preveriti z eksperimentalnimi metodami. Rezultati so bili včasih presenetljivi.

Prvič, razmislimo o trditvi, da običajni ljudje verjamejo, da človeških odločitev ne določajo deterministični zakoni. V zaporedju eksperimentov, ki so preverjali (laično) razumevanje izbire, se je izkazalo, da so bili tako otroci kot tudi odrasli nagnjeni k temu, da moralne izbire obravnavajo na indeterministični način (Nichols, 2004a). Sodelujočim v poskusu so predstavili primere dogodkov, v katerih nastopa moralna izbira (npr. dekle ukrade tablico čokolade) in fizikalnih dogodkov (npr. lonec vode zavre). Nato so jih vprašali, ali bi se, če bi bilo vse na svetu enako do tega dogodka, ta dogodek vseeno *moral* zgoditi. Tako otroci kot tudi odrasli so bili bolj nagnjeni k trditvi, da se je moral zgoditi fizikalni dogodek kot pa dogodek moralne izbire. Zdi se, da ta rezultat podpira tradicionalno trditev, da so običajni ljudje v naši kulturi prepričani, da vsaj nekatere od naših odločitev niso determinirane.

Eksperimentalna študija pa ni bila tako prizanesljiva do tradicionalne trditve, da so običajni ljudje tudi glede odgovornosti inkompatibilisti. Woolfolk, Doris in Darley (2006) so pripravili sodelujočim v poskusu zgodbo o osebi, ki so jo ugrabili in ji dali močno mamilo, ki povzroči, da oseba uboga vse ukaze in je nezmožna neizpolnitve ukaza. Ugrabitelji osebi ukažejo, naj stori neko nemoralno dejanje. Oseba pod vplivom mamil ne more drugače, kot da ukaz uboga. Osebam iz skupine z nizko stopnjo poistovetenja (angl. low identification condition) so povedali, da ugrabitelj ni želel izvesti nemoralnega dejanja, izvedel ga je samo zato,

ker je bil v to prisiljen pod vplivom mamil. Osebam iz skupine z visoko stopnjo poistovetenja (angl. high identification condition) pa so povedali, da je dejavnik vseskozi želel izvesti to dejanje in tega ni storil proti svoji volji. Rezultati so pokazali jasen vpliv poistovetenja: pri pogoju visoke stopnje poistovetenja so osebe dejavniku pripisale višjo mero odgovornosti kot pa pri pogoju z nizko stopnjo poistovetenja. Ta izid se čudovito sklada s kompatibilističnim stališčem, da je odgovornost odvisna od poistovetenja (npr. Frankfurt, 1988). Kakorkoli, v obeh primerih so sodelujoči izkazali nagnjenje k pripisovanju nižje stopnje odgovornosti dejavniku. Zato tudi ti rezultati ne predstavljajo neposredne grožnje stališču, da so ljudje glede odgovornosti inkompatibilisti.

Zadnja množica študij, ki si jih bomo ogledali, pa predstavlja večjo grožnjo stališču, da so ljudje intuitivno inkompatibilisti. Nahmias, Morris, Nadelhoffer in Turner (2005) so ugotovili, da udeleženci obravnavajo osebo kot moralno odgovorno, čeprav jim povedo, da se ta oseba nahaja v determinističnem univerzumu. Udeležencem je bil, denimo, predstavljen naslednji scenarij:

Predstavljajte si, da v naslednjem stoletju odkrijemo vse zakone narave in zgradimo superračunalnik, ki iz teh zakonov narave in trenutnega stanja vsega v svetu lahko natančno izpelje, kaj se bo dogajalo v svetu v poljubnem trenutku v prihodnosti. Pregleda lahko vse o tem, kakšen svet je in s 100 % natančnostjo napove, kakšen svet bo. Recimo, da bi takšen superračunalnik obstajal, ter da pregleda stanje univerzuma v določenem trenutku 25. marca 2150, dvajset let pred rojstvom Jeremyja Halla. Računalnik iz te informacije in zakonov narave potem izpelje, da bo Jeremy 26. januarja 2195 ob 18. uri zanesljivo oropal banko Fidelity. Kot vedno je napoved superračunalnika pravilna. Jeremy 26. januarja 2195 ob 18.00 oropa banko Fidelity.

Udeležence so potem vprašali, če se jim zdi Jeremy moralno odgovoren za rop banke. Rezultati so bili osupljivi: 83 % oseb je reklo, da je bil Jeremy moralno odgovoren za rop. V dveh dodatnih eksperimentih z različnima scenarijema so nastopili podobni učinki, kar kaže na to, da je za laike moralna odgovornost združljiva z determinizmom. Te ugotovitve so fascinantne, zato bova skušala graditi na njih v najinih lastnih eksperimentih.

Seveda je mogoče eksperimentom oporekati na metodološki osnovi. Tako na primer scenariji uporabljajo tehnično besedišče (npr. »naravni zakoni«, »trenutno stanje«) in zato bi se lahko vprašali, ali osebe dejansko razumejo te scenarije. Prav tako bi se lahko kdo pritožil, da determinizem v teh scenarijih ni dovolj poudarjen. Zgodba o superračunalniku se osredotoča na predvidljivost dogodkov v univerzumu in mnogo filozofov obravnava predvidljivost univerzuma kot manjšo grožnjo svobodni volji od vzročne neogibnosti. Zaradi takšnih metodoloških pomislekov bi kdo lahko zavrnil rezultate, a midva se k temu ne nagibava. Misliva

namreč, da so Nahmias in ostali odkrili nekaj izredno zanimivega.⁴ Navajajo tri precej različne scenarije, ki pa imajo podoben učinek. V vsakem od njihovih eksperimentov večina ljudi (60–85 %) pravi, da je oseba moralno odgovorna tudi ob predpostavki, da je determinizem resničen. Še več, rezultati sovpadajo z neodvisno psihološko študijo o kaznovanju. Viney in sodelavci so ugotovili, da študentje, ki so se izrekli kot deterministi, niso nič manj nagnjeni h kaznovanju od indeterministov (Viney et al. 1982) in tudi nič manj nagnjeni k temu, da kazen upravičijo na retributiven način (Viney et al. 1982)⁵. Domnevali bomo torej, da imajo Nahmias in ostali prav glede tega, da ljudje težijo k pripisovanju moralne odgovornosti dejavnikom, kadar ti namerno storijo slabo dejanje v determinističnem okviru.

Toda če ljudje tako konsistentno dajejo kompatibilistične odgovore na eksperimentalne vprašalnike, zakaj potem nekateri filozofi sklepajo, da so običajni ljudje inkompatibilisti?⁶ So ti filozofi zgolj preslišali mnenja svojih lastnih študentov? Sumiva, da je v ozadju nekaj bolj kompleksnega. Po najinem mnenju večina ljudi (vsaj v naši kulturi) resnično sprejema inkompatibilistične teorije moralne odgovornosti. Takšne teorije se zlahka pokažejo v filozofskih diskusijah, do katerih, denimo, pride v univerzitetnih seminarjih. Le da imajo ljudje, poleg teh teorij o moralni odgovornosti tudi takojšnje čustvene odzive na zgodbe o nemoralnih obnašanjih. To, kar vidimo v rezultatih eksperimentov Nahmiasa in kolegov, je, deloma, učinek teh čustvenih odzivov. Da bi razkrili osnovne teorije, ki jih imajo ljudje, jim moramo zastaviti vprašanja, ki zahtevajo bolj abstraktno, teoretsko kognicijo.

