

Ťažké je naozaj zmeniť názor

**Jasné rozmýšľanie o
pravde, hodnote a víťazstve**

Kniha II.

Politika zabíja myslenie

Eliezer Yudkowsky

Obsah

Predslov	iv
VI Politika zabíja myslenie	1
47 Politika zabíja myslenie	2
48 Debaty o pravidlách by nemali vyzerat jednostranne	4
49 Váhy spravodlivosti, zápisník rozumnosti	7
50 Chyba prisudzovania	9
51 Sú vaši nepriatelia od narodenia zlí?	11
52 Obrátená hlúposť nie je inteligencia	13
53 Argument zatieňuje autoritu	15
54 Pritisnite si otázku	19
55 Rozumnosť a anglický jazyk	20
56 Ľudské zlo a zahmlené myslenie	23
VII Špirály smrti a sektársky atraktor	26
57 Afektívna heuristika	27
58 Vyhodnotiteľnosť (a lacné vianočné nákupy)	30
59 Neobmedzené škály, obrovské súdne odmeny a futurizmus	33
60 Efekt svätožiary	35
61 Skreslenie superhrdinu	37
62 Iba spasitelia	39
63 Afektívne špirály smrti	41
64 Odolajte šťastnej špirále smrti	43
65 Nekritická nadkritickosť	47
66 Ochladzovanie skupinových názorov vyparovaním	50
67 Keď sa nikto neodváži naliehať na zdržanlivosť	52
68 Pokus v Robbers Cave	54
69 Každá kauza chce byť sektou	56
70 Strážcovia pravdy	58
71 Strážcovia genofondu	61
72 Strážcovia Ayn Rand	62
73 Dva koany o sektách	66
74 Aschov pokus o konformitu	68
75 O vyjadrovaní svojich obáv	71
76 Osamelý nesúhlas	73
77 Sektárske antisektárstvo	76
VIII Pozeranie čerstvými očami	82
78 Ukotvenie a prispôsobenie	83
79 Priming a kontaminácia	85
80 Veríme všetkému, čo nám povedia?	87
81 Uložené myšlienky	89
82 Kolaje "mimo vychodených kolají"	91
83 Originálny pohľad	93
84 Čudnejšie než história	95
85 Logická chyba zovšeobecňovania fiktívnej indície	96
86 Ako vyzerat (a byť) hlboký	100
87 Svoje názory meníme zriedkavejšie, než si myslíme	103
88 Odložte navrhovanie riešení	105
89 Klam pôvodu	107
IX Proti racionalizácii	109
90 Vedieť o skresleniach môže ľuďom ublížiť	110
91 Aktualizujte postupne	112
92 Jeden argument proti armáde	114
93 Spodný riadok	116
94 Filtrovaná indícia	118
95 Racionalizovanie	120
96 Rozumný argument	122

97 Vyhýbanie sa naozaj slabým miestam vašich názorov	124
98 Motivované zastavenie a motivované pokračovanie	127
99 Falošné zdôvodnenie	129
100 Je toto vaše skutočné odmietnutie?	131
101 Previazané pravdy, nákazlivé lži	134
102 O klamstvách a výbuchoch čiernych labutí	137
103 Epistemológia temnej strany	138
X Proti doublethinku	142
104 Singlethink	143
105 Doublethink (dobrovoľné skreslenie)	145
106 Nie, naozaj, ja som sa oklamal	148
107 Viera v sebaklam	149
108 Moorov paradox	152
109 Neverte, že sa sami oklamete	154
XI Príliš pohodlné výhovorky	156
110 Správne použitie pokory	157
111 Tretia možnosť	160
112 Ale stále je tu šanca, nie?	162
113 Klam sivej	164
114 Absolútna autorita	167
115 Ako ma presvedčiť, že $2 + 2 = 3$	171
116 Nekonečná istota	173
117 0 a 1 nie sú pravdepodobnosti	176
XII Zanechávanie	179
118 Dôležitosť hovorenia: "joj!"	180
119 Ponuka pomätenosti	182
120 Už to konečne vzdaj	184
121 Správne použitie pochybnosti	185
122 Dokážete čeliť skutočnosti	187
123 Meditácia o zvedavosti	188
124 Nikto vám nemôže udeliť výnimku zo zákonov rozumnosti	191
125 Ponechajte ústupovú líniu	193
126 Kríza viery	196
127 Rituál	202
XIII Medzihra	205
128 Sila inteligencie	206

Predslov

V rokoch 2006 až 2009 Eliezer Yudkowsky napísal dlhú sériu článkov na blogu (vtedy ešte Overcoming Bias).

Tieto články, pôvodne usporiadané do zbierky “postupností”, pokrývajú rozsiahle pole tém a myšlienok siahajúce od pravdy a víťazstva po umelú inteligenciu a zombie.

Kapitoly tejto knihy obsahujú drobné úpravy pôvodných článkov. Niektoré kapitoly majú oproti originálu rozsiahlejšie úpravy a sú príslušne označené. Na konci každej kapitoly sa nachádza odkaz na pôvodný článok a komentáre (po anglicky).

Anglická verzia tejto knihy je rozdelená na 6 častí; spolu obsahuje vyše 300 článkov a vyše 2000 strán textu. Kvôli veľkému rozsahu slovenský preklad vychádza postupne. Menej trpezliví čitatelia si môžu zatiaľ nepreložené články prečítať v origináli na stránke LessWrong¹.

¹ <http://lesswrong.com/>

Časť VI. Politika zabíja myslenie

Obsah

47 Politika zabíja myslenie	2
48 Debaty o pravidlách by nemali vyzerat jednostranne	4
49 Váhy spravodlivosti, zápisník rozumnosti	7
50 Chyba prisudzovania	9
51 Sú vaši nepriatelia od narodenia zlí?	11
52 Obrátená hlúposť nie je inteligencia	13
53 Argument zatieňuje autoritu	15
54 Pritisnite si otázku	19
55 Rozumnosť a anglický jazyk	20
56 Ľudské zlo a zahmlené myslenie	23

Kapitola 47. Politika zabíja myslenie

Ľuďom sa v hlavách dejú smiešne veci, keď hovoria o politike. Evolučné dôvody sú také zrejmé, že sa ich oplatí spomenúť: V pravekom prostredí bola politika otázkou života a smrti. A sexu a bohatstva a spojencov a povesti... Keď sa dnes dostanete do hádky, či by "sme" mali zvýšiť minimálnu mzdu, vykonávané adaptácie na praveké prostredie, kde byť na nesprávnej strane hádky mohlo znamenať smrť. Byť na *správnej* strane hádky mohlo *vám* umožniť zabiť svojho nenávideného soka!

Ak chcete niečo povedať o vede alebo rozumnosti, odporúčam nevyberať si príklady zo *súčasnej* politiky, ak sa tomu dá vyhnúť. Ak vaša pointa neodmysliteľne súvisí s politikou, potom hovorte o Ľudovítovi XVI. počas francúzskej revolúcie. Politika je dôležitá oblasť, v ktorej by sme jednotlivci mali používať svoju rozumnosť – ale je to otravná oblasť na *učenie sa* rozumnosti alebo na diskutovanie o rozumnosti, okrem prípadu, že všetci diskutujúci už rozumní sú.

Politika je pokračovanie vojny inými prostriedkami. Argumenty sú vojaci. Akonáhle viete, na ktorej strane ste, musíte podporovať všetky argumenty tejto strany a napádať všetky argumenty, ktoré vyzerajú v prospech nepriateľskej strany; inak je to, akoby ste vlastných vojakov bodali do chrbta a poskytovali pomoc a podporu nepriateľom. Ľudia, ktorí by s chladnou hlavou spravodlivo zvážili zo všetkých strán nejakú tému vo svojom pracovnom živote ako vedci, sa dokážu náhle premeniť na zombie skandujúce heslá, ak k danej téme existuje Modrý alebo Zelený¹ postoj.

V oblasti umelej inteligencie, špeciálne v oblasti nemonotónneho uvažovania, existuje štandardný problém: "Všetci kvakeri sú pacifisti. Žiaden republikán nie je pacifista. Nixon je kvaker a republikán. Je Nixon pacifista?"

Aký zmysel to prosím vás malo, vybrať takýto príklad? Vyburcovať politické emócie čitateľov a odpútať ich pozornosť od hlavnej otázky? Dosiahnuť, aby sa republikáni cítili nevítaní na kurzoch umelej inteligencie a odradiť ich od výberu takéhoto smeru? (Nie, skôr než sa na to niekto opýta, nie som republikán. Ani demokrat.)

Prečo by si niekto vybral takýto *rozptyľujúci* príklad na ilustráciu nemonotónneho uvažovania? Autor pravdepodobne nedokázal odolať pokušeniu dobre si štychnúť do tých nenávidených Zelených². Je to taký *dobrý* pocit uštedriť poriadny úder, viete, to je ako keby ste skúšali odolať čokoládovému koláču.

Podobne ako s čokoládovým koláčom, nie všetko, čo príjemne chutí, je aj dobré. Určite nie je dobré pre chudákov čitateľov nášho blogu, keď musia vidieť všetky nahnevané komentáre, ktoré tento článok na blogu vyvolal.

Nehovorím, že by *Overcoming Bias* [blog, na ktorom bol tento článok pôvodne uverejnený] mal byť apolitický, dokonca ani že by sme mali prijať ideál neutrálneho uhla pohľadu z Wikipédie³. Pokúsme sa však odolávať pokušeniu týchto dobrých štychancov, ak sa tomu môžeme vyhnúť. Ak vaša téma legitímne súvisí s pokusmi zakázať evolúciu zo školských osnôv, potom nech sa páči, hovorte o tom – ale neobviňujte z toho explicitne celú Republikánsku stranu;

¹ pozri: 12 Bájka o vede a politike

² pozri: 12 Bájka o vede a politike

³ http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

niektorí z vašich čitateľov môžu byť republikáni, no môžu mať pocit, že problémom je pár fanatikov, nie celá strana. Podobne ako pri neutrálnom pohľade Wikipédie, nie je podstatné, či (si myslíte, že) na vine naozaj *je* Republikánska strana. Pre duchovný rast komunity je skrátka lepšie diskutovať o téme bez privolávania politických farieb⁴.

Pôvodný článok s komentármi⁵

⁴ pozri: 12 Bájka o vede a politike

⁵ http://www.lesswrong.com/lw/gw/politics_is_the_mindkiller/

Kapitola 48. Debaty o pravidlách by nemali vyzerat' jednostranne

Robin Hanson nedávno navrhol obchody, v ktorých by sa mohli predávať zakázané výrobky¹. Existuje viacero výborných argumentov v prospech takéhoto zákona – jednotlivec má neodňateľné právo na slobodu, byrokrati majú kariérnu motiváciu zakazovať *všetko*, zákonodarcovia sú rovnako omylní ako občania. Napriek tomu (som napísal, že) do takéhoto obchodu príde aj *nejaká* chudobná, úprimná, nevelmi vzdelaná matka piatich detí a kúpi si “Nápoj z kyseliny sírovej Dr. Šarlatána” proti artritíde a zomrie, zanechajúc siroty, ktoré budú plakať vo verejnoprávnej televízii.

Iba som konštatoval jednoduchý fakt. Prečo si niektorí ľudia mysleli, že to bol argument v prospech regulácie?

Pri jednoduchých faktických otázkach (napríklad či život na Zemi vznikol prirodzeným výberom) môžeme legitímne očakávať, že dokazovanie bude jednostrannou bitkou; samotné fakty sú buď tak alebo onak a takzvaná “vyvážená argumentácia” by toto mala odrážať. Podľa bayesovskej definície indície je “silná indícia” presne ten typ indície, o ktorom očakávame, že sa bude nachádzať iba na jednej strane dôkazu.

Nie je však dôvod, aby takúto jednostrannosť vykazovali aj zložité aktivity s mnohými dôsledkami. Prečo sa teda zdá, že ľudia chcú, aby ich debaty o pravidlách boli jednostranné?

Politika zabíja myslenie.² Argumenty sú vojaci. Akonáhle viete, na ktorej strane ste, musíte podporovať všetky argumenty tejto strany a napádať všetky argumenty, ktoré vyzerajú v prospech nepriateľskej strany; inak je to, ako keby ste vlastných vojakov bodali do chrbta. Ak sa držíte tohto vzorca, budú aj vám debaty o pravidlách pripadať jednostranné – náklady a nevýhody vášho obľúbeného riešenia sú nepriateľskí vojaci, na ktorých treba útočiť všetkými možnými spôsobmi.

Človek by sa mal vyvarovať aj ďalšieho chybného vzorca, predstavy, že Hlboká Múdrost si žiada robiť dokonale vyrovnané kompromisy medzi tými dvoma názormi na pravidlá, ktoré dostávajú najviac priestoru v médiách. Zákon môže legitímne mať *nevyvážené* náklady a prínosy. Keby politické otázky neboli naklonené jedným alebo druhým smerom, nedokázali by sme sa o nich rozhodovať. Je tu však aj ľudský sklon popierať všetky náklady obľúbeného riešenia alebo popierať všetky prínosy neobľúbeného riešenia; a ľudia si teda budú myslieť, že táto nerovnováha je naklonená omnoho viac, než v skutočnosti je.

Ak povolíte obchody, v ktorých sa predávajú inak zakázané produkty, *nejaká* chudobná, úprimná, nevelmi vzdelaná matka piatich detí si tam kúpi niečo, čo ju zabije. Toto je predpoveď faktického dôsledku a ako faktická otázka sa zdá byť celkom priamočiara – príčetný človek by mal byť pripravený uznať, že je to pravda, bez ohľadu na to, aký postoj má k danej politickej téme. Môžete si *tiež* myslieť, že postaviť veci mimo zákon akurát spôsobí, že budú drahšie, že regulátori zneužívajú svoje právomoci, alebo že jej sloboda ako jednotlivca

¹ http://www.overcomingbias.com/2007/03/paternalism_is_.html

² pozri: 47 Politika zabíja myslenie

prebýja vašu túžbu zasahovať do jej života. Avšak, vecou jednoduchého faktu je, že ona napriek tomu zomrie.

Žijeme v nespravodlivom vesmíre. Ako všetci primáti, aj ľudia majú silnú negatívnu reakciu na vnímanú nespravodlivosť; preto nás tento fakt stresuje. Existujú dve obľúbené metódy, ako si s výslednou kognitívnou disonanciou poradiť. Po prvé, človek môže zmeniť svoj pohľad na fakty – poprie, že sa daná nespravodlivá udalosť stala, alebo zmení históriu, aby vyzerala spravodlivejšie. Po druhé, človek môže zmeniť svoju morálku – poprieť, že daná udalosť bola nespravodlivá.

Niektorí libertariáni by mohli povedať, že keď pôjdete do “obchodu so zakázanými výrobkami”, prejdete okolo jasných varovných nápisov hovoriacich: “VECI V TOMTO OBCHODE VÁS MÔŽU ZABIŤ” a kúpите si niečo, čo vás zabije, potom je to vaša vlastná chyba a zaslúžite si to. Ak by toto bola morálna pravda, potom by existencia obchodov predávajúcich zakázané výrobky nemala *žiadnu nevýhodu*. Nielenže by to bol v *konečnom dôsledku zisk*, ale bola by to *jednostranná zmena* bez akýchkoľvek nevýhod.

Iní argumentujú, že regulátorov možno naučiť, aby sa rozhodovali rozumne a v harmónii so záujmami spotrebiteľa; keby toto bola faktická pravda, potom (z ich morálneho pohľadu) by regulácia nemala *žiadnu nevýhodu*.

Či sa vám páči alebo nie, inteligenciu dostane človek náhodne pridelenú podľa narodenia – aj keď toto je jeden z prípadov, kde je nespravodlivosť vesmíru taká extrémna, že mnohí ľudia sa rozhodnú popierať fakty. Experimentálne indície pre čisto genetickú zložku 0,6 – 0,8 majú prevahu, ale aj keby ste toto popreli, rodičovskú výchovu alebo základnú školu si človek tiež nevyberá.

Ja som bol vychovaný, aby som veril, že popierať skutočnosť je *morálne nesprávne*. Keby som sa mal zapojiť do túžobného optimizmu, ako mi Nápoj z kyseliny sírovej pravdepodobne urobí lepšie, robil by som niečo, pred čím ma *varovali*, čo ma vychovali vnímať ako neprijateľné. Niektorí ľudia sa narodila do prostredia – neriešme teraz ich gény, pretože tá časť je príliš nespravodlivá – kde im miestny šaman povie, že je *správne* mať vieru a *nesprávne* byť skeptický. S najlepšou vôľou sa budú riadiť touto radou a zomrú. Na rozdiel od vás ich nevychovali, aby verili, že ľudia sú zodpovední za svoje individuálne rozhodnutie nasledovať smerovanie spoločnosti. Naozaj si myslíte, že vy sami ste takí chytrí, že by ste boli správne vedecky skeptickí, aj keby ste sa narodili v roku 500 nášho letopočtu? Áno, narodiť sa je lotéria, bez ohľadu na to, čo si myslíte o génoch.

Povedať: “Ľudia, ktorí si kupujú nebezpečné výrobky, si zaslúžia trpieť!” nie je zásadovosť. Je to spôsob, ako popierať, že žijeme v nespravodlivom vesmíre. Skutočná zásadovosť je povedať: “Áno, kyselina sírová spôsobuje strašnú bolestivú smrť, a nie, tá matka piatich detí si to nezaslúžila, ale napriek tomu necháme tieto obchody otvorené, pretože sme urobili takýto výpočet nákladov a prínosov.” Viete si predstaviť, že nejaký politik povie toto? Ja tiež nie. Ale keď majú ekonómovia možnosť ovplyvňovať politické rozhodnutia, pomohlo by, keby si to vedeli aspoň myslieť v súkromí – možno dokonca aj povedať v odborných článkoch, príslušne zamaskované mnohoslabičnou terminológiou, aby ich médiá nemohli citovať.

Nemyslím si, že ak niekto urobí hlúpu chybu a zomrie, že je to dôvod oslavovať. Ja to počítam za tragédiu. Nie vždy je dobré pomáhať ľuďom, zachraňovať ich pred dôsledkami vlastných činov; pri treste smrti však robím morálnu hranicu. Ak ste mŕtvi, nemôžete sa použiť z vlastných chýb.

Nanešťastie, tento vesmír so mnou nesúhlasí. Uvidíme, kto z nás zostane stáť ako víťaz, keď toto celé skončí.

DOPLNENÉ: Dve hlavné motivácie politickej jednostrannosti sú afektívna heuristika³ a klam spravodlivého sveta⁴.

Pôvodný článok s komentármi⁵

³ pozri: 57 Afektívna heuristika

⁴ http://en.wikipedia.org/wiki/Just-world_fallacy

⁵ http://www.lesswrong.com/lw/gz/policy_debates_should_not_appear_onesided/

Kapitola 49. Váhy spravodlivosti, zápisník rozumnosti

Pani Spravodlivosť¹ sa zvyčajne zobrazuje držiaca váhy. Váhy majú tú vlastnosť, že keď hocičo potiahne jednu stranu nadol, posunie tým druhú stranu nahor. Vďaka tomu sa veci veľmi pohodlne a ľahko sledujú. Je to zvyčajne aj hrubé skreslenie.

Ako ľudia máme prirodzený sklon považovať diskusiu za istú formu súboja, pokračovanie vojny, šport; a v športe stačí sledovať, koľko bodov získalo ktoré družstvo. Sú iba dve družstvá² a každý bod v neprospech jednej strany je bodom v prospech tej druhej. Každý v hľadisku si v duchu počíta, koľko bodov ktorý rečník získal voči druhému. Na konci diskusie je rečník s väčším bodovým ziskom, prirodzene, víťazom; všetko, čo povedal, musí byť teda pravda a všetko, čo povedal porazený, musí byť zle.

“Afektívna heuristika v posudzovaní rizika a úžitku³” skúmala, či pokusné osoby zmiešavajú svoje hodnotenia možných úžitkov a možných rizík technológie (napr. jadrovej elektrárne) do jedného celkového dobrého alebo zlého pocitu z danej technológie. Napríklad vám najprv poviem, že konkrétny typ jadrového reaktora vytvára menej jadrového odpadu ako iné navrhované reaktory. Ale potom vám poviem, že tento reaktor je menej stabilný než alternatívne reaktory, s väčším rizikom roztavenia, ak sa dostatočné veľké množstvo vecí pokazí naraz.

Ak má reaktor väčšiu pravdepodobnosť roztavenia, vyzerá to ako “bod proti” reaktoru, prípadne “bod proti” tomu, kto argumentuje za postavenie reaktora. A ak ten reaktor vytvára menej odpadu, je to “bod za” reaktor, prípadne “bod za” jeho postavenie. Sú teda tieto dva fakty proti sebe? Nie. V skutočnom svete, nie. Tieto dva fakty môžu byť citované rôznymi stranami tej istej diskusie, ale sú logicky nezávislé; fakty nevedia, na čej strane sú. Množstvo odpadu vytvoreného reaktorom závisí od fyzických vlastností daného typu reaktora. Iné fyzické vlastnosti reaktora spôsobujú, že je menej stabilný. Aj keby sa niektoré vlastnosti podieľali na oboch týchto výsledkoch, treba osobitne zvážiť pravdepodobnosť roztavenia a očakávané množstvo odpadu vyprodukované ročne. Sú to dve rôzne fyzikálne otázky, s dvoma rôznymi faktickými odpoveďami.

Štúdie ako horeuvedená však ukazujú, že ľudia majú sklon posudzovať technológie – a mnohé iné problémy – na základe celkového dobrého alebo zlého pocitu. Ak povieť ľuďom, že nejaký typ reaktora produkuje menej odpadu, budú rátať pravdepodobnosť jeho roztavenia ako nižšiu. To znamená, že dostanú *nesprávnu odpoveď* na fyzikálne otázky s konkrétnymi faktickými odpoveďami, pretože zmiešali logicky nezávislé otázky – vnímali fakty ako vojakov bojujúcich na rôznych stranách vojny, a mysleli si, že hocijakého vojaka na jednej strane možno použiť na boj proti hocijakému vojakovi na druhej strane.

Váhy nie sú pre pani Spravodlivosť celkom neprimerané, pokiaľ vyšetruje prísne faktickú otázku viny či neviny. Buď John Smith zabil Johna Doa, alebo nie. Učili nás (E. T. Jaynes), že všetky bayesovské indície sa skladajú z tokov pravdepodobnosti *medzi* hypotézami; že neexistuje nič také ako indícia “podporujúca” alebo “protirečiaci” nejakej hypotéze, jedine v tom, zmysle, že iné hypotézy si oproti nej pohoršia alebo prílepšia. Dokiaľ teda pani Spravodlivosť

¹ http://en.wikipedia.org/wiki/Lady_Justice

² pozri: 12 Bájka o vede a politike

³ http://www-abc.mpib-berlin.mpg.de/users/r20/finucane00_the_affect_heuristic.pdf

vyšetruje *jednu*, prísne *faktickú* otázku, s iba *dvoma* možnými odpoveďami, váhy sú primeraný nástroj. Ak chce Justitia zvážiť nejakú zložitejšiu tému, mala by sa vzdať svojich váh i svojho meča.

Nie všetky argumenty možno zredukovať na púhe hore alebo dole. Pani Rozumnosť si nosí zápisník, do ktorého si zapisuje všetky fakty, ktoré nie sú na nikoho strane.

Pôvodný článok s komentármi⁴

⁴ http://www.lesswrong.com/lw/h1/the_scales_of_justice_the_notebook_of_rationality/

Kapitola 50. Chyba prisudzovania

Chyba prisudzovania je sklon vyvodzovať závery o jedinečnej a trvalej povahe osoby na základe správania, ktoré sa dá úplne vysvetliť situáciou, v ktorej nastalo.

—Gilbert a Malone

Máme sklon vidieť príliš priamu súvislosť medzi činmi a osobnosťami druhých ľudí. Keď vidíme, ako niekto iný kope do automatu bez viditeľného dôvodu, predpokladáme, že je to „zurvalec“. Avšak keď my sami kopneme do automatu, je to preto, lebo mešká autobus, ušiel nám vlak, nestíhame prácu, a ešte aj tento prekliaty automat nám zožral peniaze už druhý deň po sebe. Myslíte si, že *v danej situácii by do toho automatu určite kopol každý*.

Svoje vlastné konanie pripisujeme svojej *situácii*, vnímame ho ako dokonale normálnu reakciu na okolnosti. Ale keď niekto iný kopne do automatu, nevidíme za ním vo vzduchu viať jeho minulosť. Vidíme len to kopnutie, *nevieme* o žiadnom dôvode a myslíme si, že to musí byť od prírody zlostný človek – veď predsa útočí bez akejkoľvek provokácie.

Zvážte však pôvodné pravdepodobnosti. Na svete je viac meškajúcich autobusov než mutantov obdarených neprirodzene vysokými hladinami hnevu, ktoré ich nútia občas spontánne nakopnúť automat. Iste, priemerný človek je v skutočnosti mutant. Ak si dobre spomínam, priemerný jednotlivec má 2 až 10 somaticky vyjadrených mutácií. Ale ľubovoľného *konkrétneho* miesta v DNA sa to týka veľmi nepravdepodobne. Podobne, ľubovoľná stránka niekoho povahy sa pravdepodobne príliš nevzdďahuje od priemeru. Naznačovať opak znamená brať si na plecia bremeno nepravdepodobnosti.

Aj keď ľudí explicitne informujeme o situačných podnetoch, zdá sa, že to nevedia správne odrátať od pozorovaného správania. Keď pokusným osobám povedali, že rečníkom *náhodne prideli*li prečítať prejav za interrupcie alebo proti, pokusné osoby si stále mysleli, že rečník má postoje smerom k náhodne pridelennej téme. (Jones a Harris 1967, „Prisudzovanie vlastností“.)

Zdá sa, že je veľmi intuitívne vysvetľovať dažď pomocou vodných duchov; vysvetľovať oheň pomocou ohnivej látky (flogistonu) unikajúcej z horiacej hmoty; vysvetľovať uspávajúci účinok lieku povedaním, že obsahuje „dormitívnu potenciu“. Skutočnosť zvyčajne zahŕňa zložitejšie mechanizmy: cyklus vyparovania a zrážania je základom dažďa; oxidačné horenie je základom ohňa; chemické interakcie s nervovým systémom sú základom uspávadla. Ale mechanizmy nám znejú omnoho zložitejšie než esencie; ťažšie sa na ne myslí, sú menej dostupné. Takže keď niekto kopne do automatu, myslíme si, že má vrodený sklon kopáť do automatu.

Okrem prípadu, keď ten „niekto“, kto kopne do automatu, sme my – v tom prípade sa správame dokonale normálne za danej situácie; určite by každý urobil to isté. Veru, preceňujeme sklon druhých reagovať rovnako ako my – „efekt falošného konsenzu“. Študenti pijúci alkohol výrazne preceňujú percento spolužiakov, ktorí pijú; abstinenti ho však výrazne podceňujú. Táto „základná chyba prisudzovania“ označuje náš sklon pripisovať správanie druhých ich povahe, zatiaľ čo u seba naopak.

Aby sme pochopili, prečo ľudia konajú tak, ako konajú, musíme si najprv uvedomiť, že každý vidí sám seba ako normálne sa správajúceho. Nepýtajte sa, s akou zvláštnou mutantskou vlastnosťou priamo zodpovedajúcou ich vonkajšiemu

správaníu sa narodili. Radšej sa pýtajte, v akej situácii sa títo ľudia podľa svojho vlastného názoru nachádzajú. Áno, ľudia majú aj vrodené vlastnosti – ale neexistuje *tolko* dedičných vrtochov, aby priamo zodpovedali každému pozorovanému vonkajšiemu správaniu.

Predstavte si, že by som vám dal ovládač s dvoma tlačidlami: červeným a zeleným. Červené tlačidlo zničí celý svet a zelené tlačidlo zabráni stlačeniu červeného tlačidla. Ktoré tlačidlo by ste stlačili? To zelené. Ktokoľvek dá odlišnú odpoveď, pravdepodobne túto otázku príliš komplikuje².

A predsa sa ma ľudia občas pýtajú, prečo chcem zachrániť svet³. Akoby som musel mať traumatické detstvo alebo niečo také. Naozaj, vyzerá to ako úplne samozrejmé rozhodnutie... ak vidíte situáciu takýmto pohľadom.

Možno mám netypické názory, ktoré si žiadajú vysvetlenie – prečo verím v takéto veci, keď väčšina ľudí nie? – ale za predpokladu týchto názorov moje *reakcie* snád' nevyžadujú ďalšie vysvetľovanie. Možno som obeťou falošného konsenzu; možno preceňujem koľko ľudí by stlačilo zelené tlačidlo, keby videli situáciu týmto pohľadom. Ale viete, stále by som sa stavil, že sú prinajmenšom vo *výraznej menšine*²²²⁷.

Väčšina ľudí sa zvnútra vníma ako úplne normálnych. Dokonca aj ľudia, ktorých nenávidíte, ľudia, ktorí urobili hrozné veci, nie sú výnimoční mutanti. Žiaľ, na toto mutácie netreba. Ak pochopíte toto, ste pripravení prestať byť prekvapení⁴ ľudskými udalosťami.

Pôvodný článok s komentármi⁵

² <http://www.yudkowsky.net/singularity/simplified>

³ <http://www.intelligence.org/AIRisk.pdf>

⁴ pozri: 196 Mysli ako skutočnosť

⁵ http://www.lesswrong.com/lw/hz/correspondence_bias/

Kapitola 51. Sú vaši nepriatelia od narodenia zlí?

Ako sme si už povedali¹, vidíme príliš priamu súvislosť medzi konaním druhých a ich vrodenu povahou. Vidíme nezvyčajnú povahu, ktorá presne zodpovedá nezvyčajnému správaniu, namiesto toho, aby sme pátrali po skutočných alebo imaginárnych situáciách, ktoré by toto správanie mohli vysvetliť. Predpokladáme mutantov.

Keď nám niekto naozaj *ublíži* – spácha skutok, ktorý (právom či neprávom) odsudzujeme – vtedy sa mi zdá, že sa chyba prisudzovania ešte zdvojnásobí. Máme zrejme *veľmi* silný sklon obviňovať za zlé skutky nepriateľovu mutantnú, zlú povahu. Nie z morálneho hľadiska, ale ako technickú otázku pôvodnej pravdepodobnosti, by sme sa mali opýtať, čo si nepriateľ môže myslieť o svojej situácii, čo by mohlo zmenšiť zdanlivú čudesnosť² jeho správania. To by nám umožnilo prísť s hypotézou menej výnimočnej povahy, a tak by sme niesli menšie bremeno nepravdepodobnosti.

11. septembra 2001 devätnásť moslimských mužov unieslo štyri lietadlá so samovražedným zámerom uškodiť Spojeným Štátom Americkým. Čo si myslíte, prečo to mohli urobiť? Pretože videli USA ako maják svetovej slobody, ale narodili sa s mutantnou povahou, ktorá spôsobuje, že nenávidia slobodu?

Realisticky, väčšina ľudí si neskladá svoj životný príbeh so sebou samým v roli záporného hrdinu. Každý je kladným hrdinom vo svojom vlastnom príbehu. V nepriateľovom príbehu, z pohľadu nepriateľa, *nebude nepriateľ vykreslený ako ten zlý*. Ak sa snažíte vyskladať motiváciu, podľa ktorej *má* nepriateľ vyzeráť ako ten zlý, budete sa úplne mýliť ohľadom toho, čo sa v nepriateľovej hlave naozaj odohráva.

Lenže politika zabíja myslenie³. Debata je vojna, argumenty sú vojaci. Akonáhle víete, na ktorej strane ste, musíte podporovať všetky argumenty tejto strany, a napádať všetky argumenty, ktoré vyzerajú v prospech nepriateľskej strany; inak je to, ako keby ste vlastných vojakov bodali do chrbta.

Keby nepriateľ mal zlú povahu, bol by to argument v prospech vašej strany. A *každý* argument v prospech vašej strany musíte podporiť, bez ohľadu na jeho hlúposť – inak prestávate tlačiť v nejakej časti frontovej línie. Každý sa snaží prekonať svojho suseda vlasteneckými prehláseniami a nikto sa neopovažuje protirečiť. Čoskoro má nepriateľ rohy, netopierie krídla, ohnivý dych a pazúre, z ktorých odkvapkáva jedovatá kyselina. Ak sa pokúsite niečo z tohto poprieť na základe púhych faktov, argumentujete za nepriateľovu stranu; ste zradca. Veľmi málo ľudí dokáže pochopiť, že nebránite nepriateľa, iba bránite pravdu.

Keby bolo treba mutantov na to, aby robili príšerné veci, história ľudského druhu by vyzerala veľmi odlišne. Mutantov by bolo málo.

Alebo je to možno strach, že porozumenie povedie k odpusteniu. Je ľahšie strieľať do zlých mutantov. O čo inšpirujúcejšie znie bojový pokrik: “Skapte, mizerní lotri!” než “Skapte, ľudia, ktorí mohli byť takí istí ako ja, ale vyrástli v odlišnom prostredí!” Zo zabíjania ľudí, ktorí *neboli* čistou temnotou, by ste mohli mať pocit viny.

¹ pozri: 50 Chyba prisudzovania

² pozri: 196 Mysli ako skutočnosť

³ pozri: 47 Politika zabíja myslenie

Toto mi pripadá ako hlboko zakorenená túžba po jednostrannej debata o pravidlách⁴, v ktorej to najlepšie rozhodnutie nemá *žiadne* nevýhody. Ak armáda prekračuje hranice, alebo k vám ide šialenec s nožom, alternatívy sú: (a) brániť sa, (b) ľahnúť si a zomrieť. Ak sa budete brániť, možno budete musieť zabiť. Ak zabijete niekoho, kto mohol, v inom svete, byť vaším priateľom, je to tragédia. Áno, *je* to tragédia. Druhá možnosť, ľahnúť si a zomrieť, je tiež tragédia. Prečo by musela existovať aj netragická možnosť? Kto hovorí, že najlepšie možné rozhodnutie nesmie mať žiadne nevýhody? Ak už niekto musí zomrieť, rovnako dobre to môže byť iniciátor útoku, aby sme tým odradili budúce násilie a tým minimalizovali celkové množstvo smrti.

