ANαliZA

časopis za kritično misel

Letnik 2, številka 1, junij 1998 · ISSN 1408-2969

PREVOD IN KOMENTAR

Fred Dretske

Ali je inteligenca lahko umetna?

prevedel Anton Mohorčič

Citiranje: Fred Dretske, »Ali je inteligenca lahko umetna?«, *Analiza: časopis za kritično misel* 2, št. 1 (1998): 129–140.

Prevod. Licenca Creative Commons, ki velja za arhiv revije, za prevode ne velja; pravice izvirnika pripadajo njegovim imetnikom.

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.

Fred Dretske
Prevedel Anton Mohorčič

ALI JE INTELIGENCA LAHKO UMETNA?

Miši niso velike živali, toda obstajajo velike miši. Zaradi istega razloga, čeprav pezdirki niso inteligentne živali - mogoče niti inteligentne ribe -, obstajajo inteligentni pezdirki. Inteligentni pezdirki so tisti, ki so nadpovprečno zmožni - zmožnejši od povprečnega pezdirka, ne pa od povprečne živali.

Inteligenco pezdirkov omenjam z namenom, da se razjasni bistvo inteligence in potemtakem umetne inteligence; torej možnost umetne inteligence. Ko premišljujemo o tem, ali je umetna inteligenca možna, lahko na inteligenco gledamo na dva različna načina. Lahko razmišljamo o inteligenci kot o nečem, kar je podobno denarju, kar skoraj vsakdo ima, vendar ga imajo eni več kot drugi. Ali pa gledamo na inteligenco kot na nekaj, kar je bolj podobno bogastvu, ki je v posesti zgolj tistih, ki imajo (najmanj) več kot povprečno vsoto denarja. Na prvi način bi pezdirki prav lahko bili inteligentni: imajo *nekaj* od tistega, kar potrebujejo, nekaj umske zmožnosti, čeprav ne veliko. Če nanje gledamo na drugi način, pa seveda niso inteligentni. Nimajo *dovolj* tistega, kar potrebujejo.

Filozofi so nagnjeni razmišljati o inteligenci na prvi način, kot bolj podobni denarju, računalniški strokovnjaki pa bolj na drugi način, kot bolj podobni bogastvu. To pogosto pripelje do nesporazuma, ko je predmet pogovora možnost umetne inteligence. Kajti če je inteligenca kot denar, nekaj, kar imajo geniji, idioti in celo pezdirki (v različnih količinah), potem so vprašanja o možnosti umetne inteligence vprašanja o možnosti strojev z *nekaj* umske zmožnosti - količina ni pomembna. Če pa po drugi strani inteligenco razumemo na primerjalen način, kot umsko bogastvo, potem je seveda za dosežke idiotov in pezdirkov zelo malo zanimanja. Možnost umetne inteligence potem postane možnost izdelave strojev, ki lahko v igri *zmagajo*, in ne strojev, ki lahko igro le *igrajo*.

Ko se kot filozof, ki ga zanima um in kaj je potrebno zanj, vprašam, ali je umetna inteligenca možna, takrat se sprašujem, kdo ima denar, ne pa kdo ga ima *veliko*. Če pezdirk, celo neumen pezdirk, plava na drugo stran bazena, ker je lačen in misli, da je tam hrana, potem je ta pezdirk inteligenten v mojem pomenu besede, v pomenu denarja. Inteligenten je zato, ker dela nekaj zato, ker misli, da je to način, da si pridobi tisto, kar si želi, in vsak organizem, čigar obnašanje vodi misel na ta način, je, *ipso facto*, inteligenten. Ni potrebno, da je misel, ki pojasnjuje obnašanje, globokoumna. Ni potrebno, da je rezultat razmišljanja, zaključek sklepanja ali rešitev problema. Celo to ni potrebno, da je resnična. Samo misel mora biti. Če je misel, potem je obnašanje, ki ga pojasnjuje, inteligentno obnašanje v tem minimalnem pomenu, v pomenu denarja, v filozofskem pomenu inteligence. To bitje

ima um - najmanj eno misel - in um, ta misel, nadzira početje živali.

Računalniški strokovnjaki, kot sem dejal, gledajo na inteligenco na primerjalen način: inteligenca kot bogastvo. Njih ne zanimajo dosežki pezdirka, celo tistega ne (če je sploh kakšen), ki je inteligenten v mojem pomenu te besede. Nekaterih od njih, tako sumim, to ne zanima zato, ker domnevajo, da je umetna inteligenca na tej stopnji že življenjsko dejstvo (ali ne obstajajo poceni igrače, ki so bistre kot pezdirki?). Namesto tega naj bi se raziskava UI ukvarjala z bolj spoštovanja vrednimi sposobnostmi bolj inteligentnih živali: človeških bitij.

To razhajanje v zanimanju ne omenjam zato, da bi omalovaževal prizadevanje pri raziskovanju UI (kakorkoli naj že, natanko, interpretiramo to raziskovanje), ampak da bi anticipiral možen nesporazum. Mene pri vprašanju umetne inteligence zanima, če je možno zgraditi inteligentne stroje v minimalnem pomenu besede, v pomenu denarja. Obstajajo očitni tržni razlogi za večje ukvarjanje s stroji z *visokim* IQ, toda mene ne zanima vprašanje, ali je ta količnik visok, ampak ali je, v primeru strojev, večji od nič - to je, ali so stroji sploh *mentalno* polnoletni. Ni vprašanje, kako bogati so, ampak ali je tisto, kar imajo, resnično denar. Če ničesar od tistega, kar delajo, ne vodi misel, potem so njihovi dosežki nepravi in njihovo mentalno bogastvo ponaredek. Nič, kar počno, ni zares pojasnjeno s tem, kar mislijo.