⁴ Ena od vrlin vprašanja, ki so ga zastavili Nahmias in kolegi o moralni odgovornosti je ta, da naj bi bil pojem 'moralna odgovornost' skupen filozofom in običajnim ljudem. Filozofi so nagnjeni k domnevi, da pojem moralne odgovornosti, ki nastopa v filozofiji, tesno sledi pojmu, ki ga običajno izražajo ljudje, ko pripisujejo moralno odgovornost. Vrh tega inkompatibilisti pogosto označijo ustrezni inkompatibilistični pojem svobodne volje natanko kot tistega, ki ga zahteva moralna odgovornost (npr. Campbell, 1951). Nahmias in kolegi prav tako postavljajo vprašanje, ali dejavnik v determinističnem scenariju »ravna po svoji lastni svobodni volji« in ugotavljajo, da ljudje dajejo odgovore, ki so skladni s kompatibilizmom. Nama se zdijo ti rezultati manj prepričljivi. Kajti izraz »svobodna volja« je postal filozofski »terminus technicus«, zato ni jasno, kako tolmačiti laične odgovore o takšnih tehničnih terminih. Še več, inkompatibilisti običajno priznajo, da obstajajo kompatibilistični pojmi svobode, ki jih uporabljajo tudi običajni ljudje. Inkompabilisti samo trdijo, da obstaja tudi vsakdanji pojem svobode, ki ni kompatibilističen.

⁵ Čeprav so rezultati Vineya in kolegov sugestivni, je merilo, po katerem so določali deterministe, preširoko, zato skupina, označena kot 'deterministi', lahko vključuje tudi indeterministe. (glej McIntyre et al., 1984 za natančne opis meril). Odprto je, ali bo ta rezultat ostal tudi, če uporabimo boljša merila za določanje indeterministov.

⁶ S tem povezan problem za inkompabiliste je zgodovina filozofije – če je inkompabilizem intuitiven, zakaj je bil kompatibilizem tako popularen med velikimi filozofi v zgodovini? Pojasnilo, ki je naklonjeno inkompabilizmu, ponuja Nichols (2007).

4. Eksperimentalna dokazila: prva faza

Izvedla sva serijo eksperimentov, da bi raziskala, ali bodo sodelujoči bolj nagnjeni k temu, da poročajo o inkompatibilističnih intuicijah, če so čustveni in motivacijski dejavniki minimalizirani. V vsakem eksperimentu je bil en pogoj, *konkretni* pogoj, oblikovan z namenom, da vzbudi večji čustveni odziv; drugi pogoj, *abstraktni* pogoj, pa naj bi sprožil abstraktno, teoretsko mišljenje. Napovedala sva, da bodo ljudje pri konkretnem pogoju bolj nagnjeni h kompatibilističnim odgovorom.

Preden podrobneje predstaviva eksperimente, naj poveva, da obstaja več načinov, kako označiti determinizem. Najbolj natančne oznake vsebujejo tehnični jezik o, denimo, zakonih narave. Meniva pa, da je uporaba tehnične terminologije pri eksperimentih te vrste napaka, zato sva se odločila, da problem predstaviva v bolj dostopnem jeziku.⁷ Seveda bo kakršen koli poskus prevoda kompleksnega filozofskega problema v enostavnejše izraze sprožil vrsto težkih vprašanj. Povsem mogoče je, da je specifičen opis determinizma, ki smo ga uporabili v naši študiji, vplival na intuicije ljudi v tej ali oni smeri. Mogoče bi bilo splošno razmerje inkompatibilističnih odgovorov nekoliko višje ali nižje, če bi uporabili malo drugačno formulacijo.

Moramo pa se zavedati, da je glavni poudarek na *razliki* med odgovori ljudi pri konkretnem pogoju in njihovimi odgovori pri abstraktnem pogoju. Kljub temu, da pri obeh pogojih uporabljava povsem enak opis determinizma, predvidevava, da bodo ljudje pri konkretnem pogoju dajali kompatibilistične, pri abstraktnem pa inkompatibilistične odgovore. Tak rezultat ne more biti zgolj artefakt najinega opisa determinizma. Če se razlika dejansko pojavi, bomo s tem imeli dobra dokazila za stališče, da afekt igra določeno vlogo pri kompatibilističnih intuicijah.

Vse naše študije so bile izvedene na dodiplomskih študentih Univerze v Utahu⁸, vse pa so se tudi začele na enak način. Sodelujočim smo dali naslednji opis determinističnega in indeterminističnega univerzuma:

⁷ V najinem determinističnem scenariju praviva, da se glede na preteklost vsaka odločitev *mora* zgoditi na tak način, kot se. Ta scenarij nama dopušča, da preveriva intuicije o vrsti kompatibilizma, ki je najbolj popularen v sodobni filozofiji. Večina sodobnih kompatibilistov dokazuje, sledeč Frankfurtu (1969), da je dejavnik lahko moralno odgovoren za svoje obnašanje, četudi je *moral* ravnati tako, kot je. (Kot bomo videli, večina vprašanih pri konkretnem pogoju daje odgovore, ki se skladajo s tem stališčem.) Za kompatibilista pa je povsem mogoče tudi, da zagovarja, da (1) nikdar ne moremo biti odgovorni za dogodek, ki se je moral pripetiti, tako kot se je in (2) četudi je določeno obnašanje določeno z zakoni narave, za dejavnika ni *nujno*, da tako ravna. Najin eksperiment ne naslavlja možnosti, da so vsakdanje intuicije skladne s to vrsto kompatibilizma. Z malo sreče bomo to možnost raziskali v prihodnosti.

⁸ Seveda bo pomembno raziskati, ali naši rezultati veljajo tudi za druge populacije. Ampak, kot vseskozi poudarjava, naju primarno zanima, kako se odgovori oseb iz iste populacije razlikujejo pri različnih pogojih.

Zamisli si vesolje (Vesolje A), v katerem je vse, kar se pripeti, popolnoma povzročeno s tem, kar se je zgodilo pred tem. To velja za vsak trenutek od začetka tega vesolja. Tisto, kar se je zgodilo na začetku vesolja, je povzročilo tisto, kar se je zgodilo potem in tako naprej, vse do sedanjega trenutka. Recimo, nekega dne se je Janez odločil, da bo imel za kosilo pečen krompirček. Tako kot vse ostalo v tem vesolju je bila ta odločitev popolnoma povzročena s tem, kar se je zgodilo pred tem. Torej, če je vse v tem vesolju bilo enako do trenutka Janezove odločitve, potem se je *moralo* zgoditi tako, da se je Janez odločil za pečen krompirček.

Zdaj pa si zamisli vesolje (Vesolje B), v katerem je *skoraj* vse, kar se pripeti, popolnoma povzročeno s tem, kar se je zgodilo pred tem. Edina izjema je človeško sprejemanje odločitev. Recimo, nekega dne se je Marija odločila, da bo imela za kosilo pečen krompirček. Ker človeške odločitve v tem vesolju niso popolnoma povzročene s tem, kar se je zgodilo pred tem, *ni bilo nujno*, da se je Marija odločila za pečen krompirček, četudi je vse v vesolju bilo natančno enako do trenutka, ko je Marija sprejela svojo odločitev. Lahko bi se odločila, da ima za kosilo kaj drugega.

Ključna razlika je, potemtakem, da je v Vesolju A vsaka odločitev popolnoma povzročena s tistim, kar se je zgodilo pred njo – glede na preteklost se vsaka odločitev *mora* zgoditi tako, kot se je. Za razliko od tega pa v Vesolju B odločitve niso popolnoma povzročene s preteklostjo, zato ni *nujno* tako, da se vsaka odločitev zgodi tako, kot se je.

1. Katero od obeh vesolj je po vašem mnenju najbolj podobno našemu? (obkroži en odgovor)

VESOLJE A

VESOLJE B

Prosim, če svoj odgovor na kratko pojasnite.

Namen tega začetnega vprašanja je bil enostavno ta, da ugotoviva, ali so subjekti prepričani, da je naš univerzum determinističen ali indeterminističen. Skoraj vsi sodelujoči (preko 90 %) so ocenili, da je indeterministični univerzum bolj podoben našemu.