Ak má nepriateľ priemernú povahu a koná podľa takého názoru na svoju situáciu, pri ktorom je násilie typickou ľudskou reakciou, to ešte neznamená, že jeho názory sú fakticky správne. Neznamená to, že sú oprávnené. Znamená to, že budete musieť zastreliť niekoho, kto je vo svojom vlastnom príbehu kladným hrdinom a v ich príbehu kladný hrdina zomrie na strane 80. To je tragédia, ale je to lepšie než alternatívna tragédia. Je to voľba, aké robí každý policajt, každý deň, aby sa naše milé malé svety nerozpustili v chaose.

Keď presne odhadujete nepriateľovu psychológiu – keď viete, čo naozaj je v nepriateľovej hlave – toto poznanie vám nedá pocit, že zasadzujete nádherný úder súperovi⁵. Nedá vám hrejivý pocit spravodlivého rozhorčenia. Nebudete mať vďaka nemu dobrý pocit zo seba samého. Ak vám váš odhad prináša neznesiteľný smútok, možno vidíte svet taký, aký naozaj je. Zriedkavejšie sa môže stať, že vám presný odhad privolá zimomriavky vážnej hrôzy, ako keď jednáte so skutočnými psychopatmi alebo s neurologicky nepoškodenými ľuďmi s názormi, ktoré naprosto zničili ich prítomnosť (scientológovia alebo Jesus Camp⁶).

Povedzme to teda priamo a nahlas – únoscovia z 11. 9. neboli zlí mutanti. Necítili nenávisť k slobode. Aj oni boli kladnými hrdinami vo svojich vlastných príbehoch a zomreli za to, čo považovali za správne – pravdu, spravodlivosť, islamský spôsob života. To, že sa tak sami vnímali, neznamená, že mali pravdu. To, že sa tak vnímali, neznamená, že musíme súhlasiť, že mali právo urobiť, čo urobili. To, že sa tak vnímali, neznamená, že pasažieri United Flight 93 mali ustúpiť a nechať im voľnú ruku. Znamená to, že v inom svete, keby vyrástli v inom prostredí, títo únoscovia mohli byť napríklad policajtmi. A to je naozaj tragédia. Vitajte na Zemi.

Pôvodný článok s komentármi⁷

⁴ pozri: 48 Debaty o pravidlách by nemali vyzeráť jednostranne

⁵ pozri: 47 Politika zabíja myslenie

⁶ http://www.youtube.com/watch?v=y_EKHK1C2IE

⁷ http://lesswrong.com/lw/i0/are_your_enemies_innately_evil/

Kapitola 52. Obrátená hlúposť nie je inteligencia

“...potom naši ľudia v tejto časovej línii prišli s nápravnou akciou. Tu.”

Utreľ obrazovku a začal zadávať kombinácie. Objavovala sa strana za stranou s údajmi o ľuďoch, ktorí tvrdili, že videli tajomné disky a každá správa bola fantastickjšia než predchádzajúca.

“Štandardná zahladzovacia technika,” uškrnul sa Verkan Vall. “Počul som iba pár zmienok o ‘lietajúcich tanieroch’ a všetko to bolo žartom. V kultúre tohto typu dokážete vždy urobiť pravdivý príbeh nedôveryhodným, ak vedľa neho postavíte desať ďalších, jasne nepravdivých.”

—H. Beam Piper, Policajná operácia

Piper má v niečom pravdu. Ja osobne neverím na žiadnych zle ukrytých mimozemšťanov v našich končinách. Ale moja neviera nemá nič spoločné so zahanbujúcou nerozumnosťou kultov lietajúcich tanierov – prinajmenšom dúfam, že nie.

Vy a ja veríme, že kulty lietajúcich tanierov vznikli v úplnej neprítomnosti lietajúcich tanierov. Kulty dokážu vďaka ľudskej hlúposti vzniknúť okolo takmer hocijakej myšlienky¹. Táto hlúposť funguje *nezávisle* na mimozemských zásahoch: Predpokladali by sme, že kulty lietajúcich tanierov vzniknú bez ohľadu na to, či lietajúce taniere existujú alebo nie. Dokonca aj keby existovalo zle ukrytí mimozemšťania, nebolo by o nič *menej* pravdepodobné, že vzniknú kulty lietajúcich tanierov. $P(\text{kulty}|\text{mimozemšťania})$ nie je menšie ako $P(\text{kulty}|\sim\text{mimozemšťania})$, pokiaľ nepredpokladáte, že zle ukrytí mimozemšťania by úmyselne potláčali kulty lietajúcich tanierov. Podľa bayesovskej definície infície², pozorovanie “existujú kulty lietajúcich tanierov” nie je indícia *proti* existencii lietajúcich tanierov. Nie je to veľmi indícia jedným ani druhým smerom.

Toto je použitie všeobecného princípu, podľa ktorého, ako hovorí Robert Pirsig: “Najväčší hlupák na svete môže povedať, že Slnko svieti, ale ono kvôli tomu nezhasne.”

Keby ste vedeli, že sa niekto mýli v 99,99 % prípadov otázok typu áno alebo nie, mohli by ste získať presnosť 99,99 % jednoducho obrátením jeho odpovedí. On by musel vynaložiť všetku prácu potrebnú na získanie dobrých indícií previazaných so skutočnosťou a systematicky tieto indície spracovať, aby dokázal takto spoľahlivo *antikorelovať*. Musel by byť superinteligentný, aby mohol byť taký hlúpy.

Auto s pokazeným motorom nebude cúvať rýchlosťou 300 km/h, ani keď je ten motor *naozaj, naozaj pokazený*.

Ak hlúposť nedokáže spoľahlivo antikorelovať s pravdou, o čo menej by malo s pravdou antikorelovať ľudské zlo? Opakom efektu svätožiary³ je efekt rohov. Všetky vnímané záporné vlastnosti korelujú. Ak je Stalin zlý, všetko, čo povie, musí byť nepravdivé. Nechceli by ste predsa súhlasiť *so Stalinom*, alebo áno?

¹ pozri: 69 Každá kauza chce byť sektou

² pozri: 178 Intuitívne vysvetlenie Bayesovej vety

³ pozri: 60 Efekt svätožiary

Stalin si myslel, že $2 + 2 = 4$. Ak však budete obhajovať akýkoľvek Stalinov výrok, vrátane “ $2 + 2 = 4$ ”, ľudia budú vidieť iba to, že “súhlasíte so Stalinom”; čiže musíte byť na jeho strane.

Dôsledky tohto princípu:

- Ak chcete poctivo argumentovať proti nejakej myšlienke, mali by ste argumentovať proti najlepším argumentom jej najsilnejších zástancov. Argumentovanie proti slabším zástancam nedokazuje *nič*, pretože aj tá najsilnejšia myšlienka bude priťahovať slabých zástancov. Ak chcete argumentovať proti transhumanizmu alebo explózii inteligencie, musíte priamo čeliť argumentom Nicka Bostroma alebo Eliezera Yudkowskeho po roku 2003. Najmenej pohodlná cesta⁴ je tá jediná správna.
- Ukazovať na smutných trápnych bláznov dohnaných k šialenstvu ich chápaním nejakej myšlienky nie je indícia proti tejto myšlienke. Mnoho priaznivcov New Age sa stalo ešte bláznivejšími vďaka svojmu osobnému výkladu kvantovej mechaniky⁵.
- Nieкто raz povedal: “Nie všetci konzervatívci sú hlúpi, ale väčšina hlúpych ľudí sú konzervatívci.” Ak sa nedokážete dostať do stavu mysle, keď vám tento výrok, či už je pravdivý alebo nie, pripadá *úplne nepodstatný* ako kritika konzervativizmu, potom nie ste pripravení rozumne rozmýšľať o politike.
- Argument ad hominem⁶ nie je správny.
- Musíte byť schopní argumentovať proti genocíde bez slov: “Hitler chcel vyvraždiť židov.” Keby Hitler *nebol chcel* genocídu, znamenalo by to, že genocída je okej?
- Hansonovskými slovami: Vaša inštinktívna ochota veriť nejakej myšlienke sa bude meniť podľa vašej ochoty *spájať sa* s ľuďmi, ktorí sú známi ako jej zástancovia – čo nesúvisí s jej skutočnou *pravdivosťou*. Niektorí ľudia môžu byť neochotní pripustiť, že Boh neexistuje, nie preto, že by existovali dôkazy, že Boh *existuje*, ale skôr preto, lebo sa nechcú pridávať k Richardovi Dawkinsovi ani tým prekliatym “provokujúcim” ateistov, ktorí chodia a nahlas hovoria: “Boh neexistuje.”
- Ak váš terajší počítač prestane fungovať, nemôžete dôjsť k záveru, že všetko na vašom terajšom systéme je pokazené a že potrebujete nový systém bez procesora AMD, bez videokarty ATI, bez pevného disku Maxtor a bez ventilátora na krabici – aj keď váš terajší počítač všetko toto má a nefunguje. Možno len potrebujete nový kábel od napájania.
- Ak sa sto vynálezcom nepodarí⁷ postaviť lietajúci stroj pomocou kovu a dreva a plátna, to neznamená, že v skutočnosti potrebujete lietajúci stroj z mäsa a kostí. Ak sa tisíc projektom nepodarí postaviť umelú inteligenciu pomocou elektronických obvodov, to nedokazuje, že zdrojom problému je elektrina. Dokiaľ nepochopíte problém, naivné obrátenie s najväčšou pravdepodobnosťou k riešeniu nepovedie.⁸

Pôvodný článok s komentármi⁹

⁴ http://lesswrong.com/lw/2k/the_least_convenient_possible_world/

⁵ http://lesswrong.com/lw/r5/the_quantum_physics_sequence/

⁶ <http://plover.net/~bonds/adhominem.html>

⁷ http://lesswrong.com/lw/vs/selling_nonapples/

⁸ pozri: 144 Umelé pridanie

⁹ http://lesswrong.com/lw/lw/reversed_stupidity_is_not_intelligence/

Kapitola 53. Argument zatieňuje autoritu

Black Belt Bayesian¹ (alias „steven“) sa pokúsil vysvetliť nesúmernosť medzi dobrými argumentmi a dobrou autoritou, ale nezdá sa mi, že by tým odpovedal na komentáre pri článku Obrátená hlúposť nie je inteligencia², takže mi dovoľte, aby som sa o to pokúsil sám:

Scenár 1: Barry je známy geológ. Charles je 14-ročný mladistvý delinkvent s dlhým záznamom trestov a občasnými psychotickými záchvatmi. Barry bez vysvetlenia povie Arthurovi nejaké kontrainatívne tvrdenie o kameňoch a Arthur tomu prisúdi pravdepodobnosť 90 %. Potom Charles povie rovnako kontrainatívne tvrdenie o kameňoch a Arthur tomu prisúdi pravdepodobnosť 10 %. Je jasné, že Arthur pri rozhodovaní, či má danému tvrdeniu veriť, zohľadňuje *autoritu* hovoriaceho.

Scenár 2: David povie kontrainatívne tvrdenie o fyzike a dá Arthurovi podrobné vysvetlenie argumentov, vrátane odkazov na odbornú literatúru. Ernie povie rovnako kontrainatívne tvrdenie, ale dá nepresvedčivé vysvetlenie obsahujúce niekoľko nepodložených skokov. Obaja tvrdia, že je to najlepšie vysvetlenie, aké dokážu poskytnúť (hocikomu, nielen Arthurovi). Po vypočutí si vysvetlení, Arthur prisúdi Davidovmu tvrdeniu pravdepodobnosť 90 % a Ernieho tvrdeniu pravdepodobnosť 10 %.

Môže sa zdať, že tieto dva scenáre sú zhruba súmerné: oba obsahujú zváženie užitočnej indicie, buď silnej a slabšej autority alebo silného a slabého argumentu.

Predstavme si však, že Arthur požiada Barryho a Charlesa o plné technické vysvetlenie vrátane odkazov na odbornú literatúru; a že Barry a Charles predložia rovnako dobré vysvetlenia a Arthur si pozrie odkazy a všetky sedia. Potom Arthur požiada Davida a Ernieho o potvrdenie ich odbornosti a ukáže sa, že David a Ernie majú zhruba rovnakú odbornosť – možno sú obaja šašovia, možno sú obaja fyzici.

Ak predpokladáme, že Arthur má dostatočné vedomosti, aby pochopil všetky technické argumenty – lebo inak je to len dobre pôsobiaci šum – zdá sa, že by Arthur mal dôjsť k záveru, že David je omnoho dôveryhodnejší než Ernie, zatiaľ čo Barry má v lepšom prípade iba maličkú výhodu oproti Charlesovi.

Veru, ak sú technické argumenty dosť dobré, Barryho výhoda voči Charlesovi možno nestojí za reč. Dobrý technický argument je taký, ktorý *odstraňuje* závislosť na osobnej autorite hovoriaceho.

Podobne, ak naozaj veríme, že Ernie dal najlepší argument, aký *mohol* dať, čo zahrňa všetky jeho inferenčné kroky a všetku podporu, ktorú zohľadnil – vrátane citátov všetkých autorít, ktoré samotný Ernie počul – potom môžeme viacmenej ignorovať informáciu o Ernieho odbornosti. Ernie môže byť fyzik alebo šašo, na tom nezáleží. (Opäť, predpokladáme, že máme dostatočné technické zručnosti na spracovanie argumentu. V opačnom prípade Ernie jednoducho vyslovuje záhadné slabiky a či týmto slabikám „veríme“, závisí do veľkej miery od jeho autority.)

Zdá sa teda, že medzi autoritou a argumentom je nesúmernosť. Ak poznáme autoritu, stále nás zaujímajú argumenty; ak však úplne poznáme argumenty, autorita nám povie už iba máličko.

¹ www.acceleratingfuture.com/steven/?p=33

² pozri: 52 Obrátená hlúposť nie je inteligencia

Zrejme (povie začiatovník) sú autorita a dôkaz dva zásadne odlišné druhy indícií³ a tento rozdiel nie je zohľadnený v nudne zjednodučených metódach bayesovskej teórie pravdepodobnosti⁴. Pretože hoci je sila indícií v oboch prípadoch rovnaká – 90 % verzus 10 % – nesprávajú sa podobne, keď ich skombinujeme. Ach, ako to len vysvetlíme?

Tu je polovica technickej ukážky, ako tento rozdiel reprezentovať pomocou teórie pravdepodobnosti. (Zvyšok môžete prijať na základe mojej osobnej autority alebo si dohľadať odbornú literatúru.)

Ak $P(H|E1) = 90\%$ a $P(H|E2) = 9\%$, aká je pravdepodobnosť $P(H|E1, E2)$? Ak na základe informácie, že E1 je pravda, priradíme H pravdepodobnosť 90 % a na základe informácie, že E2 je pravda, priradíme H pravdepodobnosť 9 %, akú pravdepodobnosť by sme mali priradiť H, ak sa dozvieme, že platí aj E1 aj E2? Toto sa podľa teórie pravdepodobnosti jednoducho nedá vypočítať iba na základe uvedených informácií. Nie, tá chýbajúca informácia nie je apriórna pravdepodobnosť H. E1 a E2 nemusia byť navzájom nezávislé.

Predpokladajme, že H je “môj chodník sa šmýka”, E1 je “môj zavražovač beží” a E2 je “je noc”. Chodník sa začne šmýkať 1 minútu po spustení zavražovača a prestane hneď ako zavražovač skončí, a zavražovač ide 10 minút. Vieme teda, že ak zavražovač beží, chodník sa šmýka s pravdepodobnosťou 90 %. Zavražovač je zapnutý 10 % nočného času, ak teda vieme, že je noc, pravdepodobnosť, že chodník sa šmýka, je 9 %. Ak vieme, že je noc a že je zavražovač zapnutý – teda ak vieme oba tieto fakty – pravdepodobnosť, že chodník sa šmýka, je 90 %.

Môžeme si to znázorniť nasledujúcim grafickým modelom:

Noc -> Zavražovač -> Šmýka sa

Či je alebo nie je noc *spôsobuje* zapnutie a vypnutie zavražovača, a či je zavražovač zapnutý *spôsobuje* šmýkanie alebo nešmýkanie chodníka.

Smer šípok je dôležitý. Keby som napísal:

Noc -> Zavražovač <- Šmýka sa

Znamenalo by to, že keby som *nevedel* nič o zavražovači, pravdepodobnosť noci a šmyklavosti by boli navzájom nezávislé. Napríklad, predstavme si, že hodím Kockou Jedna a Kockou Dva a získané čísla sčítam, aby som dostal Súčet:

Kocka Jedna -> Súčet <- Kocka Dva

Ak mi nepoviete súčet dvoch čísel, a ak mi poviete, že na prvej kocke padlo 6, to mi nič nehovorí o výsledku na druhej kocke, zatiaľ. Ak mi však teraz poviete, že súčet je 7, viem, že na druhej kocke padlo 1.

Zisťovanie či sú rôzne kúsky informácie navzájom závislé alebo nezávislé, za rôznych všeobecných podmienok, sa ukazuje ako samostatná technická téma. Vhodné knihy sú od Judeu Pearla Pravdepodobnostné uvažovanie v inteligentných systémoch: Siete dôveryhodného odvodzovania⁵ a Kauzalita⁶. (Ak máte čas iba na jednu knihu, prečítajte si tú prvú.)

Ak viete čítať kauzálne grafy, potom sa pozriete na graf hádzania kociek a hneď vidíte:

$$p(\text{kocka1}, \text{kocka2}) = p(\text{kocka1}) \times p(\text{kocka2})$$

³ pozri: 20 Čo je indícia?

⁴ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

⁵ <http://www.amazon.com/Probabilistic-Reasoning-Intelligent-Systems-Plausible/dp/1558604790/>

⁶ <http://www.amazon.com/Causality-Reasoning-Inference-Judea-Pearl/dp/0521773628/>

$$p(\text{kocka1, kocka2} \mid \text{súčet}) \neq p(\text{kocka1} \mid \text{súčet}) \times p(\text{kocka2} \mid \text{súčet})$$

Ak sa pozriete na správny diagram chodníka, uvidíte fakty ako:

$$p(\text{šmýkanie} \mid \text{noc}) \neq p(\text{šmýkanie})$$

$$p(\text{šmýkanie} \mid \text{zavlažovač}) \neq p(\text{šmýkanie})$$

$$p(\text{šmýkanie} \mid \text{noc, zavlažovač}) = p(\text{šmýkanie} \mid \text{zavlažovač})$$

To znamená, že pravdepodobnosť, že sa chodník šmýka, ak vieme o zavlažovači a noci, je rovnaká ako pravdepodobnosť, ktorú by sme priradili, keby sme vedeli iba o zavlažovači. Vedomosť o zavlažovači urobila vedomosť o noci nepodstatnou na usudzovanie o šmykľavosti.

Toto sa nazýva *zatienenie* a kritérium, ktoré nám dovoľuje vyčítať takéto podmienené nezávislosti z kauzálnych grafov sa nazýva *D-separácia*.

V prípade argumentu a autority vyzerá kauzálny diagram takto:

Pravda -> Kvalita argumentu -> Názor odborníka

Ak je niečo pravda, zvyknú v prospech toho existovať argumenty, odborníci preto vidia tieto indície a menia svoje názory. (Teoreticky!)

Ak vidíme, že si odborník niečo myslí, môžeme spätne usudzovať existenciu nejakej indície (aj keď nevieme, čo presne je táto indícia) a z existencie tejto hypotetickej indície môžeme spätne usudzovať pravdivosť výroku.

Ak však poznáme hodnotu uzla "Argument", uzol "Pravda" sa tým D-separuje od uzla "Názor odborníka" zablokovaním všetkých ciest medzi nimi, podľa určitého technického kritéria pre "zablokovanie cesty", ktoré v tomto prípade vyzerá celkom jasne. Takže aj bez kontroly presnej distribúcie pravdepodobnosti môžeme z grafu vyčítať, že:

$$p(\text{pravda} \mid \text{argument, odborník}) = p(\text{pravda} \mid \text{argument})$$

Toto nie je v rozpore s bežnou teóriou pravdepodobnosti. Je to iba kompaktnejší spôsob, ako vyjadriť určité pravdepodobnostné fakty. Dokázali by ste vyčítať tie isté rovnice a nerovnice z neozdobenej pravdepodobnostnej distribúcie – ale bolo by ťažšie to priamo uvidieť. Autorita a argument nevyžadujú dva rôzne druhy pravdepodobnosti, rovnako ako zavlažovače nemajú ontologicky odlišnú podstatu než slnečné svetlo.

V praxi sa závislosti na autorite nikdy *celkom* nezbavíte. Dobré autority budú s väčšou pravdepodobnosťou poznať existujúce indície proti niečomu a to by ste mali zohľadniť; menšia autorita ich bude poznať s menšou pravdepodobnosťou, čo robí jej argumenty menej spoľahlivé. Toto nie je faktor, ktorý môžete odstrániť púhym vypočutím si indícií, ktoré *zohľadnili*.

Je tiež veľmi ťažké redukovat argumenty na *číru* matematiku. Posudzovanie sily inferenčného kroku môže navyše závisieť na intuíciách, ktoré nedokážete duplikovať bez rovnakých tridsiatich rokov skúsenosti.

Je nezmazateľne legitímne priradiť *trochu* vyššiu pravdepodobnosť tomu, čo vám o bayesovskej pravdepodobnosti hovorí E. T. Jaynes, než priradujete celkom rovnakému výroku, ktorý povie Eliezer Yudkowsky. Päťdesiat rokov skúsenosti navyše by sa nemali počítať ako doslova *nulový* vplyv.

Ale táto o čosi silnejšia autorita je iba *ceteris paribus* a dá sa ľahko prevážiť silnejšími argumentmi. Našiel som v jednej z Jaynesových kníh menší preklep – pretože algebra prebýja autoritu.

Pôvodný článok s komentármi⁷

⁷ http://lesswrong.com/lw/lx/argument_screens_off_authority/

Kapitola 54. Pritisnite si otázku

V umení rozumnosti existuje zručnosť *blízkosti k téme* – snažiť sa sledovať indície, ktoré sú k pôvodnej otázke čo najbližšie, takže zatieňujú čo najviac iných argumentov.

Bratia Wrightovci povedia: “Moje lietadlo poletí.” Ak sa pozriete na ich autoritu (opravári bicyklov, ktorí sú zhodou okolností vynikajúci amatérski fyzici) a porovnáte ju s autoritou povedzme Lorda Kelvina, zistíte, že Lord Kelvin má väčšiu autoritu.

Ak si vypýtate na nahliadnutie výpočty bratov Wrightovcov a budete im rozumieť, a ak si vypýtate výpočty Lorda Kelvina (pravdepodobne nemá žiadne okrem vlastnej nedôveryčivosti), potom bude autorita omnoho menej dôležitá.

Ak naozaj *uvidíte, ako to lietadlo letí*, samotné výpočty sa stanú na mnohé účely nepodstatnými a Kelvinova autorita nebude stáť za zmienku.

Čím *priamejšie* sa vaše argumenty týkajú nejakej otázky, bez sprostredkujúcich vplyvov – čím bližšie sú pozorované uzly k uzlu otázky vo Veľkej Sieti Kauzality – tým *mocnejšia* je indícia. V týchto kauzálnych grafoch platí veta, že nikdy nemôžete dostať *viac* informácií zo vzdialenejších uzlov než z bližších uzlov, ktoré tie vzdialenejšie zatieňujú¹.

Jerry Cleaver povedal: “To, čo ťa zloží, nie je zabudnutie použiť nejakú zložitú neprehľadnú techniku na vysokej úrovni. Je to nevšímanie si základov. Nesledovanie lopty pohľadom.”

Rovnako ako je lepšie argumentovať fyzikou než diplomom, je lepšie argumentovať fyzikou než rozumnosťou. Kto bol rozumnejší, bratia Wrightovci alebo Lord Kevin? Ak môžeme skontrolovať ich rovnice, môže nám to byť jedno! Cnosť racionalistu nedokáže *priamo* zdvihnúť lietadlo do vzduchu.

Ak zabudnete na tento princíp, bude vám učenie sa o ďalších skresleniach škodiť², pretože vás bude rozptyľovať od priamejších argumentov. Je príliš ľahké argumentovať, že niekto vykazuje Skreslenie číslo 182 vo vašom repertoári úplne všeobecných obvinení, ale bez bližšej indície nemožno *uzavrieť* faktickú tému. Ak existujú skreslené dôvody hovoriť, že slnko svieti, to ešte nespôsobuje, že zhasne.³

Rovnako ako nemôžete vždy urobiť experiment dnes⁴, nemôžete vždy skontrolovať rovnice dnes. Niekedy dostatočne nepoznáte pozadie, niekedy sa informácia nedá odovzdať, niekedy jednoducho nie je čas. Je žiaľ veľa prípadov, keď sa oplatí hodnotiť rozumnosť hovoriaceho. Vždy by ste to však mali robiť s pocitom prázdnoty v srdci, s vedomím, že niečo tu chýba.

Vždy keď môžete, tancuje tak blízko pôvodnej otázky, ako sa dá – pritlačte sa k nej – dostaňte sa blízko a *pritisnite si ju!*

Pôvodný článok s komentármi⁵

¹ pozri: 53 Argument zatieňuje autoritu

² pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

³ pozri: 52 Obrátená hlúposť nie je inteligencia

⁴ http://lesswrong.com/lw/io/is_molecular_nanotechnology_scientific/

⁵ http://lesswrong.com/lw/ly/hug_the_query/

Kapitola 55. Rozumnosť a anglický jazyk

Včera¹ mi niekto povedal, že mu moje písanie pripomenulo Politika a anglický jazyk² od Georgea Orwella. Bolo mi čťou. Najmä preto, lebo som už mal vymyslenú dnešnú tému.

Ak *naozaj* chcete umelcov pohľad na rozumnosť³, potom si prečítajte Orwella; je to povinné čítanie pre racionalistov aj pre spisovateľov. Orwell nebol vedec, ale spisovateľ; jeho nástrojmi neboli čísla, ale slová; jeho protivníkom nebola Príroda, ale ľudské zlo. Ak chcete uväzniť ľudí na celé roky bez súdneho procesu, musíte si vymyslieť iný spôsob, ako to povedať, než: "Idem uväzniť pána Jenningsa na celé roky bez súdneho procesu." Musíte zahmlieť poslucháčovo myslenie, zabrániť tomu, aby jasný obraz pobúril jeho svedomie. Poviete: "Nespolahlivé živly boli podrobené alternatívnemu súdnemu procesu."

Orwell bol pobúreným protivníkom totalitárstva a zahmleného myslenia, do ktorého sa zlo zahaľuje – a tak Orwellove písania o jazyku skončili ako klasické racionalistické dokumenty na úrovni Feynmana, Sagana alebo Dawkinsa.

"Spisovateľom sa hovorí, aby sa vyhli používaniu pasívneho tvaru." Racionalista, ktorý je podkutý *výhradne* vo vede, si nemusí všimnúť chybu v predchádzajúcej vete; ale každý, kto trochu písal, by ju mal uvidieť hneď. Napísal som tú vetu v pasívnom tvare; nepovedal som vám, *kto* hovorí autorom, aby sa vyhli pasívnemu tvaru. Pasívny tvar odstraňuje konajúceho, zanecháva iba konané. "Nespolahlivé živly boli podrobené alternatívnemu súdnemu procesu" – podrobené *kým*? Ako sa *robí* taký "alternatívny súdny proces"? S dostatkom statických menných pomenovaní sa dokážete *vyhnúť* čomukoľvek neprijemnému.

Články v časopisoch sú často písané v pasívnom tvare. (Prepáčte; *niektorí vedci* píšu svoje články do časopisov v pasívnom tvare. Tie články sa predsa nepíšu sami.) Znie autoritatívnejšie, keď poviete "pokusným osobám bol podávaný Progenitorivox", než "dal som každému vysokoškólákovi fľašku s 20 Progenitorivoxmi a povedal som mu, nech si každý večer dá jeden, dokiaľ sa neminú." Ak odstránite vedca z popisu, zostane vám iba to najdôležitejšie, údaje. Lenže v skutočnosti tam ten vedec *je* a pokusné osoby *sú* vysokoškóľáci a Progenitorivox nebol "podávaný" ale odovzdaný s pokynmi. Pasívny tvar zastiera skutočnosť.

Súdiac podľa komentárov, ktoré dostávam na Overcoming Bias, niekto bude namietat, že používanie pasívneho tvaru v článku časopisu je sotva hriechom – napokon, keď sa nad tým *zamyslíte*, uvedomíte si, že vedec je tam. Nevyzerá to ako logická chyba. A to je dôvod, prečo racionalisti potrebujú čítať Orwella, nielen Feynmana či Jaynesa.

Literatúra faktu sprostredkováva *poznanie*, fikcia sprostredkováva *zážitok*. Medicína dokáže extrapolovať, čo by sa stalo nechránenému človeku vo vzduchoprázdne. Fikcia dokáže, aby ste si to prežili.

Niektorí racionalisti sa pokúsia analyzovať zavádzajúcu vetu⁴, pokúsia sa vidieť, či by tam *mohlo byť* niečo zmysluplné, pokúsia sa *zostrojiť* logickú interpretáciu. Budú zhovievaví, v prípade pochybnosti sa priklonia na stranu autora. Na druhej

¹ http://www.lesswrong.com/lw/jb/applause_lights/flt

² <http://www.k-1.com/Orwell/index.cgi/work/essays/language.html>

³ http://www.lesswrong.com/lw/ja/we_dont_really_want_your_participation/

⁴ pozri: 16 Signály na potlesk

strane, spisovatelia sa cvičia v tom, aby k sebe *neboli* zhovievaví. Čokoľvek si obecenstvo *myslí*, že ste povedali, to ste *povedali*, či už ste to chceli povedať alebo nie; nemôžete sa hádať s obecenstvom bez ohľadu na to, aké chytré sú vaše zdôvodnenia.

Spisovateľ vie, že čitateľ sa *nezastaví*, aby chvíľu porozmýšľal. Zážitok z fikcie je nepretržitý prúd prvých dojmov. Spisovateľ-racionalista venuje pozornosť *zážitkom*, ktoré slová vytvárajú. Ak vyhodnocujete verejnú rozumnosť nejakej vety, ak zámerne analyzujete slová, preformulováate tvrdenia, skúšate rôzne významy, hľadáte zrnká pravdy, potom strácate zo zreteľa prvý dojem – čo obecenstvo *vidí* alebo skôr, čo *cíti*.

Románopisec by si všimol krikľavú nesprávnosť vety: “Pokusným osobám bol podávaný Progenitorivox”. Kde je tam niečo živé, čo môže čitateľ prežívať? Táto veta vytvára vzdialený pocit autoritatívnosti a to je *všetko* – *jediným* zážitkom je pocit, že vám niekto povedal niečo spoľahlivé. Románopisec by videl podstatné mená ako príliš abstraktné na to, aby ukázali, čo sa naozaj stalo – postdoktorand s fľaškou v ruke, snažiaci sa tváriť prísne; študent počúvajúci s nervóznym úškrnom.

Nechcem tým povedať, že by sa články do časopisov mali písať ako romány, ale že racionalista by si mal začať uvedomovať, aké *zážitky* ktoré slová vytvárajú. Racionalista musí chápať myseľ a ako ju ovládať. To zahŕňa aj prúd vedomia, tú časť vás, ktorá sa rozvíja v jazyku. Racionalista si musí začať uvedomovať skutočný, skúsenostný dopad⁵ viet, nielen ich púhu výrokovú sémantiku.

Alebo, stručnejšie povedané: *Význam neospravedľňuje dopad!*

Nezaujíma ma, aká racionálne interpretácie dokážete *zostrojiť* zo signálov na potlesk⁶ ako: “UI by mala byť vyvinutá demokratickým procesom”. To neospravedľňuje iracionálny dopad tejto vety, signalizovanie obecenstvu, aby zatlieskalo, nehovoriac o jej pochybnej nejasnosti.

Tu je Orwell, karhajúci *dopad* fráz, ich účinok na zážitok myslenia:

Keď človek sleduje nejakého unaveného kecala na pódiu, ako mechanicky opakuje známe frázy – *zverstvá, zločiny, železná čižma, krvavá tyrania, slobodní ľudia celého sveta, stojme bok po boku* – často má čudný pocit, že nesleduje živú ľudskú bytosť, ale akúsi bábkú... Rečník, ktorý používa tento druh frazeológie, sa do istej miery sám premenil na stroj. Z jeho hrtana vychádzajú príslušné zvuky, ale jeho mozog sa na tom nezúčastňuje, nie ako keby tie slová vyberal sám...

Predovšetkým treba nechať význam, nech si vyberie slovo, a nie naopak. Najhoršia vec, ktorú v próze môžete urobiť so slovami, je poddať sa im. Keď myslíte na konkrétny predmet, myslíte bez slov a potom, keď chcete popísať tú vec, ktorú ste si predstavovali, asi lovíte, dokiaľ nenájdete tie správne slová, ktoré vyzerajú primerane. Keď myslíte na niečo abstraktné, máte väčší sklon začať od slov a pokiaľ nevynaložíte vedomé úsilie, aby ste tomu zabránili, existujúci dialekt vtrhne a vykoná svoje dielo za vás, na úkor zahmlenia alebo priam zmenenia toho, čo ste mysleli. Je pravdepodobne lepšie odkladať používanie slov tak dlho, ako sa len dá, a čo najviac si vyjasniť svoj význam pomocou obrazov a pocitov.

⁵ pozri: 35 Sémantické stopky

⁶ pozri: 16 Signály na potlesk

Posledný odsek akoby napísal Peirce⁷. K Ceste vedie viac než jedna cesta.

Pôvodný článok s komentármi⁸

⁷

http://books.google.com/books?id=NK2dLn48zWIC&pg=PA338&lpg=PA338&ots=vDJyvFZotS&sig=5Z1ZG8_yfTFIUpClq_YSUgR749M#PPA338,M1

⁸ http://www.lesswrong.com/lw/jc/rationality_and_the_english_language/

Kapitola 56. Ľudské zlo a zahmlené myslenie

George Orwell¹ videl pád civilizovaného sveta do totalitárstva, premenu alebo úpadok jednej krajiny za druhou; čižma dupajúca na ľudskú tvár, naveky a pamätaj, že naveky. Vy ste sa narodili príliš neskoro na to, aby ste si pamätali² časy, keď rast totalitárstva vyzeral nezastaviteľný, keď jedna krajina za druhou padala do rúk tajnej polície, hromové klopanie o polnoci, zatiaľ čo profesori na univerzitách slobodného sveta oslavovali čistky v Sovietskom Zväze ako pokrok. Pripadá vám to rovnako cudzie ako fikcia³; je ťažké brať to vážne⁴. Pretože vo vašej vetve času berlínsky múr padol. A ak Orwellovo meno nie je vytesané do jedného z tých kameňov, malo by byť.