V skladu s tem razumevanjem inteligence je potem vprašanje o možnosti umetne inteligence pravzaprav vprašanje, ali lahko zgradimo umetne sisteme, katerih obnašanje, vsekakor pa del tega, naj bi bilo razumljeno v okviru tistega, kar mislijo. To ni, kot bomo videli, samo vprašanje, ali lahko zgradimo stroje, ki mislijo. Tudi če bi jih lahko, to še ne bi bil odgovor na vprašanje, ali je umetna inteligenca mogoča. Kajti inteligenca zahteva *več* kot misel, kot bomo videli, pa je ravno ta "več" tisto, zaradi česar je možnost inteligence sporna.

1. Inteligenca: delovanje, ki ga vodi misel

Dejal sem, da je inteligenca v minimalnem smislu, v smislu denarja, stvar tega, ali tisto, kar mislite,[1] vodi to, kar počnete, ali pa vsaj del tega. Potrebno je, da povem nekaj več o tem, ker vse implikacije morda niso očitne. Vse implikacije pa so potrebne za oceno možnosti umetne inteligence.

Reči, da inteligenca zahteva, da misel vodi delovanje, je reči, da sama misel ni dovolj. Misli morajo nekaj *storiti*. Za to, da bi bil inteligenten, ni dovolj biti mislec ali biti dejaven. Mišljenje mora - vsaj včasih - pojasniti delovanje. Znojim se ob otoplitvi in drgetam ob ohladitvi. Zaspim ob gledanju televizije, smrčim in se zbudim. Mežikam, kolcam, rigam, diham, zardevam, kiham, sanjam, motim se, pozabljam imena in izgubljam ključe. Vse to počnem. Prav tako imam temu primerne misli. Toda dejstvo, da sem oboje, mislec in dejaven, ne prispeva ničesar k inteligenci. To pa zato ne, ker čeprav počnem vse to, moje početje ni pojasnjeno z ničimer, kar mislim. Drgetam, če to hočem ali ne in ne glede na to, če imam to za

[1] To naj se razume tako, da "kar počnete", ne potrebuje telesne dejavnosti. To je lahko sklep, reševanje problema ali seštevanje števil v glavi.

koristen način obnašanja na mrazu ali ne. Ko naredite nenaden gib proti mojim očem, mežiknem. Ne morem si pomagati. Seveda si ne želim vašega prsta v očesu. Prav tako je res, da mislim, da je eden izmed načinov, da se izognem prstu v očesu ta, da zaprem oko, ko sunete proti njemu. Toda čeprav mislim tako in čeprav zaprem oko, ko sunete proti njemu, ga ne zaprem zato, ker mislim tako. Oko bi zaprl, če bi takšna prepričanja imel ali ne. Mehanizmi teh refleksov so globoko vsajeni. Sprožijo se dosti prej, preden ima zavestna misel čas za delovanje. Imam razloge, da zaprem oči, toda oči so zaprte, preden imajo ti razlogi čas za delovanje.

Kljub dejstvu, da mežikam in da je takšno obnašanje koristno, kljub dejstvu, da prav tako mislim, da je to koristna stvar, to obnašanje ni razodevanje inteligence. To ni razodevanje inteligence zato, ker obnašanja, čeprav ustreza misli, ne pojasnjuje misel. Ne vodi ga misel. Tega ne počnem zaradi tistega, kar mislim. Lahko se, če hočemo, zahvalimo evoluciji za oblikovanje samodejnih refleksov, ki v takšnih primerih preprečujejo poškodbe, toda ne moremo se za to zahvaliti sebi. Če naj to šteje kot inteligentno ukrepanje, gre zasluga materi Naravi in ne meni. Če v takšnih okoliščinah pripisujem inteligenco *sebi*, je to napačno pripisovanje zaslug. To bi bilo podobno, kot če bi protipožarni napravi pripisali inteligenco, ko ta pogasi ogenj, ki bi jo sicer uničil. Da bi bili inteligentni, morate narediti več kot to, kar bi naredil nek inteligenten sistem. Kajti protipožarna naprava počne to. Naredi prav to, kar bi naredil sam, če bi mi, nezmožnemu pobega, grozilo uničenje zaradi požara. Pogasi požar. Za inteligenco ni potrebno samo to, ampak obnašanje, ki ga vodijo razlogi, ki ga delajo razumnega. Zahteva početje nečesa (če je koristno ali ne), *ker* je to bilo mišljeno kot sredstvo za dosego nekega želenega cilja. To je nekaj, kar sam včasih počnem in kar je storil ustvarjalec protipožarne naprave, ne pa protipožarna naprava sama.

Žival, ki se povsem po naključju ustavi za balvanom, ki jo zakrije pred pogledom bližnjega plenilca, se ne vede inteligentno samo zato, ker ji to dejanje rešuje življenje. Delati nekaj, kar vam rešuje življenje, je inteligentno, toda to šteje za inteligentno obnašanje samo, če obnašanje nastopi zato, *ker* je bilo mišljeno, da ima tak učinek; samo če misel, da bo to izid ali verjeten izid, nastopa pri povzročitvi obnašanja. Če misel ne igra nobene vloge pri povzročitvi obnašanja, potem - kakorkoli bi takšno obnašanje lahko bilo koristno - se ne šteje kot inteligentno, celo če ga spremlja misel. Molji, ki pristajajo na površinah, ki ustrezajo njihovi lastni obarvanosti, počno nekaj, kar povečuje možnost njihovega lastnega preživetja, niso pa zaradi tega inteligentni. Samo če so počeli to z namenom, da bi se zaščitili pred ostrookimi plenilci, ker so mislili, da bodo zaradi tega težje opazni, bi počeli nekaj, kar bi kazalo inteligenco. Tega pa vsekakor ne počnejo zaradi tega. Čeprav je to pametno narediti, to niso pametni molji. Obnašanje nima pravega izvora. Ne vodi ga misel.