Po tem, ko so odgovorili na začetno vprašanje, so osebe dobile vprašanje, namenjeno preverjanju intuicij o kompatibilizmu in inkompatibilizmu. Naključno razporejeni so dobili v presojo *konkretni* ali *abstraktni* pogoj. Preizkusili smo več različic, a se bomo osredotočili na najpomembnejše. V enem izmed konkretnih pogojev smo vprašali naslednje:

V vesolju A je moškega z imenom Bill pritegnila njegova tajnica. Odloči se, da je edini način, da bi bil z njo ta, da ubije svojo ženo in tri otroke. Ve, da je v primeru požara nemogoče pobegniti iz njegove hiše. Preden odide

na poslovno pot, v svoji kleti aktivira napravo, ki požge hišo in ubije njegovo družino.

Ali je Bill popolnoma moralno odgovoren za uboj svoj žene in otrok?

DA

NE

Pri tem primeru je večina subjektov (72 %) podala kompatibilistični odgovor, da je Bill popolnoma moralno odgovoren. To lahko primerjamo z rezultati eksperimentov Nahmiasa in njegovih sodelavcev. Sedaj pa pogledjmo enega od abstraktnih primerov:

Ali je v vesolju A mogoče, da je oseba popolnoma moralno odgovorna za svoja dejanja?

DA

NE

V tem primeru pa je večina vprašanih (86 %) dala *inkompatibilističen* odgovor!

Na kratko, večina ljudi da kompatibilističen odgovor v konkretnem primeru, ampak velika večina da *inkompatibilističen* odgovor v abstraktnem primeru. Le kaj bi lahko pojasnilo tako veliko razliko? Najprej pogledjmo deflacijsko možnost. Mogoče je konkreten primer tako dolg in kompleksen, da subjekti izgubijo izpred oči dejstvo, da je dejavnik v determinističnem vesolju. To je povsem razumna razlaga. Da bi videli, če zares pojasni razliko, smo naredili vnovičen poskus s konkretnim primerom, le da je tokrat scenarij bil kratek in preprost. Subjektom smo dali iste začetne opise in jim postavili naslednje vprašanje:

V vesolju A Bill zabode svojo ženo in otroke do smrti zato, da bi lahko bil s svojo tajnico. Ali je mogoče, da je Bill popolnoma moralno odgovoren za smrt svoje družine?

DA

NE

Tudi v tem preprostem scenariju se je 50 % vprašanih izreklo za kompatibilistični odgovor, kar se še vedno precej razlikuje od zelo majhnega števila kompatibilističnih odgovorov v abstraktnem primeru.⁹

Kot sva že omenila, obstaja več načinov, kako opisovati determinizem in splošno razmerje inkompatibilističnih odgovorov bi bilo morda višje ali nižje, če bi uporabila nekoliko drugačen opis. Vseeno pa ne moremo kar tako odpisati visokega števila inkompatibilističnih odgovorov pri abstraktnem pogoju, kot da gre le za rezultat nekega subtilnega predsodka v našem opisu determinizma. Konec koncev smo pri konkretnem pogoju uporabili natanko enak opis, vendar so vprašanci pri tem pogoju v pomembni meri bolj verjetno dajali kompatibilistične odgovore.

⁹ $\chi^2(1, N = 41) = 6.034, p < .05$, dvostranski test.

Ti začetni eksperimenti so ponovili ugotovitev (v originalu Nahmias et al.), da imajo ljudje kompatibilistične intuicije, kadar so soočeni s kratkimi vinjetami, ki sprožijo čustvene odgovore. Dali pa so nam tudi nov in presenetljiv rezultat. Kadar so subjekti dobili abstraktno vinjeto, so večinoma imeli *inkompatibilistične* intuicije. Tak vzorec rezultatov kaže na to, da čustva igrajo ključno vlogo pri tvorbi kompatibilističnih intuicij.¹⁰

5. Psihološki modeli

Do zdaj sva podajala dokazila za stališče, da so različne vsakdanje intuicije o odgovornosti proizvedene s strani različnih vrst psiholoških procesov. Ampak, če se res izkaže, da ena vrsta procesov vodi do kompatibilističnih intuicij, druga pa do inkompatibilističnih, katero vrsto procesov naj bi potemtakem uporabili kot najboljše vodilo glede resničnega odnosa med moralno odgovornostjo in determinizmom?

Preden načnemo to vprašanje, moramo vedeti nekaj več o posameznih psiholoških procesih, ki morda stojijo za različnimi tipi vsakdanjih intuicij. Zato bova obravnavala vrsto možnih modelov. Začela bova z vpogledom v tri izjemno enostavne modele in potem nadaljevala z obravnavo načinov, kako bi elemente teh enostavnih modelov lahko povezali v bolj kompleksne modele.

Model napake v izvedbi

Morda najbolj očiten način pojasnila dobljenih podatkov bi bil ta, da lahko močni čustveni odzivi vplivajo na naše sodbe in jih popačijo. Po tem stališču ljudje običajno pripisujejo odgovornost tako, da se opirajo na implicitno ozadno teorijo, ko pa so soočeni z neko zares nezaslišano kršitvijo moralnih norm (kot v naših konkretnih primerih), občutijo močno čustveno reakcijo, ki jim onemogoči, da bi teorijo pravilno aplicirali. Na kratko, ta hipoteza postulira *čustveno napako v izvedbi* (angl. affective performance error). Z drugimi besedami, razlikujemo med notra-

¹⁰ Izvedla sva tudi poskus, ki je uporabljal bolj življenjski primer kot pa deterministični okvir, opisan v glavnih eksperimentih. K temu so naju navedle nekatere dojemljive opazke Daniela Batsona, ki nam je tudi pomagal z zelo koristnimi predlogi pri načrtovanju študije. Ideja je bila v tem, da preveriva, ali abstraktni pogoji v večji meri tvorijo inkompatibilistične intuicije kot pa s čustvi obložene konkretne situacije. Vsem subjektom smo predstavili genetsko pogojenost, ki vodi do tega, da oseba izvede grozljiva dejanja, povedali pa smo jim tudi, da obstaja ne preveč draga tableta, ki izniči to pogojenost in da zdaj vsakdo, ki ima tako pogojenost, prejme tableto. V abstraktnem pogoju smo subjekte naprosili, da naj naznačijo, ali bi ljudje, ki so imeli tako pogojenost, preden je bila tableta odkrita, lahko bili moralno odgovorni za svoja dejanja. Pri konkretnem pogoju smo subjektom povedali, da je Bill imel tako pogojenost, preden je bila tableta odkrita, ter da je Bill ubil svojo ženo in otroke, da bi bil s svojo tajnico. Subjekti so potem morali naznačiti, ali je Bill bil moralno odgovoren za svoje dejanje. Rezultati so bili jasni in so se skladali z našimi predhodnimi ugotovitvami. Subjekti so v pomembno nižji meri pripisovali odgovornost pri abstraktnem pogoju kot subjekti, ki so dobili konkretno vprašanje. Ključni učinek tako lahko dobimo tudi, če uporabljamo povsem različne materiale.

njimi reprezentacijami kriterijev za moralno odgovornost ljudi in sistemi izvedbe, ki jim omogočajo, da aplicirajo te kriterije v posameznih primerih. Po tem modelu čustvene reakcije zmotijo normalno delovanje sistemov izvedbe.

Model napake v izvedbi podpira širok spekter literature iz socialne psihologije o interakciji med čustvi in teoretičnim mišljenjem. Ta literatura je razkrila številne načine, na katere lahko čustveni odzivi vplivajo na sposobnost pravilnega razmišljanja. Zaradi vpliva čustvenih ali motivacijskih predsodkov je manj verjetno, da se bodo ljudje spomnili določenih vrst relevantnih informacij, manj verjetno, da bodo prepričani v neželena dokazila in manj verjetno, da bodo uporabili kritične vire proti sklepom, ki so motivacijsko nevtralni (*glej* Kunda, 1990 za pregled). Glede na to, da takšne predsodke najdemo še pri mnogih drugih aspektih mišljenja, je naravno sklepati, da jih lahko najdemo tudi pri sodbah o moralni odgovornosti.