Orwell videl osud ľudského druhu a vynaložil krčovité úsilie, aby ho vytrhol z jeho cesty. Orwellovou zbraňou bolo jasné písanie. Orwell vedel, že zmätený jazyk je zmätené myslenie; vedel, že ľudské zlo a zmätené myslenie sa prepletajú ako konjugované vlákna DNA:

V našej dobe je politická reč a písanie prevažne obhajovaním neobhájiteľného. Veci ako pokračovanie britskej nadvlády v Indii, ruské čistky a deportácie, zhodenie atómových bomb na Japonsko sa síce dajú obhajovať, ale iba argumentmi, ktoré väčšine ľudí pripadajú príliš brutálne a ktoré nie sú v súlade s vyhlasovanými cieľmi politických strán. Preto sa politický jazyk musí skladať prevažne z eufemizmov, podsúvania odpovedí a čírej hmlistej nejasnosti. Bezbranné dediny sú bombardované zo vzduchu, ich obyvatelia sú vyhnaní do polí, dobytok postrieľaný samopalmi, chyže podpálené zápalnými strelami: toto sa nazýva *pacifikovanie*...

Orwell bol priamy ohľadom cieľa svojej jasnosti:

Ak zjednodušíte svoju angličtinu, oslobodíte sa od najhorších bláznovstiev pravovernosti. Nemôžete hovoriť žiadnym z jej nevyhnutných nárečí a keď urobíte hlúpu poznámku, jej hlúposť bude zrejmá, dokonca aj vám samotným.

Urobiť našu hlúposť zrejmou, dokonca aj nám samotným – toto je srdce blogu *Overcoming Bias*.

Zlo sa zakráda, skryté, po neosvetlených tieňoch mysle. My sa obzeráme s jasnosťou histórie a plačeme, keď spomíname⁵ na hladomory naplánované Stalinom a Maom, ktoré zabili desiatky miliónov⁶. Označujeme to ako zlo, pretože to bolo vykonané vedomým ľudským úmyslom spôsobiť bolesť a smrť nevinným ľudským bytostiam. Označujeme to ako zlo kvôli odporu, ktorý voči tomu cítime, keď sa obzeráme s jasnosťou histórie. Aby sa páchatelia zla vyhli tejto prirodzenej opozícii, tento odpor musí zostať skrytý. Jasnosti sa treba vyhnúť za každú cenu. Ako majú ľudia s jasným pohľadom sklón vzpierať sa zlu, ktoré vidia, tak má ľudské zlo, kdekoľvek existuje, v pláne zahmlieť myslenie.

1984 to ukazuje ostro: Orwellovi najväčší zločinci sú vystrihovači a premalovávači fotografií (založené na historickom vystrihovaní a fotografovaní v

¹ <http://www.k-1.com/Orwell/index.cgi/work/essays/language.html>

² pozri: 42 Sprístupniť históriu

³ pozri: 42 Sprístupniť históriu

⁴ pozri: 42 Sprístupniť históriu

⁵ pozri: 42 Sprístupniť históriu

⁶ pozri: 283 Necitlivosť k rozsahu

Sovietskom Zväze). A na vrchole všetkej temnoty na Ministerstve lásky, O'Brien mučí Winstona, aby pripustil, že dva plus dva sa rovná päť:

"Pamätáš sa," pokračoval, "ako si si napísal do denníka: 'Sloboda je slobodou povedať, že dva plus dva sa rovná štyri'?"

"Áno," povedal Winston.

O'Brien zdvihol ľavú ruku, chrbátom k Winstonovi, so skrytým palcom a štyrmi vystretými prstami.

"Koľko prstov ukazujem, Winston?"

"Štyri."

"A keď strana povie, že to nie sú štyri, ale päť – potom koľko?"

"Štyri."

Toto slovo skončilo zhĺknutím od bolesti. Ručička na stupnici vyskočila na štyridsaťpäť. Po celom Winstonovom tele vyrazil pot. Vzduch sa mu dral do pľúc a potom zase von v hlbokých stonoch, ktoré ani zaťatím zubov nedokázal zastaviť. O'Brien ho sledoval, štyri prsty stále vystreté. Potiahol páku naspäť. Tentokrát bolesť ustúpila iba mierne.

Som opakovane zdesený, ako napohľad rozumní ľudia – napríklad Robinov kolega Tylen Cowen⁷ – si nemyslia, že prekonávať skreslenia je dôležité. Toto je vaša *myseľ*, o čom hovoríme. Vaša ľudská inteligencia⁸. To, čo vás oddeľuje od opice. To, čo vybudovalo tento svet. Nemyslíte si, že je dôležité, ako *myseľ* funguje? Nemyslíte si, že systematické poruchy mysle sú dôležité? Myslíte si, že by Inkvizícia mučila bosorky, keby všetci boli ideálni bayesiáni?

Tyler Cowen si zrejme myslí, že prekonávanie skreslení je rovnakým skreslením ako samotné skreslenia: "Vnímam Robinov blog ako príklad skreslenia, a teda ako ukážku, že skreslenie môže byť veľmi užitočné." *Dúfam*, že je to len výsledok príliš abstraktného myslenia v snahe znieť chytro. Môže si Tyler vážne myslieť, že necitlivosť k rozsahu pri hodnote ľudského života⁹ je na rovnakej úrovni ako snaha vytvoriť plány, ktoré *naozaj* zachránia toľko životov, koľko sa len dá?

Orwell¹⁰ bol nútený bojovať proti podobnému postoju – že pripustiť hocikaký rozdiel je mladistvou naivnosťou:

Stuart Chase a ďalší už takmer tvrdia, že všetky abstraktné slová sú bezvýznamné a používajú to ako zámienku na presadzovanie istého druhu politického kvietizmu. Pokiaľ neviete, čo je to fašizmus, ako môžete bojovať proti fašizmu?

Možno prekonávanie skreslení nevyzerá dostatočne vzrušujúco, ak je podávané ako zápas s púhymi náhodnými chybami. Možno je ťažšie sa pre niečo zapáliť, ak proti nám nestojí nejaké jasné zlo. Spravme si teda absolútne jasno v tom, že kde je na tomto svete ľudské zlo, kde je krutosť a mučenie a úmyselné vraždenie, tam sú skreslenia, ktoré to zahaľujú. Kde ľudia s jasným pohľadom odporujú týmto skresleniam, zahalené zlo útočí naspäť. Pravda *má* svojich nepriateľov. Keby bol Overcoming Bias časopisom v bývalom Sovietskom Zväze, každý autor a diskutér na tomto blogu by bol poslaný do pracovného tábora.

⁷ <http://marginalrevolution.com/marginalrevolution/2007/08/how-important-i.html>

⁸ <http://intelligence.org/2007/07/10/the-power-of-intelligence/>

⁹ pozri: 283 Necitlivosť k rozsahu

¹⁰ <http://www.k-1.com/Orwell/index.cgi/work/essays/language.html>

V celej ľudskej histórii, každý krok dopredu bol poháňaný novou jasnosťou myslenia. Okrem niekoľkých prírodných katastrof, každá veľká pohroma bola poháňaná hlúposťou. Naším posledným nepriateľom sme my sami; toto je vojna a my sme vojaci.

Pôvodný článok s komentármi¹¹

¹¹ http://lesswrong.com/lw/jd/human_evil_and_muddled_thinking/

Časť VII. Špirály smrti a sektársky atraktor

Obsah

57 Afektívna heuristika	27
58 Vyhodnotiteľnosť (a lacné vianočné nákupy)	30
59 Neobmedzené škály, obrovské súdne odmeny a futurizmus	33
60 Efekt svätožiary	35
61 Skreslenie superhrdinu	37
62 Iba spasitelia	39
63 Afektívne špirály smrti	41
64 Odolajte šťastnej špirále smrti	43
65 Nekritická nadkritickosť	47
66 Ochladzovanie skupinových názorov vyparovaním	50
67 Keď sa nikto neodváža naliehať na zdržanlivosť	52
68 Pokus v Robbers Cave	54
69 Každá kauza chce byť sektou	56
70 Strážcovia pravdy	58
71 Strážcovia genofondu	61
72 Strážcovia Ayn Rand	62
73 Dva koany o sektách	66
74 Aschov pokus o konformitu	68
75 O vyjadrovaní svojich obáv	71
76 Osamelý nesúhlas	73
77 Sektárske antisektárstvo	76

Kapitola 57. Afektívna heuristika

Afektívna heuristika je keď subjektívne vnímanie niečoho ako dobrého alebo zlého funguje ako heuristika – zdroj rýchlych vnemových úsudkov. Príjemné a nepríjemné pocity sú v strede ľudského uvažovania a afektívna heuristika prichádza s krásnymi skresleniami – niektorými z mojich najobľúbenejších.

Začnime jedným z tých pomerne menej bláznivých skreslení. Idete sa sťahovať do nového mesta a musíte prepraviť starožitné kyvadlové hodiny. V prvom prípade sú tieto starožitné hodiny darom od vašich starých rodičov k vašim 5. narodeninám. V druhom prípade sú hodiny darom od vzdialeného príbuzného a nemáte k nim žiadne zvláštne pocity. Kolko by ste boli ochotní zaplatiť za poistku, ktorá vám vyplatí 100 dolárov ak sa tieto hodiny počas prevozu stratia? Podľa Hsee a Kunreuthera (2000) pokusné osoby uvádzali ochotu zaplatiť viac než dvojnásobnú sumu v prvom prípade. To môže znieť rozumne – prečo neplatiť viac za ochranu cennejšieho majetku? - kým si neuvedomíte, že poistka tie hodiny *nechráni*, iba vyplatí peniaze, ak sa hodiny stratia, a vyplatí rovnakú sumu v oboch prípadoch. (A áno, bolo uvedené, že poistka je uzavretá mimo dopravej firmy, takže nedáva sťahovákovi žiadnu zvláštnu motiváciu.)

Dobre, to však *neznie* príliš šialene. Možno by ste sa z toho dostali tvrdením, že pokusné osoby poistovali svoje citové výsledky, nie finančné výsledky – kupovali si útechu.

A čo toto? Yamagishi (1997) ukázal, že pokusné osoby hodnotili chorobu ako nebezpečnejšiu, keď bola opísaná ako zabíjajúca 1 286 ľudí z každých 10 000 nakazených, v porovnaní s chorobou, ktorá mala úmrtnosť 24,14 %. Zdá sa, že predstava tisíc mŕtvol je omnoho znepokojujúcejšia než jedna osoba, ktorá má väčšiu šancu prežiť než umrieť.

Však počkajte, bude to horšie.

Predstavte si, že sa letisko musí rozhodnúť, či použije peniaze na nákup nového zariadenia, kým kritici argumentujú, že by sa peniaze mali radšej použiť na iné stránky bezpečnosti letiska. Slovic a kol. (2002) predložili dvom skupinám pokusných osôb argumenty za a proti nákupu zariadenia, pričom reakciu merali na škále od 0 (vôbec by nepodporil) po 20 (veľmi silná podpora). Jedna skupina mala zariadenie opísané tak, že zachráni 150 životov. Druhá skupina mala zariadenie opísané tak, že zachráni 98 % zo 150 životov. Hypotéza za týmto experimentom znela, že 150 životov znie neurčito dobre – je to veľa? málo? - zatiaľ čo zachrániť 98 % niečoho je jasne veľmi dobré, pretože 98 % je veľmi blízko hornej hranici percentuálnej škály. A hľa, záchrana 150 životov mala priemernú podporu 10,4, kým záchrana 98 % zo 150 životov mala priemernú podporu 13,6.

Alebo si vezmite správu od Denes-Raja a Epsteina (1994): Pokusné osoby, ktoré dostali možnosť vyhrať 1 dolár vždy, keď si náhodne vytiahnu červený žele cukrík z misky, často dávali prednosť miske, kde bolo červených cukríkov viac ale v menšom pomere. Napríklad 7 zo 100 malo prednosť pred 1 z 10.

Podľa Denes-Raja a Epsteina tieto pokusné osoby neskôr uvádzali, že hoci vedeli, že pravdepodobnosti sú proti nim, cítili, že majú väčšiu šancu, keď je viac červených cukríkov. To môže pripadať bláznivé vám, ó Štatisticky Sofistikovaný Čitateľ, ale keď sa nad tým pozornejšie zamyslíte, uvedomíte si, že to dokonale dáva zmysel. Pravdepodobnosť 7 % oproti pravdepodobnosti 10 % môže byť zlá správa, ale je viac než vyvážená väčším počtom červených cukríkov. Je to horšia

pravdepodobnosť, to áno, ale aj tak máte väčšiu šancu vyhrať, jasné? Mali by ste nad touto myšlienkou meditovať, dokiaľ nedosiahnete osvietenie ohľadom toho, ako väčšina tejto planéty uvažuje o pravdepodobnosti.

Finucane a kol. (2000) testoval teóriu, že ľudia budú miešať svoje úsudky o konkrétnych dobrých alebo zlých stránkach niečoho s celkovým dobrým alebo zlým pocitom z danej veci. Napríklad informácie o možných rizikách alebo možných výnosoch jadrových elektrární. Logicky informácia o riziku nemá žiadnu súvislosť s informáciou o výnose. Ak je fyzický fakt o dizajne reaktora, že je pasívne bezpečný (nedostane sa do nadkritického stavu ani keď sa okolitý chladiaci systém pokazí), to ešte neznamená, že reaktor bude nevyhnutne vytvárať menej odpadu alebo bude produkovať elektrinu lacnejšie atď. Všetky tieto veci sú dobré, ale nie je to tá istá vec. Napriek tomu, Finucane a kol. zistil, že pre jadrové reaktory, zemný plyn a konzervačné látky v potravinách predloženie informácie o vyššom výnose spôsobilo, že ľudia vnímali riziko ako menšie; predloženie informácie o vyššom riziku spôsobilo, že ľudia vnímali výnos ako nižší; a tak ďalej cez všetky kvadranty.

Finucane a kol. ďalej zistili, že časový tlak výrazne zvyšoval nepriamo úmerný vzťah medzi vnímaným rizikom a vnímaným výnosom, čo je v súlade so všeobecným zistením, že časový tlak, slabé informácie alebo rozpyľovanie všetky zvyšujú dominanciu vnemových heuristik nad analytickou rozvahou.

Ganzach (2001) našiel ten istý efekt v oblasti financií. Podľa bežnej ekonomickej teórie by výnos a riziko mali korelovať *pozitívne* – inými slovami, ľudia si priplácajú za bezpečnejšie investície, čo znižuje ich výnos; akcie dávajú väčšie výnosy než dlhopisy, ale majú primerane vyššie riziko. Pri hodnotení akcií *známych* firiem mali úsudky analytikov ohľadom rizika a výnosov priamu koreláciu, ako sa bežne predpovedalo. Pri hodnotení akcií *neznámych* firiem však analytici mali sklon hodnotiť akcie akoby boli všeobecne dobré alebo všeobecne zlé – nízke riziko a vysoký výnos, alebo vysoké riziko a nízky výnos.

Na ďalšie čítanie odporúčam dobrú záverečnú kapitolu v Slovic a kol 2002: "Rozumní konatelia alebo rozumní hlupáci: Dôsledky afektívnej heuristiky pre behaviorálnu ekonómiu."¹

Denes-Raj, V., & Epstein, S. (1994). Konflikt medzi intuitívnym a rozumovým spracovaním: Keď sa ľudia správajú v rozpore so svojím úsudkom. *Journal of Personality and Social Psychology*, 66, 819-829.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). Afektívna heuristika pri posudzovaní rizík a výnosov.² *Journal of Behavioral Decision Making*, 13, 1-17.

Ganzach, Y. (2001). Hodnotenie rizika a výnosov finančných aktív. *Organizational Behavior and Human Decision Processes*, 83, 353-370.

Hsee, C. K. & Kunreuther, H. (2000). Účinok afektu pri poistných rozhodnutiach.³ *Journal of Risk and Uncertainty*, 20, 141-159.

Slovic, P., Finucane, M., Peters, E. and MacGregor, D. 2002. Rozumní konatelia alebo rozumní hlupáci: Dôsledky afektívnej heuristiky pre behaviorálnu ekonómiu.⁴ *Journal of Socio-Economics*, 31: 329-342.

Yamagishi, K. (1997). Kedy je úmrtnosť 12,86 % nebezpečnejšia než 24,14 %: Dôsledky pre komunikáciu rizika. *Applied Cognitive Psychology*, 11, 495-506.

¹ <http://heuristics.behaviouralfinance.net/affect/Slov02.pdf>

² http://www-abc.mpib-berlin.mpg.de/users/r20/finucane00_the_affect_heuristic.pdf

³ <http://faculty.chicagogsb.edu/christopher.hsee/vita/Papers/AffectEffectInInsurance.pdf>

⁴ <http://heuristics.behaviouralfinance.net/affect/Slov02.pdf>

Pôvodný článok s komentármi⁵

⁵ http://lesswrong.com/lw/lq/the_affect_heuristic/

Kapitola 58. Vyhodnotiteľnosť (a lacné vianočné nákupy)

Ako sa blíži *drahá* časť Helovínskomiculášovskovianočného¹ obdobia, väčšinu čitateľov musí trápiť otázka:

Drahý *Overcoming Bias*, existujú skreslenia, ktoré môžem využiť, aby som *vyzeral* štedro a nemusel *naozaj* minúť veľa peňazí?

Rád vám oznamujem, že odpoveď je áno! Podľa Hsee (1998) – v článku nazvanom “Menej je lepšie: Keď sú lacné možnosti hodnotené vyššie než drahšie možnosti” – ak niekomu kúpite šatku za 45 dolárov, budete pravdepodobne vnímaný ako štedrejší než keď mu kúpite kabát za 55 dolárov.

Toto je špeciálny prípad všeobecnejšieho javu. V predchádzajúcom pokuse, Hsee (1996), sa pokusných osôb pýtali, koľko by boli ochotné zaplatiť za hudobný slovník zo second handu:

- Slovník A, z roku 1993, obsahuje 10 000 hesiel, vyzerá ako nový.
- Slovník B, z roku 1993, obsahuje 20 000 hesiel, má roztrhnutú obálku, ale inak vyzerá ako nový.

Trik bol v tom, že niektoré pokusné osoby videli oba slovníky vedľa seba, kým iné pokusné osoby videli iba *jeden* slovník...

Osoby, ktoré videli iba *jednu* z týchto možností boli ochotné zaplatiť v priemere 24 dolárov za Slovník A a v priemere 20 dolárov za Slovník B. Osoby, ktoré videli *obe* možnosti vedľa seba, boli ochotné zaplatiť 27 dolárov za Slovník B a 19 dolárov za Slovník A.

Samozrejme, počet hesiel v slovníku je dôležitejší než či má roztrhnutú obálku, prinajmenšom ak ho niekedy plánujete na niečo použiť. Ale ak vám ukážu iba jeden slovník a ten má 20 000 hesiel, číslo 20 000 veľa nepovie. Je to málo? Veľa? Kto vie? *Nedá sa to vyhodnotiť*. Na druhej strane, roztrhnutá obálka – tá bije do očí. To má jasnú afektívnu hodnotu²: konkrétne, zlú.

Keď ich vidíme vedľa seba, počet hesiel sa zmení z *nevyhodnotiteľného* na *vyhodnotiteľný*, pretože môžeme porovnať dva porovnateľné údaje. A akonáhle sa počet hesiel stane porovnateľným, tento faktor prebije dôležitosť roztrhnutej obálky.

Podľa Slovic a kol. (2002): Chceli by ste radšej:

1. Šancu 29/36 vyhrať 2 doláre
2. Šancu 7/36 vyhrať 9 dolárov

Zatiaľ čo priemerné *ceny* (ekvivalentné hodnoty) odhadované pri týchto možnostiach boli 1,25 a 2,11 dolára, ich priemerné hodnotenie príťažlivosti bolo 13,2 a 7,5. Ceny aj hodnotenia príťažlivosti boli získané tak, že sa pokusným osobám povedalo, že sa náhodne vyberú dve z hodnotených stávk a oni dostanú tú s vyššou cenou alebo s vyšším hodnotením príťažlivosti. (Pokusné osoby mali motív hodnotiť stávky ako príťažlivejšie alebo cenu ako vyššiu než by v skutočnosti boli ochotné zaplatiť.)

¹ <http://www.overcomingbias.com/2007/11/merry-hallowmas.html>

² pozri: 57 Afektívna heuristika

Stávka hodná viac peňazí vyzerala menej príťažlivo, klasické prevrátenie preferencií. Výskumníci si mysleli, že dolárové výhry sú skôr v súlade s odhadovanou cenou, zatiaľ čo pravdepodobnosť výhry je skôr v súlade s príťažlivosťou. Takže (povedali si výskumníci), prečo neskúsiť urobiť výhru zo stávky emocionálne ešte vypuklejšiu – afektívne hodnotiteľnejšiu – príťažlivejšiu?

Ako to urobili? Pridaním malej straty do stávky. Pôvodná stávka mala šancu 7/36 vyhrať 9 dolárov. Nová stávka mala šancu 7/36 vyhrať 9 dolárov a 29/36 prehrať 5 centov. Pri starej stávke ste implicitne hodnotili príťažlivosť 9 dolárov. Nová stávka vás nabáda hodnotiť príťažlivosť vyhrania 9 dolárov v *porovnaní s* prehraním 5 centov.

“Výsledky,” hovorí Slovic a kol., “prekonali naše očakávania.” V novom experimente mala jednoduchá stávka so šancou 7/36 vyhrať 9 dolárov priemerné hodnotenie príťažlivosti 9,4, zatiaľ čo zložitejšia stávka, ktorá mala navyše šancu 29/36 prehrať 5 centov mala priemerné hodnotenie príťažlivosti 14,9.

Nasledujúci pokus testoval, či pokusné osoby dajú prednosť starej stávke pred istým ziskom 2 dolárov. Iba 33 % študentov dalo prednosť starej stávke. V druhej skupine, ktorá si mala vybrať medzi istými 2 dolármi a novou stávkou (s pridanou možnosťou prehry 5 centov), až 60,8 % dalo prednosť stávke. Napokon, 9 dolárov nie je príliš príťažlivá suma peňazí, ale 9 dolárov oproti 5 centom je *mimoriadne* príťažlivý pomer výhry k prehre.

Môžete urobiť stávku atraktívnejšou tým, že k nej čisto pridáte stratu! Nie je psychológia zábavná? To je dôvod prečo nikto, to naozaj oceňuje zázračnú³ zložitost ľudskej inteligencie, nechce dizajnovať UI, ktorá by bola ako človek.

Samozrejme, toto funguje iba ak pokusné osoby nevidia tie dve stávky vedľa seba.

Podobne, ktorej z nasledujúcich zmrzlín podľa vás pokusné osoby v Hsee (1998) dávali prednosť?

- obrázok -

Prirodzene, odpoveď závisela na tom, či pokusné osoby videli iba jednu zmrzlinu alebo obe vedľa seba. Pokusné osoby, ktoré videli jednu zmrzlinu, boli ochotné zaplatiť 1,66 dolára Predajcovi H a 2,26 dolára Predajcovi L. Pokusné osoby, ktoré videli obe zmrzliny, boli ochotné zaplatiť 1,85 dolára Predajcovi H a 1,56 dolára Predajcovi L.

Čo to znamená pre vaše vianočné nákupy? Že ak miniete 400 dolárov na 16GB iPod Touch, obdarovaný uvidí najdrahší MP3 prehrávač. Ak miniete 400 dolárov na Nintendo Wii, obdarovaný uvidí najlacnejšie herné zariadenie. Čo je lepšia hodnota za vaše peniaze? Ach, ale táto otázka dáva zmysel, iba keď vidíte obe vedľa seba. Vy na ne budete myslieť vedľa seba, kým budete nakupovať, ale obdarovaný uvidí iba to, čo dostane.

Ak máte pevne určené množstvo peňazí na minútie – a vašim cieľom je ukázať priateľstvo, nie naozaj *pomôcť* obdarovanému – lepšie je úmyselne nekupovať hodnotu. Rozhodnite sa, koľko miniete na urobenie dojmu na obdarovaného a potom nájdite tej najmenej hodnotný predmet, ktorý tolko stojí. Čím lacnejšia je daná *kategória* predmetov, tým drahšie bude *ten konkrétny* predmet vyzeráť, ak máte pevne určenú sumu na minútie. Čo si viac zapamätáte, tričko za 25 dolárov alebo sviečku za 25 dolárov?

To dáva celkom nový zmysel japonskému zvyku kupovať 50-dolárové melóny, však? Pozriete sa na to a potrasiete hlavou so slovami: “Čo to je s tými

³ pozri: 130 Zázrak evolúcie

Japoncami?" A predsa sú vďaka tomu vnímaní ako neuveriteľne štedrý, priam rozhadzovační, hoci minú iba 50 dolárov. Môžete minúť 200 dolárov na luxusnú večeru a nebudete vyzeráť tak bohato ako by ste mohli minúť 50 dolárov za melón. Keby tak bol zvyk kupovať špáratká za 25 dolárov alebo zrnká prachu za 10 dolárov, to by sa nám mohlo piecť minúť ešte menej.

PS: Ak tento trik naozaj použijete, chcem vedieť, čo ste kúpili.

Hsee, C. K. (1996). Hypotéza vyhodnotiteľnosti: Vysvetlenie prevrátenie preferencií medzi spoločným a samostatným hodnotením alternatív.⁴ *Organizational Behavior and Human Decision Processes*, 67, 242-257.

Hsee, C. K. (1998). Menej je lepšie: Keď sú lacné možnosti hodnotené vyššie než drahšie možnosti.⁵ *Journal of Behavioral Decision Making*, 11, 107-121.

Slovic, P., Finucane, M., Peters, E. and MacGregor, D. (2002.) Rozumní konatelia alebo rozumní hlupáci: Dôsledky afektívnej heuristiky pre behaviorálnu ekonómiu.⁶ *Journal of Socio-Economics*, 31: 329-342.

Pôvodný článok s komentármi⁷

⁴ <http://faculty.chicagogsb.edu/christopher.hsee/vita/Papers/EvaluabilityHypothesis.pdf>

⁵ <http://faculty.chicagogsb.edu/christopher.hsee/vita/Papers/LessIsBetter.pdf>

⁶ <http://heuristics.behaviouralfinance.net/affect/Slov02.pdf>

⁷ http://lesswrong.com/lw/lh/evaluability_and_cheap_holiday_shopping/

Kapitola 59. Neobmedzené škály, obrovské súdne odmeny a futurizmus

“Psychofyzika” napriek svojmu názvu je úctyhodná oblasť, ktorá spája fyzické účinky so zmyslovými účinkami. Ak vypustíte akustickú energiu do vzduchu – vydáte zvuk – *ako hlasno* to bude znieť človeku, ako funkcia akustickej energie? O koľko viac akustickej energie musíte napumpovať do vzduchu, aby to človeku znelo dvakrát hlasnejšie? Nie je to dvakrát viac; skôr asi osemkrát viac.

Akustická energia a fotóny sa merajú priamočiaro. Keď chcete zistiť, ako hlasno nejaký akustický podnet *znie*, ako jasne nejaký svetelný zdroj *vyzerá*, zvyčajne sa opýtate počúvajúceho alebo pozerajúceho. To sa dá pomocou obmedzenej škály od “veľmi tichý” po “veľmi hlasný” alebo “veľmi slabé” až “veľmi jasné”. Môžete použiť aj neobmedzenú škálu, kde nula je “vôbec nepočuť” alebo “vôbec nevidno”, ale odtiaľ možno zvyšovať bez ohraničenia. Keď používate neobmedzenú škálu, pozorovateľ zvyčajne dostane konštantný podnet, *modul*, ktorý dostane pevne dané hodnotenie. Napríklad zvuk, ktorému sa priradí číslo 10. Pozorovateľ potom môže označiť zvuk dvakrát hlasnejší ako modul napísaním 20.

A toto sa ukazuje ako celkom spoľahlivá technika. Ale čo sa stane, ak dáte pokusným osobám neobmedzenú škálu, ale žiaden modul? Od 0 do nekonečna a žiaden pomocný údaj s pevne danou hodnotou? Potom si vymyslia svoj vlastný modul, samozrejme. *Pomery* medzi podnetmi budú medzi pokusnými osobami naďalej spoľahlivo korelovať. Osoba A povie, že zvuk X má hlasitosť 10 a zvuk Y má hlasitosť 15. Ak osoba B povie, že zvuk X má hlasitosť 100, potom je dobrý odhad, že osoba B priradí zvuku Y hlasitosť okolo 150. Ak však neviete, čo používa osoba C ako svoj modul – svoju mierku – potom sa nedá odhadovať, čo osoba C povie na zvuk X. Možno to bude 1. Možno to bude 1000.

Ak pokusné osoby hodnotia *jediný* zvuk, na *neobmedzenej* škále, *bez* pevne daného štandardu na porovnanie, takmer *všetok* rozptyl je spôsobený svojvoľným výberom modulu, nie samotným zvukom.

“Hm,” pomyslíte si, “to znie dosť ako keď porota určuje výšku odškodného. Nečudo, že je tam taký rozptyl!” Zaujímavá analógia, ale ako by ste to dokázali pokusom?

Kahneman a kol., 1998 a 1999, predložili 867 pokusným osobám spôsobilým sedieť v porote opisy súdnych prípadov (napr. dieťa, ktorému sa zapálili šaty) a požiadal ich o jedno z nasledujúceho:

1. Ohodnotili nehoráznosť činov obvineného, na ohraničenej škále
2. Ohodnotili stupeň nakoľko by mal obvinený byť potrestaný, na ohraničenej škále, alebo
3. Určili výšku odškodného v dolároch

A hľa, zatiaľ čo pokusné osoby navzájom veľmi dobre korelovali v hodnoteniach nehoráznosti a hodnoteniach trestu, ich odškodné bolo od buka do buka. Napriek tomu *poradové hodnotenia* odškodného od danej osoby – poradové číslo od najnižšieho odškodného po navyššieho odškodné – dobre korelovali medzi osobami.

Ak chcete vedieť, akú časť rozptylu na škále "trestu" možno vysvetliť konkrétnym scenárom – konkrétnym súdnym prípadom, ktorý bol predložený viacerým pokusným osobám – potom odpoveď, ešte aj pre hrubé skóre, bola 0,49. Pre *poradia* dolárových odpovedí bolo predpovedané množstvo rozptylu 0,51. Pre *hrubé dolárové* údaje bol vysvetlený rozptyl iba 0,06!

Čo znamená: keby ste poznali predložený scenár – už spomínané dieťa, ktorému sa zapálili šaty – mohli by ste dobre odhadnúť hodnotenie trestu, dobre odhadnúť *poradie* odškodného v dolároch oproti iným prípadom, avšak samotný výška odškodného v dolároch by bola celkom nepredvídateľná.

Vziať medián dvanástich náhodne vybraných odpovedí tiež veľmi nepomohlo.

Určenie odškodného porotou teda nie je natoľko ekonomickým ocenením ako vyjadrením postoja – psychologickou mierou rozhorčenia vyjadrenou na neobmedzenej škále bez štandardného modulu.

Všímam si, že mnohé *futuristické predpovede* je podobne najlepšie vnímať ako vyjadrenia postoja. Vezmite si otázku: "Ako dlho potrvá, kým budeme mať UI na úrovni človeka?" Odpovede, ktoré som na toto videl, siahajú od buka do buka. Pri jednej pamätihodnej príležitosti mi bežný človek z oblasti UI poveda: "Päťsto rokov." (!!)

Samotný dôvod, prečo sa čas do UI *nedá dobre predpovedať*, je téma sa samostatnú dlhú diskusiu. Ale neznamená to, že ten človek, ktorý povedal "Päťsto rokov", videl túto odpoveď v budúcnosti. Ani to číslo nemohol dostať štandardnou chabou metódou pomocou Moorovho zákona. Čo teda toto číslo 500 *znamenalo*?

Podľa môjho odhadu to je, ako keby ste sa pýtali: "Na škále, kde nula znamená 'vôbec nie zložité', aký zložitý sa vám *zdá* problém UI?" Keby to bola obmedzená škála, každý príčetný respondent by označil odpoveď "extrémne ťažký" na pravom konci. Všetko vám *pripadá* extrémne ťažké, keď nevíete, ako sa to robí. Ale namiesto toho je tu neobmedzená škála bez štandardného modulu. Ľudia si teda vymyslia číslo, ktoré predstavuje "extrémne zložité", čo môže byť 50, 100 alebo dokonca 500. Potom pridajú na koniec slovo "rokov" a to je ich futuristická predpoveď.

"Aký ťažký sa vám zdá problém UI?" nie je jediná náhradná otázka. Iní odpovedajú, ako keby som sa pýtal: "Ako pozitívne sa cítite ohľadom UI?", kde nižšie čísla znamenajú pozitívnejšie pocity a tiež pridajú na koniec slovo "rokov". Ale ak tieto "časové odhady" predstavujú niečo iné než vyjadrenia postojov na neobmedzenej škále bez modulu, nedokázal som určiť, čo.

Kahneman, D., Schkade, D. A., and Sunstein, C. 1998. Spoločné rozhorčenie a náhodné odškodné: Psychológia odškodného.¹ *Journal of Risk and Uncertainty* 16, 49-86.

Kahneman, D., Ritov, I. and Schkade, D. A. 1999. Ekonomické preferencie alebo vyjadrenia postoja? Analýza dolárových reakcií na verejné témy.² *Journal of Risk and Uncertainty*, 19: 203-235.