Tako se kljub mojemu govorjenju o inteligenci pezdirkov izkaže, da pezdirki mogoče sploh niso inteligentni. Celo tisti ne, ki so nadpovprečno zmožni. Vse je odvisno od tega, ali je ta zmožnost (pa naj bo podpovprečna ali nadpovprečna) pod nadzorom misli.

Do te točke sem govoril o obnašanju, ki ga nadzira misel, o obnašanju, ki ga vodi misel, nedoločeno. Dejal sem, da je to nujno za inteligenco, pravzaprav, da je to njeno bistvo. Toda pri tem, kako razumemo ta odnos nadzora ali vodenja, moramo biti previdni. Da bi bilo obnašanje inteligentno, mora tisto, *kar nekdo misli* - vsebina misli, ne pa, kot se to dogaja, nosilec te vsebine - nadzirati obnašanje. To je subtilna, toda nič manj pomembna razlika. Dovolite, da jo poskusim pojasniti z analogijo.

Rečem nekaj v mikrofon. Rečem: "Zatresi se hitro!" Z odzivom na moj ukaz se opna mikrofona hitro zatrese. To tresoče obnašanje naprave je določeno s frekvenco in amplitudo zvoka, ki ga proizvajam. Pri proizvajanju teh zvokov nekaj povem. Zvoki, ki jih proizvajam, imajo pomen. Ker zvoki, ki jih proizvajam, pomenijo *zatresi se hitro*, mi je s proizvajanjem teh zvokov (z določenimi nameni) uspelo reči *zatresi se hitro*. Kljub dejstvu, da mikrofon dela natančno to, kar mu rečem, in kljub dejstvu, da moje govorjenje, naj naredi to, nekako povzroča, da se obnaša na ta način, obnašanja ne vodi tisto, kar rečem. Tisto, *kar rečem*, ni to, kar ima nadzor nad obnašanjem mikrofona - to, kar povzroči tresenje mikrofona -, ampak, vsekakor, zvoki, ki jih proizvajam z govorjenjem tistega, kar rečem. To je mogoče pokazati s tem, da si predstavljamo, da isti zvoki pomenijo nekaj drugega. Predpostavimo, da ti zvoki pomenijo *bodi tiho*. V tem primeru mikrofon prezre moj ukaz in se zatrese hitro. Očitno je, da mikrofon ne vidi razlike med mojim govorjenjem, naj se zatrese hitro, in govorjenjem, naj bo tiho. Vse, na kar je ta naprava občutljiva (prav tako kot človeško uho), je zvok, ne pomen. Ne odziva se na to (tako kot ljudje), kar govorec reče, temveč na zvoke, ki jih proizvaja govorec, ko to govori.

In tako bi utegnilo biti z mislijo. Celo če neka žival ima misli in te misli povzročijo, da nekaj naredi, lahko da žival ne dela tega, kar dela, zato, ker misli to, kar misli. Kar misli, vsebina ali pomen njenih misli, je lahko vzročno nepomembno. Predstavljajmo si robota, imenujmo ga Clyde, opremljenega z umetnimi čutili, ki ga vodijo skozi njegovo okolje. Vsebuje baterije, ki se morajo polniti vsakih osem ur. Clydeovi ustvarjalci so nekatera njegova notranja električna stanja spremenili v misli. Ni pomembno, kako so to storili. To je skrivnost in ni potrebno, da jo raztrobim, da bi pokazal tisto, kar nameravam. Ustvarjalci so spremenili notranje stanje # 42, kadarkoli se pojavi, v misel, da je naravnost spredaj električni priključek. Stanje # 43, več ali manj stalno notranje stanje, je misel, da če se nekdo priključi na električni priključek, se mu bodo napolnile baterije. Ko stanji # 42 in # 43 nastopita skupaj, nekaj, kar se zgodi, ko Clyde "vidi" električni priključek, in če je minilo najmanj šest ur od zadnjega polnjenja, potem stanji # 42 in # 43 povzročita, da se Clyde premakne naprej, se priključi na priključek in počaka, da se njegove baterije napolnijo.

To se gotovo zdi kot neka vrsta inteligentnega obnašanja. Clyde počne nekaj, kar mu je potrebno narediti, napolni se, preden mine osem ur, in misel, da to lahko naredi s priključitvijo na električni priključek, je tisto, kar je povzročilo, da se je priključil. Tako se zdi, da imajo misli, ki naredijo obnašanje razumno, nadzor nad njegovim obnašanjem. Clyde se priključi, ker *misli*, da s tem lahko napolni svoje baterije, ravno to, kar mora narediti, da bi nadaljeval z delovanjem.