Bolj specifično, obstajajo dokazila, da čustva včasih vplivajo na pripise odgovornosti. Lerner in kolegi so ugotovili, da kadar se pri osebah vzbudijo negativna čustva, imajo druge za bolj odgovorne in bolj zaslužne za kazni, *celo kadar negativna čustva vzbudi povsem nepovezan dogodek* (Lerner et al., 1998). V njihovi študiji so osebe v stanju *jeze* gledale video posnetek nasilneža, ki je pretepal najstnika, medtem pa so sodelujoči pri *čustveno-nevtralnem* pogoju gledali posnetek abstraktnih figur (Lerner et al., 1998: 566). Vsem sodelujočim so potem predstavili to, za kar so jim rekli, da je drugačen eksperiment, ki preverja, kako ljudje ocenjujemo odgovornost za nepazljivost. Subjekti pri pogoju jeze (oz. tisti, ki so gledali posnetek nasilneža) so potem pripisovali večjo odgovornost kot subjekti v čustveno-nevtralnem pogoju. Čeprav so torej čustva subjektov bila inducirana z video posnetkom, so njihova čustva vplivala na sodbe o odgovornosti v povsem nepovezanih scenarijih. Najbolj naraven način za interpretacijo tega rezultata je, da so čustva res vplivala na razmišljanje ljudi v njihovih ocenah odgovornosti.

Zagovorniki modela napake v izvedbi bi lahko rekli, da je pri eksperimentih, o katerih smo tu poročali, na delu podoben pojav. Priznali bi, da ljudje dajejo kompatibilistične odgovore v določenih okoliščinah, vendar pa bi zanikali, da ljudje v pravem smislu zares zagovarjajo kompatibilističen nazor glede moralne odgovornosti. Namesto tega bi trdili, da si lahko kompatibilistične odgovore v naših konkretnih situacijah razlagamo kot napake v izvedbi, ki jih sprožijo čustveni odzivi. Pri abstraktnem pogoju pa se razkrije ozadna teorija takšna kot v resnici je – inkompatibilizem.

Model čustvene kompetence

Obstaja še en možen način razumevanja vloge čustev v ocenjevanju moralne odgovornosti. Namesto predpostavke, da čustva služijo samo predsodkom in izkrivljanju naših teoretskih sodb, bi lahko predlagali, da naše čustvene reakcije dejansko ležijo v jedru procesa, s katerim pripisujemo odgovornost. Morda ljudje obi-

čajno postavljajo sodbe o odgovornosti tako, da izkusijo čustveno reakcijo, ki, v kombinaciji z določenimi drugimi procesi, omogoča oceno moralne odgovornosti. Seveda bi težko zanikali, da imajo nekateri ljudje tudi razvite teorije o moralni odgovornosti, ki jih uporabljajo v določenih primerih (npr. pri pisanju filozofskih spisov). Vendar pa bi zagovorniki tega stališča zanikali, da naše hladne kognitivne teorije o odgovornosti igrajo kako resnejšo vlogo v procesu, s katerim običajno tvorimo sodbe o odgovornosti. Rekli bi, da ta proces v glavnem vodijo čustva.

Temu stališču 'čustvene kompetence' dajejo nekaj podpore nedavne študije o ljudeh, ki imajo zaradi psiholoških bolezni težave s čustvenim procesiranjem. Ko takim ljudem postavimo vprašanja, ki zahtevajo moralne sodbe, večkrat dajejo bizarne vzorce odgovorov (Blair, 1995; Blair et al., 1997; Hauser et al., 2006). Z drugimi besedami, ko pri osebi odstranimo sposobnost čustvenih reakcij, se zdi, da nimamo opravka z osebo, ki bi lahko aplicirala osnovna merila morale na nepristranski in neizkrivljen način. Namesto tega dobimo nekoga, ki ima težave z razumevanjem same morale. Rezultati takšnih študij so nekatere raziskovalce privedli do sklepa, da morajo čustva igrati pomembno vlogo v temeljni kompetenci, ki je podlaga moralni presoji (Blair, 1995; Haidt, 2001; Nichols, 2004b; Prinz, 2007).

Zagovorniki tega stališča bi lahko trdili, da je edini način, da bi zares razumeli človeško sposobnost moralne presoje ta, da pogledamo njihove odgovore v primerih, ki izzovejo čustvene reakcije. Ko pregledamo te primere, se zdi, da ljudje kažejo jasno tendenco k kompatibilističnim odgovorom, zato bi lahko predlagali, da subjekte v naših poskusih obravnavamo kot kompatibiliste. Seveda pa naši podatki kažejo tudi na to, da ti subjekti dajejo inkompatibilistične odgovore, ko dobijo teoretska vprašanja. A lahko bi dobili občutek, da nam preučevanje teoretskih prepričanj nič ne pove o tem, kako oseba *dejansko* pride do moralnih sodb. (Pomislite, kaj bi se zgodilo, če bi skušali preučevati človeško jezikovno sposobnost tako, da bi ljudem postavljali teoretska vprašanja o sintaksi!) Teoretiki čustvene kompetence bi lahko trdili, da je najboljši način za opis naših ugotovitev ta, da je naša osnovna moralna kompetenca kompatibilistična, toda nekateri ljudje sprejemajo teorijo, ki je v protislovju s to osnovno kompetenco.

Model konkretne kompetence

Na koncu moramo pretehtati še možnost, da na odgovore ljudi na noben način ne vplivajo čustva. Morda odgovore pri konkretnih pogojih dejansko generira čisto kognitivni proces. Četudi predpostavljamo, da lahko ta proces apliciramo zgolj na konkretne primere, iz tega ne sledi nujno, da so za proces bistvena čustva. Lahko bi se izkazalo, da imamo povsem kognitivni proces, brez vsakih čustev, ki pa ga, zaradi takšnih ali drugačnih razlogov, lahko apliciramo na konkretna vprašanja, ne pa tudi na abstraktna.

Zanimiva verzija te hipoteze je, da človeške intuicije pri konkretnih pogojih tvorijo notranji »modul moralne odgovornosti.«¹¹ Vhodni podatki za modul bi lahko bili podatki o dejavniku in njegovem ali njenem obnašanju, izhodni rezultat pa bi bila intuicija o tem, ali je dejavnik moralno odgovoren ali ne. Modul najverjetneje ne bi uporabljal enake vrste procesov kot so tisti, ki so uporabljeni pri zavestnem razmišljanju. Namesto tega bi uporabljal proces, ki je hiter, avtomatski in povsem nezaveden.

Ključna ideja je v tem, da je med tem modulom in ostalimi deli psihe možna zgolj omejena komunikacija. Modul dane osebe dobi kot vhodne podatke določene zelo specifične vrste informacij o dejavniku (dejstvo, da je dejavnik človeško bitje, dejstvo, da ve, kaj dela, ipd.), velika večina prepričanj te osebe pa bi bila popolnoma nedostopna procesom znotraj modula. Tako modul ne bi mogel dostopati do teorij, ki jih ima oseba o razmerju med determinizmom in moralno odgovornostjo. Morda celo ne bi mogel uporabiti prepričanja osebe, da je dejavnik v determinističnem univerzumu. Ker bi bila ta prepričanja nedosegljiva znotraj modula, bi se lahko sklepi modula dramatično razlikovali od sklepov, do katerih bi oseba prišla v procesu zavestnega razmišljanja.