Pôvodný článok s komentármi³

¹ http://ist-socrates.berkeley.edu/~maccoun/LP_KahnemanSchkadeSunstein1998.pdf

² <http://link.springer.com/article/10.1023%2FA%3A1007835629236>

³ http://lesswrong.com/lw/li/unbounded_scales_huge_jury_awards_futurism/

Kapitola 60. Efekt svätožiary

Afektívna heuristika¹ je, keď celkový pocit dobrého alebo zlého prispieva k mnohým ďalším úsudkom, či už je to logické alebo nie, či si to uvedomujete alebo nie. Pokusné osoby, ktorým sa povedalo o výnosoch jadrovej elektrárne majú sklon hodnotiť ju ako menej rizikovú; burzovní analytici hodnotiaci akcie neznámych firiem ich hodnotia ako všeobecne dobré alebo všeobecne zlé – nízke riziko a vysoké výnosy, alebo vysoké riziko a nízke výnosy – v rozpore s bežnou ekonomickou teóriou, ktorá hovorí, že riziko a výnosy by mali pozitívne korelovať.

Efekt svätožiary je prejav afektívnej heuristiky² v sociálnej psychológii. Robert Cialdini v knihe *Vplyv: Veda a prax* zhrňa:

Výskum ukázal, že dobre vyzerajúcim jednotlivcom automaticky pripisujeme také priaznivé čery ako talent, láskavosť, česťnosť a inteligencia (prehľadová štúdia týchto indícií vid' Eagly, Ashmore, Makhijani, & Longo, 1991). Navyše, tieto úsudky robíme bez vedomia, že fyzická príťažlivosť hrá v tomto procese rolu. Niektoré dôsledky tohto nevedomého predpokladu, že "dobro vyzerajúci rovná sa dobrý" ma desia. Napríklad štúdia kanadských federálnych volieb v roku 1974 zistila, že atraktívni kandidáti získali vyše dva-a-pol-krát viac hlasov ako neatraktívni kandidáti (Efran & Patterson, 1976). Napriek takýmto indíciám uprednostňovania pekných politikov, nasledujúci výskum ukázal, že voliči si neuvedomujú svoje skreslenie. V skutočnosti 73 percent oslovených kanadských voličov odmietlo najsilnejším možným spôsobom, že by ich hlasy boli ovplyvnené fyzickým výzorom; iba 14 percent vôbec pripustilo možnosť takéhoto vplyvu (Efran & Patterson, 1976). Voliči môžu popierať vplyv príťažlivosti na zvoliteľnosť koľko len chcú, ale indície naďalej potvrdzujú jej problematickú prítomnosť (Budesheim & DePaola, 1994).

Podobný účinok sa našiel pri prijímacích pohovoroch. Podľa jednej štúdie mala dobrá upravenosť uchádzačov v simulovanom pracovnom pohovore väčší vplyv na rozhodnutie zamestnať než pracovná kvalifikácia – napriek tomu, že osoby robiace rozhovor tvrdili, že výzor hral v ich rozhodovaní malú rolu (Mack & Rainey, 1990). Výhoda príťažlivých uchádzačov sa týkala nielen zamestnania, ale aj výplaty. Ekonómovia skúmajúci americké a kanadské vzorky zistili, že príťažliví zamestnanci dostávajú v priemere o 12 až 14 percent vyššie výplaty než ich nepríťažliví spolupracovníci (Hammermesh & Biddle, 1994).

Rovnako znepokojujúci výskum naznačuje, že náš súdny proces podobne podlieha vplyvu telesných rozmerov a štruktúry kostí. Vyzerá to, že dobre vyzerajúci ľudia majú väčšiu šancu, že sa k nim právny systém zachová veľmi výhodne (vid' Castellow, Wuensch, & Moore, 1991; a Downs & Lyons, 1990, prehľadové štúdie). Napríklad v štúdii z Pennsylvanie (Stewart, 1980) výskumníci hodnotili fyzickú príťažlivosť 74 mužov nezávisle obžalovaných na začiatku súdneho procesu. Keď si omnoho neskôr výskumníci pozreli súdne záznamy výsledkov z týchto prípadov, zistili, že atraktívni muži dostali výrazne ľahšie tresty. V skutočnosti mali príťažliví obžalovaní dvakrát väčšiu šancu vyhnúť sa väzeniu ako nepríťažliví obžalovaní. V inej štúdii – táto sa týkala

¹ pozri: 57 Afektívna heuristika

² pozri: 57 Afektívna heuristika

odškodného prideleného vo fingovanom súdnom procese ublíženia z nedbanlivosti – obvinený, ktorý vyzeral lepšie ako jeho obeť, bol odsúdený zaplatiť v priemere 5 623 dolárov; ale ak bola obeť príťažlivejšia, priemerné odškodné bolo 10 051 dolárov. Navyše, sklon uprednostňovať príťažlivejších vykazovali aj muži aj ženy v porote (Kulka & Kessler, 1978).

Iné pokusy ukázali, že príťažliví ľudia skôr získajú pomoc v núdzi (Benson, Karabenic, & Lerner, 1976) a sú presvedčivejší pri menení názorov obecnstva (Chaiken, 1979).

Tento vplyv príťažlivosti na hodnotenie inteligencie, čestnosti alebo láskavosti je jasným príkladom skreslenia – najmä keď tie druhé posudzujete na základe daného textu – pretože by sme nečakali, že sa naše úsudky o čestnosti a príťažlivosti budú spájať z nejakého legitímneho dôvodu. Na druhej strane, aká časť mojej vnímanej inteligencie je vďaka mojej čestnosti? Aká časť mojej vnímanej čestnosti je vďaka mojej inteligencii? Nájsť pravdu a povedať pravdu nie je vo svojej podstate až tak výrazne odlišné ako vyzerať pekne a vyzerať bystro...

Ale tieto štúdie efektu svätožiary u príťažlivosti by v nás mali vzbudiť podozrenie, že podobný efekt svätožiary môže byť aj u láskavosti alebo inteligencie. Povedzme, že poznáme niekoho, kto vyzerá nielen veľmi inteligentne, ale aj čestne, altruisticky, milo a pokojne. Mali by ste sa zamyslieť, či niektoré z týchto vnímaných vlastností nie sú ovplyvnené naším vnímaním iných. Možno je tá osoba naozaj inteligentná, čestná a altruistická, ale nie až taká milá³ alebo pokojná⁴. Mali by ste sa zamyslieť, ak sa ľudia, ktorých poznáte, delia príliš jasne na anjelov a diablov.

A – viem, že si nemyslíte, že to práve vy musíte urobiť, ale možno by ste mali – budte o trochu skeptickejší pri príťažlivých politických kandidátoch.

Cialdini, R. B. 2001. *Vplyv: Veda a prax*. Boston, MA: Allyn and Bacon.

Pôvodný článok s komentármi⁵

³ <http://ozyandmillie.org/2003/03/24/ozy-and-millie-1134/>

⁴ <http://ozyandmillie.org/2006/11/16/ozy-and-millie-1770/>

⁵ http://lesswrong.com/lw/lj/the_halo_effect/

Kapitola 61. Skreslenie superhrdinu

Predstavte si, že je tu ťažko ozbrojený sociopat, únosca s rukojemníkmi, ktorý práve odmietol všetky pokusy o vyjednávanie a oznámil svoj úmysel začať zabíjať rukojemníkov. V skutočnom živote kladní hrdinovia zvyčajne nevykopávajú dvere, ak má záporný hrdina rukojemníkov. Ale niekedy – *veľmi* zriedkavo, ale predsa – život napodobňuje Hollywood do tej miery, že skutoční kladní hrdinovia musia vylamovať dvere.

Predstavte si dve široko oddelené skutočnosti, dvoch hrdinov, ktorí vtrhnú do miestnosti, aby ako prví čelili lotrovi.

V jednej skutočnosti je hrdina taký silný, že dokáže hádzať autá, môže z nosa vyfukovať ohnivé strely, má rentgenový zrak a jeho koža nielenže *odráža* guľky, ale ich vyparí pri dotyku. Lotor sa opevnil v základnej škole a vzal si ako rukojemníkov vyše dvesto detí; ich rodičia čakajú vonku a plačú.

V druhej skutočnosti je hrdina policajt v New Yorku a rukojemníci sú tri prostitútky, ktoré lotor zobral z ulice.

Zvážte veľmi pozorne túto otázku: Kto je väčší hrdina? A kto s väčšou pravdepodobnosťou bude mať svoj vlastný komix?

Efekt svätožiary¹ je, že vnímanie všetkých kladných vlastností koreluje. Profily hodnotené vysoko na škále príťažlivosti sú zároveň hodnotené vyššie na škále talentu, láskavosti, úprimnosti a inteligencie.

A tak komixové postavy, ktoré vyzerajú silné a nezraniteľné, čo sú obe kladné vlastnosti, vyzerajú, že majú aj viac odvahy a hrdinstva. A predsa:

Aké ťažké to vôbec môže byť správať sa statočne a odvážne, keď ste viacmenej nezraniteľní?

—Empowered, 1. časť

Nepamätám sa, či som nasledujúce niekde čítal alebo či som to vymyslel sám: Zdá sa, že napríklad *sláva* sa sčítava so všetkými ostatnými osobnostnými črtami. Vezmime si Gándhího. Bol Gándhí *najväčším altruistom* 20. storočia, alebo iba *najznámejším* altruistom? Gándhí čelil policajtom s obuškami a vojakom s puškami. Lenže Gándhí bol celebrita a to ho chránilo. Čo zvyšní účastníci pochodu, ktorí čelili obuškom a puškám aj keď by neboli žiadne medzinárodné správy, keby skončili v nemocnici alebo zastrelení?

Čo si Gándhí myslel o tom, že dostal titulky, stal sa celebritou, získal slávu a miesto v dejinách, *stal sa archetypom* nenásilného odporu, keď riskoval menej než hocikto z tých, ktorí pochodovali s ním? Ako sa cítil, keď niektorý z týchto anonymných hrdinov prišiel k nemu s rozžiarеныmi očami a povedal Gándhímu, aký je úžasný. Videl vôbec Gándhí niekedy svoj svet z tohto uhla pohľadu? Neviem; nie som Gandhi.

Toto nie je v žiadnom zmysle kritika Gándhího. Pointa nenásilného odporu nie je predvádzať svoju odvahu. To sa dá urobiť omnoho ľahšie, keď sa spustíte z niagarského vodopádu v sude. Gándhí nemal ma výber ohľadom toho, že bol trochu, aj keď nie úplne, chránený ako celebrita. A Gándhího činy si vyžadovali odvahu – nie toľko odvahy ako pochodovať anonymne, ale aj tak dosť odvahy.

¹ pozri: 60 Efekt svätožiary

Skreslenie, na ktoré chcem ukázať je, že Gandhiho body za slávu sa v našom vnímaní *pripočítajú* k jeho oprávnené získaným bodom za altruizmus. Keď myslíte na nenásilie, myslíte na Gándhího – nie na anonymnú demonstrantku v jednom z Gándhího pochodov, ktorá čelila obuškom a puškám, dostala bitku, musela ísť do nemocnice a celý zvyšok svojho života krívala *a nikto si nikdy nespomenie na jej meno*.

Podobne, čo je viac – riskovať svoj život, aby ste zachránili dvesto detí alebo riskovať svoj život, aby ste zachránili troch dospelých?

Odpoveď závisí od toho, čo myslíme slovom *viac*. Ak si niekedy budete musieť *vybrať* medzi záchranou troch dospelých a záchranou dvesto detí, vyberte si tú druhú možnosť. “Kto zachráni jeden život, je to akoby zachránil celý svet²” môže byť pekný signál na potlesk³, ale je to hrozná morálna rada, ak si máte vybrať jedno alebo druhé. Ak teda myslíte “viac” v zmysle “Ktoré je dôležitejšie?” alebo “Ktoré je želanejší výsledok?” alebo “Ktoré by som si mal vybrať, keby som si musel vybrať jedno alebo druhé?” potom je viac zachrániť dvesto než troch.

Ale ak sa pýtate v zmysle, čo je väčšia prejavená cnosť, potom niekto, kto je ochotný riskovať svoj život, aby zachránil iba tri životy, prejavil viac odvahy než niekto, kto je ochotný riskovať svoj život, aby zachránil dvesto, ale nie tri.

To neznamená, že si môžete vedome vybrať ísť riskovať svoj život kvôli záchrane troch dospelých a nechať dvesto školákov osudu, pretože chcete prejsť viac cnosti. Niekto, kto riskuje svoj život, *pretože chce byť cnostný*, prejavil omnoho menej cnosti ako niekto, kto riskuje svoj život, *pretože chce zachrániť druhých*. Niekto, kto si vyberie zachrániť tri životy namiesto dvesto životov, pretože si myslí, že tým prejaví viac cnosti, je tak sebecky fascinovaný svojou vlastnou “veľkosťou”, akoby spáchal morálny ekvivalent zabitia.

Je to jeden z tých scenárov *wu wei*: Nemôžete prejsť cnosť snahou prejsť cnosť. Ak dostanete na výber medzi bezpečnou metódou ako zachrániť svet, ktorá nevyžaduje žiadne osobné obete ani nepohodlie a metódou, pri ktorej riskujete svoj život a musíte podstúpiť mnohé príkoria, nemôžete sa stať hrdinom tým, že si úmyselne vyberiete druhú možnosť. Nie je nič hrdinské na tom chcieť byť hrdinom. Je to stratený cieľ⁴.

Skutočne cnostní ľudia, ktorí sa naozaj snažia zachrániť životy, namiesto snahy prejsť cnosť, budú vytrvalo hľadať, ako zachrániť viac životov s menšou námahou, čo znamená, že sa prejaví menej ich cnosti. Môže to byť mätúce, ale nie je to rozpor.

Nemôžeme si však vždy vybrať, že budeme nezraniteľní guľkami. Potom, čo sme spravili, čo bolo v našich silách na zníženie rizika a zvýšenie rozsahu, všetko *zostávajúce* hrdinstvo sa dobre a pravdivo prejaví.

Policajt, ktorý nasadzuje svoj život v bojovej línii bez superschopností, bez rentgenového pohľadu, bez super sily, bez schopnosti lietať, a najmä bez odolnosti voči guľkám, prejavuje omnoho viac cnosti ako Superman – ktorý je *iba superhrdina*.

Pôvodný článok s komentármi⁵

² pozri: 284 Jeden život proti celému svetu

³ pozri: 16 Signály na potlesk

⁴ pozri: 149 Stratené ciele

⁵ http://lesswrong.com/lw/lk/superhero_bias/

Kapitola 62. Iba spasitelia

Včera som diskutoval, ako efekt svätožiary¹, ktorý spôsobuje, že ľudia vidia všetky kladné vlastnosti ako korelované – napríklad príťažlivejší jednotlivci sú zároveň vnímaní ako láskavejší, úprimnejší a inteligentnejší – spôsobuje, že viac obdivujeme hrdinov, ak sú super silní a odolní voči gulkám. Hoci logicky si vyžaduje omnoho viac odvahy byť hrdinom, ak *nie* ste odolní voči gulkám. Ďalej, že prejavuje viac cnosti konať odvážne, aby ste zachránili jeden život, než aby ste zachránili svet. (Hoci ak máte na výber jedno alebo druhé, samozrejme by ste mali zachrániť svet².)

“Policajt, ktorý nasadí svoj život v bojovej línii bez superschopností,” povedal som, “prejaví omnoho viac cnosti než Superman, ktorý je *iba superhrdina*.”

Ale budme konkrétnejší.

John Perry³ bol policajt v New Yorku, ktorý bol zároveň Extropián a transhumanista, kde som ho spoznal po mene. John Perry mal o chvíľu ísť do dôchodku a založiť si právnu prax, keď sa dozvedel, že lietadlo narazilo do World Trade Center. Zomrel pri zrútení severnej veže. Nepoznal som Johna Perryho osobne, takže toto nemôžem potvrdiť z prvej ruky; ale veľmi málo Extropiánov verilo v Boha a predpokladám, že aj John Perry bol ateista.

Čo znamená, že Perry vedel, že riskuje svoju samotnú existenciu, každý týždeň v práci. Nie je to tak, že by ako väčšina ľudí v dejinách vedel, že má na výber iba ako zomrie, a tak si vybral zmysluplnú možnosť – Perry bol totiž transhumanista; mal skutočnú nádej. A Perry tam šiel a napriek tomu riskoval svoj život. Nie preto, že by čakal nejakú nadprirodzenú odmenu. Nie preto, že by očakával vôbec niečo po svojej prípadnej smrti. Ale pretože tam boli v nebezpečenstve iní ľudia a oni tiež nemali nehmotné duše, a jeho nádej na život nemala väčšiu cenu než ich.

Nepoznal som Johna Perryho. Neviem, či vnímal svet týmto spôsobom. Ale fakt, že ateista a transhumanista môže stále byť policajtom, môže vbehnúť do čakárne v horiacej budove, hovorí o ľudskom duchu viac než všetci mučeníci, ktorí kedy dúfali v nebo.

Takže toto je jeden konkrétny policajt...

...a teraz ten superhrdina.

Ako ten príbeh rozprávajú kresťania, Ježiš Kristus vedel chodiť po vode, tíšiť búrky, vyháňať slovom démonov. Musel to byť pohodlný život: Hrozí hlad? Nakopírujem si trochu chleba. Nepáči sa mi strom? Preklajem ho. Rimania robia problém? Pošlem na nich Otca. Nakoniec tento čarovný život skončil a Ježiš sa dobrovoľne ponúkol na ukrižovanie. Byť priklinovaný na kríž nie je pohodlný spôsob smrti. Ale ako ten príbeh rozprávajú kresťania, Ježiš to urobil s vedomím, že o tri dni neskôr ožije a potom pôjde do Neba. Čo bola tá hrozba, ktorá pohla Ježiša čeliť tomuto dočasnému utrpeniu nasledovanému večnosťou v Nebi? Bol to život jedného človeka? Bola to skorumpovanosť cirkvi v Judei alebo útlak Ríma? Nie: ako ten príbeh rozprávajú kresťania, v ohrození bol večný osud každého človeka, kým sa Ježiš nechal dočasne priklinovať na kríž.

Ja nechcem až takto odsudzovať človeka, ktorý nie je naozaj taký vinný. Čo ak Ježiš – nie, vyslovujme jeho meno korektne: Ješu – čo ak Ješu z Nazaretu nikdy nechodil po vode a *predsa* vzdoroval cirkvi v Judei podporovanej mocou Ríma?

¹ pozri: 60 Efekt svätožiary

² pozri: 284 Jeden život proti celému svetu

³ <http://www.nleomf.org/911heroes/Perry.html>

Nezaslúžilo by si toto viac úcty než na akú si môže nárokovať Ježiš Kristus, ktorý bol iba spasiteľ?

Žiaľ, akosi sa zdá väčším hrdinom ten, kto má oceľovú kožu a božskú moc. Akosi sa zdá, že prejavuje viac cnosti zomrieť dočasne pre záchranu celého sveta než zomrieť natrvalo v konflikte so skorumpovanou cirkvou. Zdá sa to také *všedné*, akoby mnoho ďalších ľudí v dejinách urobilo to isté.

Pohodlne usadení dvetisíc rokov v budúcnosti môžeme Jeišua zahrnúť všetkou možnou kritikou, ale Jeišu robil to, o čom veril, že je správne, postavil sa proti cirkvi, o ktorej veril, že je skazená, a zomrel za to. Bez výhody spätného pohľadu⁴ od neho sotva možno očakávať, že by predpovedal skutočný dopad svojho života na svet. V porovnaní s inými prorokmi svojej doby bol pravdepodobne pomerne úprimnejší, pomerne menej násilný, a pomerne odvážnejší. Ak neberieme do úvahy neplánované dôsledky, najhoršia vec, ktorú možno o Jeišuovi povedať je, že iní to v dejinách urobili lepšie. (Napadajú mi Epikuros, Buddha, Markus Aurelius.) Jeišu zomrel navždy a - z jedného pohľadu - to urobil kvôli úprimnosti. Pätnásť storočí pred vedou náboženská úprimnosť nebola oxymoron.

Ako povedal Sam Harris:

Nestačí, že Ježiš bol človek, ktorý sa premenil natolko, že Kázeň na hore mohla byť spoveďou jeho srdca. Musel byť ešte aj Synom Boha, narodený z panny a predurčený vrátiť na zem oblaky slávy. Dôsledkom takejto dogmy je umiestnenie Ježišovho príkladu navždy mimo náš dosah. Jeho učenie prestane byť množinou empirických tvrdení o vzťahu medzi etikou a duchovným vhlľadom a namiesto toho sa stane samoučelnou a trochu morbidnou rozprávkou. Podľa kresťanských dogiem stať sa ako Ježiš je nemožné. Človek môže iba vymenovávať svoje hriechy, veriť v neuveriteľné a čakať na koniec sveta.

Vážne pochybujem, že Jeišu niekedy hovoril Kázeň na hore. Napriek tomu si Jeišu zaslúži úctu. Zaslúži si viac úcty než by mu kresťania ponechali.

Ale keďže Jeišu pravdepodobne očakával⁵, že jeho duša prežije, nezaslúži si viac úcty než John Perry.

Pôvodný článok s komentármi⁶

⁴ http://lesswrong.com/lw/il/hindsight_bias/

⁵ pozri: 13 Viera vo vieru

⁶ http://lesswrong.com/lw/ll/mere_messiahs/

Kapitola 63. Afektívne špirály smrti

Súvisí s kapitolami: Afektívna heuristika¹, Efekt svätožiary²

Mnoho³, mnoho⁴, mnoho⁵ je chýb v ľudskom uvažovaní, ktoré nás vedú preceňovať ako dobre naša obľúbená teória vysvetľuje fakty. Chemická teória flogistonu mohla vysvetliť takmer všetko, pokiaľ nemusela niečo predpovedať vopred. A na vysvetlenie čím väčšieho množstva javov ste použili svoju obľúbenú teóriu, tým pravdivejšia vaša obľúbená teória vyzerá – nebola vari potvrdená tými mnohými pozorovaniami? Čím pravdivejšia vyzerá táto teória, tým skôr budete spochybňovať indície, ktoré jej protirečia. Čím všeobecnejšia sa zdá táto obľúbená teória, na tým viac vysvetlení ju budete používať.

Ak poznáte niekoho, kto verí, že Belgicko tajne ovláda americký bankový systém alebo že môžu použiť neviditeľnú modrú duchovnú silu na nájdenie voľného parkovacieho miesta, takto nejako to asi začalo.

(Zostaňte v strehu a uvidíte tolko vecí, ktoré budú vyzeráť, že potvrdzujú túto teóriu...)

Tento cyklus pozitívnej spätnej väzby dôverčivosti a potvrdenia je naozaj obávaný a zodpovedný za mnoho chýb, vo vede aj v bežnom živote.

Ale to nie je nič v porovnaní so špirálou smrti, ktorá začne nábojom pozitívneho citu – myšlienkou, ktorá vyzerá *naozaj dobre*.

Nový politický systém, ktorý môže zachrániť svet. Veľký vodca, silný a vznešený a múdry. Zázračný nápoj, ktorý vylieči boľavé brucho aj rakovinu.

Sakra, prečo nie všetky tri? Veľká úloha si vyžaduje veľkého vodcu. Veľký vodca by mal dokázať namiešať pár čarovných nápojov.

Efekt svätožiary⁶ je, že každá vnímaná kladná vlastnosť (napríklad príťažlivosť alebo sila) zvyšuje vnem ľubovôlej inej kladnej vlastnosti (napríklad inteligencie alebo odvahy). Dokonca aj keď to nedáva zmysel, alebo je to proti zmyslu⁷.

Kladné vlastnosti zvyšujú vnímanie každej ďalšej kladnej vlastnosti? To znie podobne ako keď rozpadajúci sa atóm uránu vystrelí neutróny, ktoré rozbijú ďalšie atómy uránu.

Slabý kladný cit je podkritický; nezačne sa vymykať spod kontroly. Príťažlivá osoba vyzerá úprimnejšia, čo ju povedzme robí ešte trochu atraktívnejšou; ale účinný násobiteľ neutrónov je menší ako 1. Metaforicky povedané. Rezonancia trochu zamieša veci ale potom odumrie.

Keď sa intenzívny kladný cit pripojí k Veľkej Veci, rezonancia zasiahne všetko. Veriaci komunista vidí Marxovu múdrosť v každom hamburgeri kúpenom v McDonalde; v každom povýšení, ktoré im zamietli a ktoré by dostali v skutočnom robotníckom raji; v každých voľbách, ktoré nedopadnú podľa ich vkusu, v každom “nesprávne orientovanom” novinovom článku. Vždy keď použije Veľkú

¹ pozri: 57 Afektívna heuristika

² pozri: 60 Efekt svätožiary

³ http://lesswrong.com/lw/il/hindsight_bias/

⁴ pozri: 34 Falošná kauzalita

⁵ pozri: 34 Falošná kauzalita

⁶ pozri: 60 Efekt svätožiary

⁷ pozri: 61 Skreslenie superhrdinu

Myšlienku na interpretovanie ďalšej udalosti, Veľká Myšlienka vyzerá byť ešte potvrdenejšia. Je to dobrý pocit – pozitívne posilnenie – a samozrejme, keď máme z niečoho dobrý pocit, žiaľ, o to viac tomu *chceme* veriť.

Keď máte z Veľkej Veci taký dobrý pocit, že *idete hľadať* ďalšie príležitosti ako mať z Veľkej Veci ešte lepší pocit, používate ju na vysvetlenie nových udalostí každý deň, rezonancia kladného citu je ako miestnosť plná pascí na myši, do ktorej nasypete pingpongové loptičky.⁸

Môžete to nazvať “šťastný atraktor”, “príliš pozitívna spätná väzba”, “uzavretá slučka chvály” alebo “zábavné noviny”. Ja sobne dávam prednosť slovnému spojeniu “afektívna špirála smrti”.

Nabudúce: Ako odolať afektívnej špirále smrti. (Nápoveda: Nie tak, že odmietnete kedykoľvek čokoľvek obdivovať, ani že budete držať obdivované veci v malých bezpečných ohraničených magistériách.)

Pôvodný článok s komentármi⁹

⁸ http://www.youtube.com/watch?v=ORqc1x3_Evg

⁹ http://lesswrong.com/lw/lm/affective_death_spirals/

Kapitola 64. Odolajte šťastnej špirále smrti

Kde bolo, tam bolo, bol jeden človek, ktorý bol presvedčený, že objavil Velkú Myšlienku. Skutočne, ako ten človek rozmýšľal o tejto Velkej Myšlienke viac a viac, uvedomoval si, že je to nielen *nejaká* veľká myšlienka, ale tá *najúžasnejšia myšlienka všetkých čias*. Táto Velká Myšlienka odhalí tajomstvá vesmíru, nahradí autoritu skorumpovaného systému plného chýb, dá svojim nositeľom takmer magickú moc, nasýti hladných, uzdraví chorých, urobí z celého sveta lepšieho miesta, atď. atď. atď.

Ten človek bol Francis Bacon a jeho Velká Myšlienka bola vedecká metóda a bol to jediný fanatik v dejinách, ktorý si pripisoval taký stupeň úžitku pre ľudstvo a ukázalo sa, že má úplnú pravdu.

(Bacon samozrejme nevynašiel vedu sám, ale prispel a možno bol prvý, kto si uvedomil jej moc.)

Toto je problém rozhodnutia, že nikdy nebudete nič tak veľmi obdivovať: Niektoré myšlienky naozaj sú také dobré. Aj keď nikto *nesplnil* sľuby odvážnejšie než tie Baconove; prinajmenšom zatiaľ.

Ako však potom môžeme odolať šťastnej špirále smrti¹ voči Vede samotnej? Šťastná špirála smrti² začína vtedy, keď veríte, že niečo je *také* úžasné, že vás efekt svätožiari³ vedie k hľadaniu *ďalších* a *ďalších* pekných vecí, ktoré o tom môžete povedať, vďaka čomu to vidíte ako *ešte* úžasnejšie a tak ďalej, po špirále smerom do priepasti. Čo ak je Veda *naozaj* taká úžasná, že nemôžeme pripustiť jej skutočnú slávu a zachovať si pritom pričetnosť? To znie ako pekná vec, nie? *Ach nie, už to začína, utekajteee...*

Ak vyberiete štandardnú⁴ uloženú⁵ hlbokú múdrosť⁶ na *nepreháňanie obdivu vedy*, nájdete myšlienky ako: "Veda nám dala klimatizáciu, ale aj vodíkovú bombu" alebo "Veda nám môže povedať o hviezdach a biológii, ale nevie dokázať ani vyvrátiť⁷ draka v mojej garáži⁸". Lenže ľudia, od ktorých *pochádzajú* takéto myšlienky sa *nepokúšali* odolať šťastnej špirále smrti. Neznepokojovali sa, že sa ich vlastný obdiv vedy vymkne spod kontroly. Pravdepodobne sa im nepáčilo niečo, čo veda povedala o ich obľúbenom názore a hľadali, ako podkopať jej autoritu.

Štandardné negatívne myšlienky, ktoré sa hovoria o vede, pravdepodobne nebudú pôsobiť na niekoho, kto naozaj cíti nadšenie z vedy – to nie je cieľová skupina. Musíme teda namiesto toho hľadať iné zlé veci, ktoré možno povedať.

Ak však selektívne hľadáte niečo zlé, čo môžete povedať o vede – hoci aj v snahe odolať šťastnej špirále smrti – neusvedčujete automaticky sami seba z racionalizovania⁹? Prečo by ste mali venovať pozornosť svojim myšlienkach, keď viete, že sa snažíte sami seba manipulovať¹⁰?

¹ pozri: 63 Afektívne špirály smrti

² pozri: 63 Afektívne špirály smrti

³ pozri: 60 Efekt svätožiari

⁴ pozri: 82 Koláže "mimo vychodených kolají"

⁵ pozri: 81 Uložené myšlienky

⁶ pozri: 86 Ako vyzerá (a byť) hlboký

⁷ pozri: 19 Nárok náboženstva na nevyvrátenosť

⁸ pozri: 13 Viera vo vieru

⁹ pozri: 95 Racionalizovanie

¹⁰ pozri: 105 Doublethink (dobrovoľné skreslenie)

Vo všeobecnosti som skeptický voči ľuďom, ktorí tvrdia, že jedno skreslenie možno použiť na vyváženie druhého. Znie mi to ako automechanik, ktorý povie, že na pravom stierači máte pokazený motor, ale namiesto opravy vám jednoducho zlomí ľavý stierač, aby sa tým veci vyrovnali. Takýto druh chytrosti vedie k strielaniu vlastných gólov. Nech je riešením čokoľvek, mali by sme v rámci neho veriť pravdivým veciam, nie veriť, že veríme veciam, o ktorých veríme, že sú nepravdivé.

Dokážete zabrániť šťastnej špirále smrti tým, že svoj obdiv vedy obmedzíte na úzku oblasť? Časť šťastnej špirály smrti je vidieť Velkú Myšlienku všade – myslieť na to, ako by komunizmus vyliečil rakovinu, keby len dostal príležitosť. Pravdepodobne najspôhlivejším znakom vodcu sekty je, že predstiera odbornosť nie v jednej oblasti, ani v skupine súvisiacich oblastí, ale vo *všetkom*. Vodca vie, čo by členovia sekty mali jesť, čo si obliekať, čím sa živiť; s kým by mali mať sex; na aké umenie by sa mali pozerieť; akú hudbu by mali počúvať...

Nanešťastie pre tento plán, väčšina ľudí biedne zlyháva, keď sa snažia opísať tú peknú malú krabičku, vo vnútri ktorej by veda mala zostať. Zvyčajný trik: "Pozri, veda nedokáže vyliečiť rakovinu" tu nezaberie. "Veda nedokáže nič povedať o láske rodiča k dieťaťu" - pardon, to je jednoducho nepravda¹¹. Ak sa pokúsíte odseknúť vedu napríklad od rodičovskej lásky, tým nepopierate iba kognitívnu vedu a evolučnú psychológiu. Popierate aj, že Martine Rothblatt založila United Therapeutics s cieľom nájsť liek na pľúcnu hypertenziu svojej dcéry. (Dodávam, že úspešne.) Veda je legitímne spojená, jedným či druhým spôsobom, s takmer každou dôležitou stránkou ľudskej existencie.

No dobre, tak čo je príkladom *nepravdivého* pekného tvrdenia, ktoré nemôžete povedať o vede?

Podľa môjho skromného názoru jedno nepravdivé tvrdenie je, že veda je taká úžasná, že vedci by sa nemali ani len snažiť prevziať etickú zodpovednosť za svoju prácu¹², pretože to automaticky skončí dobre. Toto tvrdenie podľa mňa nerozumie podstate procesu, ktorým veda pomáha ľudstvu. Vedci sú ľudia, majú prosociálny záujem tak ako väčšina ľudí, a toto je prinajmenšom *časť* toho, prečo veda v konečnom dôsledku robí viac dobra než zla.

Ale tento bod očividne nie je mimo debaty. Tu je teda jednoduchšie nepravdivé pekné tvrdenie: "Pacienta s rakovinou možno uzdraviť púhym publikovaním dostatočného množstva článkov v odborných časopisoch." Alebo: "Sociopati sa môžu stať celkom normálni, ak si dajú záväzok nikdy neveriť ničomu bez indície replikovaných experimentov s $p < 0,05$."

Viere v takéto tvrdenia sa nevyhýbame tým, že si stanovíme citovú hranicu a rozhodneme sa, že veda je iba trochu pekná. Ani hľadaním dôvodov veriť, že uverejňovanie odborných článkov *spôsobuje* rakovinu. Ani vierou, že veda nedokáže o rakovine povedať nič dobré ani zlé.

Namiesto toho, ak viete dostatočne podrobne¹³ ako veda funguje, potom viete, že hoci je možné, aby "veda vyliečila rakovinu", pacient s rakovinou píšuci články do odborných časopisov nezažije zázračnú remisiu. Táto *konkrétna* reťaz príčin a následkov nebude fungovať.

Šťastná špirála smrti je emocionálnym problémom iba kvôli problému vnímania, efektu svätožiary¹⁴, ktorý nás robí ochotnejšími prijať ďalšie pozitívne tvrdenia, keď už sme prijali úvodné pozitívne tvrdenie. Tohto efektu sa nezabavíme

¹¹ pozri: 137 Evolučná psychológia

¹² <http://intelligence.org/2007/10/21/should-ethicists-be-inside-or-outside-a-profession/>

¹³ pozri: 26 Cnosť presnosti

¹⁴ pozri: 60 Efekt svätožiary

tým, že si to jednoducho budeme želať; pravdepodobne nás vždy bude trochu ovplyvňovať. Ale dokážeme spomaliť, zastaviť sa, zvážiť každé dodatočné pekné tvrdenie ako dodatočnú prirážajúcu podrobnosť¹⁵, a zamerať sa na konkrétne časti tvrdenia bez ohľadu na jeho pozitívnosť.