Če zavoljo poskusa domnevamo, da sta stanji # 42 in # 43 misli, ali lahko rečemo, da se Clyde obnaša inteligentno? Ali njegovo obnašanje ali vsaj del tega obnašanja vodijo misli? Ni nujno. Nismo še slišali, kaj moramo vedeti, da bi odgovorili na to vprašanje.[2] Celo če vzamete mojo besedo kot dejstvo, da sta stanji # 42 in # 43 pristni misli, ne morete zaključiti, da Clyde počne karkoli inteligentnega, da je karkoli od tega, kar počne, pod nadzorom tistega, *kar misli*. Mogoče to, da ima Clyde misel, da je pred njim neki električni priključek - misel s tem pomenom ali vsebino -, prav tako ni v nobeni zvezi s tem, kar počne, kakor tudi dejstvo, da moje besede pomenijo *zatresi se*, ni povezano z obnašanjem bližnjega mikrofona. Če je to tako, potem - čeprav Clyde dela to, kar mora početi, in pristna misel povzroča, da to dela - obnašanje ne kaže nobene resnične inteligence. Obnašanje je povzročeno z mislijo, ki ima pravo vsebino, toda ne zato, ker ima to vsebino. Clydeu v teh okoliščinah pripisati inteligenco je prav tako prenagljeno kot sklep, da mikrofon razume tisto, kar mu povem, ker naredi tisto, kar mu rečem. Resnica zadeve je, da čeprav so zvoki, ki jih proizvedem, pomembni za obnašanje mikrofona, pa njihov pomen, tisto, kar mu rečem, naj naredi, ni relevanten, in *to* je tisto, kar je potrebno za pristno razumevanje. Enako velja za Clydea. Ali s priključitvijo samega sebe na električni priključek kaže inteligenco, ni odvisno samo od tega, ali Clydeove misli povzročijo, da se tako obnaša, temveč tudi od tega, ali tisto, *kar misli*, vsebina njegovih misli, igra vlogo v pojasnilu, zakaj to počne.

Ta zgodba ima nauk. Lažje je narediti, da stroj misli, težje pa ga je narediti inteligentnega. Če je, na primer, vse, kar je potrebno, da bi sistem mislil, le zavzeti pravilno držo do njega, pravilno pojasnjevalno držo, potem postane otročje lahko, da iz Clydea naredimo misleca misli. Zavzemite le pravilno držo do njega! Toda s takšnim pojmovanjem misli je zelo težko narediti Clydea inteligentnega. Da bi bil inteligenten, mora dejstvo, da Clyde misli, postati del pojasnila tega, zakaj se obnaša tako, kot se; skrivnost pa je, kako lahko naše zavzemanje ali nezavzemanje določene pojasnjevalne drže do Clydea pojasni, zakaj se Clyde obnaša tako, kot se. To pojasni, zakaj se mi (do Clydea) obnašamo tako, kot se, toda ne pojasni Clydeovega obnašanja. To spominja na H-delce Jerryja Fodorja.[3] Vse, kar mora delec biti, da bi bil eden od Fodorjevih H-delcev, je to, da je delec v času, ko kovanec v Fodorjevi dlani leži s številko navzgor. Če je to tako, potem je zelo lahko (vsaj za Jerryja Fodorja) spreminjati stvari v H-delce. Vse, kar mora narediti, je, da obrne kovanec v svoji dlani s številko navzgor. Toda s takšnim pojmovanjem H-delcev je zelo težko videti, kakšna je pojasnjevalna vrednost tega, da je nekaj H-delec. Nič, kar H-delci počnejo, ni pojasnjeno s tem, da so H-delci. Enako velja za misel. Karkoli je potrebno, da nekaj postane misel, ne more biti prelahko, gotovo ne tako lahko, kot nekateri razglašajo. Kajti če bi bilo tako, potem dejstvo - enako kot pri delcih Jerryja Fodorja -, da so misli, ne pojasnjuje, zakaj se je sistem, ki ima te misli, obnašal tako, kot se je. In brez povezave med mislijo in obnašanjem ni inteligence.

2. Načini pojasnjevanja obnašanja

[2] Lahko, seveda, vemo dovolj, da sklepamo, da so Clydeovi ustvarjalci zelo pametni, toda vprašanje o umetni inteligenci je, kot domnevam, vprašanje o Clydeovi inteligenci in ne o inteligenci njegovih ustvarjalcev.

[3] Jerry Fodor, *Psychosemantics*, MIT Press, Boston, Mass. 1987, str. 33.

Dejal sem, da inteligenca zahteva, da to, *kar* mislimo, nadzira naše obnašanje. Če to, kar mislimo, ni pomembno za to, kar počnemo, potem ne glede na to, kako inteligentni se lahko zdimo, ne glede na to, kako sijajen je bil naš stvarnik pri načrtovanju tega, da se obnašamo, kakor se obnašamo, ne kažemo nobene resnične kvalitete inteligence, ko se tako obnašamo.

Toda kako lahko misel, v tem pomenu, nadzira obnašanje? Če smo vsi fizični sistemi, če smo, v osnovi, vsi stroji - eni mokri, drugi suhi -, potem misli (primerki misli) morajo biti materialna stanja našega živčnega sistema. Zato mora biti naše obnašanje, v obsegu, ki je pod nadzorom misli, pod nadzorom fizičnih stanj našega živčnega sistema. Kako bi lahko bilo drugače? Kako lahko obstaja fizični sistem, katerega fizični učinek lahko pojasnimo s čimerkoli drugim kot pa z notranjimi fizičnimi stanji, ki povzročajo ta učinek?

Kot materialist sem docela pripravljen priznati - celo poudaril bom -, da so posamezne misli fizična (domnevno živčna) stanja misleca. Toda ni vsaka lastnost teh notranjih stanj in zlasti ne njihova vsebina (kar nekdo misli, ko misli te misli), dejstvo, ki ima opraviti izključno (ali celo predvsem) z notranjim ustrojem. Dovolite, da to pojasnim tako, da spregovorim o sistemih, ki prav nič ne mislijo: o rastlinah.

Rastlina škrlatna gilija poleti spreminja barvo. To je nekaj, kar rastlina počne, del rastlinskega obnašanja. Rastlina nima misli in želja, namenov in načrtov, toda počne stvari, včasih zelo zanimive stvari, in botanike zanima pojasnilo, zakaj se rastline obnašajo tako, kot se. Zakaj to počne škrlatna gilija, zakaj se spreminja iz rdeče v belo v sredini julija? Dobro, eno pojasnilo je, da rastlina to počne, da bi privabila opraševalce.[4] V zgodnjem času cvetenja so glavni opraševalci kolibriji in kolibrije bolj privlačijo rdeči cvetovi. V poznejšem obdobju se kolibriji odselijo in osnovni opraševalci postanejo vešče, ki imajo raje bele cvetove. Cvetlica spreminja barvo, "da bi" izkoristila to sezonsko spremembo v njenih okoliščinah. S spreminjanjem barve obrodi več sadežev in to je tisto, z besedami botanikov, zakaj to počne.