Hibridni modeli

Do zdaj smo obravnavali tri enostavne modele pripisovanja odgovornosti. Lahko pa bi konstruirali tudi bolj kompleksne modele, in sicer z združevanjem elementov navedenih treh enostavnih modelov. Tako bi se, na primer, lahko izkazalo, da naše sodbe o moralni odgovornosti podpira modul, toda delovanje tega modula je včasih okrnjeno z napakami v čustveni izvedbi, ali pa da temeljna kompetenca, ki je osnova za sodbe o odgovornosti, na bistven način uporablja čustva, vendar pa to čustvo na nek način deluje kot vhodni podatek za modul. Možnih je še več takih hibridov.

Ker ne moreva obravnavati vseh možnih hibridnih modelov, se bova osredotočila na tistega, ki se nama zdi še posebej plavzibilen. Pri hibridnem modelu, o katerem bova govorila, igra čustvo dve različni vlogi pri pripisovanju moralne odgovornosti. Natančneje, čustvo je *obenem* del kompetence, ki je osnova sodb o odgovornosti, kot *tudi* dejavnik, ki lahko včasih vodi k napakam v izvedbi. Da bi dobili občutek, o čem govoriva, si predstavljajte, da poskušate določiti, katere pesmi naj bi obravnavali kot 'ganljive', in recimo, ugotovite, da je eno od teh pesmi napisal vaš najboljši prijatelj. Tukaj se zdi, da bi vaša osnovna kompetenca, ki je osnova vaše presoje, vključevala določeno vrsto čustev (vaša občutja glede pesmi), toda sisteme izvedbe, ki omogočajo vašo presojo, bi lahko zmotila druga vrsta čustev

¹¹ Kolikor vemo, do sedaj nobena predhodna raziskava ni postulirala modula moralne odgovornosti, je pa precej entuziazma za bolj splošno zamisel, da z mnogimi osnovnimi kognitivnimi kapacitetami upravljajo moduli (Fodor, 1983; Leslie, 1994) in vrsta avtorjev je predlagala, da bi določene aspekte moralnih sodb lahko podpirali modulom podobni mehanizmi (Dwyer, 1999; Harman, 1999; Hauser, 2006).

(vaša čustva do prijatelja). Po hibridnem modelu, ki ga obravnavamo, bi podoben proces bil na delu pri sodbah o moralni odgovornosti. Kompetenca, ki je osnova za te sodbe, uporablja čustva, toda čustva so prav tako lahko implicitna v procesih, ki na koncu vodijo do napak v izvedbi.

Zagovorniki tega modela bi lahko trdili, da čustva igrajo pomembno vlogo v kompetenci, ki je osnova za sodbe o moralni odgovornosti, vendar pa bi morali rezultate zgoraj navedenih eksperimentov še vedno obravnavati kot napake v izvedbi.¹² Z drugimi besedami, četudi predpostavljamo, da ima čustvo pomembno vlogo pri sodbah o moralni odgovornosti, lahko še vedno sklepamo, da je temeljna kompetenca v osnovi teh sodb inkompatibilistična ter da so odgovori v naših konkretnih primerih rezultat napačne aplikacije te kompetence.

6. Eksperimentalna dokazila: druga faza

Zdaj, ko smo opisali nekatere psihološke modele, ki bi lahko pojasnili naše rezultate, lahko malo globlje raziščemo, ali eksperimentalna dokazila nasprotujejo kateremu od modelov. Eno od ključnih vprašanj je, ali so kompatibilistični odgovori v naših eksperimentih res posledica čustvenega odziva. Primerjali smo konkretne pogoje z abstraktnimi primeri in predlagali, da konkretni pogoji sprožijo bolj čustven odziv, ki je subjekte potisnil proti kompatibilističnim odgovorom. Možno pa je, da je bila ključna 'konkretnost' sama po sebi, ne pa čustva, povezana s konkretnostjo. Možno je torej, da na kompatibilistične odgovore niso vplivala čustva, ampak so bili izzvani enostavno zato, ker je scenarij vseboval določeno dejanje določenega posameznika. Prav res, natanko takšne vrste razlago bi pričakovali od pojasnila posebnega modula odgovornosti. Na srečo pa obstaja neposreden način, da ta predlog preverimo.

Da bi raziskali, ali lahko konkretnost sama po sebi razloži kompatibilistične odgovore, sva izvedla še en eksperiment, v katerem se je pomembnost čustvenosti spreminjala v dveh vprašanjih, konkretnost pa je ostala konstantna. Spet so subjekti dobili začetna opisa dveh vesolj, A in B, in vse smo vprašali, katero vesolje se jim zdi najbolj podobno našemu. Subjektom smo naključno dodelili pogoj *močnega čustva* ali *šibkega čustva*. V primeru *močnega čustva* smo subjekte vprašali naslednje:

Kot že mnogokrat prej v preteklosti Bill zasleduje in posili neko tujko. Ali je mogoče, da je Bill popolnoma moralno odgovoren za posilstvo tujke?

V pogoju *šibkega čustva* pa smo subjekte vprašali:

¹² Zahvaljujemo se Jesseu Prinzu, ki je predlagal to možnost.

Kot že mnogokrat prej v preteklosti Mark goljufa pri svojih davkih. Ali je mogoče, da je Mark popolnoma moralno odgovoren za goljufanje pri davkih?

Poleg tega je pri vsakem pogoju polovica vprašanih izvedela, da je dejavnik v vesolju A, druga polovica pa, da je v vesolju B. Vsak subjekt je bil tako naključno razporejen v eno od celic tabele:

	Dejavnik v indeterminističnem vesolju	Dejavnik v determinističnem vesolju
Močna čustva		
Šibka čustva		

Kaj smo odkrili? Čeprav smo uporabili te izredno konkretne scenarije, se je pojavila jasna razlika med primeroma z močnimi in šibkimi čustvi. Osebe, ki smo jih spraševali o dejavnikih v *determinističnem* vesolju, so z veliko večjo verjetnostjo dajale inkompatibilistične odgovore pri pogoju šibkih čustev kot v primeru z močnimi čustvi. Večina ljudi je rekla, da *ni* mogoče, da bi davčni goljuf bil popolnoma moralno odgovoren, velika večina pa je tudi potrdila, da *je* mogoče, da je posiljevalec popolnoma moralno odgovoren. V nasprotju s tem pa so subjekti, ki smo jih spraševali o dejavniku v indeterminističnem okolju, večinoma dejali, da je mogoče, da je dejavnik popolnoma moralno odgovoren ne glede na to, ali je davčni goljuf ali posiljevalec.¹³ Rezultati so v tabeli:

¹³ Kot v prejšnjih poskusih je velika večina vprašanih rekla, da je naše vesolje najbolj podobno indeterminističnemu. Zdi se nama, da bi sprejemanje determinizma dejansko lahko vodilo ljudi k temu, da bi imeli bolj kompatibilistične nazore (prim. Nichols, 2006), zato sva se odločila, da iz statističnih analiz predhodno izločiva manjšino, ki je dala deterministični odgovor. Statistični podatki so naslednji. Kontrast med pogojema močnih in šibkih čustev za pogoj determinizma je bil statistično pomemben ($\chi^2(1, N = 44) = 8,066, p < .01$). Se pravi, ljudje so z večjo verjetnostjo trdili, da je mogoče, da je posiljevalec popolnoma moralno odgovoren. Kontrast med pogojema z močnimi čustvi je bil prav tako statistično pomemben ($\chi^2(1, N = 45) = 7,204, p < .01$). Se pravi, ljudje so z večjo verjetnostjo trdili, da je mogoče, da je posiljevalec popolnoma moralno odgovoren v indeterminističnem vesolju. Kontrast med obema pogojema s šibkimi čustvi je bil zelo visok ($\chi^2(1, N = 45) = 26,492, p < 0.0001$). Osebe so z dramatično večjo verjetnostjo rekle, da je mogoče, da je davčni goljuf popolnoma moralno odgovoren v indeterminističnem univerzumu.