Čo ak konkrétne pekné tvrdenie “nemožno vyvrátiť”, ale existujú argumenty “za aj proti” tomuto tvrdeniu? V skutočnosti si na tieto slová treba dávať pozor vo všeobecnosti, pretože toto ľudia často hovoria, keď si precvičujú indície¹⁶ alebo sa vyhýbajú naozaj slabým miestam¹⁷. Pri nebezpečenstve šťastných špirál smrti dáva zmysel skúsiť sa vyhýbať radosti z *neuzavretých* tvrdení – aby sa nestali zdrojom ďalších ešte pozitívnejších citov voči niečomu, čo už máte radi.

Šťastná špirála smrti je *veľký* emocionálny problém iba kvôli príliš pozitívnej spätnej väzbe, kvôli schopnosti procesu prekročiť kritické množstvo. Možno nebudete vedieť úplne odstrániť efekt svätožiary, ale môžete použiť dosť kritického myslenia, aby ste svätožiaru udržali pod kritickým množstvom – aby ste zabezpečili, že rezonancia vyhasne namiesto vybúchnutia.

Mohli by ste priam povedať, že celý problém začína, keď sa ľuďom nechce kriticky skúmať každú prirážajúcu podrobnosť navyše¹⁸ – vyžadovať dostatočnú¹⁹ indíciu na vyváženie zložitosti²⁰, hľadať²¹ chyby, nielen podporu, používať zvedavosť²² – akonáhle prijali nejaký základný predpoklad. Bez klamu konjunkcie²³ by stále mohol byť efekt svätožiary²⁴, ale nebola by šťastná špirála smrti²⁵.

Ešte aj pri tých najkrajších Pekných Veciach v známom vesmíre by dokonaly racionalista, ktorý by presne vyžadoval potrebné indície na každé ďalšie (pozitívne) tvrdenie, nezažil afektívnu rezonanciu. Vy toto nedokázate, ale môžete sa držať dosť blízko rozumnosti, aby sa vaše šťastie nevymklo spod kontroly.

Naozaj nebezpečné prípady sú tie, kde *hocijaká kritika nejakého pozitívneho tvrdenia o Veľkej Veci vyzerá ako zlá alebo spoločensky neprijateľná*. Argumenty sú vojaci, každé pozitívne tvrdenie je vojak na našej strane, bodat vlastných vojakov do chrbta je vlastizrada.²⁶ Potom sa reťazová reakcia stane *nadkritickou*. Viac o tomto zajtra.

Dodatok: Stuart Armstrong ponúka súvisiacu radu²⁷:

Rozdeľ svoju Veľkú Vec na menšie nezávislé myšlienky a *ber ich ako nezávislé*.

Napríklad marxista by mohol rozdeliť Marxovu Veľkú Vec na teóriu hodnoty práce, teóriu politických vzťahov medzi triedami, teóriu mzdy, teóriu konečného politického stavu ľudstva. Potom by mal každé z toho vyhodnotiť nezávisle a ich pravdivosť alebo nepravdivosť by nemala presakovať do zvyšných. Ak dokážeme urobiť toto, mali by sme byť v bezpečí pred špirálou, keďže každá teória je príliš úzka na to, aby začala svoju vlastnú špirálu.

¹⁵ pozri: 6 Prirážajúce podrobnosti

¹⁶ pozri: 92 Jeden argument proti armáde

¹⁷ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

¹⁸ pozri: 6 Prirážajúce podrobnosti

¹⁹ pozri: 22 Koľko indície treba?

²⁰ pozri: 23 Occamova britva

²¹ pozri: 94 Filtrovaná indícia

²² pozri: 123 Meditácia o zvedavosti

²³ http://lesswrong.com/lw/ji/conjunction_fallacy/

²⁴ pozri: 60 Efekt svätožiary

²⁵ pozri: 63 Afektívne špirály smrti

²⁶ pozri: 47 Politika zabíja myslenie

²⁷ http://lesswrong.com/lw/lm/affective_death_spirals/gp5

Metaforicky je to ako držať podkritické množstvá plutónia oddelene od seba. Tri Veľké Myšlienky vás omnoho ťažšie doženú do šialenstva než jedna Veľká Myšlienka. Armstrongova rada zároveň propaguje konkrétnosť. Akonáhle niekto povie: "Publikovanie dostatočného množstva odborných článkov môže vyliečiť vašu rakovinu", opýtajte sa: "Je toto dôsledkom experimentálnej metódy, a ak áno, v ktorom kroku experimentálneho procesu sa tá rakovina vylieči? Alebo je to dôsledok vedy ako spoločenského procesu, a ak áno, závisí to od toho, či jednotlivý vedec chce vyliečiť rakovinu, alebo môže jednať v sebeckom záujme?" Dúfajme, že toto povedie preč od dobrých či zlých pocitov, smerom k uvedomeniu si zmätenia a nedostatku podpory.

Dodatok 2: Aby som to zhrnul, šťastnej špirále smrti sa *vyhnete* (1) rozdelením Veľkej Myšlienky na časti (2) považovaným každého dodatočného detailu za priťažujúci (3) rozmýšľaním o konkrétnych častiach kauzálnej reťaze namiesto o dobrých či zlých pocitoch (4) neprecvičovaním si indícií (5) nepridávaním šťastia z tvrdení, o ktorých: "nemôžete *dokázať*, že sú nesprávne"; ale *nie* (6) odmietaním čokoľvek príliš obdivovať (7) podniknutím skresleného hľadania zlých stránok, dokiaľ sa opäť nezačnete cítiť nešťastne (8) násilným strčením myšlienky do bezpečnej krabice.

Pôvodný článok s komentármi²⁸

²⁸ http://lesswrong.com/lw/ln/resist_the_happy_death_spiral/

Kapitola 65. Nekritická nadkritickosť

Z času na čas vidíte, ako sa ľudia hádajú, či je ateizmus “náboženstvo”. Ako som naznačil v článku Cieľ a pragmatizmus¹, hádanie sa o význame slova takmer vždy znamená, že ste stratili z očí pôvodnú otázku. Ako vôbec môže takáto hádka začať?

Ateista reční, obviňuje “náboženstvo” za inkvizíciu, krížové výpravy a rôzne konflikty s islamom alebo vnútri neho. Veriaci môže odpovedať: “Ale ateizmus je tiež náboženstvo, pretože aj vy máte vieru ohľadom Boha; veríte, že Boh neexistuje.” Potom ateista povie: “Ak ateizmus je náboženstvo, potom nezbieranie známok je hobby,” a začne hádka.

Prípadne možno namietat: “Ale rovnako veľké hrôzy spôsobil Stalin, ktorý bol ateista a potláčal cirkvi v mene ateizmu; preto je nesprávne obviňovať z násilia náboženstvo.” Na to môže mať ateista chuť povedať: “Stalinovým náboženstvom bol komunizmus.” Na to veriaci odpovie: “Ak komunizmus je náboženstvo, potom fandom Star Wars je vláda” a začne hádka.

Mal by “náboženský” názor byť definovaný ako definitívny názor na existenciu aspoň jedného Boha, napríklad priradenie pravdepodobnosti nižšej ako 10 % alebo vyššej ako 90 % existencii Zeusa? Alebo by “náboženský” názor mal byť definovaný ako pozitívny názor, povedzme pravdepodobnosť vyššia ako 90 % v prospech existencie aspoň jedného Boha? V tom prvom prípade bol Stalin “nábožný”; v tom druhom prípade bol Stalin “nie nábožný”.

Ale práve toto je nesprávny spôsob, ako pozerat na problém. To, čo naozaj chcete vedieť – o čom tá hádka bola na začiatku – je prečo boli v niektorých bodoch ľudských dejín veľké skupiny ľudí zabíjané a mučené, napohľad v mene nejakej myšlienky. Predefinovanie slova nezmení historické fakty jedným ani druhým smerom.

Komunizmus bola zložitá katastrofa a možno neexistuje jeden konkrétny dôvod pre to celé, žiadna jedna kritická linka v reťazi kauzality. Ale ak by som mal naznačiť základnú chybu, bolo by to... nuž, nech to Boh povie za mňa:

Ak tvoj brat, syn tvojho otca alebo tvojej matky, alebo tvoj syn alebo dcéra, alebo tvoja manželka, alebo tvoj najlepší priateľ, sa ťa pokúsi tajne zvieť slovami: Podme a slúžme iným bohom, neznámym tebe a tvojim predkom, bohom okolitých národov, či už blízkych alebo vzdialených, kdekoľvek na svete, nesmieš súhlasiť, **nesmieš ho počúvať**; nesmieš mu prejavíť lútosť, nesmieš ho ušetriť ani skryť jeho vinu. Nie, **musíš ho zabiť**, tvoja ruka musí prvý úder vedúci k jeho smrti a ruky ostatných ľudí musia nasledovať. Musíš ho ukameňovať na smrť, pretože sa ťa pokúsil odkloniť od Jahveho, tvojho Boha.

—Piata kniha Mojžišova 13:7-11, pridané zvýraznenie

Toto bolo aj pravidlo, ktoré Stalin určil pre komunizmus a Hitler pre nacizmus: ak ti tvoj brat skúša povedať, prečo sa Marx mýli, ak ti tvoj syn skúša povedať, že židia neplánujú svetovládu, nediskutuj s ním ani nepredkladaj svoje indície; nerob replikovatelné pokusy ani neskúmaj históriu; ale okamžite ho udaj tajnej polícii.

¹ http://lesswrong.com/lw/lf/purpose_and_pragmatism/

Včera som naznačil, že jedným z kľúčov odolávania² afektívnej špirále smrti³ je princíp "priťahujúcich podrobností"⁴ - len si *pamätať*, že treba spochybňovať konkrétne podrobnosti každého dodatočného pekného tvrdenia o Veľkej Myšlienke. (To nie je triviálna rada. Ľudia na toto často zabúdajú, keď počujú, ako im futurista vykresľuje úžasne podrobný obraz zázrakov budúcnosti, alebo keď myslia na svoju vlastnú najobľúbenejšiu myšlienku.) Toto by nás nezbavilo efektu svätožiary⁵, ale snád' by to obmedzilo rezonanciu pod kritickú hladinu, takže by jedno pekne znejúce tvrdenie vyvolalo v priemere menej ako 1,0 ďalšieho pekne znejúceho tvrdenia.

Pravým opakom tejto rady, ktorý posúva efekt svätožiary *nad* kritickú hladinu, je keď ľuďom pripadá nesprávne argumentovať proti *hocijakému* pozitívnemu tvrdeniu o Veľkej Myšlienke. Politika zabíja myslenie.⁶ Argumenty sú vojaci. Akonáhle viete, na ktorej ste strane, musíte podporovať všetky tvrdenia v jej prospech a argumentovať proti všetkým tvrdeniam v jej neprospech. Inak by to bolo ako pomáhať nepriateľovi alebo bodat vlastných vojakov do chrbta.

Ak...

- ...máte pocit, že protirečiť niekomu, kto urobí nesprávne pekné tvrdenie v prospech evolúcie⁷, by znamenalo pomáhať kreacionistom;
- ...máte pocit, že dostávate duchovný kladný bod za každú peknú vec, ktorú poviete o Bohu a že argumentovať proti nej by ohrozilo váš vzťah s Bohom;
- ...ak máte silný pocit, že zvýšni ľudia v miestnosti by vás nemali radi za to, že "nepodporujete naše vojská", keby ste argumentovali proti najnovšej vojne;
- ...povedanie hocičoho proti komunizmu spôsobí, že vás ukameňujú zastrelením;

...potom je afektívna špirála smrti už nadkritická. Už je to Super Šťastná Špirála Smrti.

Teda nie náboženstvo ako také je kľúčovou kategóriou voči našej pôvodnej otázke: "Čo spôsobuje to zabíjanie?" Najlepšie rozlíšenie, aké som počul⁸ medzi "nadprirodzeným" a "naturalistickým" svetonázorom je, že nadprirodzený svetonázor tvrdí existenciu ontologicky základných myšlienkových podstát, ako sú duchovia, zatiaľ čo naturalistický svetonázor redukuje myšlienkové javy na nemyšlienkové časti. Sústreď sa na toto ako na zdroj problémov znamená akceptovať výnimočnosť náboženstva. Nadprirodzené tvrdenia sa oplatí rozlišovať, pretože sa vždy ukážu ako nesprávne z celkom zásadných⁹ dôvodov. Ale aj tak to je iba jeden druh chyby.

Afektívna špirála smrti sa môže vytvoriť okolo nadprirodzeného názoru; najmä okolo monoteizmu, ktorého vrcholom je Super Šťastný Činiteľ, definovaný v prvom rade tým, že súhlasíme s každým pekným názorom o ňom; najmä ak komplex mémov narastie na dostatočne sofistikovaný, aby tvrdil, že za nevieru bude nasledovať nadprirodzený trest. Špirála smrti však môže vzniknúť aj okolo politickej inovácie, charizmatického vodcu, viery v osud rasy, alebo ekonomickej hypotézy. Lekcia z dejín znie, že afektívne špirály sú nebezpečné bez ohľadu na to, či obsahujú alebo neobsahujú nadprirodzeno. Náboženstvo nie je dosť špeciálne ako typ chyby, aby bolo kľúčom k tomuto problému.

² pozri: 64 Odolajte šťastnej špirále smrti

³ pozri: 63 Afektívne špirály smrti

⁴ pozri: 6 Priťahujúce podrobnosti

⁵ pozri: 60 Efekt svätožiary

⁶ pozri: 47 Politika zabíja myslenie

⁷ pozri: 130 Zázrak evolúcie

⁸ <http://richardcarrier.blogspot.sk/2007/01/defining-supernatural.html>

⁹ pozri: 36 Tajomné odpovede na tajomné otázky

Sam Harris sa dostal bližšie, keď ukázal obviňujúcim prstom na *vieru*. Ak neukladáte primerané bremeno dôkazu každému jednému dodatočnému peknému tvrdeniu, afektívna rezonancia začne veľmi ľahko. Pozrite sa na chudákov z new age. Kresťanstvo si vyvinulo ochrany proti kritike, argumentujúc za zázraky viery; new age kultúrne zdedil uloženú myšlienku¹⁰, že viera je dobrá, ale chýba mu vylučujúce sväté písmo kresťanstva na bránenie sa konkurujúcim mémom. Priaznivci new age končia v šťastných špirálach smrti okolo hviezd, stromov, magnetov, diét, kúziel, jednorozcov...

Afektívne špirály smrti sa však stávajú omnoho vražednejšie, keď sa kritika stane hriechom, prešľapom, alebo zločinom. Na tomto svete sú veci, ktoré sú hodné veľkej chvály a nemôžete *jednoznačne* povedať, že chvála za istú hranicu je zakázaná. Ale *nikdy* neexistuje Myšlienka natoľko pravdivá, že je nesprávne kritizovať akýkoľvek argument, ktorý ju podporuje. Nikdy. Nikdy nikdy nikdy. *Toto* je jednoznačné. Prevažná väčšina¹¹ možných názorov v netriviálnom priestore odpovedí je nesprávna, a podobne aj prevažná väčšina možných *podporujúcich argumentov* pre pravdivý názor je nesprávna, a toto nezmení ani tá najšťastnejšia myšlienka.

A je trojnásobne ultra zakázané reagovať na kritiku násilím. Je iba pár prikázaní v ľudskom umení rozumnosti, ktoré nemajú žiadne "ak", "a", "alebo" a iné únikové doložky. Toto je jedno z nich. Zlý argument má dostať protiargument. Nemá dostať guľku. Nikdy. Nikdy nikdy nikdy.

Pôvodný článok s komentármi¹²

¹⁰ pozri: 81 Uložené myšlienky

¹¹ pozri: 22 Koľko indicie treba?

¹² http://lesswrong.com/lw/lo/uncritical_supercriticality/

Kapitola 66. Ochladzovanie skupinových názorov vyparovaním

Prvé štúdie siekt s prekvapením zistili, že keď sekta utrpí väčší šok – keď sa nenaplnilo proroctvo, keď sa odhalilo morálne zlyhanie zakladateľa – často z toho vyjde silnejšia než predtým, so zvýšenou vierou a fanatizmom. Svedkovia Jehovovi určili Armageddon na rok 1975 na základe biblických výpočtov; rok 1975 prišiel a odišiel. Sekta Unariánov, dodnes silná, prežila nezjavenie sa medzigalaktickej vesmírnej flotily¹ 27. septembra 1975. (Je zaujímavé, že článok vo Wikipédii o Unariánoch² spomína neúspešné proroctvo v roku 2001, ale nespomína predchádzajúci neúspech v roku 1975.)

Prečo by sa skupinová viera mala stať *silnejšou* po stretnutí sa so zdrvivúcou protiindíciou?

Zvyčajné vysvetlenie tohto javu je založené na kognitívnej disonancii. Keď ľudia urobia “nezvratné” činy v službe danej viery – rozdajú všetok svoj majetok v očakávaní pristátia lietajúcich tanierov – nemôžu si pripustiť, že sa mýlili. Tlak na ich názor vytvára obrovskú kognitívnu disonanciu; musia nájsť posilňujúce myšlienky, aby tento šok prekonali, a tak sa stanú ešte väčšími fanatikmi. Podľa tohto výkladu je silnejší skupinový fanatizmus výsledkom zvýšenia fanatizmu jednotlivcov.

Pozeral som na Java applet, ktorý znázorňoval využitie ochladzovania vyparovaním na tvorbu Boseho-Einsteinovho kondenzátu³, keď mi napadlo, že na zvýšenie fanatizmu môže pôsobiť celkom iná sila. Ochladzovanie vyparovaním vytvára bariéru potenciálnej energie okolo skupiny horúcich atómov. Tepelná energia je v prírode v podstate štatistická – nie všetky atómy sa pohybujú rovnakou rýchlosťou. Kinetická energia jednotlivých atómov sa mení ako do seba atómy navzájom narážajú. Ak nastavíte bariéru potenciálnej energie len o trochu vyššie než je priemerná teplotná energia, jednotlivé atómy z času na čas získajú dostatočne vysokú kinetickú energiu na únik z pasce. Keď unikne nezvyčajne rýchly atóm, odnáša so sebou nezvyčajne veľké množstvo kinetickej energie, takže priemerná energia klesne. Skupina sa postupne stáva chladnejšou než je bariéra potenciálnej energie okolo nej. Hranie sa s Java appletom⁴ mi to objasnilo.

Vo Festingerovej klasike: “Keď proroctvo zlyhá” jeden z členov sekty vyšiel z dverí okamžite po tom, čo lietajúci tanier nepristál. Kto sa nahnevá a odíde prvý? *Priemerný* člen sekty? Alebo relatívne skeptickejší člen, ktorý mohol dovtedy konať ako hlas umiernenosti, brzda voči fanatickejšími členom?

Keď členovia s najväčšou kinetickou energiou uniknú, zvyšné diskusie budú medzi extrémnymi fanatikmi na jednej strane a o trochu menej extrémnymi fanatikmi na druhej strane, pričom konsenzus skupiny bude niekde “uprostred”.

A čo by bola analógia kolapsu, ktorý vytvorí Boseho-Einsteinov kondenzát? Nuž, nemusíme túto analógiu ťahať až tak ďaleko. Ale možno si spomeniete, že som použil analógiu štiepnej jadrovej reakcie na afektívnu špirálu smrti; keď skupina vylúči všetky svoje umiernené hlasy, všetci zostávajúci ľudia sa budú navzájom

¹

² http://en.wikipedia.org/wiki/Unarius_Academy_of_Science

³ http://www.colorado.edu/physics/2000/bec/evap_cool.html

⁴ http://www.colorado.edu/physics/2000/bec/evap_cool.html

podporovať a potláčať nesúhlas, čo môže zvýšiť priemerný fanatizmus vo vnútri. (Nemám tu žiadnu termodynamickú analógiu, dokiaľ niekto nevyvinie jadrovú bombu, ktorá vybuchne, keď sa ochladí.)

Keď sa Objektivistí dozvedeli o dlhodobej afére Ayn Rand s Nathanielom Brandenom, významná časť Objektivistov sa odtrhla a pridala sa k Brandenovi vyhlasujúc "otvorený systém" Objektivismu, ktorý nebude natoľko pevne spojený s Ayn Rand. Kto zostal s Ayn Rand aj po vypuknutí škandálu? Tí, ktorí v ňu *naozaj*, *naozaj* verili – a možno pár nerozhodných, ktorí po odchode umiernených hlasov počuli argumenty iba z jednej strany. To môže vysvetľovať, preto je Inštitút Ayn Rand (údajne) po rozchode ešte fanatickejší než pôvodná zakladajúca skupina Objektivistov okolo Brandena a Rand.

Pred pár rokmi som bol v transhumanistickej mailovej diskusii, kde malá skupina presadzovala "sociálne demokratický transhumanizmus" žľčovito urážajúc každého libertariána v diskusii. Väčšina libertariánov opustila mailovú diskusiu a väčšina zostávajúcich prestala písať. V dôsledku toho sa zostávajúca skupina posunula výrazne doľava. Bolo to úmyselné? Pravdepodobne nie, pretože si nemyslím, že páchatelia poznali psychológiu tak dobre. (Keď už sme pri tom, nespomínam si, že by som niekde inde videl analógiu s ochladzujúcim odparovaním, hoci to neznamena, že si to niekto nevšimol predtým.) Nanajvýš sa mohli pokúšať stať "väčšími rybami v malom rybníku".

To je dôvod, prečo je dôležité byť naklonený v prospech tolerovania nesúhlasu. Počkajte ešte dosť dlho *potom*, čo sa vám zdá oprávnené vylúčiť člena zo skupiny, než ho naozaj vylúčite. Keď sa zbavíte starých extrémov, pozícia skupiny sa posunie a niekto iný sa stane čudákom. Ak vylúčite aj jeho, ste na dobrej ceste, aby ste sa stali Boseho-Einsteinovým kondenzátom a, ehm, vybuchli.

Odvrátená strana: Thomas Kuhn veril, že veda sa musí stať "paradigmou" so spoločným technickým jazykom vylučujúcim nečlenov, kým niečo naozaj dokáže urobiť. V počiatočných štádiách vedy, podľa Kuhna, sa prívrženci snažia zo všetkých síl, aby ich práca bola zrozumiteľná mimo akadémie. Lenže (podľa Kuhna) veda môže urobiť skutočný pokrok ako technická disciplína iba keď sa vzdá podmienky dostupnosti zvonka, a keď vedci pracujúci v danej paradigme začnú vo svojej komunikácii predpokladať oboznámenosť s množstvom základného technického materiálu. Znie to cynicky v porovnaní s tým, čo sa zvyčajne hovorí⁵ o vnímaní vedy verejnosťou, ale rozhodne tu vidím zrnko pravdy.

Moja vlastná teória moderovania internetu znie, že musíte byť ochotní vylúčiť trollov a spam, aby vznikla konverzácia. Musíte byť dokonca ochotní vylúčiť milých alebo technicky neinformovaných ľudí z technickej mailovej diskusie, ak chcete, aby sa urobilo niečo užitočné. Skutočne otvorená diskusia na internete rýchlo zdegeneruje. Podľa tejto teórie by ste sa však mali vyvarovať vylučovania *sčítaných* trollov – plnia skrytú funkciu legitimizovania menej extrémneho nesúhlasu. Nemali by ste však mať toľko sčítaných trollov, že sa začnú hádať jeden s druhým alebo začnú dominovať v diskusii. Ak máte jedného človeka, ktorý je známy Chlapík, Ktorý Nesúhlasí So Všetkým, každý s rozumnejším a umiernenejším nesúhlasom nebude vyzeráť ako jediný vyčnievajúci klinec. Táto teória internetového moderovania mi možno v praxi až tak dobre nefungovala, takže ju berte s nadhľadom.

Pôvodný článok s komentármi⁶

⁵ pozri: 16 Signály na potlesk

⁶ http://lesswrong.com/lw/lr/evaporative_cooling_of_group_beliefs/

Kapitola 67. Keď sa nikto neodváži naliehať na zdržanlivosť

Jedného dňa som ráno vstal z postele, zapol som si počítač a môj mailový klient Netscape mi automaticky stiahol prehľad správ na daný deň. V ten konkrétny deň boli správy o tom, že dve unesené lietadlá narazili do World Trade Center.

Toto boli moje prvé tri myšlienky, v uvedenom poradí:

Tuším naozaj žijem v Budúcnosti.

Ako dobre, že to nebol jadrový útok.

a potom

Prehnaná reakcia na toto bude desaťkrát horšia než pôvodná udalosť.

Samotný činiteľ “desaťkrát horšia” sa ukázal ako hrubé podcenenie. Ani ja som neuhádol, ako zle sa veci vyvinú. V tom je tá náročnosť pesimizmu: je *naozaj ťažké* mieriť dostatočne nízko, aby ste boli príjemne prekvapení zhruba rovnako často ako nepríjemne prekvapení.

Napriek tomu som si hneď uvedomil, že každý všade bude hovoriť o tom, aká hrozná, aká príšerná táto udalosť bola; a že sa nikto neopováži byť hlasom zdržanlivosti, primeranej reakcie. Prvé odhady 11. septembra boli, že zomrelo šesť tisíc ľudí. Každý politik, ktorý by povedal: “6000 smrť je 1/8 každoročných amerických obetí dopravných nehôd” by musel do hodiny podať demisiu.

Nie, 11. september nebol dobrý deň. Ale ak *každý* dostane hviezdičku za zdôrazňovanie, ako veľmi to bolí, a *nikto* sa neopováži naliehať na zdržanlivosť pri protiútoky, potom bude reakcia väčšia, než je primerané, bez ohľadu na to, aká presne by bola primeraná úroveň.

Toto je ešte temnejší zrkadlový obraz šťastnej špirály smrti¹ – špirála nenávisti. Každý, kto napadne Nepriateľa je vlastenec; ktokoľvek sa snaží analyzovať čo len jediné negatívne tvrdenie o Nepriateľovi, je zradca. Ale tak ako je prevažná väčšina všetkých zložitých výrokov nepravdivá, aj prevažná väčšina negatívnych vecí, ktoré môžete o hocikom povedať, dokonca aj o tom najhoršom človeku na svete, je nepravdivá.

Myslím si, že najlepšou ilustráciou bude: “samovražední únoscovia boli zbabelí²”. Trocha zdravého rozumu, prosím? Vyžaduje to trocha odvahy dobrovoľne napáliť svojím lietadlom do budovy. Pri všetkých ich hriechoch, zbabelosť na to zozname nebola. Ale predpokladám, že hocičo zlé, čo poviete o teroristovi, bez ohľadu na to, aké je to hlúpe, musí byť pravda. Dostal by som ešte väčšiu hviezdičku, keby som obvinil Al Kaidu, že zavraždila Kennedyho? Alebo keby som ich obvinil, že sú stalinisti? Vážne, *zbabelosť*?

Áno, záleží na to, že únoscovia z 11. septembra neboli zbabelci. Nie iba kvôli realistickému porozumeniu nepriateľovej psychológie. Špirály nenávisti jednoducho spôsobujú priveľa škody. Je skrátka príliš nebezpečné, aby niekde

¹ pozri: 63 Afektívne špirály smrti

² pozri: 51 Sú vaši nepriatelia od narodenia zlí?

na svete existoval cieľ, či už sú to Židia alebo Adolf Hitler, o ktorom *povedať negatívne veci* je dôležitejšie než *povedať presné veci*.

Keď obranné sily obsahujú tisíce lietadiel a stovky tisícov ťažko ozbrojených vojakov, človek by mal vziať do úvahy, že samotný tento imunitný systém je schopný napáchať omnoho väčšie škody než 19 chlapov a štyri civilné lietadlá. USA minuli miliardy dolárov a tisíce životov vojakov, aby sami sebe uškodili účinnejšie, než by mohla snívať ľubovoľná teroristická skupina.

Keby USA úplne ignorovali útok 11. septembra – iba by pokrčili plecami a znovu postavili tú budovu – bolo by to lepšie než skutočný vývoj histórie. Ale takáto politická možnosť nebola. Aj keby sa každý v súkromí dovätínil, že imunitná reakcia bude škodlivejšia než choroba, americkí politici nemali inú možnosť ako si udržať kariéru, než kráčať priamo do pasce Al Kaidy. Ktokoľvek argumentuje za väčšiu reakciu, je vlastenec. Ktokoľvek analyzuje vlastenecké tvrdenie, je zradca.

Na začiatku boli múdrejšie reakcie na 11. september, než by som odhadoval. Videl som niekoho v kongrese – nepamätám sa, koho – povedať pred kamerami: “Zabudli sme, že prvoradou úlohou vlády nie je ekonomika, ani zdravotná starostlivosť, ale ochrana krajiny pred útokom.” Až sa mi oči rozšírili, že politik dokázal povedať niečo, čo nebol signál na potlesk³. Daná osoba z kongresu musela byť vo veľkom emocionálnom šoku, že povedala niečo také... skutočné.

Ale za dva dni skutočný šok pominul a starosť o imidž získala úplnú kontrolu nad politickou debatou. Potom špirála stupňovania celkom prevládla. Akonáhle sa zdržanlivosť stala nevysloviteľnou, bez ohľadu na to, kde diskusia začala, úroveň zúrivosti a hlúposti mohla časom iba rásť.

Pôvodný článok s komentármi⁴

³ pozri: 16 Signály na potlesk

⁴ http://lesswrong.com/lw/lw/when_none_dare_urge_restraint/

Kapitola 68. Pokus v Robbers Cave

Mali ste niekedy ako dieťa podozrenie, či váš nejaký "letný tábor" v skutočnosti neslúži nejakému prefikánu skrytému účelu – napríklad, či to celé nie je vedecký pokus a "táboroví vedúci" nie sú v skutočnosti výskumníci pozorujúci vaše správanie?

Ani ja.

Ale boli by sme paranoidnejší, keby sme čítali Konflikt a spolupráca v skupine: Pokus v Robbers Cave¹ od Sherifa, Harveya, Whitea a Sherifa (1954/1961). V tejto skupine pokusné osoby – prepáčte, "táborníci" - boli 22 chlapci medzi piatym a šiestym ročníkom, vybraní z 22 rôznych škôl v Oklahoma City, zo stabilných protestantských rodín strednej triedy, s dobrým prospechom, mediánové IQ 112. Boli natoľko dobre vychovaní a navzájom podobní, ako len výskumníci dokázali zabezpečiť.

Tento pokus, urobený po zmätenom závere druhej svetovej vojny, mal za úlohu skúmať príčiny – a možné lieky – konfliktu medzi skupinami. Ako vyvolali konflikt medzi skupinami, ktorý mohli skúmať? Nuž, rozdelili týchto 22 chlapcov na dve skupiny po 11 táborníkov a...

...a ukázalo sa, že to úplne stačí.

Výskumníci pôvodne plánovali, že pokus vykonajú v troch etapách. V Etape 1 by sa každá skupina táborníkov usadila zvlášť, nevediac o sebe navzájom. Na konci Etapy by sa skupiny navzájom postupne dozvedeli o svojej existencii. V Etape 2 by pomocou série súťaží a pretekov vyvolali rozpor medzi oboma skupinami.

Nebola potrebná žiadna Etapa 2. Vzájomná nevraživosť existovala prakticky od okamihu, keď sa každá skupina dozvedela o existencii tej druhej: Oni používajú *naše* táborisko, *naše* baseballové ihrisko. Pri prvom stretnutí si obe skupiny začali navzájom nadávať. Dali si mená Štrkáči a Orli (dokiaľ si mysleli, že sú jedinou skupinou v táborisku, nepotrebovali mať meno).

Keď boli v súlade s vopred určenou experimentálnou procedúrou ohlásené súťaže a preteky, súperenie medzi skupinami dosiahlo horúčkovitú úroveň. Prvé dva dni bolo v súťaži jasné športové správanie, ale rýchlo sa vytratilo.

Orli ukradli Štrkáčom vlajku a spálili ju. Štrkáči sa vlámali do chaty Orlov a ukradli modré rifle vodcu skupiny, nafarbili ich na oranžovo a ďalší deň niesli ako vlajku s nápisom "Posledný Orol". Orli podnikli odvetné vlámanie k Štrkáčom, prevrátili im posteľ a rozhádzali špinu. Potom sa vrátili do svojej chaty, kde sa zabarikádovali a pripravili si zbrane (ponožky naplnené kameňmi) pre prípad protiútoku. Keď Orli vyhrali záverečný pretek naplánovanej Etapy 2, Štrkáči sa vlámali do ich chaty a ukradli cennosti. Toto prerástlo do pástného súboja, ktorý musel personál zastaviť v obave zo zranení. Orli si tento príbeh medzi sebou rozprávali a premenili celú záležitosť na veľkolepé víťazstvo – hnali Štrkáčov "až za polcestu do ich chaty" (čo nebola pravda).

Každá skupina si vytvorila negatívny stereotyp o Cudziach a kontrastujúci pozitívny stereotyp o Našich. Štrkáči veľa nadávali. Keď Orli vyhrali jednu hru, usúdili, že vyhrali vďaka svojim modlitbám a že Štrkáči prehrali pretože stále nadávali. Orli sa rozhodli, že sami prestanú nadávať. Ďalej usúdili, že keďže Štrkáči stále

¹ <http://psychclassics.yorku.ca/Sherif/>

naďavajú, bude múdrejšie nerozprávať sa s nimi. Orli si vytvorili sebaobraz správnych a slušných; Štrkáči si vytvorili sebaobraz silných a drsných.

Členovia skupín sa držali za nos, keď okolo nich išli členovia druhej skupiny.

V Etape 3 sa výskumníci pokúšali zredukovať napätie medzi dvoma skupinami.

Púhy kontakt (prítomnosť bez súťaženia) nezmenšoval napätie medzi skupinami. Spoločná účasť na príjemnej udalosti – napríklad strieľanie ohňostrojov na štvrtého júla – nezmenšila napätie; namiesto toho sa začali ohadzovať jedlom.

Čo myslíte, že *zabralo*?

(Priestor pred prezradením pointy...)

Chlapcom povedali, že možno bude v celom tábore nedostatok vody, pretože vodný systém má záhadné ťažkosti – možno kvôli vandalom. (Vonkajší nepriateľ, jeden z najstarších trikov.)