To pojasnilo obnašanja rastline se sklicuje na dejavnike v evolucijski zgodovini rastline, na dogodke in okoliščine, ki so obstajali v davni preteklosti in najbrž daleč stran. To je navzlic temu popolnoma zadovoljivo pojasnilo, zakaj rastlina spreminja barvo.[5]

Sedaj pa premislimo o fizično povsem enaki rastlini, taki, ki je molekularna dvojčica škrlatne gilije. Dvojčica se je razvila v povsem drugačnem okolju, v okolju, v katerem v sredini obdobja cvetenja prispejo požrešni hrošči, njih pa privlačijo rdeči cvetovi. Kot rezultat tega selektivnega pritiska je nastopila počasna sprememba: rastlina se je spremenila v obliko, v kateri je v sredini vsakega obdobja cvetenja spremenila barvo iz rdeče v belo. Hrošči sovražijo bele cvetove. Izognili so se dvojčici in ta je docvetela.

[4] To pojasnilo so podali botaniki, od katerih sem vzel primer: K. N. Paige in T. G. Whitham, "Report of research", objavljeno v *Science*, izšlo v *Scientific American*, 252, (4/1985), str. 74.

[5] Ne mislim, da so ta evolucijska pojasnila pojasnila *posameznega* obnašanja (rastline ali živali), toda to je tehnična poanta, ki jo tu preskakujem. Mislim, da je ta analogija koristna, celo če selekcijska pojasnila obnašanja niso povsem enaka kot razvojna.

Čeprav je dvojčica fizično povsem istovetna s škrlatno gilijo in čeprav se zato obnaša na povsem enak način, je pojasnilo, zakaj se obnaša na ta način, povsem drugačno. Škrlatna gilija spreminja barvo, da bi privabila vešče, dvojčica spreminja barvo, da bi odgnala hrošče. Leta in leta lahko preučujete ti dve rastlini pod mikroskopom, veste vse, kar se da vedeti o njuni trenutni fizični zgradbi, vendar nikoli ne ugotovite, da obstajajo povsem različni razlogi, da se obnašata tako, kot se. Čeprav je to obnašanje, spreminjanje barve, povsem fizični dogodek in čeprav ga povzročajo dobro razumljene kemične spremembe v rastlini, pojasnilo za obnašanje rastline zahteva, da gremo izven rastline, k nečemu, kar se je zgodilo v zgodovini rastline (ali v zgodovini te *vrste* rastline). Zatorej, celo če veste, kateri fizični dogodki v rastlini povzročajo spremembo barve, ni nujno, da veste, zakaj rastlina spreminja barvo. Pri obeh, škrlatni giliji in dvojčici, so spremembo povzročile povsem enake kemične spremembe v rastlini. Vendar pa ne spremenita barve zaradi istega razloga. Pri spraševanju, zakaj se sistem, vključno z rastlino, obnaša tako, kot se, pogosto zastavljamo vprašanje, ki zahteva, da odgovor iščemo zunaj sistema.

Lahko si celo, če želimo, zamislimo tretjo rastlino, imenujmo jo ponarejenka, ki se je ne da razločiti od prvih dveh, ki pa nima *nobenega* razloga za spreminjanje barve vsako leto. Lahko si, če želite, zamislite ponarejenko, sestavljeno v laboratoriju iz enostavnih celičnih sestavin, uporabljajoč škrlatno gilijo kot načrt. Botaniki sestavijo ponarejenko, jo vsadijo v tla in opazujejo njeno obnašanje. Prav gotovo, v sredini julija spremeni barvo. Zakaj? Zakaj postane bela? Ali naredi to zato, da bi pregnala hrošče? Ali naredi to zato, da bi privabila kolibrije? Ne. Ponarejenka, čeprav fizično istovetna z drugima dvema rastlinama in čeprav se zaradi tega obnaša povsem na enak način, nima enakih razlogov za tako obnašanje. Nekako nima *nobenih* razlogov za tako obnašanje.

Ponarejenka je botanična analogija živali, ki počne nekaj brez razuma. Ko smrčim ali kolcam, seveda nekaj počnem, prav tako kot ponarejenka nekaj počne, ko julija spreminja barvo, vendar tako ponarejenka kakor tudi jaz nimava razloga za tako obnašanje. To ne pomeni, da ne obstajajo vzročna pojasnila za najino tovrstno obnašanje. To pomeni le, da v mojem primeru ni nobenega razumskega, nobenega smotrnega, nobenega psihološkega pojasnila. V primeru ponarejenke pa obnašanje nima nobenega prilagoditvenega smotra, nobenega teleološkega pojasnila. Vendar bi ponarejenka, čeprav nima smotra, prestala botanični Turingov test. Obnašala bi se natanko na enak način, kot se je obnašala neka “inteligentna” rastlina, škrlatna gilija ali dvojčica. Toda v nasprotju s tema dvema rastlinama pri ponarejenki ne bi bilo nobenega smotra za tako obnašanje. V tem smislu za to, da ponarejenka spremeni barvo, ne obstaja noben razlog. Če njena sprememba barve prežene zablodelega hrošča ali privabi nekaj tavajočih kolibrijev, je to nenamerno. To ni tisto, zaradi česar to počne. Botanični Turingov test, prav tako kot originalni, ne preizkuša prave stvari. Ne pove nam tistega, kar bi želeli vedeti: ali je obnašanje (v primeru rastlin) smotrno ali (v primeru živali) inteligentno.