	Dejavnik v indeterminističnem vesolju	Dejavnik v determinističnem vesolju
Močna čustva	95 %	64 %
Šibka čustva	89 %	23 %

Ti rezultati nam pomagajo razjasniti vlogo čustev v pripisovanju odgovornosti. Tudi ko preverjamo konkretnost, vidimo, da čustva vplivajo na človeške intuicije o odgovornosti v determinističnem okolju. Splošni vzorec rezultatov torej nakazuje, da imajo čustva pomembno vlogo v procesu, ki generira kompatibilistične intuicije.

Zdaj imamo dobra dokazila, da igrajo čustva neko vlogo v kompatibilističnih sodbah. Vseeno pa ostaja težko vprašanje, ali je to, kar se kaže v odgovorih, rezultat čustvene kompetence ali čustvene napake v izvedbi. Poglejmo, ali kateri od teh modelov bolje pojasni eksperiment, o katerem sva poročala.

Misliva, da model napake v izvedbi precej dobro pojasni najine rezultate. Kar vidimo v primeru davčne goljufije je, da ljudje, kadar je vloga čustev minimizirana, dajejo dramatično različne odgovore glede na to, ali je dejavnik v determinističnem ali indeterminističnem univerzumu. Po hipotezi o napaki v izvedbi ti odgovori razkrivajo pravo kompetenco glede pripisovanja odgovornosti, saj je v primeru, ko so čustva šibka, čustveni predsodek minimiziran. V primeru močnih čustev, kot je primer serijskega posiljevalca, pa čustva popačijo normalno kompetenco pripisovanja odgovornosti. To pojasni, zakaj nastopa tako velika razlika med primeri s šibkimi in močnimi čustvi pri pogoju determinizma.

Oglejmo si zdaj model čustvene kompetence. Mnogo manj jasno je, da bi teoretik čustvene kompetence lahko imel dobro razlago za rezultate. Predvsem se zdi težko videti, kako lahko model čustvene kompetence pojasni, zakaj odzivi v primeru šibkih čustev strmo padejo pri pogoju determinizma, v primeru močnih čustev pa temu ni tako. Morda bi teoretik čustvene kompetence lahko rekel, da primeri šibkih čustev, kot je davčna goljufija, ne sprožijo naše kompetence pripisovanja odgovornosti, zato teh odgovorov ne bi smeli šteti za odraz naše običajne kompetence. Očitno pa je, da bi bilo potrebne veliko dela, da bi pokazali, da takšni vsakdanji primeri domnevnega pripisovanja odgovornosti v resnici ne štejejo kot primeri, v katerih deluje naša kompetenca pripisovanja odgovornosti. Zato na prvi

pogled model napake v izvedbi ponuja boljšo razlago teh rezultatov kot pa pojasnilo čustvene kompetence.

Seveda, četudi je morda res, da najine rezultate najbolje razloži model napake v izvedbi, to ne pomeni, da so čustva nepomembna za normalno kompetenco. Kot je bilo rečeno v prejšnjem razdelku, je ena od plavzibilnih možnosti hibridni model, pri katerem (i) je naša običajna kompetenca v pripisovanju odgovornosti odvisna od čustvenih sistemov, vendar pa (ii) čustva prav tako generirajo predsodek, ki vodi do kompatibilističnih odgovorov v naših eksperimentih.

Čeprav najin eksperiment ponuja nekaj razlogov v podporo modelu napake v izvedbi, pa se zdi jasno, da odločitev med modelom napake v izvedbi in modelom čustvene kompetence kot pojasnilom kompatibilističnih odgovorov ne more temeljiti na enem samem ključnem eksperimentu. Tukaj bi potrebovali globlje razumevanje vloge čustev v moralni kogniciji nasploh. (Če bi imeli globlje razumevanje tega bolj splošnega področja, bi lažje ugotovili, kako bi lahko empirične študije obravnavale bolj specialno vprašanje o vlogi čustev v sodbah o moralni odgovornosti.) Ampak naša nezmožnost takojšnje razrešitve vseh relevantnih vprašanj še ni razlog za pesimizem. Nasprotno, lahko smo optimistični glede perspektiv raziskovanja na tem področju. V zadnjih letih je zanimanje za načine, katere čustva vplivajo na moralno mišljenje, močno naraslo – z novimi empiričnimi študijami in stalnim teoretskim razvojem – zelo verjetno je, da bo naslednjih nekaj let prineslo pomembne nove vpoglede v to vprašanje.

7. Filozofske implikacije

Najine ugotovitve nam pomagajo razložiti, zakaj je razprava med kompatibilisti in inkompatibilisti tako trmasto vztrajna. Zdi se, da določeni psihološki procesi tvorijo kompatibilistične intuicije, spet drugi pa inkompatibilistične. Tako se vsako od teh dveh stališč sklicuje na določen element naše psihološke zgradbe.

Eksperimentalni rezultati pa nam ne dajo zgolj vpogleda v vzročni izvor določenih filozofskih stališč, pomagajo nam tudi pri oceni nekaterih argumentov, s katerimi so ta stališča podprta. Navsezadnje se mnogo teh argumentov opira na eksplicitno sklicevanje na intuicije. Če ugotovimo, da so različne intuicije posledica različnih psiholoških mehanizmov, bomo mogoče prišli do zaključka, da bi morale imeti nekatere od teh intuicij večjo težo kot druge.

Kar moramo zdaj vedeti je, katere intuicije moramo jemati resno in katere opustiti, ker so proizvodi mehanizmov, ki zavajajo.

Jasno je, da bo odgovor delno odvisen od tega, kateri, če sploh kateri, od treh navedenih modelov se bo izkazal za pravega. In ker zaenkrat še nimamo podatkov, potrebnih za odločitev med tekmujočimi modeli, na tem mestu še ne moreva ponuditi končnega sklepa. Najin pristop bo torej v tem, da obravnavava vsakega od

modelov posebej in se vprašava, kakšne implikacije bi imel (če bi se izkazalo, da je pravilen) za širša filozofska vprašanja o vlogi intuicij v razpravi o moralni odgovornosti.

Model napake v izvedbi

Če kompatibilistične intuicije razlagamo s pomočjo modela napake v izvedbi, tem intuicijam ne bi smeli pripisati velike teže. Kajti v tem modelu so, kot smo opisali, kompatibilistične intuicije rezultat izkrivljajočih učinkov čustev in motivacije. Če bi napake v izvedbi lahko odstranili, bi morale kompatibilistične intuicije izginiti.

Bodite pozorni, da model napake v izvedbi ne trdi, da so kompatibilistične intuicije dejansko *nepravilne*. Vse, kar trdi je, da proces, ki te intuicije ustvarja, vsebuje napako določene vrste. Seveda je mogoče, da kljub temu, da proces vsebuje to napako, na koncu vodi do pravilnega sklepa. Še vedno pa se nam zdi, da ima model napake v izvedbi pomembne filozofske implikacije. Najmanj kar je, po tem modelu dejstvo, da imajo ljudje včasih kompatibilistične intuicije, samo po sebi ni razlog za predpostavko, da je kompatibilizem pravilen.

Filozofske implikacije modela napake v izvedbi imajo poseben pomen, saj se zdi, da do sedaj zbrana eksperimentalna dokazila kažejo, da je osnovna ideja tega modela dejansko resnična. No, končne sodbe še ni. Nadaljnje raziskave bi lahko pokazale, da je kateri od drugih modelov bolj natančen, zato moramo razmisliti tudi o filozofskih posledicah teh modelov.

Model čustvene kompetence

Odgovori ljudi v konkretni situaciji v našem začetnem eksperimentu so pristni odraz njihove notranje kompetence. Po tem modelu kompatibilistični odgovori ljudi v teh pogojih niso zamegljeni z nikakršno napako v izvedbi. Pač pa odgovori odražajo uspešno implementacijo sistema, ki ga običajno uporabljamo za tvorbo sodb o odgovornosti, in ta sistem bi tako morali obravnavati kot kompatibilistični sistem.