V oblasti medzi táborom a nádržou mali strážiť štyri hliadky. (Zo začiatku sa hliadky skladali iba z členov rovnakej skupiny.) Všetky hliadky sa mali stretnúť pri nádrži, ak nič nenašli. Keďže nič nenašli, skupiny sa stretli pri nádrži a uvideli, že z kohútika netečie voda. Obe skupiny chlapcov diskutovali, v čom môže byť problém, klopali na boky nádrže, objavili rebrík na vrch, skontrolovali, že vodná nádrž je plná, a nakoniec našli vrece napchané v kohútiku. Všetci chlapci sa zhromaždili okolo kohútika, aby ho vyčistili. Padali návrhy od členov oboch skupín a chlapci z oboch skupín sa ich snažili realizovať.

Keď bol kohútik konečne vyčistený, Štrkáči, ktorí mali poľné fľaše, nenamietali, aby sa Orli napili z kohútika ako prví (Orli so sebou nemali fľaše). Nepadali žiadne urážky, dokonca ani zvyčajné “dámy majú prednosť”.

To nebol koniec súperenia. Ďalšie ráno bolo ďalšie ohadzovanie jedlom s urážkami. Ale niekoľko ďalších spoločných úloh, vyžadujúcich spoluprácu medzi oboma skupinami – napríklad roztlačanie nákladného auta – to dosiahlo. Na konci tábora Štrkáči použili 5 dolárov získaných v pretekoch hádzania fazulí na nákup cukríkov pre všetkých chlapcov z oboch skupín.

Pokus v Robbers Cave vykresľuje psychológiu bánd lovcov-zberačov, opakujúcu sa v čase², najlepšie zo všetkých pokusov, ktoré kedy sociológovia vymysleli.

Akúkoľvek podobnosť s modernou politikou si iba predstavujete.

(Niekedy si myslím, že druhá najdôležitejšia vec, ktorú ľudstvo potrebuje, je superzloduch. Možno sa podujmem na túto úlohu, keď dokončím svoju terajšiu prácu.)

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. 1954/1961. *Štúdia o pozitívnych a negatívnych vzťahoch medzi experimentálne vytvorenými skupinami: štúdia Robbers Cave*. University of Oklahoma.

Pôvodný článok s komentármi³

² pozri: 137 Evolučná psychológia

³ http://lesswrong.com/lw/lt/the_robbers_cave_experiment/

Kapitola 69. Každá kauza chce byť sektou

Súvisí s kapitolami: Chyba prisudzovania¹, Afektívne špirály smrti²

Cade Metz v *The Register* nedávno³ tvrdil⁴, že tajná mailová diskusia hlavných administrátorov Wikipédie začala byť posadnutá blokováním všetkých kritikov a možných kritikov Wikipédie. Vráťane zablokovania produktívneho používateľa, keď jeden administrátor – iba kvôli jeho produktivite – získal presvedčenie, že tento používateľ je špiónom poslaným z *Wikipedia Review*. A že ľudia navrchu Wikipédie uzavreli svoje rady, aby sa tak bránili. (Zatiaľ som sa tieto obvinenia neskúmal.)

Existuje nejaká hlboká morálna chyba v snahe systematizovať poznanie sveta, ktorý by viedla prívržencov tejto Kauzy do šialenstva? Možno iba ľudia, ktorí majú vrodené totalitárske sklony, sa pokúšajú stať svetovou autoritou na všetko...

Pozor, sklon prisudzovať⁵! (Sklon prisudzovať: robenie záverov o niekoho jedinečných predpokladoch na základe správania, ktoré sa dá plne vysvetliť situáciou, v ktorej nastalo. Keď vidíme, ako niekto kope do predajného automatu, myslíme si, že je to "zurvalec", ale keď my kopneme do predajného automatu, je to preto, lebo nám zmeškal autobus, ušiel vlak, a tento stroj nám zožral peniaze.) Ak sú obvinenia Wikipédie pravdivé, dajú sa vysvetliť *obyčajnou* ľudskou povahou, nie *mimoriadnou* ľudskou povahou.

Rozdelenie na našich a cudzích⁶ je súčasťou bežnej ľudskej povahy. Rovnako ako šťastné špirály smrti⁷ a špirály nenávisti⁸. Vznešená Kauza nevyžaduje skrytú chybu svojich prívržencov, aby vytvorila sektársku skupinu. Stačí, že jej prívrženci sú ľudia. Všetko ostatné nasleduje prirodzene, štandardný úpadok, ako keď sa vám chladničke pokazí jedlo, keď vypadne elektrina.

V tom istom zmysle, ako sa každý teplotný rozdiel chce vyrovnáť a každý počítačový program sa chce stať zbierko ad-hoc záplat, každá Kauza sa *chce* stať sektou. Je to stav vysokej entropie, ku ktorému má systém sklon, atraktor v ľudskej psychológii. Nemusí to mať nič spoločné s tým, či je Kauza naozaj Vznešená. Možno si myslíte, že Dobrá Kauza nalepí svoje dobro na všetky stránky ľudí, ktorí sú s ňou spojení – že nasledovníci tejto Kauzy budú mať menší sklon k súbojom o spoločenské postavenie, k rozlišovaniu našich a cudzích, k afektívnym špirálam, k uctievaniu vodcov. Ale veriť jednej pravdivej myšlienke nevypína efekt svätožiary⁹. Vznešená kauza neurobí zo svojich stúpencov niečo iné ako ľudí. Existuje veľa zlých myšlienok, ktoré môžu narobiť mnoho škody – ale to nemusí byť nutne to, o čo tu ide.

Každá skupina ľudí s nezvyčajným cieľom – dobrým, zlým, či hlúpym – sa bude blížiť sa k sektárskemu atraktoru, pokiaľ nevyvinú sústavné úsilie zabrániť tomu. Môžete udržať svoj byt chladnejší než vonkajšie počasie, ale musíte mať stále zapnutú klimatizáciu, a akonáhle vypnete elektrinu – vzdáte boj proti entropii – veci sa vrátia do "normálneho" stavu.

¹ pozri: 50 Chyba prisudzovania

² pozri: 63 Afektívne špirály smrti

³ http://www.theregister.co.uk/2007/12/04/wikipedia_secret_mailing/

⁴ http://www.theregister.co.uk/2007/12/06/wikipedia_and_overstock/

⁵ pozri: 50 Chyba prisudzovania

⁶ pozri: 68 Pokus v Robbers Cave

⁷ pozri: 63 Afektívne špirály smrti

⁸ pozri: 67 Keď sa nikto neodváža naliehať na zdržanlivosť

⁹ pozri: 60 Efekt svätožiary

V jednom zaujímavom prípade bola skupina, z ktorej sa stala polosekta, ktorej bojový pokrik bol: "Racionalita! Rozum! Objektívna skutočnosť!" (Viac o tomto v nasledujúcich článkoch.) Dať svojej Veľkej Myšlienke nápis "rozum" vám nepomôže o nič viac než nalepiť na svoj dom nápis "Zima!" Aj tak musíte zapnúť klimatizáciu - vynaložiť požadovanú energiu na jednotku času, aby ste obrátili prirodzený úpadok do sektárstva. Uctievanie rozumu vás neurobí príčetnejšími o nič viac, než vám uctievanie gravitácie dovolí lietať. Nemôžete sa rozprávať s termodynamikou, ani sa modliť k teórii pravdepodobnosti. Môžete ich *použiť*, ale nie stať sa ich členmi.

Sektárstvo je kvantitatívne, nie kvalitatívne. Otázka neznie: "Sekta, áno alebo nie?" ale "Koľko sektárstva a kde?" Ešte aj vo Vede, ktorá je archetypálnou Pravou Naozaj Vznešenou Kauzou, môžeme pohotovo ukázať na súčasnú bojovú líniu vojny proti sektárskej entropii, kde sa bojová línia pomaly posúva dopredu a dozadu. Sú odborné časopisy ochotnejšie prijať články so známejším autorským rukopisom alebo od neznámeho autora z dobre známej inštitúcie, v porovnaní s neznámym autorom z neznámej inštitúcie? Nakolko článku veríme na základe autority a nakolko na základe experimentu? Ktoré odborné časopisy používajú anonymných recenzentov, a nakolko efektívne je anonymné recenzovanie?

Citujem tento príklad, namiesto štandardného¹⁰ hmlistého obvinenia: "Vedci nie sú otvorení novým myšlienkam," pretože ukazuje *bojovú líniu* - miesto, kde ľudskú psychológiu aktívne vytláčame, kde je zhromaždená sektárska entropia pumpovaná von. (Čo samozrejme vyžaduje nejaké tepelné straty.)

Tento článok nie je katalógom techník na aktívne pumpovanie proti sektárstvu. Niektoré¹¹ také¹² techniky som už spomenul, iné spomeniem neskôr. *Dnes* som chcel ukázať na to, že hodnota samotnej Kauzy neznamená, že môžete vynaložiť *menej* úsilia na odolávanie sektárskemu atraktor. A že ak dokážete ukázať prstom na bojovú líniu, neznamená to, že priznávate nehodnosť svojej Vznešenej Kauzy. Možno si myslíte, že keby otázka znela: "Sekta, áno alebo nie?", boli by ste povinní odpovedať: "Nie", inak by ste zradili svoju milovanú Kauzu. Ale to je ako keby ste si mysleli, že máte rozdeliť stroje na "dokonale efektívne" a "neefektívne" namiesto merania ich strát.

Naopak, ak veríte, že Nečistá Podstata všetkých Hlúpych Iných Káuz spôsobila, že dopadli zle, ak sa smejete na hlúposti "obetí siekt", ak si myslíte, že zakladatelia a členovia siekt sú mutanti, potom nebudete vynakladať dostatočné úsilie na to, aby ste pumpovali proti entropii - aby ste odolávali svojej ľudskosti.

Pôvodný článok s komentármi¹³

¹⁰ pozri: 86 Ako vyzerá (a byť) hlboký

¹¹ pozri: 64 Odolajte šťastnej špirále smrti

¹² pozri: 123 Meditácia o zvedavosti

¹³ http://lesswrong.com/lw/lv/every_cause_wants_to_be_a_cult/

Kapitola 70. Strážcovia pravdy

Súvisí s kapitolami: Tsuyoku Naritai¹, Obrátená hlúposť nie je inteligencia²

Voči racionalistom je občas namierená táto kritika: “Inkvizítori si mysleli, že *oni* majú pravdu! Je jasné, že táto vec s ‘pravdou’ je nebezpečná.”

Existuje veľa jasných odpovedí, napríklad: “Ak si myslíš, že poznať pravdu *by* ti dalo právo mučiť a zabíjať, robíš chybu, ktorá nijako nesúvisí s epistemológiou.” Alebo: “A toto historické tvrdenie, ktoré si práve povedal o inkvizícii – to je pravda³?”

Obrátená hlúposť nie je inteligencia:⁴ “Ak váš terajší počítač prestane fungovať, nemôžete dôjsť k záveru, že všetko na vašom terajšom systéme je pokazené a že potrebujete nový systém bez procesora AMD, bez videokarty ATI... aj keď váš terajší počítač všetko toto má a nefunguje. Možno len potrebujete nový kábel od napájania.” Na dôjdenie k zlému záveru stačí jeden chybný krok, nemusí byť každý krok chybný. Inkvizítori verili, že $2 + 2 = 4$, ale to nebolo zdrojom ich šialenstva. Možno teda ani epistemologický realizmus nebol ten problém?

Vyzerá vieryhodne, že keby sa inkvizícia skladala z relativistov vyznávajúcich, že nič nie je pravda a na ničom nezáleží, že by mali menej nadšenia pre svoje mučenie. Takisto by mali menej nadšenia, keby boli po lobotómii. Myslím, že je to férové prirovnanie.

Ale predsa... si myslím, že vzťah inkvizície k pravde tam zohrával rolu. Inkvizícia verila, že existuje niečo ako pravda, a že je to dôležité; nuž, podobne aj Richard Feynman. Ale inkvizítori neboli hľadači pravdy. Oni boli *strážcovia pravdy*.

Raz som čítal tvrdenie (neviem nájsť zdroj), že dôležitou zložkou *ducha doby* je, či umiestňuje svoje ideály do svojej budúcnosti alebo minulosti. Takmer všetky kultúry pred osvietením verili v Pád z Milosti – že veci boli kedysi vo vzialenej minulosti dokonalé, ale potom nastala katastrofa a odvtedy ide všetko pomaly dole kopcom:

V dobe, keď bol život na Zemi úplný... Milovali jeden druhého a nevedeli, že to je “lásky k blížnemu”. Nepodvádzali jeden druhého a pritom nevedeli, že sú “dôveryhodní ľudia”. Boli spoľahliví a nevedeli, že to je “dobrá viera”. Žili slobodne spolu, dávajúc a prijímajúc, a nevedeli, že sú štedrý. Z tohto dôvodu ich činy neboli zaznamenané. Nepísali dejiny.

—Cesta Čuang C'

Dokonalá doba v minulosti podľa našich najlepších antropologických indícií nikdy neexistovala. Ale kultúra, ktorá vidí život ako neúprosne upadajúci, sa veľmi líši od kultúry, v ktorej možno dosiahnuť bezprecedentné výšky.

(Hovorím “kultúra” a nie “spoločnosť”, lebo v jednej spoločnosti môže byť viacero subkultúr.)

Mohli by ste povedať, že rozdiel medzi napríklad Richardom Feynmanom a inkvizíciou bol ten, že inkvizícia verila, že *má* pravdu, zatiaľ čo Richard Feynman pravdu *hľadal*. To však nie je celkom obhájitelné, pretože nepochybne existovali pravdy, o ktorých si Richard Feynman tiež myslel, že ich *má*. Napríklad: “Obloha je modrá” alebo: “ $2 + 2 = 4$ ”.

¹ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

² pozri: 52 Obrátená hlúposť nie je inteligencia

³ pozri: 46 Jednoduchá pravda

⁴ pozri: 52 Obrátená hlúposť nie je inteligencia

Áno, aj vo vede existujú prakticky isté pravdy. Všeobecnú relativitu môže vyvrátiť nejaká budúca fyzika – nie však spôsobom, ktorý by predpokladal, že Slnko bude obiehať okolo Jupitera; nová teória musí starej teórii ukradnúť jej úspešné predpovede, nie im protirečiť. Ale evolučná teória sa nachádza na vyššej organizačnej úrovni než atómy, a nič, čo objavíme o kvarkoch, nevyvráti darwinizmus, alebo biologickú teóriu buniek, alebo chemickú teóriu atómov, alebo stovky ďalších skvelých objavov, ktorých pravda je potvrdená za hranicou *rozumnej pochybnosti*.

Sú toto “absolútne pravdy”? Nie v zmysle, že by mali pravdepodobnosť doslova 1,0. Ale sú to prípady, kde si veda v zásade myslí, že našla pravdu.

A predsa vedci nemučia ľudí, ktorí spochybňujú chemickú teóriu atómov. Prečo nie? Pretože si nemyslia, že im ich istota dáva právo mučiť druhých? No áno, to je rozdiel *na povrchu*; ale prečo si to tí vedci *nemyslia*?

Pretože chémia netvrdí, že existuje nejaký nadprirodzený trest za pochybovanie o chemickej teórii atómov? Ale opäť podme hlbšie a opýtajme sa: “Prečo?” Prečo si chemici *nemyslia*, že pôjdete do pekla, ak pochybujete o teórii atómov?

Pretože odborné časopisy by neuverejnili váš článok, dokiaľ by ste nemali solídne experimentálne pozorovanie pekla? Ale príliš mnohí vedci dokážu vôľou potlačiť svoj skeptický reflex⁵. Prečo nemajú chemici súkromné sekty tvrdiace, že nechemici pôjdu do pekla, najmä keď mnohí z nich sú zároveň kresťanmi?

Otázky ako táto nemajú jednoduchú jednofaktorovú odpoveď. Ale tvrdil by som, že *jeden* z faktorov sa týka zastávania *obranného* postoja voči pravde, verzus *produktívneho* postoja voči pravde.

Keď ste Strážcom Pravdy, nemôžete k Pravde prispieť ničím užitočným *okrem* toho, že ju strážite. Keď sa snažíte získať Nobelovu cenu za chémiu objavom nového benzénu alebo fullerénu, niekto, kto spochybňuje teóriu atómov nie je pre váš svetonázor ani tak hrozbou ako stratou času.

Keď ste Strážcom Pravdy, jediné, čo môžete robiť, je odstrkávať nevyhnutný pokles do entropie mlátením do všetkého, čo sa odchyľi od Pravdy. Ak existuje nejaký spôsob, ako pumpovať proti entropii, generovať nové pravdivé názory pri malej tepelnej strate, táto pumpa môže držať pravdu nažive aj bez tajnej polície. V chémii môžete pousy zopakovať a vidieť sami – a toto udržiava vzácnu pravdu nažive bez potreby násilia.

A nie je to také strašná hrozba, ak sa raz niekde pomýlime – ak na chvíľu uveríme niečomu nepravdivému – pretože *zajtra* môžeme opäť nájsť stratenú zem.

Ale celý tento trik funguje iba preto, lebo experimentálna metóda je “kritériom dobra”, ktoré nie je púhym “kritériom zhody”. Pretože experimenty dokážu obnoviť pravdu bez potreby autority, dokážu aj *prekonať* autoritu a vytvoriť nové pravdivé názory tam, kde predtým žiadne neboli.

Kde existujú kritériá dobra, ktoré nie sú kritériami zhody, tam môžu existovať *zmeny*, ktoré sú *vylepšenia* namiesto hrozieb. Kde existujú *iba* kritériá zhody, kde neexistuje spôsob, ako prekonať autoritu, tam neexistuje ani spôsob, ako vyriešiť spor medzi autoritami. Okrem vyhladenia. Väčšia palica vyhráva.

Nechcem týmto ponúkať veľký všeobjímajúci jednofaktorový pohľad na dejiny. Chcem tým ukázať na hlboký psychologický rozdiel medzi vnímaním svojej veľkej úlohy v živote ako *ochranu*, *stráženie*, *uchovávanie*, verzus *objavovanie*, *tvorenie*, *vylepšovanie*. Ukazuje smer “hore” na stupnici času do minulosti alebo do budúcnosti? Je to rozdiel, ktorý podfarbuje všetko, zapúšťa úponky všade.

⁵ pozri: 183 Mimo laboratória

Preto som vždy trval na tom, napríklad, že ak chcete začať hovoriť o “etike UI”, mali by ste hovoriť o tom, ako chcete *vylepšiť* dnešnú situáciu pomocou UI, namiesto iba chránenia rôznych vecí pred pokazením. Akonáhle prijmete kritérií púheho súhlasu, začnete strácať zo zretela svoje ideály – prestanete vidieť nesprávne a správne, a začnete vidieť jednoducho “iné” a “rovnaké”.

Ešte by som tvrdil, že tento základný psychologický rozdiel je jedným z dôvodov, prečo akademické kruhy, ktoré prestanú robiť aktívny pokrok, majú sklon stať sa *nevraživými*. (Prinajmenšom podľa jemných štandardov vedy. Ničenie *povesti* je podľa historických štandardov mierne; väčšina defenzívnych systémov viery šla priamo po krku.) Ak veľké otrasy neprichádzajú dosť často na to, aby pravidelne vyzdvihovali mladých vedcov na základe zásluh namiesto konformity, daný odbor prestane odolávať štandardnej degenerácii⁶ do autority. Keď sa nerobí veľa objavov, nie je celý deň čo robiť okrem lovenia kacírov.

Aby vaše duševné zdravie malo čo najväčší prospech z postoja objavovať / tvoriť / vylešovať, musíte *naozaj robiť pokrok*, nielen dúfať, že príde.

Pôvodný článok s komentármi⁷

⁶ pozri: 69 Každá kauza chce byť sektou

⁷ http://lesswrong.com/lw/lz/guardians_of_the_truth/

Kapitola 71. Strážcovia genofondu

Ako všetci vzdelaní obyvatelia 21. storočia, asi ste počuli o druhej svetovej vojne. Možno si pamätáte, že Hitler a nacisti mali v pláne uskutočniť romantizovaný proces evolúcie, vyšľachtiť novú nadradenú rasu, nadľudí, silnejších a chytrejších než všetko, čo existovalo dovtedy.

V skutočnosti je toto rozšírený omyl. Hitler veril, že árijskí nadľudia *už v minulosti existovali* – severský stereotyp, blondavé modrooké šelmy – ale boli *znečistení* miešaním sa s nečistými rasami. Bol to rasový Pád z Milosti.

Vypovedá to niečo o tom, nakoľko je západná civilizácia presiaknutá predstavou *pokroku*, keď človek počúva o nacistickej eugenike a počuje: “Pokúšali sa vyšľachtiť nadčloveka.” Vy, milí čitatelia – keby ste na tom vy boli tak biedne, že by ste podporovali násilnú eugeniku – vy by ste sa pokúsili stvoriť nadčloveka. Pretože umiestňujete svoje ideály do budúcnosti, nie do minulosti. Pretože ste *tvoriví*. Myšlienka vrátiť sa šľachtením späť k nejakému severskému archetypu pred tisíc rokov by vám ani nenapadlo ako možnosť – čo, iba *Vikingovia*? To je *všetko*? Keby ste na tom boli tak biedne, že by ste boli ochotní zabíjať, tak by ste sa sakra pokúsili dosiahnuť výšku, ktorú nikto pred vami nedosiahol, inak by to celé bolo zbytočné, že? Nuž, to je jeden z dôvodov, prečo nie ste nacisti, drahí čitatelia.

Vypovedá to niečo o tom, ako ťažké je pre relatívne zdravého predstaviť si sám seba v koži relatívne chorého, keď počujeme o nacistoch, a prekrútime ten príbeh tak, že z nich urobíme vadných transhumanistov.

Komunisti boli tí vadní transhumanisti. “Nový sovietsky človek” a také veci. Nacisti boli v tomto príbehu celkom jednoznačne biokonzervatívci.

Pôvodný článok s komentármi¹

¹ http://lesswrong.com/lw/m0/guardians_of_the_gene_pool/

Kapitola 72. Strážcovia Ayn Rand

Skeptikom predstava, že rozum môže viesť k sekte, pripadá absurdná. Príznaky sekty sú o 180 naopak oproti rozumu. Ale ja vám predvediem, ako sa to nielenže môže stať, ale sa aj stalo, a to skupine, ktorú by ste mohli považovať za tú najnepravdepodobnejšiu sektu v dejinách. Je to lekcia, čo sa stane, keď sa pravda stane dôležitejšou než hľadanie pravdy.

—Michael Shermer, Najnepravdepodobnejšia sekta v dejinách

Myslím si, že Michael Shermer vysvetľuje Objektivizmus príliš zložito. Dostanem sa k tomu.

Romány Ayn Rand oslavujú technológiu, kapitalizmus, odpor jednotlivca voči Systému, obmedzenú vládu, súkromné vlastníctvo, sebeckosť². Jej vrcholný literárny hrdina John Galt bol <PREZRADENIE ZÁPLETKY> vedec, ktorý vynášiel novú formu lacnej obnoviteľnej energie, ale potom ju odmietol dať svetu, pretože zisk by bol rozkradnutý na podporu skorumpovanej vlády. </PREZRADENIE ZÁPLETKY>

A potom sa – nejako – toto celé zvrátilo na morálny a filozofický “uzavretý systém” s Ayn Rand v strede. Pojem “uzavretý systém” nie je moje obvinenie; je to pojem, ktorý používa Inštitút Ayn Rand na opísanie Objektivizmu. Objektivizmus je definovaný dielami Ayn Rand. Teraz, keď je Rand mŕtva, Objektivizmus je uzavretý. Ak v ľubovoľnom ohľade nesúhlasíte s dielami Rand, nemôžete byť Objektivistami.

Max Gluckman raz povedal: “Veda je každá disciplína, v ktorej hlupáci tejto generácie dokážu prekročiť hranicu, ktorú dosiahli génovia predchádzajúcej generácie.” Veda sa posúva dopredu ničením svojich hrdinov, Newton padol za obeť Einsteinovi. Každý mladý fyzik sníva o tom, že bude novým šampiónom, ktorého sa budúci fyzici budú snažiť zvrhnúť z trónu.

Filozofickým vzorom Ayn Rand bol Aristoteles. No, možno bol Aristoteles nádejným mladým matematickým talentom pred 2350 rokmi, ale od jeho čias matematika urobila badateľný pokrok. Bayesovská teória pravdepodobnosti je kvantitatívna logika, v ktorej je Aristotelova kvalitatívna logika iba špeciálnym prípadom; nie je však žiaden náznak, že by Ayn Rand vedela o bayesovskej teórii pravdepodobnosti, keď písala svoj magnum opus, *Atlas pokrčil plecami*. Rand písala o “rozumnosti”, ale neoboznámila sa s moderným výskumom heuristiky a skreslení. Ako môže hocikto tvrdiť, že je majster racionalista, a pritom nevedieť nič o takýchto základných predmetoch?

“Počkaj chvíľu,” namietla čitateľ, “to nie je fér! *Atlas pokrčil plecami* bol uverejnený v roku 1957! Vtedy ešte prakticky nikto nevedel o Bayesovi.” Ba. Nabudúce mi povie, že Ayn Rand zomrela v roku 1982 a nemala šancu prečítať si knihu *Úsudok v neistote: Heuristika a skreslenia*, ktorá vyšla v tom istom roku.

Veda nie je fér. To som tým chcel povedať. Ašpirujúci racionalista v roku 2007 začína s obrovskou výhodou oproti ašpirujúcemu racionalistovi z roku 1957. Podľa toho poznáme, že nastal pokrok.

Mne osobne myšlienka dobrovoľného prijatia systému vyslovene zviazaného s názormi jedného človeka, ktorý je mŕtvy, pripadá niekde medzi hlúposťou a samovražednosťou. Počítač nemá ani päť rokov a už je zastaralý.

² pozri: 261 Falošná morálka

Tá živosť, ktorú Rand obdivovala na vede, na obchode, na každej železnici, ktorá nahradila cestu pre konské záprahy, na každom mrakodrape postavenom s *novou* architektúrou – to všetko vychádza z princípu *prekonania dávnych majstrov*. Ako môže existovať veda, ak najmúdrejší vedec, ktorý kedy bude, už žil? Kto by mohol postaviť siluetu New Yorku, ktorú Rand tak obdivovala, keby najvyššia budova, aká bude kedy existovať, už bola postavená?

Napriek tomu Ayn Rand neuznávala nikoho väčšieho, ani v minulosti, ani v prichádzajúcej budúcnosti. Rand, ktorá začala obdivom rozumu a jednotlivca, skončila vyhánaním každého, kto sa jej opovážil protirečiť. Shermer³: “[Barbara] Branden si spomenula na večer, keď jeden z priateľov Rand poznamenal, že sa mu páči hudba Richarda Straussa. ‘Keď koncom večera odišiel, Ayn povedala, reagujúc svojím čoraz typickejším spôsobom: “Teraz rozumiem, prečo on a ja nikdy nemôžeme byť spriaznené duše. Rozdiel v našich zmysloch života je príliš veľký.” Často ani nepočkala, kým priateľ odišiel, než urobila takúto poznámku.”

Zdá sa, že Ayn Rand sa časom zmenila.

Rand vyrástla v Rusku, a videla bolševickú revolúciu na vlastné oči. Vo veku 21 rokov dostala víza na návštevu príbuzných v Amerike a nikdy sa nevrátila. Je ľahké nenávidieť autoritárstvo, keď ste obeťou. Je ľahké držať palce slobode jednotlivca, keď vy sami ste utláčaní.

Vyžaduje si omnoho silnejšiu povahu báť sa authority, keď máte moc. Keď sa na vás ľudia obracajú s otázkami, je ťažšie povedať?: “Čo do čerta viem ja o hudbe? Som spisovateľka, nie skladateľka,” alebo: “Je ťažké vidieť, ako obľuba hudobného diela môže byť nepravdivá⁴.”

Keď ste vy tá, ktorá utláča tých, čo sa vás opovážia uraziť, uplatňovanie moci vyzerá akosi omnoho *zdôvodniteľnejšie* než keď ste boli tá utláčaná. Akosi vám sami napadajú všetky možné výborné zdôvodnenia⁵.

Michael Shermer ide do detailov, ako si myslí, že Randovej filozofia skončila úpadkom do sektárstva. Shermer napríklad hovorí (pôsobí to tak), že Objektivizmus zlyhal, pretože Rand si myslela, že istota je možná, zatiaľ čo veda si nikdy nie je istá. V tomto so Shermerom nesúhlasím. Chemická teória atómov je sakra istá. Ale chemici sa nestali sektou.

Ja si vlastne myslím, že Shermer sa stal obeťou sklonu prisudzovať⁶, keď predpokladá, že existuje nejaký konkrétny vzťah medzi Randovej filozofiou a spôsobom, ako z jej nasledovníkov stala sekta. Každá kauza sa chce stať sektou.⁷

Ayn Rand ušla zo Sovietskeho Zväzu, napísala knihu o individualizme, ktorá sa mnohým ľuďom páčila, dostala veľa komplimentov a vytvorila si kliku obdivovateľov. Jej obdivovatelia nachádzali krajšie a krajšie veci, ktoré o nej hovorili (šťastná špirála smrti⁸) a jej sa to príliš páčilo, než aby im povedala, nech sklapnú. Zistila, že má moc utláčať tých, ktorí sa jej nepáčili a neodolala pokušeniu moci.

Ayn Rand a Nathaniel Branden mali tajný mimomanželský pomer. (So súhlasom oboch svojich partnerov, čo v mojich očiach znamená veľa. Ak chcete z tohto urobiť “problém”, museli by ste upresniť, že partneri boli *nešťastní* – a ešte aj tak do toho ostatných nič nie je.) Keď sa ukázalo, že Branden “podvádzal” Rand

³ http://www.2think.org/02_2_she.shtml

⁴ pozri: 2 Rozumné čítanie

⁵ pozri: 135 Falošné optimalizačné kritériá

⁶ pozri: 50 Chyba prisudzovania

⁷ pozri: 69 Každá kauza chce byť sektou

⁸ pozri: 63 Afektívne špirály smrti

ešte s inou ženou, Rand sa rozčertila a exkomunikovala ho. Mnohí Objektivistu sa odtrhli, keď sa správa o afére stala verejnou.

Kto zostal s Rand, namiesto nasledovania Brandena alebo celkového opustenia Objektizmu? Jej *najsilnejší* podporovatelia. Kto odišiel? Predchádzajúce umiernené hlasy. (Ochladzovanie skupinových názorov vyparovaním.⁹) Potom už mala Rand svoju zvyšnú kliku absolútne v hrsti a nebolo dovolené žiadne spochybňovanie.

Jediná nevšedná vec na celej tejto udalosti je, aké všedné to bolo.

Možno si myslíte, že myšlienkový systém, ktorý chváli "rozum" a "racionalitu" a "individualizmus" tým nejakú získa nejakú špeciálnu imunitu...?

Nuž, nestalo sa.

Fungovalo to asi rovnako dobre ako dať nápis "Zima" na chladničku, ktorá nie je zapnutá.

Aktívna snaha potrebná na odolanie sklzu do entropie tam nebola a prirodzene nasledoval rozpad.

A ak toto nazývate "najnepravdepodobnejšia sekta v dejinách", akurát urážate skutočnosť¹⁰.

Nech je to lekcia pre nás všetkých: Oslavovať¹¹ "rozumnosť" nič neznamená. Dokonca aj povedať: "Musíš zdôvodniť svoje názory Rozumom, nie iba súhlasiť s Veľkým Vodcom" akurát spustí automatický programček, ktorý vezme všetko, čo Veľký Vodca povedal a vygeneruje zdôvodnenie, ktoré vaši spolunasledovníci budú považovať za Rozum-né.

Kde sa teda dá nájsť skutočné umenie rozumnosti? Matematickým štúdiom teórie pravdepodobnosti a teórie rozhodovania. Vstrebaním kognitívnych vied ako je evolučná psychológia, alebo heuristik a skreslení. Čítaním historických kníh...

"Študujte vedu, nie iba mňa!" je asi tá najdôležitejšia rada, ktorú Ayn Rand mohla dať svojim nasledovateľom, ale nedala. Niet takého človeka, ani nikdy nebolo, ktorého plec by uniesli celú váhu skutočnej vedy s mnohými prispievateľmi.

Myslím si, že stojí za zmienku, že literárni hrdinovia Ayn Rand boli architekti a inžinieri; John Galt, jej vrcholná postava, bol fyzik; a predsa samotná Ayn Rand nebola veľká vedkyňa. Pokiaľ viem, nebola zvlášť dobrá v matematike. Nemohla dúfať, že bude súperom svojim vlastným hrdinom. Možno práve tam začala strácať zmysel pre Tsuyoku Naritai¹².

Zatiaľ čo ja, viete, obdivujem drzosť Francisa Bacona¹³, ale ponechávam si schopnosť chvastavo priznať: "Keby som sa mohol vrátiť v čase a nejakú dosiahnuť, že by Francis Bacon pochopil problém, na ktorom práve pracujem¹⁴, oči by mu vyskočili z jamiek ako zátky od šampanského a vybuchli by."

Obdivujem Newtonove úspechy. Môj postoj k volebnému právu pre ženy mi však bráni prijať Newtona ako vzor morálky. Rovnako ako moja znalosť bayesovskej pravdepodobnosti mi bráni vnímať Newtona ako najvyšší neporaziteľný zdroj matematického poznania. A moja znalosť špeciálnej relativity, hoci je slabá a málo používaná, mi bráni vidieť Newtona ako najvyššiu autoritu vo fyzike.

⁹ pozri: 66 Ochladzovanie skupinových názorov vyparovaním

¹⁰ pozri: 196 Mysli ako skutočnosť

¹¹ pozri: 33 Veda ako uniforma

¹² pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

¹³ pozri: 64 Odolajte šťastnej špirále smrti

¹⁴ <http://intelligence.org/AIRisk.pdf>

Newton reálne nemal šancu objaviť žiadnu z myšlienok, ktorými sa nad ním vyvyšujem – *ale pokrok nie je fér! To je pointa!*

Veda má hrdinov, ale nemá bohov. Velké Mená nie sú naši nadriadení, dokonca ani naši súper, sú to mílniky, okolo ktorých sme prešli na našej ceste; a tým najdôležitejším mílnikom je hrdina, ktorý ešte len príde.