Pri opisovanju teh rastlin sem si kajpada dovolil svobodo. Zaradi dramatičnega učinka sem o njih govoril, kot da imajo razloge, da so se tako obnašale. O njihovem obnašanju sem govoril kot o smotrnem. Rastline seveda nimajo razlogov v

enakem smislu, kot jih imamo mi, da se obnašamo tako, kot se. Rastline nimajo smotrov, misli, želja in namenov. Nobena od njih ni inteligentna. Čeprav zavajajoče, sem uporabil ta način govorjenja, da bi poudaril pojasnitev obnašanja. Pri pojasnjevanju obnašanja sistema pogosto ne iščemo notranjega fizičnega vzroka, temveč iščemo zunanje dogodke in okoliščine, ki so oblikovali to notranjo strukturo in jo preoblikovali tako, da so naredili te vzroke za vzroke tovrstnih sprememb. Iščemo tisto, kar sem nekje drugje že imenoval *strukturirajoči*, ne pa *prožilni* vzroki obnašanja.[6]

3. Vloga učenja

Da bi razumeli, kako se ustvari pristna inteligenca, kako sistem pride do tega, da ga vodi tisto, kar misli, pretehtajte naslednjo zgodbo o generični živali, ki jo bom imenoval Buster.

Buster se začenja vesti inteligentno, in to na način, ki ga prej ni zmogel. Kar pride na dan, je, mislim, proces, proces učenja, ki to omogoča. Buster živi na področju, na katerem so tudi kožuhovinasti črvi. Imenoval jih bom (dovolj naravno) kožrvi. Kožrvi so vseh velikosti in barv, vendar so vsi kožuhovinasti (K) in vsi so podobni črvom (Č). Ker so kožrvi edina bitja v Busterjevem okolju, ki so obenem kožuhovinasta in oblikovana kot črvi, lastnosti K in Č zadostujeta za določitev nekega objekta v tem okolju kot kožrva.

Buster lahko opazuje kožrve in njihove vidne značilnosti. S tem mislim to, da Buster - zahvaljujoč evoluciji - pride na svet opremljen s čutilnimi mehanizmi, ki zaznavajo take stvari, kot so oblika in barva, gibanje in orientacija, tekstura in velikost pri srednje velikih stvareh, kakršni so kožrvi. Čeprav se bo mogoče moral naučiti, kje naj jih išče, se Busterju ne bo treba učiti, da bi jih videl. Ker so vsepovsod okrog njega, je vse, kar mora storiti, da bi jih videl, odpreti svoje oči.

Čeprav je Buster od rojstva tako opremljen, da vidi kožrve, nanje ne reagira instinktivno, gensko programirano. Ti v tem oziru niso stvari, tako kot so to svetle luči ali vroče površine, za katere je instinktivno dovzeten. Buster se ne umakne ali skrije pred kožrvi, ne renči ali bulji vanje in jih ne napada ali grize. Če pa to naredi, potem tega ne naredi kožrvu *kot* kožrvu, temveč kot oviri na njegovi poti, kot nečemu, kar mu zapira pogled, ali pa kot nečemu, kar se premika v travi. Buster kožrva lahko zaradi radovednosti povoha, dregne vanj, stopi nanj ali ga gleda, če počne kaj neobičajnega, toda kožrv je za Busterja, vsaj prvotno, bolj tako, kot je za nas drevo, grm ali skala: samo ena med mnogimi nekategoriziranimi stvarmi v našem okolju, pri katerih si ne moremo pomagati, da jih ne bi videli, ko so naše oči odprte, ki jim pa, pred učenjem, posvečamo le malo pozornosti.

Kar vas prosim, da si zamislite, je seveda neka žival, ki lahko izkuša in tudi izkuša predmete določene vrste (kožrve), toda taka žival, ki še ni razvila zmožnosti odziva na njih *kot* na predmete določenega tipa. Buster še nima pojma KOŽRV. Čeprav lahko doživlja kožrve, čeprav si jih v tem smislu (če to je smisel) lahko notranje predstavlja, še nima takšnih prepričanj: to je kožrv. Obstaja razlika med vi-

[6] V *Explaining Behavior*, MIT Press, Boston 1988. K tej razliki se bom vrnil kasneje.

deti kožrva in imeti prepričanje, da je to (kar vidite) kožrv, razlika, ki jo lahko premostimo z učenjem. Pomislite, kaj je potrebno, da bi od tega, da vidimo petunije, nekaj, kar celo otroci in živali lahko počno, prišli do tega, da jih vidimo *kot* petunije, da vidimo, *da so* petunije - do nečesa, kar zmore le izkušen opazovalec. Da bi Buster razvil zmožnost misliti *za* kožrve, ki jih vidi, da so kožrvi, in da bi jih videl *kot* kožrve, mora razviti *nov* način za njihovo predstavljanje. V nasprotju z Busterjevim doživljanjem kožrvov, za kar ima prirojene čutilne mehanizme, pa narava Busterja ni opremila, da bi prepoznaval kožrve. To si mora pridobiti sam.

Na tej stopnji njegovega razvoja morda velja opozoriti, da - čeprav se Buster kožrvov zaveda, zaznavno zaveda (jih vidi) - obstaja drugi smisel (ki ga lahko hitro zamenjamo s prvim), v katerem se jih ne zaveda. Ne zaveda se jih še *kot* kožrvov, ne zaveda se še, *da* so kožrvi, ali vsekakor, da so posamezna *vrsta* predmetov.