V mnogih pogledih model čustvene kompetence spominja na stališče, ki ga P. F. Strawson zagovarja v svojem klasičnem članku »Svoboda in zamera« (1962). Po tem stališču bi bil zmoten poskus razumevanja pojma moralne odgovornosti, ki bi temeljil na neke vrste metafizični teoriji. Najboljše je začeti z raziskavo povratnih odzivov (angl. reactive attitudes, krivda, obžalovanje, hvaležnost, itd.) in njihove vloge pri našem vsakdanjem pripisovanju odgovornosti.

Kljub očitnim podobnostim med modelom čustvene kompetence in Strawsonovo teorijo pa je pomembno, da opozorimo, da je v določenih ozirih model čustvene kompetence zavezan substancialno šibkejšim trditvam. Predvsem model ne trdi, da ljudje *pravilno* presojujejo v konkretnih pogojih. Vse, kar trdi, je le to, da odgo-

vori ljudi pri teh pogojih odražajo uspešno implementacijo njihovega notranjega sistema za tvorbo sodb o odgovornosti. Ta trditev torej pušča odprto vprašanje, ali so kriteriji, uporabljeni v tem notranjem sistemu, sami po sebi pravilni ali ne.

Za primerjavo si oglejmo načine, na katere ljudje po navadi tvorijo sodbe o verjetnosti. Izkaže se, da verjetnostne sodbe velikokrat vsebujejo nepravilna sklepanja, zato bi morda domnevali, da ljudje nepravilno aplicirajo svoje lastne notranje kriterije za verjetnostna sklepanja. Mnogo psihologov pa to stališče zavrača. Trdijo namreč, da ljudje dejansko pravilno aplicirajo svoje notranje kriterije, zmotna verjetnostna sklepanja pa nastanejo, ker so ti kriteriji sami po sebi zmotni (glej npr. Tversky in Kahneman, 1981; 1983).

Jasno je, da bi lahko podoben pristop uporabili v primeru sodb o odgovornosti. Četudi kompatibilistične intuicije ljudi odražajo uspešno implementacijo njihovega notranjega sistema za tvorbo sodb o odgovornosti, bi še vedno lahko rekli, da je sistem sam po sebi zmoten. Zato bi model čustvene kompetence lahko potrjeval predpostavko, da so temeljni pogledi ljudi na odgovornost kompatibilistični, vendar pa bi bilo napačno obravnavati ta model kot neposredni zagovor teh intuicij.

Model konkretne kompetence

Implikacije modela konkretne kompetence so na ključen način odvisne od natančnih podrobnosti obravnavane kompetence. Ker ne moremo reči ničesar zelo splošnega o vseh modelih v tej osnovni kategoriji, se bomo osredotočili na implikacije trditve, da so pripisovanja odgovornosti zajeta v posebnem modulu.

Kot je omenilo mnogo avtorjev, modularnost vsebuje neke vrste kompromis. Ključne prednosti modulov so, da običajno delujejo avtomatsko, nezavedno in izredno hitro. A te prednosti imajo svojo ceno. Razlog, zakaj so moduli zmožni tako hitrega delovanja, je, da preprosto ignorirajo določene vire potencialno relevantnih informacij. Tudi ko vemo, da sta črti v Müller-Lyerjevi iluziji enako dolgi, smo še vedno žrtve optične iluzije. Morda imamo pri pripisovanju moralne odgovornosti opravka s podobnim fenomenom – z 'moralno iluzijo'. Morda imajo ljudje kompleksno in sofisticirano teorijo o odnosu med determinizmom in moralno odgovornostjo, toda relevantni modul enostavno nima dostopa do te teorije. Vedno znova podaja sodbe, da je dejavnik kriv, četudi so te sodbe v nasprotju z zavestno teorijo, ki jo imamo nekje drugje v umu.

Seveda bi zagovorniki kompatibilizma lahko opozorili, da ta argument lahko tudi obrnemo. Lahko bi trdili, da modul sam po sebi vsebuje kompleksno in sofisticirano teorijo, do katere pa preostali um nima dostopa. Sklep bi potem bil, da bomo prišli do osiromašenega in neustreznega razumevanja, razen če ne uporabimo tega modula za presojo odnosa med determinizmom in moralno odgovornostjo. Ta vrsta argumenta se na nekaterih področjih gotovo zdi plavzibilna (npr. področje

slovnične teorije). Ni pa jasno, ali kaj podobnega velja v domeni pripisovanja odgovornosti.¹⁴

Refleksivni ekvilibrij

V tem delu sva se ukvarjala s filozofskimi vprašanji o tem, ali nam védenje o določenih mentalnih procesih lahko omogoči koristen vpogled v kompleksna moralna vprašanja. Seveda moramo ta filozofska vprašanja skrbno razlikovati od čistega psihološkega vprašanja, ali ljudje *mislijo*, da jim določeni mentalni procesi dajejo vpogled v te zadeve. Četudi ljudje mislijo, da jim dani proces omogoča koristen moralni vpogled, se lahko izkaže, da je ta proces dejansko popolnoma nezanesljiv in da bi bilo zanje bolje, če bi k zadevi pristopili na povsem drugačen način.

Vseeno pa se nama je zdelo, da bi bilo zanimivo izvedeti, kako ljudje sami razrešujejo napetost med svojimi nasprotujočimi se intuicijami, zato sva izvedla še zadnji eksperiment. Vsem sodelujočim sva dala kratek opis rezultatov naših prejšnjih študij in jih nato pozvala, da razsodijo o konfliktu med kompatibilističnimi ter inkompatibilističnimi intuicijami. Glede na to, da intuicije ljudi v konkretnih primerih nasprotujejo njihovim intuicijam v abstraktnih primerih, ali bi še vztrajali pri konkretni sodbi, da je Bill moralno odgovoren ali pri abstraktni sodbi, da v determinističnem univerzumu nihče ne more biti odgovoren?¹⁵ Rezultati so pokazali, da nobena stran ni imela večine. Približno polovica sodelujočih je izbrala sodbo, da je dejavnik moralno odgovoren, medtem ko je druga polovica izbrala

¹⁴ Razlika med hipotezo o modularnosti in hipotezo o čustvih prvič nastopi v filozofski literaturi v kontekstu razprave o vlogi moralnih razmislekov pri intencionalnih dejanjih (Knobe, 2006; Malle in Nelson, 2003; Nadelhoffer, 2004; Young et al., 2006). V tem kontekstu hipotezo o modularnosti običajno obravnavajo kot obrambo vsakdanjih intuicij. Vendar pa obstaja ključna razlika med tem kontekstom in sedanjim. Razlika je v tem, da bi informacija o moralnem statusu dejanj lahko bila dostopna modulu intencionalnih dejanj, za informacijo o determinizmu pa ni verjetno, da bi bila dostopna v modulu moralne odgovornosti.

¹⁵ Načrt pilotske študije je bil narejen po modelu začetnih eksperimentov, opisanih v razdelku 3. Udeležence smo vprašali tako o primeru močnih čustev (Bill zabode svojo ženo) kot abstraktnih primerih (zaradi uravnoteženja). Potem so morali odgovoriti na vprašanje o refleksivnem ekvilibriju:

Predhodne raziskave kažejo, da ko ljudje dobijo zgornje vprašanje 3, pogosto rečejo, da je Bill popolnoma moralno odgovoren za umor svoje družine. Ko pa ljudje dobijo zgornje vprašanje 2, jih večina reče, da ni mogoče, da bi bili ljudje v veselju A popolnoma moralno odgovorni za svoja dejanja. Jasno je, da so ti odgovori nekonsistentni. Ker ni mogoče, da bi bili v veselju A popolnoma moralno odgovorni, Bill ne more biti popolnoma moralno odgovoren.