Byť ďalším mílnikom na ceste ľudstva je to najlepšie, čo možno o niekom povedať; ale toto sa Ayn Rand zdalo príliš slabé. A tak sa z nej stal iba Najvyšší Prorok.

Pôvodný článok s komentármi¹⁵

¹⁵ http://lesswrong.com/lw/m1/guardians_of_ayn_rand/

Kapitola 73. Dva koany o sektách

Súvisí s kapitolou: Každá kauza chce byť sektou¹

Nováčika racionalistu študujúceho u majstra Ougiho pokarhal priateľ slovami: “Celý čas tráviš počúvaním svojho majstra a rozprávaním o ‘rozumnom’ tomto a ‘rozumnom’ tamtom – dostal si sa do sekty!”

Nováčika to hlboko znepokojilo. Slová: “Dostal si sa do sekty!” mu zneli v ušiach, keď si večer ľahol spať, a dokonca aj vo sne.

Ďalší deň nováčik prišiel k Ougimu, vyrozprával mu, čo sa stalo a povedal: “Majster, stále ma trápi obava, že toto celé je naozaj sekta, a že tvoje učenie je iba dogma.”

Ougi odpovedal: “Ak nájdeš na ceste ležať kladivo a predáš ho, môžeš si zapýtať nízku cenu alebo vysokú cenu. Ak si však to kladivo necháš a používaš ho na zatĺkanie klinec, kto môže pochybovať o jeho hodnote?”

Nováčik odpovedal: “Vidíš, toto je presne jedna z tých vecí, ktoré ma znepokojujú – tvoje tajomné zenové odpovede.”

Ougi povedal: “Dobre teda. Budem hovoriť jasne a predložím dokonale rozumné argumenty, ktoré dokážu, že si sa nedostal do sekty. Ale najprv si musíš nasadiť tento hlúpy klobúk.”

Ougi podal nováčikovi obrovský štyridsaťlitrový kovbojský klobúk.

“Ehm, majster...” povedal nováčik.

“Keď ti všetko vysvetlím,” povedal Ougi, “uvidíš, prečo je to potrebné. Alebo môžeš v noci naďalej bdieť a uvažovať, či je toto sekta.”

Nováčik si nasadil kovbojský klobúk.

Ougi povedal: “Ako dlho budeš opakovať moje slová a ignorovať ich zmysel? Neusporiadané myšlienky začínajú ako pocit pripútanosti k obľúbeným záverom. Si príliš úzkostlivý ohľadom svojho sebaobrazu racionalistu. Prišiel si ku mne hľadať upokojenie. Keby si bol naozaj zvedavý² a nevedel by si, či je to tak alebo onak, napadli by ti nejaké spôsoby, ako vyriešiť svoje pochybnosti³. Pretože si si potreboval vyriešiť svoju kognitívnu disonanciu, bol si ochotný nasadiť si hlúpy klobúk. Keby som bol zlým človekom, mohol som ťa nechať zaplatiť sto strieborných mincí. Keď sa budeš sústrediť na otázky ohľadom skutočného sveta, cena alebo bezcennosť tvojho chápania bude čoskoro zrejmá. Si ako šermiar, ktorý sa stále obzerá, aby videl, či sa mu niekto neposmieva...”

“To stačí,” povedal nováčik.

“Ty si chcel dlhšiu verziu,” povedal Ougi.

Tento nováčik neskôr nahradil Ougiho a stal sa známym ako Ni no Tachi. Odvtedy nikdy nedovolil svojim žiakom citovať jeho slová v diskusiách, hovoriac: “Používajte techniky ale nehovorte o nich.”

¹ pozri: 69 Každá kauza chce byť sektou

² pozri: 123 Meditácia o zvedavosti

³ pozri: 121 Správne použitie pochybnosti

* * *

Nováčik racionalista prišiel k majstrovi Ougimu a povedal: "Majster, obávam sa, že naše racionalistické dódžó je... no... trochu sektárske."

"To je vážna obava," povedal Ougi.

Nováčik chvíľu čakal, ale Ougi nepovedal nič viac.

Nováčik teda opäť prehovoril: "Myslím tým, mrzí ma to, ale keď musíme nosiť tieto rúcha a kapucne - skrátka to vyzerá akoby sme boli nejakí slobodomurári alebo také čosi."

"Aha," povedal Ougi, "rúcha a kapucne."

"No *áno*, rúcha a kapucne," povedal nováčik. "Skrátka to vyzerá hrozne nerozumne."

"Zodpoviem všetky tvoje obavy," povedal majster, "ale najprv si musíš nasadiť tento hlúpy klobúk." Ougi vytiahol čarodejnicky klobúk, vyšívany mesiacmi a hviezdami.

Nováčik si vzal klobúk, pozrel naň a potom frustrrovane vybuchol: "*A ako môže toto pomôcť?*"

"Keďže sa tak zaujímaš o interakcie medzi odevom a teóriou pravdepodobnosti," povedal Ougi, "nemalo by ťa prekvapiť, že musíš mať špeciálny klobúk, aby si to pochopil."

Keď nováčik dosiahol hodnosť doktoranda, prijal meno Bouzo a bol ochotný diskutovať o rozumnosti iba keď nosil šašovský oblek.

Pôvodný článok s komentármi⁴

⁴ http://lesswrong.com/lw/m4/two_cult_koans/

Kapitola 74. Aschov pokus o konformite

Solomon Asch, ktorého pokusy začali v 1950-tych rokoch a odvtedy sa mnohokrát replikovali, zvýraznil jav dnes nazývaný “konformita”. V klasickom pokuse pokusná osoba vidí úlohu ako na tomto obrázku: Ktorá z čiar A, B, C je rovnako dlhá ako čiara X? Máte chvíľu na to, aby ste určili svoju odpoveď...

Trik je v tom, že pokusná osoba sedí vedľa ďalších ľudí pozerajúcich na obrázok – napohľad ďalšie pokusné osoby, v skutočnosti experimentátorovi spolupracovníci. Ostatné “pokusné osoby” v tomto pokuse jedna za druhou hovoria, že čiara C vyzerá rovnako dlhá ako čiara X. Skutočná pokusná osoba sedí ako predposledná. Koľko ľudí by v takejto situácii povedalo “C” - jednoznačne nesprávnu odpoveď, ktorá súhlasí s jednohlasnou odpoveďou ostatných pokusných osôb? Aké percento si myslíte, že by to bolo?

Tri štvrtiny pokusných osôb v Aschovom pokuse dalo “konformnú” odpoveď aspoň raz. Tretina pokusných osôb odpovedala konformne na viac ako polovicu otázok.

Interview po pokuse ukázalo, že väčšina pokusných osôb tvrdila, že naozaj neveria svojom konformným odpovediam, niektorí však povedali, že si naozaj myslia, že konformná odpoveď bola tá správna.

Ascha tieto výsledky znepokojili:

Že sme v našej spoločnosti našli taký silný sklon ku konformite... je znepokojujúce. Vyvoláva to otázky o našom spôsobe vzdelávania a o hodnotách, ktorými sa riadi naše správanie.

Nie je to triviálna otázka, či sa pokusné osoby v Aschovom pokuse správali *nerozumne*. Robertova Aumannova veta o súhlase ukazuje, že úprimní bayesovci sa nemôžu zhodnúť, že sa nezhodnú – ak navzájom vedia o svojich odhadoch pravdepodobnosti, majú rovnaký odhad pravdepodobnosti. Aumannova veta o súhlase bola dokázaná vyše dvadsať rokov po Aschovom pokuse, ale iba formalizuje a posilňuje intuitívne jasnú vec – že názory iných ľudí sú často legitímne indície.

Keby ste sa pozerali na horeuvedený diagram a vedeli by ste *ako fakt*, že zvyšní ľudia v pokuse sú úprimní a vidia ten istý diagram ako vy, a že traja ďalší povedali, že C je rovnako veľké ako X, aká je šanca, že *iba* vy ste jediný, ktorý má pravdu? Ja si nemyslím, že mám v zrakovom hodnotení nejakú výhodu – nemyslím si, že by som bol lepší než priemerný človek v posudzovaní, či sú dve čiary rovnako dlhé. Z hľadiska individuálnej rozumnosti dúfam, že by som si všimol svoj veľký zmätok¹, a potom priradil pravdepodobnosť > 50 % väčšinovému hlasu.

Z hľadiska skupinovej rozumnosti mi pripadá, že úprimný racionalista by mal povedať: “Aké prekvapujúce, že mne pripadá B rovnako dlhé ako X. Ale ak sa pozeráme na rovnaký diagram a odpovedáme úprimne, nemám dôvod veriť, že môj odhad je lepší než váš.” Tá posledná veta je dôležitá – je to omnoho slabšie tvrdenie nesúhlasu než: “Aha, ja vidím optický klam – samozrejme rozumiem, prečo si myslíte, že je to C, ale správna odpoveď je B.”

Konformné pokusné osoby v týchto pokusoch nie sú *automaticky* usvedčené z nerozumnosti na základe toho, čo som doteraz opísal. Ale ako ste mohli

¹ pozri: 27 Tvoja sila racionalistu

čakať, čert je skrytý v detailoch výsledkov pokusu. Podľa meta-analýzy vyše sto replikácií od Smitha a Bonda (1996):

Konformita sa prudko zvyšuje po 3 spolupracovníkoch, ale potom sa ďalej nezvyšuje po 10-15 spolupracovníkoch. Keby boli ľudia rozumne konformní, potom by názor 15 ďalších ľudí mal byť podstatne silnejšou indíciou než názor ďalších 3 ľudí.

Pridanie jediného nesúhlasiaceho – iba jednej osoby, ktorá dá správnu odpoveď, alebo hoci len nesprávnu odpoveď odlišnú od nesprávnej odpovede skupiny – znižuje konformitu veľmi prudko, na 5-10 %. Ak používate nejakú intuitívnu verziu Aumannovej vety o zhode a myslíte si, že ak 1 človek nesúhlasí s 3 ľuďmi, tak tí 3 majú pravdepodobne pravdu, potom by ste si vo väčšine prípadov mali myslieť to isté keď 2 ľudia nesúhlasia so 6 ľuďmi. (Nie je to automaticky pravda, ale pravda *ceteris paribus*.) Na druhej strane, ak sú ľudia emocionálne nervózni z toho, že by sami vyčnievali z davu, potom je ľahké ukázať, ako jediný človek, ktorý s vami súhlasí, alebo hoci len jediný človek, ktorý nesúhlasí so skupinou, vás urobí omnoho menej nervóznymi.

Nie je prekvapivé, že pokusné osoby v situácii s jedným nesúhlasiacim si nemysleli, že ich nekonformitu ovplyvnil alebo umožnil tento nesúhlasiaci. Tak ako 90 % šóferov, ktorí si myslia, že patria medzi lepších 50 %, niektorí z nich môžu mať pravdu, ale nie všetci. Ľudia si neuvedomujú príčiny svojej konformity alebo nesúhlasu, čo spochybňuje argument, že sú to prejavy rozumnosti. Napríklad pri hypotéze, že ľudia sa spoločensky-rozumne rozhodujú klamať, aby nevyčnievali z davu, zdá sa, že (aspoň niektoré) pokusné osoby v podmienkach jedného nesúhlasiaceho nepredpokladajú, že by použili takúto “vedomú stratégiu”, keby sa stretli s jednohlasnou opozíciou.

Keď osamelý nesúhlasiaci náhle prepol do *konformity so skupinou*, miera konformity pokusných osôb sa vrátila do rovnakej výšky ako v prípade bez nesúhlasiaceho. Byť prvým nesúhlasiacim je hodnotná (a drahá!) spoločenská služba, ale musíte v nej vydržať.

Konzistentne v rámci pokusov a medzi nimi, skupiny zložené zo samých žien (aj pokusná osoba aj spolupracovníčky) sú omnoho konformnejšie než skupiny zložené zo samých mužov. Asi polovica žien je konformná vo viac než polovici prípadov, oproti tretine mužov. Ak by ste argumentovali, že priemerná pokusná osoba je rozumná, potom sú zrejme ženy príliš prispôsobivé a muži príliš neprispôsobiví, takže žiadna skupina nie je naozaj *rozumná*...

Pokusy na rôznych skupinách (napríklad telesne postihnutá pokusná osoba vedľa telesne postihnutých spolupracovníkov) podobne ukázali, že konformita je výrazne silnejšia medzi členmi rovnakej skupiny.

Konformita je nižšia pri očividných diagramoch, ako ten navrchu tejto stránky, oproti diagramom, kde je chyba menej zrejmalá. To je ťažké vysvetliť, ak (všetky) pokusné osoby robia spoločensky rozumné rozhodnutie nevyčnievať z davu.

Dodatok: Paul Crowley ma upozornil, že keď pokusné osoby môžu odpovedať spôsobom, ktorý skupina nevidí, konformita tiež klesne, čo je tiež argument proti Aumannovskej interpretácii.

Asch, S. E. (1956). Štúdie o nezávislosti a konformite: Menšina jedného proti jednohlasnej väčšine. *Psychological Monographs*, 70.

² http://www.radford.edu/~jaspelme/_private/gradsoc_articles/individualism_collectivism/conformity%20and%20culture.pdf

Bond, R. a Smith, P. B. (1996.) Kultúra a konformita: Meta-analýza štúdií požívajúcich Aschovu (1952b, 1956) úlohu hodnotenia čiary.² *Psychological Bulletin*, 119, 111-137.

Pôvodný článok s komentármi³

³ http://lesswrong.com/lw/m9/aschs_conformity_experiment/

Kapitola 75. O vyjadrovaní svojich obáv

Tá strašidelná vec na Aschovom pokuse s konformitou¹ je, že môžete primäť veľa ľudí, aby povedali, že čierne je biele, ak ich dáte do miestnosti plnej ľudí, ktorí hovoria to isté. Nádejná vec na Aschovom pokuse s konformitou je, že jediný nesúhlasiaci výrazne znížil mieru konformity, dokonca aj keď ten nesúhlasiaci dával inú nesprávnu odpoveď. A *zdrvujúca* vec je, že nesúhlas sa počas pokusu *nenaučil* – keď tento osamelý nesúhlas začal súhlasieť so skupinou, miera konformity sa opäť vrátila.

Byť hlasom nesúhlasu môže skupine priniesť skutočný úžitok. Ale zároveň (je povestné, že) niečo stojí. A potom sa toho musíte držať. Navyše sa môžete myliť.

Nedávno som mal zaujímavú skúsenosť, keď som začal diskutovať o nejakom projekte s dvoma ľuďmi, ktorí predtým sami robili nejaké plány. Myslel som si, že sú príliš optimistickí² a dal som k projektu kopec návrhov typu “pridajme sem pre istotu rezervu”. Čoskoro šiel okolo štvrtý chlap, ktorý mal jedného z tých dvoch odviesť domov, a začal dávať návrhy. V tomto bode som mal náhle vhlád, ako sa skupiny stávajú príliš sebaisté, pretože vždy, keď som vzniesol možný problém, ten štvrtý chlap povedal: “Neboj sa, to určite zvládneme!” alebo niečo podobne ubezpečujúce.

Jednotlivci pracujúci osamote budú mať prirodzene pochybnosti. Budú rozmýšľať: “Dokážem naozaj urobiť XYZ?”, pretože nie je nič neslušné na tom, keď človek spochybňuje *vlastné* schopnosti. Ale keď dvaja neistí ľudia vytvoria skupinu, je slušné hovoriť pekné a ubezpečujúce veci, a neslušné spochybňovať schopnosti toho druhého. Spolu sa stanú optimistickjšími než by bol každý z nich osobitne, pochybnosti každého z nich sa uhasia napohľad dôveryhodným ubezpečením toho druhého, neuvedomujúc si, že ten druhý mal na začiatku rovnaké pochybnosti.

Najdesivejšia možnosť, ktorú naznačuje Aschov pokus s konformitou je príznak, že každý súhlasí so skupinou, presvedčený sebaistými hlasmi druhých, dávajúc si pozor, aby nedal najavo vlastné pochybnosti – neuvedomujúc si, že ostatní potláčajú podobné obavy. Toto sa nazýva “pluralistická ignorancia”.

Robin Hanson a ja máme dlhodobú diskusiu o tom, kedy presne by sa aspirujúci racionalisti mali odvážiť nesúhlasieť. Ja sa prikláňam k všeobecne obľúbenému postoju, že nemáte na výber a musíte si tvoriť svoje vlastné názory. Robin Hanson odporúča omnoho rebelskejší názor, že vy – nie iba iní ľudia – by ste mali zvážiť, že druhí môžu byť múdrejší. Bez ohľadu na naše rôzne debaty obaja súhlasíme, že Aumannova veta o zhode naznačuje, že spoločné vedomie faktického³ nesúhlasu ukazuje, že *niekto* musí byť nerozumný⁴. Bez ohľadu na to, aké čudné pohľady na nás vrhajú, stojíme si ohľadom skromnosti za svojím: Zabudnite na to, čo vám ľudia hovoria o individualizme, *mali* by ste venovať pozornosť tomu, čo si druhí ľudia myslia.

Ehm. Pointa je, že pre racionalistu je nesúhlas so skupinou vážna vec. Nemôžete to zmietnuť slovami: “Každý má právo na svoj vlastný názor.”

¹ pozri: 74 Aschov pokus o konformite

² pozri: 7 Klam plánovania

³ pozri: 2 Rozumné čítanie

⁴ http://lesswrong.com/lw/gr/the_modesty_argument/

Myslím si, že najdôležitejšia lekcia, ktorú si môžeme vziať z Aschovho pokusu je rozlišovať medzi “vyjadrením obavy” a “nesúhlasom”. Upozorniť na vec, ktorú ostatní nevyslovili, neznamená sľúbiť nesúhlas so skupinou na konci debaty.

Ideálny bayesovský proces konvergenencie zahŕňa výmenu indícií, ktorá je pre počúvajúceho nepredvídateľná. Aumannova zhoda sa týka iba *spoločných vedomostí*, kde vy viete, ja viem, vy viete, že ja viem, atď. Hansonov článok “Nemôžeme predvídať nesúhlas”⁵ poskytuje obraz, ako zvláštne by vyzeralo sledovať ideálnych racionalistov, ako konvergujú k odhadu pravdepodobnosti; nevyzerá to ako keď sa dvaja obchodníci na trhu dohadujú na cene.

Žiaľ, *spoločensky* príliš nerozlišujeme medzi “vyjadrením obavy” a “nesúhlasom”. Skupina racionalistov môže súhlasiť, že budú predstierať, že je tam rozdiel, ale ľudia takto nie sú naozaj nastavení. Keď raz prehovoríte, spáchali ste spoločensky neodvolateľný čin; stali ste sa vytrčajúcim klincom, nesúlalom v pohodlnej harmónii skupiny, a to nemôžete odvolať. Každý, koho urazilo vaše vyjadrenie obáv o ich schopnosti úspešne dokončiť úlohu XYZ, bude voči vám dodatočne pravdepodobne chovať rovnaké množstvo trpkosti aj keď nakoniec poviete: “Nie je problém, súhlasím so skupinou.”

Aschov pokus ukazuje, že schopnosť nesúhlasu inšpirovať druhých je skutočná. Aschov pokus ukazuje, že moc konformity je skutočná. Ak sa každý zdrží vyjadrenia svojich súkromných obáv, to skupinu iste povedie do šialenstva. Ale dejiny sú plné príkladov, čo to stojí byť prvým, či dokonca druhým, ktorý povie, že cisár je nahý. Ľudia nemajú ani vrozený sklon rozlišovať medzi “vyjadrením obavy” a “nesúhlasom so všeobecnou vedomosťou”; tento rozdiel je umelým výtvorom racionalistu. Ak si čítate cynickejší prúd kníh o sebazdokonaľovaní (napríklad Machiavelliho *Vládár*), budú vám radiť celkom maskovať svoju nekonformitu, nevyjadrovať svoje obavy na začiatku a súhlas na konci. Ak preukážete skupine službu, že budete tým, ktorý dá hlas zrejým problémom, nečakajte, že vám za to skupina poďakuje.

Nesúhlas má svoje ceny a prínosy – či už “nesúhlasíte” alebo “vyjadrujete obavy” - a to rozhodnutie je na vás.

Pôvodný článok s komentármi⁶

⁵ http://www.overcomingbias.com/2007/01/we_cant_foresee.html

⁶ http://lesswrong.com/lw/ma/on_expressing_your_concerns/

Kapitola 76. Osamelý nesúhlas

Súvisí s článkom a kapitolami: Argument skromnosťou¹, Kolaje “mimo vychodených kolají”², Aschov pokus s konformitou³

Aschov pokus s konformitou⁴ ukázal, že prítomnosť jediného nesúhlasiaceho dramaticky znížila výskyt “konformných” nesprávnych odpovedí. Individualizmus je ľahký, ukazuje pokus, keď máte vo svojej odlišnosti spoločnosť. Každá ďalšia osoba v miestnosti, okrem jednej, povie, že čierne je biele. Vy sa stanete druhou osobou, ktorá povie, že čierne je čierne. A cítite sa úžasne: vy dvaja, osamelí a vzdorovití rebeli, proti celému svetu! (Následné interview ukázali, že pokusné osoby v situácii s jedným nesúhlasiacim vyjadrili silný pocit kamarátstva s nesúhlasiacim – hoci si samozrejme nemysleli, že jeho prítomnosť ovplyvnila ich vlastnú nekonformnosť.)

Lenže k rebelii sa môžete *pripojiť* až keď sa niekto niekde stal *prvým* rebelom. Niekto musí povedať, že čierne je čierne po tom, čo počul *všetkých* ostatných, jedného za druhým, povedať, že čierne je biele. A toto – ukazuje pokus – je *omnoho* ťažšie.

Osamelý nesúhlas vám nepripadá ako chodiť do školy oblečený v čiernom⁵. Pripadá vám to ako ísť do školy v šašovskom kostýme.

To je ten rozdiel medzi *pridaním sa k vz bure* a *opustením svorky*.

Ak je nejaká vec, ktorú neznesiem, je to falošnosť – ak nejaký ten čas čítate *Overcoming Bias*, asi ste si to všimli. Nuž, osamelý nesúhlas musí byť jedna z najčastejšie a najostentatívnejšie predstieraných vlastností navôkol. Každý chce byť rebel.

Nechcem tu zhadzovať akt pridania sa k vz bure. Existujú vzbury, ku ktorým sa oplatí pridať. Vyžaduje si odvahu riskovať nesúhlas svojej skupiny rovesníkov, či ešte horšie, ich pokrčenie plecami. Treba však povedať, že ísť na rockový koncert nie je rebelstvo. Ale napríklad vegetariánstvo áno. Ja osobne nie som vegetarián, ale vážim si ľudí, ktorí sú, pretože si to vyžaduje výrazné množstvo tichej odvahy povedať ľuďom, že hamburger na večeru nemôže byť. (Aj keď v okolí San Francisca sa ľudia už bežne opýtajú.)

Ale aj tak, ak poviete ľuďom, že ste vegetarián, budú si myslieť, že rozumejú vašim motívom (aj keď nerozumejú) Môžu nesúhlasiť. Môžu sa uraziť, ak sa vám to podarí oznámiť príliš hrdo, alebo sa môžu uraziť skrátka preto, lebo sú veľmi urážliví. Ale vedia, ako sa k vám majú správať.

Keď niekto chodí do školy v čiernom, učitelia a ostatné deti rozumejú, akú rolu tým chce zaujať v ich spoločnosti. Je to Mimo Systému – veľmi štandardným spôsobom, ktorý každý pozná a chápe. Teda nie, viete, *naozaj* mimo systému. Je to Výzva Štandardnému Mysleniu štandardného typu, takže ľudia pohoršene povedia: “Nerozumím prečo...” ale nemusia naozaj myslieť na nič, na čo už nemysleli predtým. Ako sa hovorí: “Už niekedy nejaká ‘podvratná literatúra’, ktorú ste čítali, spôsobila, že ste zmenili nejaký politický názor?”

Naozajstnú odvahu si vyžaduje priniesť medzi ľudí okolo seba úplné *nepochopenie*, keď urobíte niečo, čo *nie je* iba Štandardná Vz bura číslo 37, niečo,

¹ http://lesswrong.com/lw/gr/the_modesty_argument/

² pozri: 82 Kolaje “mimo vychodených kolají”

³ pozri: 74 Aschov pokus o konformite

⁴ pozri: 74 Aschov pokus o konformite

⁵ pozri: 82 Kolaje “mimo vychodených kolají”

na čo nemajú pripravený scenár. Nehnevajú sa na vás ako na rebela, iba si myslia, že ste akýsi divný a odvrátia sa. Táto predstava vytvára omnoho hlbší strach. To je rozdiel medzi vysvetľovaním vegetariánstva a vysvetľovaním kryoniky⁶. Na svete existujú aj iní kryonici, niekde, ale nie sú tam vedľa vás. Musíte to vysvetliť sám, ľuďom, ktorí si myslia, že ste skrátka *divný*. Nie je to zakázané, ale je to mimo hraníc, o ktorých ľudia ani nerozmýšľajú. Vy si idete nechať zmraziť hlavu? Vy si myslíte, že vás to zachráni pred smrťou? Čo tým myslíte, že informácia v mozgu? He? Čo? *Šibe* vám?

Som v pokušení napísať post facto vysvetlenie pomocou evolučnej psychológie⁷: Mohli ste dať dokopy skupinku kamarátov a odísť od svojej bandy lovcov-zberačov, ale musieť ísť do lesa *sám* bol pravdepodobne rozsudok smrti – prinajmenšom z hľadiska rozmnožovania. Nemyslíme na to explicitne, ale to nie je v povahe evolučnej psychológie. Pridať sa k vzbure, o ktorej každý vie, je strašidelné, ale zďaleka nie také strašidelné ako robiť niečo naozaj odlišne. Nieco, čo v pradávných časoch mohlo skončiť nie rozdelením skupiny, ale vyhnaním vás samého.

Ako dokladá prípad kryoniky, strach myslieť naozaj odlišne je silnejší než strach zo smrti. Lovci-zberači museli byť pripravení čeliť smrti každý deň pri love veľkých cicavcov alebo len pri prechádzaní sa vo svete plnom dravcov. Potrebovali tú odvahu aby mohli žiť. Odvaha vzdorovať kmeňovým štandardným spôsobom myslenia, myslieť si veci, ktoré vyzerali naozaj divne – nuž, toto pravdepodobne svojim nositeľom rovnako neprospievalo. Nerozmýšľame nad tým explicitne; tak evolučná psychológia⁸ nefunguje. My ľudia sme skrátka postavení takým spôsobom, že viacerí z nás idú skákať padákom než sa prihlásiť na kryoniku.

A to ešte nie je tá najväčšia odvaha. Na svete existuje viac ako jeden kryonicista. Iba Robert Ettinger to musel povedať ako *prvý*.

Aby ste boli revolučným *vedcom*, musíte byť prvým človekom, ktorý bude protirečiť tomu, čo si myslí každý, koho poznáte. To nie jediná cesta k vedeckej sláve; je to pomerne vzácné aj medzi tými veľkými. Nikto sa nemôže stať vedeckým revolucionárom tým, že sa bude snažiť napodobňovať revolucionárov. Môžete sa tam dostať iba nasledovaním správnych odpovedí vo všetkých veciach, či je správna odpoveď revolučná alebo nie. Ale ak postupom času – ak po vstrebaní všetkej moci a múdrosti z vedomostí, ktoré už boli zhromaždené – ak po tom všetkom a s dávkou čistého šťastia zistíte, že vás vaše hľadanie čistej správnosti vedie na nové územie... *vtedy* máte príležitosť prejavíť odvahu zlyhať.

Toto je skutočná odvaha osamelého nesúhlasu, ktorú sa každá poondená rocková kapela na svete pokúša predstierať.

Samozrejme nie všetko, čo si vyžaduje odvahu, je dobrý nápad. Chce to odvahu skočiť z útesu, ale potom sa iba rozpleštíte.

Strach z osamelého nesúhlasu je brzdou k dobrým myšlienkam, ale nie každá nesúhlasná myšlienka je dobrá. Pozrite si od Robina Hansona Proti voľnomyšlienkárom⁹. Väčšina problému ako mať nové pravdivé vedecké myšlienky sa týka časti "pravdivé".

Naozaj nie je *potrebné* byť odlišný kvôli samotnej odlišnosti. Ak robíte veci odlišne iba keď vidíte mimoriadne dobrý dôvod, stále budete mať viac než dostatok problémov na celý zvyšok života.

⁶ <http://www.alcor.org/>

⁷ pozri: 137 Evolučná psychológia

⁸ pozri: 137 Evolučná psychológia

⁹ http://www.overcomingbias.com/2007/06/against_free_th.html

Existuje okolo nás niekoľko naozaj nefalšovaných skupín obrazoborcov. Cirkev SubGénia má napríklad asi ako skutočný cieľ *zmiast* bežných ľudí, nie iba uraziť ich. Existujú aj ostrovy skutočnej tolerance vo svete, napríklad stretnutia science fiction. *Existujú* istí ľudia, ktorí sa neboja oddeliť od svorky. Takých ľudí je v skutočnosti mnoho menej než tých, čo sa považujú za rebelov; ale existujú. A predsa sú vedeckí revolucionári omnoho zriedkavejší. Rozmýšľajte o tom.

No a *ja*, viete, ja *naozaj som* obrazoborec. Každý si to o sebe myslí, ale ja mám *pravdu*, viete. Ja by som *určite* nosil šašovský kostým do školy. Moje vážne rozhovory boli s knihami, nie s inými deťmi.

Ale ak si myslíte, že by ste *určite* nosili ten šašovský kostým, nebudte príliš hrdí ani na to! Znamená to iba, že musíte vynaložiť úsilie *opačným smerom*, aby ste sa vyhli príliš ľahkému nesúhlasu. To je to, čo musím robiť ja, aby som napravil svoju povahu. Iní ľudia majú dôvody, prečo si myslia to, čo si myslia, a celkom ich ignorovať je rovnako zlé ako báť sa im protirečiť. Nemali by ste chcieť skončiť ako voľnomyšlienkar¹⁰. To totiž nie je *cnosť* – iba skreslenie iným smerom.

Pôvodný článok s komentármi¹¹

¹⁰ http://www.overcomingbias.com/2007/06/against_free_th.html

¹¹ http://lesswrong.com/lw/mb/lonely_dissent/

Kapitola 77. Sektárske antisektárstvo

Súvisí s kapitolou: Každá kauza chce byť sektou¹

Vstúpiť do sekty je pravdepodobne jedna z najhorších vecí, ktoré sa vám môžu stať v modernom svete. V najlepšom prípade skončíte v skupine úprimných ale pomýlených ľudí, ktorí robia úprimnú chybu, ale inak sa správajú dobre, miniete kopec času a peňazí a nakoniec vám nezostane nič. Toto je vlastne opis neúspešnej začínajúcej firmy v Silicon Valley. Čo by vraj mala byť čertovsky otrasná skúsenosť, keď sa nad tým zamyslím. Takže áno, veľmi strašidelné.

Skutočné sekty sú nepomerne horšie. "Bombardovanie láskou" ako náborová technika, zameraná na ľudí, ktorí prekonávajú osobnú krízu. Spánková deprivácia. Vyčerpanosť vyvolaná ťažkou prácou. Ďalšie cestovanie, aby sa členovia izolovali od priateľov a rodiny. Denné stretnutia, kde sa spovedajú nečisté myšlienky. Nie je nezvyčajné, že sekta oberie člena o *všetky* peniaze – životné úspory aj mesačnú výplatu – čím ich núti do závislosti v jedle aj oblečení. Hladovka ako test za neposlušnosť. Vážne vymývanie mozgu a vážne ubližovanie.

Keď zvážim toto všetko, mal by som asi viac súcitiť s ľuďmi, ktorí sú hrozne nervózni z pridania sa k nejakému čudne vyzerajúcemu správaniu, že *možno vstupujú do sekty*. Nemalo by mi to liezť na nervy. Lenže lezie.

Bod číslo jedna: "Sekty" a "nie sekty" nie sú prirodzene oddelené druhy ako psy a mačky. Ak sa pozriete na hociktorý zoznam charakteristík sekty², nájdete tam prvky, ktoré by mohli ľahko opísovať politické strany a firmy – "členovia skupiny sú povzbudzovaní, aby nedôverovali kritike zvonku, lebo sú za ňou skryté motívy", "hierarchické autoritatívna štruktúra". Už som písal o skupinových zlyhaniach ako je polarizácia skupiny³, šťastné špirály smrti⁴, nekritickosť⁵, a ochladzovanie vyparovaním⁶, pričom každé z toho asi prispieva ku všetkým ostatným. Keď sa tieto zlyhania pritočia k sebe a spoja, ich spojením vznikne Super-Zlyhanie hlúpejšie než každá z jeho častí, ako Voltron⁷. Ale to nie je sektárska *podstata*; to je sektársky *atraktor*.

Psy sa rodia so psou DNA a mačky sa rodia s mačacou DNA. V dnešnom svete nie je nič medzitým. (Ani s genetickými manipuláciami by nebolo také ľahké vytvoriť organizmus s polovicou psích génov a polovicou mačacích génov.) Neexistuje žiadna navzájom sa povzbudzujúca množina psích vlastností, ku ktorým sa jedna mačka môže napoly zatúlať a stať sa polopsom.

Ľudská myseľ, ktorá rozmýšľa v kategóriách, asi dáva prednosť podstatám pred atraktormi. Človek chce povedať: "Toto je sekta" alebo "Toto nie je sekta" a tým je úloha klasifikácie navždy uzavretá. Ak pozorujete, že Sokrates má desať prstov, nosí šaty a hovorí plynulo po grécky, môžete povedať: "Sokrates je človek" a z toho odvodiť, že "Sokrates je zraniteľný boľhlavom" bez toho, že by ste robili konkrétne krvné testy na potvrdenie jeho smrteľnosti. Rozhodli ste Sokratovu ľudskosť raz a navždy.