Nekega dne Buster iz čiste radovednosti povoha velikega rdečega kožrva, nakar izkusi boleče pekoč občutek v svojem nosu. Od takrat se Buster izogiba velikih rdečih kožrvov. Enako se mu zgodi z malim zelenim kožrvom. Kmalu vidimo Busterja, da se obnaša na način, za katerega bi se vsi strinjali, da je inteligenten: izogiba se kožrvov. Nič več jih ne povoha. Ko se približajo, se hitro umakne. Zakaj? Ali Busterjevo obnašanje nadzirajo misli, in če je tako, katere misli?

Dovolite, da postavim to pojasnjevalno vprašanje v poseben kontekst, da bi pojasnil, kaj natanko iščemo, ko sprašujemo po pojasnilu Busterjevega obnašanja. Predstavljajmo si, da postavimo Busterja v umetno okolje, v katerem so ponarejeni kožrvi (gosenice ali, če hočete, mehanični kožrvi), predmeti, ki so, čeprav niso kožrvi, prav tako kožuhovinasti in črvasti. Buster opazi približevanje ponarejenega kožrva in se hitro umakne. Pravkar prispeli opazovalec bi hotel vedeti, zakaj to naredi. Zakaj se Buster umakne? Tega prej ni počel. Kaj je pojasnilo Busterjevega obnašanja?

Prvi poskus: Buster se je umaknil, ker je videl, da mu se približuje kožrv.

To ne more biti pravilen odgovor, kajti Buster ni videl kožrva, ki se mu približuje. Tisto ni bil kožrv.

Drugi poskus: Buster se je umaknil, ker je videl, da mu se približuje nekaj, kar je podobno kožrvu (nekaj, kar je tako K kot Č).

To prav tako ne more biti res. Vsaj popolno pojasnilo ne more biti, ker je Buster že pred svojim bolečim srečanjem videl približevanje predmetov, podobnih kožrvom (v tem primeru so to res bili kožrvi), ki so bili tako K kot Č, Buster pa se ni umaknil. Imel je natanko enako izkušnjo pred učenjem, kot jo ima sedaj, po učenju, vrsto izkušnje, ki pomeni videnje in približevanje kožrva. Sedaj se umakne, tedaj se pa ni. Zakaj? Kakšna je razlika? To je gotovo vprašanje, na katerega je potrebno odgovoriti, da bi razumeli, zakaj se Buster umakne.

Poskus, da bi pojasnili Busterjevo obnašanje z omenjanjem tistega, kar vidi, je poskus pojasnjevanja njegovega obnašanja v okviru prožilnega vzroka. V takšni situaciji je - značilno - prožilni vzrok obnašanja nek dogodek v okolju, ki povzroči, da je oseba ali žival prepričana o nečem, kar (skupaj z željo) pomaga pojasniti

nastalo obnašanje. Lačen lev se približuje. Vidim ga. Vedoč, kaj so to levi in kakšni so, sem prisiljen verjeti, da je to lev. Prepoznam ga - vidim, kaj je. Stečem. Približevanje leva ali moje videnje približevanja je prožilni vzrok mojega bega. Lev povzroči, da sem prepričan o nečem - da se lev približuje -, kar, skupaj z mojimi strahovi, pojasnjuje, zakaj sem zbežal.

Ta primer razjasni, da prožilni vzrok mojega bega pojasnjuje moje bežanje samo na način vmesnega prepričanja. Če približevanje leva ne bi povzročilo prepričanja, da se lev približuje, potem, celo če bi me to nagnalo v beg, vi ne bi vedeli, *zakaj* sem bežal. Če ne bi verjel, da je lev bil lev, zakaj sem potem zbežal, ko sem ga videl? Za kaj drugega sem imel leva? Opisovanju prožilnega vzroka obnašanja uspe pojasniti obnašanje samo, če mu uspe predložiti ali nakazati neko vmesno pojasnjevalno prepričanje.[7]

Zato drugi poskus pojasnitve ni uspešen. Ne pove nam, kaj, če sploh kaj, Buster misli o tem, kar vidi. Pove nam samo to, kaj vidi. Temu se torej posreči pojasniti Busterjevo obnašanja samo takrat, ko nekaj predloži ali nakaže o vmesnem prepričanju.

Tretji poskus: Buster se umakne, ker misli, da je to (kar vidi) kožrv (ali kožuhovinasto-črvasto-bodeča stvar) in ker se boji kožrvov (ali ne želi, da ga kaj piči).

Stopili smo na novo področje. Sedaj podajamo racionalna pojasnila Busterjevega obnašanja, pojasnila v okviru tega, kaj Buster misli in kaj Buster želi. Buster je imel motiv, da se izogne bolečini veliko pred njegovimi bolečimi srečanji s kožrvi. Lahko predpostavljamo, da je to bolj ali manj trajni instinkt ali želja. Ne moremo pa predpostavljati, da je njegovo prepričanje, prepričanje, ki dopolnjuje to racionalno pojasnitev, bilo Busterju na razpolago pred njegovo učečo se izkušnjo s kožrvi. Busterjeva zmožnost, da je o kožrvih, ki jih vidi, prepričan, da so kožrvi, njegova zmožnost, da jih vidi *kot* kožrve (ali vsaj kot nekaj, kar lahko piči), je zmožnost, ki si jo je pridobil v svojih srečanjih s kožrvi.

Vse to se mi kaže kot povsem očitno. Za vse to gre pri tvorjenju pojma. Kratka tovrstna epizoda učenja je dovolj, da pripišemo taki živali, kot je Buster, osnovni pojem, pojem, ki je bistven za posedovanje take vrste prepričanja, ki bo, ko se ustrezni primer obnašanja pojavi, in v povezavi z ustreznimi motivacijskimi stanji in drugimi prepričanji, pojasnil, zakaj Buster počne to, kar počne.