Zanima nas, kako bodo ljudje razrešili ta nesklad. Zato vas prosimo, da ne glede na to, kako ste odgovorili na vprašanji 2 in 3, označite, s čim od navedenega se najbolj strinjate:

- i. V veselju A *ni* mogoče, da bi ljudje bili moralno odgovorni za svoja dejanja.
- ii. Bill, ki je v veselju A, je popolnoma moralno odgovoren za uboj svoje družine.

sodbo, da ne more biti v determinističnem univerzumu nihče odgovoren.¹⁶ Očitno o teh zadevah ni pri običajnih ljudeh nič več konsenza kot pri filozofih.

8. Zaključek

Kot sva rekla na začetku, so se sodelujoči v debati glede moralne odgovornosti sklicevali na velikansko raznovrstnost argumentov. Tako ali drugače naj bi bile relevantne teorije iz metafizike, moralne filozofije, filozofije duha in celo kvantne mehanike. Raziskovalci stalno odkrivajo nove načine, kako bi na videz nepovezane razmisleke vključili v razpravo o tem problemu. Pričujoči esej ni obravnaval celotnega obsega te razprave. Omejila sva se samo na en tip dokazil – dokazila, ki temeljijo na intuicijah ljudi.

Filozofi, ki so obravnavali laične intuicije na tem področju, trdijo, da so intuicije ljudi bodisi kompatibilistične ali inkompatibilistične. Najine ugotovitve so bile bistveno bolj kompleksne in morda bolj zanimive. Zdi se, da imajo ljudje *tako* kompatibilistične *kot* inkompatibilistične intuicije. Še več, zdi se, da različne vrste intuicij tvorijo različni psihološki procesi. Da pa bi ocenili pomen teh ugotovitev za debato o moralni odgovornosti, bi morali točno vedeti, katere vrste psihološki proces je ustvaril določen tip intuicij in kolikšno težo naj pripišemo rezultatu vsakega od teh procesov. Tu sva le začela z obravnavo teh problemov, jasno pa je, da je še veliko več dela pred nami.

Literatura

- Blair, R. (1995). "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition*, 57.
- Blair, R., Jones, L., Clark, F., Smith, M., and Jones, L. (1997). "The Psychopathic Individual: A Lack of Responsiveness to Distress Cues?" *Psychophysiology*, 34.
- Campbell, C. A. (1951). "Is 'Free Will' a Pseudo-problem?" *Journal of Philosophy*, 60.
- Dwyer, S. (1999). "Moral Competence". V K. Murasugi and R. Stainton (ur.), *Philosophy and Linguistics*. Westview Press.
- Fischer, J. (1999). "Recent Work on Moral Responsibility". *Ethics*, 110.
- Fodor, J. (1983). *Modularity of Mind*. MIT Press.

¹⁶ Sodelovalo je 19 oseb. Od teh jih je 10 dalo inkompatibilistični odgovor na vprašanje o refleksivnem ekvilibrju, 9 pa kompatibilistični.

- Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility". *Journal of Philosophy*, 66.
- Frankfurt, H. (1988). *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Greene, J., Sommerville, R., Nystrom, L., Darley, J., and Cohen, J. (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment". *Science*, 293.
- Haidt, J. (2001). "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment". *Psychological Review*, 108.
- Haidt, J., Koller, S.H., Dias, M.G. (1993). "Affect, Culture, and Morality, or Is it Wrong to Eat Your Dog?". *Journal of Personality and Social Psychology*, 65.
- Harman, G. (1999). "Moral Philosophy and Linguistics". V K. Brinkmann (ur.), *Proceedings of the 20 th World Congress of Philosophy: Volume 1: Ethics*. Philosophy Documentation Center.
- Hauser, M. (2006). *Moral Minds: The Unconscious Voice of Right and Wrong*. NY: Harper Collins.
- Hauser, M., Young, L., and Cushman, F. (forthcoming). "Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions". V W. Sinnott-Armstrong (ur.). *Moral Psychology*. Cambridge, MA: MIT Press.
- Hume, D. (1740/1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Kane, R. (1999). "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism". *Journal of Philosophy*, 96.
- Knobe, J. (2003a). "Intentional Action and Side-Effects in Ordinary Language". *Analysis*, 63.
- Knobe, J. (2003b). "Intentional Action in Folk Psychology: An Experimental Investigation". *Philosophical Psychology*, 16.
- Knobe, J. (2006). "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology". *Philosophical Studies*, 130.
- Kunda, Z. (1990). "The Case for Motivated Reasoning". *Psychological Bulletin*, 108.
- Lerner, J., Goldberg, J., and Tetlock, P. (1998). "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility". *Personality and Social Psychology Bulletin*, 24.

- Leslie, A. (1994). "ToMM, ToBY and Agency: Core Architecture and Domain Specificity". V L. Hirschfeld and S. Gelman (ur.) *Mapping the Mind*. Cambridge: Cambridge University Press.
- Malle, B. and Nelson, S. (2003). "Judging Mens Rea: The Tension Between Folk Concepts and Legal Concepts of Intentionality". *Behavioral Sciences and the Law*, 21.
- McIntyre, R., Viney, D., and Viney, W. (1984). "Validity of a scale designed to measure beliefs in free will and determinism". *Psychological Reports*, 54.
- Nadelhoffer, T. (2004). "Praise, Side Effects, and Folk Ascriptions of Intentional Action". *The Journal of Theoretical and Philosophical Psychology*, 24.
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2005). "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility". *Philosophical Psychology*, 18.
- Nichols, S. (2002). "Norms with Feeling." *Cognition*, 84.
- Nichols, S. (2004a). "The Folk Psychology of Free Will: Fits and Starts". *Mind & Language*, 19.
- Nichols, S. (2004b). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- Nichols, S. (2006). "Folk Intuitions about Free Will". *Journal of Cognition and Culture*, 6.
- Nichols, S. (2007). "The Rise of Compatibilism: A Case Study in the Quantitative History of Philosophy". *Midwest Studies in Philosophy*, 31.
- Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Pylyshyn, Z. (1999). "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception". *Behavioral and Brain Sciences*, 22.
- Smart, D., and Loewenstein, G. (2005). "The Devil You Know: The Effects of Identifiability on Punitiveness". *Journal of Behavioral Decision Making*, 18.
- Strawson, G. (1986). *Freedom and Belief*. Oxford: Oxford University Press.
- Strawson, P. (1962). "Freedom and Resentment". *Proceedings of the British Academy*, 48. Slovenski prevod »Svoboda in zamera«. V Šuster, D. (2007), (ur.) *O svobodni volji. Od Leibniza do Libeta*. Maribor: Aristej.
- Tversky, A., Kahneman, D. (1981). "The Framing of Decisions and the Psychology of Choice". *Science*, 211.

- Tversky, A., Kahneman, D. (1983). "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probabilistic Reasoning". *Psychological Review*, 90.
- Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.
- Vargas, M. (2006). "On the Importance of History for Responsible Agency". *Philosophical Studies*, 127.
- Viney, W., Waldman, D., and Barchilon, J. (1982). "Attitudes toward Punishment in Relation to Beliefs in Free Will and Determinism". *Human Relations*, 35.
- Viney, W., Parker-Martin, P., and Dotten, S. D. H. (1988). "Beliefs in Free Will and Determinism and Lack of Relation to Punishment Rationale and Magnitude". *Journal of General Psychology*, 115.
- Watson, G. (1987). "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme". V *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, F. Schoeman (ur.), Cambridge University Press.
- Weinberg, J., Nichols, S., and Stich, S. (2001). "Normativity and Epistemic Intuitions". *Philosophical Topics*, 29.
- Woolfolk, R., Doris, J., and Darley, J. (2006). "Identification, Situational Constraint, and Social Cognition". *Cognition*, 100.
- Young, L., Cushman, F., Adolphs, R., Tranel, D., and Hauser, M. (2006). "Does Emotion Mediate the Effect of an Action's Moral Status on its Intentional Status? Neuropsychological Evidence". *Journal of Cognition and Culture*, 6.