To, kar Buster razvije med kratkim, toda bolečim srečanjem s kožrvi, je način predstavljanja le-teh, ki presega čutilni mehanizem, ki ga je dobil ob rojstvu, in presega vse tisto, kar omogoča videnje kožrvov. To, kar ima sedaj, je predstava, ki, za razliko od njegove izkušnje s kožrvi, pomaga pojasniti, zakaj se jim sedaj izogiba. To je skratka prepričanje.[8] Busterjevo kasnejše obnašanje, njegov umik pred kožuhovinastimi črvi, kaže inteligenco, ker je to obnašanje, ki ga vodijo notranja stanja (prepričanja ali misli), ki upodobijo predmet misli (kožrve) na način,

[7] V *Explaining Behavior* sem dokazoval, da so prepričanja (kar sem imenoval) strukturirajoči vzroki obnašanja - vzroki, da ena stvar (prožilni vzrok) povzroči drugo.

[8] Tisti holisti, ki bi ugovarjali, da se taka posamezna predstavna struktura imenuje prepričanje, naj mislijo o tem (če to kaj pomaga) kot o proto-prepričanju, kot o takšni vrsti stvari, ki si bo (ali bi si lahko) pridobila status pristnega prepričanja s kopičenjem zadostnega števila predstavnih sosedov enake vrste.

ki ustreza obnašanju - *kot* nekaj, kar bo pičilo nekogaršnji nos, če ta pride preblizu. V Bustru pred učenjem ni bilo takšnih predstav in od tod tudi nobene možnosti za inteligentno obnašanje do kožrvov (vsaj ne do kožrvov *kot* kožrvov). Učenje je to spremenilo. S preoblikovanjem nadzornih okrožij je preoblikovalo notranjo predstavo o kožrvih v določilo obnašanja in to je storilo natanko zaradi tega, *kar* predstavlja ta notranja predstava.[9]

Tega inteligentnega obnašanja se ne da umetno ustvariti v laboratoriju. Takšno obnašanje lahko "vgradimo" v umetnega Busterja. Lahko si zamislimo, tako kot smo naredili pri škrlatni giliji, stvarjenje umetnega Busterja v laboratoriju, ki se bo, ker se ga fizično ne da razločiti od našega izkušenega Busterja, obnašal na povsem enak način: pri srečanju s kožuhovinastimi črvi se bo hitro umaknil. Toda čeprav se bo ponarejeni Buster obnašal na povsem enak način, se ne bo, nič bolj kot ponarejena rastlina, tako obnašal zaradi istega razloga kot prototip. Čeprav se bo ponarejeni Buster, enako kot pravi, izogibal kožuhovinastim črvom, pa se jim ne bo izogibal zaradi istega (če sploh kakšnega) razloga kot pravi Buster. Pravi Buster je v svojih srečanjih s kožrvi razvil način njihovega predstavljanja (kot tistih, ki boleče pikajo), ki (glede na njegove strahove) ne le racionalizira njegove reakcije nanje, ampak prek procesa učenja vzročno sodeluje v proizvajanju reakcije (izogibanju), ki to predstavo (zaradi njene *vsebine*) racionalizira. Tako imamo v pravem Busterju, ne pa tudi v kopijah, ki nimajo enake zgodovine, notranje stanje, ki ni samo predstava (kožrvov), ampak je predstava, ki pomaga pojasniti - v okviru svoje predstavne vsebine (to je nekaj, kar pika) -, zakaj se žival, v kateri se ta pojavlja, obnaša tako, kot se. Z drugimi besedami, imamo nekaj, kar se približuje resnični inteligenci - misel, ki ne pomaga samo pri upravljanju obnašanja, ampak to počne v okviru tistega, o čemer je misel ali *kaj* ta predstavlja.

Iz vsega tega sem izpeljal sklep, da je kaj malo upanja za umetno inteligenco. Če z inteligenco mislimo delovanje, ki ga vodi misel, potem moramo - da bi ustvarili sistem, ki se ne obnaša samo na enak način kot nek inteligentni sistem, ampak se obnaša tako zaradi razlogov in s tem kaže inteligenco - dati temu sistemu neko zgodovino, v kateri je tako vodenje možno. Brez takšne zgodovine imamo zagotovo lahko obnašanje, ki ga povzročajo notranje predstave. To se dogaja pri ponarejenem Busterju. Ponarejenega Busterja smo *ustvarili* (uporabljajoč pravega Busterja kot model) tako, da bodo notranje predstave kožrvov (enako kot pri pravem Busterju) povzročile gibanje umikanja. Toda tega, česar ne moremo dati ponarejenemu Busterju, vsaj v laboratoriju ne, so pojasnjevalni odnosi med tistim, *kar* je predstavljeno (vsebina misli), in obnašanjem, ki ga te predstave povzročajo. Brez tega pa ne obstaja nobeno delovanje, ki bi ga vodilo tisto, *kar je mišljeno*, in torej nobena pristna inteligenca.

Umetna inteligenca se torej sooča z enakimi ovirami kot umetna botanika: lahko da ste zmožni narediti sisteme, ki se obnašajo kot resnični, toda ne bodo se tako obnašali zaradi enakih razlogov. Zaradi tega izdelek ne bo inteligenten. Za pristno inteligenco potrebujemo pravo vrsto zgodovine, takšno, ki bo vpeljala pojasnjevalno povezavo med predstavljenim (vsebino) in obnašanjem, ki ga ta vsebina pomaga pojasniti. Toda vse, kar ima

[9] To, kar sem podal, je površni povzetek moje razlage (v *Explaining Behavior*), kako nastajajo pojmi - in od tod zmožnost za prepričanje in inteligentno delovanje - iz procesa učenja. Čeprav površno, upam, da je to dovolj za sporočilo občutka glavne strategije.

ustrezno zgodovino, ni več umetno.