¹ pozri: 69 Každá kauza chce byť sektou

² <http://www.prem-rawat-talk.org/forum/uploads/CultCharacteristics.htm>

³ pozri: 68 Pokus v Robbers Cave

⁴ pozri: 63 Afektívne špirály smrti

⁵ pozri: 65 Nekritická nadkritickosť

⁶ pozri: 66 Ochladzovanie skupinových názorov vyparovaním

⁷

Ale ak pozorujete, že istá skupina ľudí asi vykazuje polarizáciu na našich a cudzích⁸, a vidí kladný efekt svätožiary⁹ okolo svojej Úplne Najoblúbenejšej Veci – čo môže byť Objektivizmus¹⁰ alebo vegetariánstvo alebo neurónové siete¹¹ – nemôžete *na základe zatiaľ zhromaždenej indicie* odvodiť, či dosiahli nekritickosť¹². Nemôžete odvodiť, či je ich hlavná myšlienka pravdivá, nepravdivá, alebo naozaj užitočná no nie až taká užitočná, ako si myslia. Z *doteraz zhromaždených informácií* nemôžete odvodiť, či sú inak slušní alebo či vás odľakajú do izolácie a budú vám odopierať spánok a jedlo. Vlastnosti sekty nie sú všetky prítomné alebo všetky neprítomné.

Ak si pozriete online hádky “X je sekta”, “X nie je sekta”, jedna strana prechádza cez internetový zoznam vlastností sekty, nájde jednu, ktorá sedí a povie: “Preto je to sekta!” A obranca nájde vlastnosť, ktorá nesedí a povie: “Preto to nie je sekta!”

Nedokážete zostaviť presný obraz dynamiky uvažovania skupiny pomocou takéhoto esencionalizmu. Musíte venovať pozornosť jednotlivým vlastnostiam jednotlivo.

Navyše, obrátená hlúposť nie je inteligencia¹³. Ak vás zaujíma ústredná *myšlienka*, nie iba skupina, ktorá ju realizuje, potom múdre myšlienky môžu mať hlúpych nasledovníkov. Mnoho priaznivcov new age rozpráva o “kvantovej fyzike”, ale to nie je dôkaz proti kvantovej fyzike. Samozrejme aj hlúpe myšlienky majú hlúpych nasledovníkov. S binárnym esencionalizmom sa spája predstava, že ak odvodíte, že skupina je “sekta”, potom jej názory musia byť nepravdivé, pretože nepravdivé názory sú vlastnosťou siekt, tak ako mačky majú srst'. Ak vás zaujíma myšlienka, pozerajte na myšlienku, nie na ľudí¹⁴. Sektárstvo je skôr vlastnosť *skupín* než *hypotéz*.

Druhá chyba je, že keď sa ľudia nervózne pýtajú: “Toto nie je sekta, však?”, znie mi to akoby hľadali *ubezpečenie o rozumnosti*. Myšlienka, že racionalista by sa nemal príliš pripútať ani k svojmu sebaobrazu racionalistu si zaslúži svoj vlastný článok (ale pozrite si toto¹⁵, toto¹⁶, a toto¹⁷). Ale aj bez zachádzania do detailov môže byť človeku jasné, že *nervózne hľadanie ubezpečenia* nie je najlepší stav mysle na vyhodnocovanie otázok o rozumnosti. Nebudete skutočne zvedaví¹⁸ ani myslieť na to, ako naplniť svoje pochybnosti¹⁹. Namiesto toho si nájdete nejaký internetový zdroj, ktorý povie, že sekty používajú spánkovú depriváciu na ovládanie ľudí, vy si všimnete, že Vaša Oblúbená Skupina nepoužíva spánkovú depriváciu, a dôjdete k záveru: “Nie je to sekta. Uf!” Keď to nemá srst', nemôže to byť mačka. Veľmi ubezpečujúce.

Lenže každá kauza *sachce* stať sektou²⁰, či už je samotná kauza múdra alebo hlúpa. Rozdeľovanie na našich a cudzích²¹ atď. je súčasť ľudskej povahy, nie špeciálna kliatba²² mutantov²³. Rozumnosť je výnimka, nie pravidlo. Musíte vynakladať sústavné úsilie na udržanie rozumnosti proti prirodzenému sklzu do

⁸ pozri: 68 Pokus v Robbers Cave

⁹ pozri: 60 Efekt svätožiary

¹⁰ pozri: 72 Strážcovia Ayn Rand

¹¹ http://thedailywtf.com/Articles/No_We_Need_a_Neural_Network.aspx

¹² pozri: 65 Nekritická nadkritickosť

¹³ pozri: 52 Obrátená hlúposť nie je inteligencia

¹⁴ pozri: 54 Pritisnite si otázku

¹⁵ pozri: 258 Dvanásť cností rozumnosti

¹⁶ pozri: 3 Načo pravda? A...

¹⁷ pozri: 73 Dva koany o sektách

¹⁸ pozri: 123 Meditácia o zvedavosti

¹⁹ pozri: 121 Správne použitie pochybnosti

²⁰ pozri: 69 Každá kauza chce byť sektou

²¹ pozri: 68 Pokus v Robbers Cave

²² pozri: 50 Chyba prisudzovania

²³ pozri: 51 Sú vaši nepriatelia od narodenia zlí?

entropie. Ak sa rozhodnete: "Nie je to sekta!" a s úľavou si vydýchnete, potom nebudete vynakladať priebežné úsilie potláčať *prirodzené* sklony k sektárstvu. Rozhodnete sa, že sektárska podstata je neprítomná a prestanete pumpovať proti entropii atraktoru sekty.

Ak ste hrozne nervózni zo siekt, potom budete chcieť popierať každý náznak hocijakej vlastnosti, ktorá pripomína sektu. Lenže *každá* skupina, ktorá má cieľ a vidí ho v pozitívnom svetle, je ohrozená efektom svätožiari²⁴ a bude musieť pumpovať proti entropii, aby sa vyhla afektívnej špirále smrti²⁵. Toto platí aj pre bežné organizácie ako sú politické strany – ľudia, ktorí si myslia, že "liberálne hodnoty" alebo "konzervatívne hodnoty" dokážu vyliečiť rakovinu atď. Platí to pre začínajúce firmy v Silicon Valley, neúspešné aj úspešné. Platí to pre používateľov Macu aj pre používateľov Linuxu. Efekt svätožiari²⁶ nezačne byť okej len preto, že to robí každý; keby každý skočil z útesu, vy by ste neskočili. Proti chybám v uvažovaní treba bojovať, nie tolerovať ich. Ale ak ste príliš nervózni ohľadom: "Ste si istí, že toto nie je sekta?" potom sa budete zdráhať vidieť *akékoľvek* znaky sektárstva, pretože to by naznačovalo, že ste v sekte a *Toto nie je sekta!!!* Nebudete teda vidieť aktuálne bojiská, kde sa bežné sklony k sektárstvu plížia vpred alebo sú zatláčané naspäť.

Tretia chyba nervózneho pýtania sa: "Toto nie je sekta, však?" je to, že nervozita je tam, podľa môjho silného podozrenia, z celkom nesprávnych dôvodov.

Čím to je, že skupiny, ktoré vychvaľujú svoju Šťastnú Vec až ku hviezdám, povzbudzujú členov, aby darovali všetky svoje peniaze a pracovali v dobrovoľnom nevolníctve, a majú súkromné pozemky, kde sú ich členovia prísne izolovaní, nazývame "náboženstvá" a nie "sekty", ak tu už boli niekoľko storočí?

Prečo väčšina ľudí, ktorí sa na kryoniku nervózne pýtajú: "Toto nie je sekta, však?" by nebola rovnako nervózna z účasti na republikánskom alebo demokratskom politickej mítingu? Rozdelenie na našich a cudzích²⁷ a šťastné špirály smrti²⁸ sa môžu vyskytnúť v politickej diskusii, v bežnom náboženstve, v športovom fanení. Keby *nervozita* pochádzala zo strachu z *chýb rozumu*, ľudia by sa pýtali: "Toto nie je rozdelenie na našich a cudzích²⁹, však?" na demokratických alebo republikánskych politických mítingoch rovnako bojazlivým hlasom.

Existujú legitímne dôvody báť sa libertariánov menej než kultu lietajúcich tanierov, pretože libertariáni nemajú povest', že by používali spánkovú depriváciu na konverziu ľudí. Lenže ani kryonici nemajú povest' používania spánkovej deprivácie. Prečo by ste sa teda mali viac báť toho, že vám zmrazia hlavu, keď prestanete dýchať³⁰?

Mám podozrenie, že táto *nervozita* nie je strach z nesprávneho názoru, ani strach z fyzickej ujmy. Je to strach z osamelého nesúhlasu³¹. Ten nervózný pocit, ktorí mali pokusné osoby v Aschovom pokuse s konformitou³², keď všetky ostatné pokusné osoby (v skutočnosti spolupracovníci) jedna za druhou hovorili, že čiara C je rovnako dlhá ako čiara X, a pokusnej osobe sa zdalo, že čiara B je rovnako dlhá ako čiara X. Strach z opustenia svorky.

To je dôvod, prečo skupiny, ktorých názory tu boli dostatočne dlho, aby vyzerali "normálne" nevyvolávajú rovnakú nervozitu ako "sekty", hoci oficiálne

²⁴ pozri: 60 Efekt svätožiari

²⁵ pozri: 63 Afektívne špirály smrti

²⁶ pozri: 60 Efekt svätožiari

²⁷ pozri: 68 Pokus v Robbers Cave

²⁸ pozri: 63 Afektívne špirály smrti

²⁹ pozri: 68 Pokus v Robbers Cave

³⁰ <http://www.alcor.org/>

³¹ pozri: 76 Osamelý nesúhlas

³² pozri: 74 Aschov pokus o konformitu

náboženstvá vás môžu tiež obrať o všetky peniaze a poslať vás do kláštora. To je dôvod, prečo skupiny ako politické strany, ktoré sú silne zodpovedné za chyby v rozumnosti, nevyvolávajú takú nervozitu ako "sekty". Slovo "sekta" sa nepoužíva, aby symbolizovalo chyby v rozumnosti, používa sa ako nálepka na niečo, čo vyzerá čudne.

Nie každá zmena je zlepšenie, ale každé zlepšenie je nevyhnutne zmena. Ak chcete niečo robiť lepšie, nemáte inú možnosť ako robiť to inak. Všeobecná múdrosť obsahuje veľké množstvo skutočnej múdrosti; áno, má zmysel vyžadovať pre podivnosť bremeno dôkazu navyše. Ale táto *nervozita* nie je nejaké úmyselné, rozumné zvažovanie. Je to strach veriť niečomu, čo spôsobí, že na vás budú vaši priatelia pozeráť naozaj čudne. A tak sa ľudia pýtajú: "Toto nie je sekta, však?" tónom, ktorý by nikdy nepoužili pri návšteve politického mítingu ani pri postavení gigantickej vianočnej výstavy.

To je tá časť, ktorá má štvu.

Je to akoby, akonáhle veríte niečomu, čomu neverili vaši predkovia, priletela z oblohy Sektárska Víla a naplnila vás Sektárskou Esenciou, a ani sa nenazdáte a už všetci nosíte rúcha³³ a spievate³⁴. Akoby "čudné" názory boli *priamou príčinou* problémov, a nie spánková deprivácia a bitky. Škoda spôsobená sektami – samovražda Heaven's Gate a tak ďalej – iba ukazuje, že každý s čudnými názormi je blázon; prvá a hlavná vlastnosť "člena sekty" je, že je to Outsider s Podivnými Spôsobmi.

Áno, spoločensky nezvyčajný názor vystavuje skupinu nebezpečenstvu kvôli mysleniu na našich a cudzích³⁵ a ochladzovaniu vyparovaním³⁶ a ďalším problémom. Ale nezvyčajnosť je rizikový faktor, nie samotná choroba. To isté platí pre cieľ, o ktorom si myslíte, že stojí za dosiahnutie. Či je tento názor pravdivý alebo nie, mať pekný cieľ vás vždy vystavuje riziku šťastnej špirály smrti³⁷. Ale to robí úžasné ciele rizikovým faktorom, nie chorobou. Niektoré ciele naozaj stoja za nasledovanie³⁸.

Na druhej strane, nevidím legitímny dôvod na spánkovú depriváciu alebo vyhrážanie sa nesúhlasiacim bitkou, bodka³⁹. Keď skupina robí toto, či už ju voláte "sekta" alebo "nie sekta", už ste priamo odpovedali⁴⁰ na pragmatickú otázku ohľadom toho, či do nej vstúpiť.

Problém číslo štyri: Strach z osamelého nesúhlasu je niečo, čo *samotné sekty* zneužívajú. Strach z toho, že sa na vás vaši priatelia budú pozeráť odsudzujúco je *presne ten efekt, ktorý skutočné sekty používajú na konvertovanie a udržanie si členov* – obklopa vás jednohlasnou zhodou medzi veriacimi sekty.

Strach zo zvláštnych myšlienok, impulz ku konformite⁴¹, nepochybne varoval mnoho potenciálnych obetí pred kultmi lietajúcich tanierov. Keď ste mimo, drží vás mimo. Ale keď ste *vnútri*, drží vás *vnútri*. Konformita vás iba prilepí tam, kde práve ste, či je to dobré alebo zlé miesto.

Človek by chcel, aby existoval nejaký spôsob, ako vedieť *naisto*, že nie ste v "sekte". Nejaká jednoznačná, zdrvivajúca replika pre ľudí, ktorí sa naňho čudne pozerajú. Nejaký spôsob, ako raz a navždy vedieť, že robíte správnu vec, bez

³³ pozri: 73 Dva koany o sektách

³⁴ http://lesswrong.com/lw/m2/the_litany_against_gurus/

³⁵ pozri: 68 Pokus v Robbers Cave

³⁶ pozri: 66 Ochladzovanie skupinových názorov vyparovaním

³⁷ pozri: 63 Afektívne špirály smrti

³⁸ pozri: 64 Odolajte šťastnej špirále smrti

³⁹ pozri: 65 Nekritická nadkritickosť

⁴⁰ pozri: 54 Pritisnite si otázku

⁴¹ pozri: 74 Aschov pokus o konformite

tých ustavičných pochybností. Myslím, že sa to volá “potreba uzavretia”. A – samozrejme – aj toto sekty zneužívajú.

Preto to volám “sektárske antisektárstvo”.

Žiť s pochybnosťami nie je cnosť – cieľom každej pochybnosti je zničiť samu seba⁴² úspechom alebo zlyhaním, a pochybnosť, ktorá len tak čaká, nedosiahne nič. Ale niekedy pochybnosti chvíľu trvá, než sa zničí. Žiť so zásobníkom momentálne nevyriešených pochybností je pre racionalistu nevyhnutný životný fakt. Pochybností by sme sa nemali báť. Inak si budete musieť vybrať medzi veľmi strašidelným a veľmi hlúpym životom.

Ak naozaj, úprimne neviete zistiť, či daná skupina je “sekta”, potom si musíte vybrať v podmienkach neistoty. O tom je celá teória rozhodovania.

Problém číslo päť: Nedostatok strategického myslenia

Poznám ľudí, ktorí si dávajú pozor pri singularitárstve⁴³ a sú opatrní *aj* pri politických stranách a tradičných náboženstvách. Sú *opatrní*, nie nervózni ani defenzívni. Títo ľudia na prvý pohľad vidia, že singularitárstvo očividne nie je plnohodnotná sekta so spánkovou depriváciou atď. Ale boja sa, že sa singularitárstvo sektou *stane*, kvôli rizikovým faktorom ako premena predstavy mocnej UI na Super Šťastného Činiteľa⁴⁴ (činiteľa definovaného v prvom rade tým, že súhlasíte s každou peknou vecou, ktorá sa o ňom povie). Že niečo dnes nie je sekta, ešte neznamená, že sa to nestane sektou v budúcnosti. Sektárstvo je atraktor, nie podstata.

Či ma otravuje *tento* druh opatrnosti? Vôbec nie. Ja sám trávim dosť času znepokojovaním sa nad týmto scenárom. Pokúšam sa svoje kamienky go umiestňovať tak, aby som vopred blokoval pohyb týmto smerom. Preto som napríklad napísal sériu článkov o sektárskych zlyhaniach rozumu.

Ľudia, ktorí hovoria o “rozumnosti” tiež majú rizikový faktor navyše. Dávať ľuďom rady, ako majú rozmýšľať, je svojou podstatou nebezpečná vec. Ale je to *rizikový faktor*, nie *choroba*.

Obe moje obľúbené Kauzy majú riziko sektárstva. Napriek tomu sa ma pýtajú: “Si si istý, že toto nie je sekta?” omnoho častejšie, keď hovorím o mocných UI, než keď hovorím o teórii pravdepodobnosti a kognitívnej vede. Nevieť, či je jeden rizikový faktor vyšší než druhý, ale viem, ktoré z toho *znie čudnejšie*.

Problém číslo 6 otázky: “Toto nie je sekta, však?”...

Už len tá otázka samotná ma stavia do veľmi otravnej situácie typu Hlava 22. Skutočný Zlý Guru by samozrejme použil nervozitu dotyčného proti nemu samotnému a vytvoril by dôveryhodný prepracovaný argument vysvetľujúci Prečo Toto Nie Je Sekta, a dotyčný by ho ochotne prijal. Občas mám dojem, že presne toto od mňa ľudia *žiadajú!* Kedykoľvek sa pokúšam písať o sektárstve a ako sa mu vyhnúť, mám pocit, akoby som sa poddával tejto pomýlenej túžbe – akoby som v konečnom dôsledku dával ľuďom *ubezpečenie*. Ešte aj keď hovorím ľuďom, že je nevyhnutný ustavičný boj proti entropii.

Cítim sa, akoby som sa robil prvým nesúhlasiacim v Aschovom teste konformity, hovoriac ľuďom: “Áno, čiara X naozaj je rovnaká ako B, je okej ak to poviete aj vy.” Oni by sa nemali potrebovať pýtať! Alebo ešte horšie, cítim sa, akoby som predkladal prepracovaný argument Prečo Toto Nie Je Sekta. Je to *nesprávna otázka*.

⁴² pozri: 121 Správne použitie pochybnosti

⁴³ <http://intelligence.org/AIRisk.pdf>

⁴⁴ pozri: 65 Nekritická nadkritickosť

Však sa sami pozrite na myšlienkové procesy skupiny a rozhodnite sa sami, či je to niečo, na čom sa chcete zúčastniť, keď sa konečne zbavíte svojho strachu z nezvyčajnosti. Je to vaša vlastná zodpovednosť prestať myslieť sektársky, bez ohľadu na to, v ktorej skupine momentálne fungujete.

Keď sa niekto opýta: "Toto nie je sekta, však?", potom bez ohľadu na to, ako odpoviem, sa vždy cítim, akoby som niečo obhajoval. Ja nemám rád tento pocit. Nie je úlohou Bayesovského Majstra⁴⁵ dávať ubezpečenia, ani nie je úlohou racionalistu niečo brániť.

Sekty žijú so skupinového myslenia, nervozity, túžby po ubezpečení. Nemôžete nervozitu poslať preč púhym želaním, a falošná sebadôvera je ešte horšia. Ale dokiaľ niekto potrebuje ubezpečenie – hoci len ubezpečenie, že je racionalistom – vždy to bude diera v jeho brnení. Šikovný šermiar sa sústreďí na cieľ⁴⁶ a neobzerá sa, či sa mu niekto náhodou nevysmieva. Keď viete, čo sa pokúšate urobiť a prečo, viete, či to robíme alebo nie, a či vám skupina pomáha alebo vás brzdí.

(PS: Ak niekto príde za vami a opýta sa: "Si si istý, že toto nie je sekta?", nepokúšajte sa mu všetky tieto pojmy vysvetliť jednou vetou. Podceňujete inferenčné vzdialenosti.⁴⁷ Dotyčný povie: "Aha, takže *pripúšťaš*, že ste sekta!" alebo "Počkaj, ty hovoríš, že by som sa nemal báť vstúpiť do sekty?" alebo "Takže... báť sa sekty je sektárstvo? To mi znie hodne sektársky." Takže posledný faktor mrzutosti – číslo sedem, ak počítate so mnou – je, že toto všetko sa tak dlho vysvetľuje.)

Pôvodný článok s komentármi⁴⁸

⁴⁵ pozri: 73 Dva koany o sektách

⁴⁶ pozri: 54 Pritisnite si otázku

⁴⁷ pozri: 9 Očakávame krátke inferenčné vzdialenosti

⁴⁸ http://lesswrong.com/lw/md/cultish_countercultishness/

Časť VIII. Pozeranie čerstvými očami

Obsah

78 Ukotvenie a prispôsobenie	83
79 Priming a kontaminácia	85
80 Veríme všetkému, čo nám povedia?	87
81 Uložené myšlienky	89
82 Koláže “mimo vychodených kolají”	91
83 Originálny pohľad	93
84 Čudnejšie než história	95
85 Logická chyba zovšeobecňovania fiktívnej indicie	96
86 Ako vyzeráť (a byť) hlboký	100
87 Svoje názory meníme zriedkavejšie, než si myslíme	103
88 Odložte navrhovanie riešení	105
89 Klam pôvodu	107

Kapitola 78. Ukotvenie a prispôsobenie

Predstavte si, že pred vaším zrakom zatočím kolesom šťastia a vyjde číslo 65. Potom sa opýtam: Myslíte si, že percento afrických krajín, ktoré sú v OSN, je väčšie alebo menšie než toto číslo? Aké percento afrických krajín je podľa vás v OSN? Zamyslite sa nad týmito otázkami na chvíľu sami, ak chcete, ale prosím nepoužívajte Google.

Ďalej, skúste odhadnúť, počas 5 *sekúnd*, hodnotu nasledujúceho matematického výrazu. 5 sekúnd. Pripravení? Pozor... *Štart!*

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

Tversky and Kahneman (1974) zapisovali odhady pokusných osôb, ktoré videli na kolese šťastia rôzne čísla. Mediánový odhad osôb, ktoré videli na kolese číslo 65 bol 45%; mediánový odhad osôb, ktoré videli číslo 10 bol 25%.

Súčasná teória pre tento a podobné pokusy je, že pokusné osoby prijímú počiatkové neinformatívne číslo ako svoj štartový bod alebo *kotvu*; a potom sa *prispôsobia* smerom nahor alebo nadol od svojho počiatkového odhadu, kým nedosiahnu odpoveď, ktorá “znie prijateľne”; a potom sa prestanú prispôsobovať. Typickým výsledkom je nedostatočné prispôsobenie od kotvy – vzdialenejšie čísla by tiež mohli byť “prijateľné”, ale človek sa zastaví pri prvej prijateľne znejúcej odpovedi.

Podobne, študenti, ktorým ukážeme “ $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ ” dávajú mediánový odhad 512, zatiaľ čo študenti, ktorým ukážeme “ $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ ” dávajú mediánový odhad 2 250. Motivujúca hypotéza bola, že študenti skúsia vynásobiť (alebo odhadom skombinovať) prvých pár činiteľov a potom výsledok prispôsobia nahor. V oboch prípadoch bolo prispôsobenie nedostatočné, v porovnaní so skutočnou hodnotou 40 320, ale prvá množina odhadov bola omnoho nedostatočnejšia, pretože začínali z nižšej kotvy.

Tversky a Kahneman hlásia, že ponúknutie finančnej odmeny za presnosť nezmenšilo efekt ukotvenia.

Strack and Mussweiler (1997) sa pýtali na rok, kedy Einstein prvýkrát navštívil Spojené Štáty. Úplne neprijateľné kotvy ako 1215 alebo 1992 spôsobili rovnako veľký efekt ukotvenia ako prijateľnejšie kotvy ako 1905 alebo 1939.

Toto má zrejmé použitie napríklad pri dohadovaní o plate, alebo nákupe auta. Neradím vám, aby ste to zneužívali, ale aby ste si dali pozor na tých, čo chcú zneužiť vás.

A tiež: Všímajte si svoje vlastné myslenie a snažte sa všimnúť si, kedy *prispôsobujete* číslo pri hľadaní odhadu.

Pokusy o zbavenie sa skreslenia ukotvením sa vo všeobecnosti ukázali ako neúčinné. Odporučil by som tieto dva: Po prvé, ak vám pôvodný odhad pripadá neprijateľný, skúste ho úplne zahodiť a prísť s novým odhadom, a nie posúvať sa od kotvy. Ale toto samotné nemusí byť dostatočné – pokusné osoby, ktoré dostali pokyn vyháňať sa ukotveniu, to naďalej robili (Quattrone a kol. 1981). Takže po druhé, aj keď ste vyskúšali tú prvú metódu, skúste ešte pomyslieť na kotvu v opačnom smere – kotvu, ktorá je očividne príliš malá alebo príliš veľká (ak tá prvá bola príliš veľká alebo príliš malá) – a chvíľku na nej zotrvať.

Quattrone, G.A., Lawrence, C.P., Finkel, S.E., & Andrus, D.C. (1981). Skúmania pri ukotvovaní: Efekt prvotného rozsahu, krajnej kotvy, a sugestívnych náznakov. Rukopis, Stanford University.

Strack, F. & Mussweiler, T. (1997). Vysvetlenie záhadného efektu ukotvenia: Mechanizmus výberovej dostupnosti. *Journal of Personality and Social Psychology*, 73, 437-446.

Tversky, A. a Kahneman, D. 1974. Usudzovanie pri neistote: Heuristiky a skreslenia. *Science*, 185:1124-1131.

Pôvodný článok s komentármi¹

¹ http://lesswrong.com/lw/j7/anchoring_and_adjustment/

Kapitola 79. Priming a kontaminácia

Predstavte si, že požiadate pokusné osoby, aby stlačili jedno tlačidlo, ak postupnosť písmen tvorí slovo a druhé tlačidlo, ak písmená netvorí slovo. (Napri. "banack" verzus "banán".) Potom im ukážete postupnosť "voda". Neskôr budú rýchlejšie identifikovať postupnosť "piť" ako slovo. Toto je známe ako "kognitívny priming"; táto konkrétna forma by bola "sémantický priming" alebo "pojmový priming".

Fascinujúca vec na primingu je, že sa odohráva na takej nízkej úrovni – zvyšuje rýchlosť *rozoznania písmen ako tvoriacich slovo*, čo sa zrejme stane *skôr* než sa zamyslíte nad významom daného slova.

Priming odhaľuje aj masívnu paralelizáciu šírenia aktivácie: ak videnie "voda" aktivuje slovo "piť", pravdepodobne aktivuje aj "riečka" alebo "pohár" alebo "špliechať"... a táto aktivácia sa šíri cez významové súvislosti pojmov až naspäť k rozoznávaniu postupností písmen.

Priming je podvedomý a nezastaviteľný, artefakt ľudskej nervovej architektúry. Pokúšať sa zastaviť svoj vlastný priming je ako pokúšať sa zastaviť šírenie aktivácie svojich vlastných nervových obvodov. Skúste nahlas povedať farbu – nie význam, ale farbu – nasledujúcej postupnosti písmen: "ZELENÁ".

Mussweiler a Strack (2000) sa pýtali pokusných osôb nasledujúcu ukotvujúcu otázku¹: "Je priemerná ročná teplota v Nemecku vyššia alebo nižšia ako 5 stupňov Celzia / 20 stupňov Celzia?" Neskôr, pri úlohe identifikovania slov, pokusné osoby, ktoré dostali kotvu 5 stupňov rýchlejšie identifikovali slová ako "zima" a "sneh", kým pokusné osoby s vysokou kotvou rýchlejšie identifikovali "teplo" a "slnko". Toto ukazuje mechanizmus neprispôsobivosti pri ukotvení: priming kompatibilných myšlienok a spomienok.

Všeobecnejší záver je, že *úplne neinformujúce, vedome nepravdivé* alebo *úplne nesúvisiace* "informácie" môžu ovplyvniť odhady a rozhodnutia. V oblasti heuristik a skreslení je tento všeobecnejší jav známy ako *kontaminácia*. (Chapman a Johnson 2002.)

Každý výskum heuristik a skreslení objavil efekty ukotvenia², napríklad pokusné osoby dávajú nižšie (vyššie) odhady percenta krajín OSN nachádzajúcich sa v Afrike podľa toho, či sa ich predtým pýtali, či je toto percento vyššie alebo nižšie ako 10 (65). Tento efekt sa pôvodne pripisoval tomu, že sa pokusné osoby posúvajú od kotvy ako počiatočného bodu, a zastavia sa akonáhle dosiahnu prijateľný interval; a posúvajú sa nedostatočne, pretože sa zastavia na jednom konci intervalu spoľahlivosti. (Tversky a Kahneman 1974.)

Raná hypotéza Tverskeho a Kahnemana sa stále zdá byť v niektorých prípadoch správnym vysvetlením, obzvlášť keď pokusné osoby sami vytvoria úvodný odhad (Epley a Gilovich 2001). Moderný výskum však naznačuje, že väčšina ukotvenia je v skutočnosti kvôli kontaminácii, nie prispôbovaniu sa posúvaním.

Váš obchod s potravinami asi má otravné nápisy ako "Maximum 12 kusov na zákazníka" alebo "5 kusov za 10 dolárov". Sú tieto nápisy efektívne pri ovplyvňovaní zákazníkov, aby kupovali väčšie množstvá? Asi si myslíte, že vás to neovplyvňuje³. Ale *niekoho* to musí ovplyvňovať, pretože tieto nápisy

¹ Anchoring and Adjustment

² pozri: 78 Ukotvenie a prispôsobenie

³ http://en.wikipedia.org/wiki/Bias_blind_spot

dokázateľne fungujú, čo je dôvod, prečo ich obchody stále vyvesujú. (Wansink a kol. 1998.)

Najhroznejšou stránkou kontaminácie je, že⁴ slúži⁵ ako⁶ ďalšia⁷ z⁸ tisícich⁹ tvárí¹⁰ sklonu¹¹ potvrdzovať¹². Akonáhle sa nejaká myšlienka dostane do vašej hlavy, primuje informácie, ktoré sú s ňou kompatibilné – a tým si zabezpečuje svoju trvalú existenciu. Nielenže máme selekčný tlak na vyhrávanie politických hádok; sklon potvrdzovať je zabudovaný priamo do nášho hardwaru, asociatívne siete primujú kompatibilné myšlienky a spomienky. Nešťastný vedľajší účinok našej existencie ako tvorov s nervovou sústavou.

Jediný letmý obrázok môže stačiť na to, aby primoval rozoznávanie súvisiacich slov. Nemyslíte si, že treba viac na to, aby sa sklon potvrdzovať rozbehol. Chce to iba ten jeden rýchly záblesk a spodný riadok je už určený¹³, pretože svoje názory meníme omnoho zriedkavejšie než si myslíme¹⁴...

Chapman, G.B. a Johnson, E.J. 2002. Zahŕňanie nepodstatného: Kotvy pri hodnotení názorov a hodnoty.¹⁵ In Gilovich et. al. (2003).

Epley, N., & Gilovich, T. (2001). Vrátenie prispôsobenia do heuristiky ukotvenia a prispôsobenia: Diferenciálne spracovanie sebou vytvorených a experimentátorom poskytnutých kotiev. *Psychological Science*, 12, 391–396.

Mussweiler, T. a Strack, F. Porovnávať znamená veriť: model selektívnej dostupnosti hodnotiaceho ukotvenia. *European Review of Social Psychology*, 10, 135-167.

Tversky, A. a Kahneman, D. 1974. Usudzovanie pri neistote: Heuristiky a skreslenia. *Science*, 185:1124--1131.

Wansink, B., Kent, R.J. a Hoch, S.J. 1998. Model ukotvenia a prispôsobenia pri rozhodovaní o množstve nákupov. *Journal of Marketing Research*, 35(February): 71-81.

Pôvodný článok s komentármi¹⁶

⁴ pozri: 39 Pozitívne skreslenie: Pozeraj do tmy

⁵ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

⁶ pozri: 111 Tretia možnosť

⁷ http://lesswrong.com/lw/il/hindsight_bias/

⁸ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁹ pozri: 34 Falošná kauzalita

¹⁰ pozri: 92 Jeden argument proti armáde

¹¹ pozri: 95 Racionalizovanie

¹² pozri: 94 Filtrovaná indícia

¹³ pozri: 93 Spodný riadok

¹⁴ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

¹⁵ <http://web.archive.org/web/20040325202602/http://cebiz.org/ejj/PDF%20Papers/Incorporating%20the%20Irrelevant.PDF>

¹⁶ http://lesswrong.com/lw/k3/priming_and_contamination/

Kapitola 80. Veríme všetkému, čo nám povedia?

Niektoré rané experimenty o ukotvovaní¹ a prispôsobovaní² testovali, či *rozptyľovanie* pokusných osôb – ich kognitívne “vyťaženie” požiadavkou, aby sledovali, či sa v postupnostiach číslíc vyskytuje “5” alebo také čosi – zníži prispôsobovanie, a tým zvýši vplyv kotvy. Väčšina experimentov asi potvrdzuje myšlienku, že kognitívna vyťaženosť zvyšuje ukotvovanie a kontamináciu³ vo všeobecnosti.

Všímajúc si hromadiace sa experimentálne výsledky – viac a viac zistení o kontaminácii zvýraznenej kognitívnou vyťaženosťou – Daniel Gilbert uvidel, ako sa vynára naozaj šialený vzor: Veríme všetkému, čo nám povedia?

Človek by si mohol prirodzene myslieť, že keď nám niekto povie nejaký výrok, najprv *pochopíme*, čo ten výrok znamená, potom sa nad tým výrokom *zamyslíme*, a nakoniec ho buď *prijmeme* alebo *odmietneme*. Tento napohľad samozrejmy model toku kognitívnych procesov sa ťahne už od Descarta. Ale Descartov súper, Spinoza, nesúhlasil. Spinoza naznačil, že najprv *pasívne prijmeme* výrok v procese jeho pochopenia, a až dodatočne *aktívne prestaneme veriť* výrokom, ktoré po zvážení odmietneme.

Posledných pár storočí filozofi viacmenej držali s Descartom, keďže jeho pohľad vyzeral viac, však viete, logicky a intuitívne⁴. Ale Gilbert videl spôsob, ako otestovať Descartovu a Spinozovu hypotézu experimentálne.

Ak má pravdu Descartes, potom by rozptyľovanie pokusných osôb mali prekážať aj prijímaniu pravdivých tvrdení aj odmietaniu nepravdivých. Ak má pravdu Spinoza, potom by rozptyľovanie pokusných osôb malo spôsobiť, že si nepravdivé tvrdenia budú pamätať ako pravdivé, ale nemalo by spôsobiť, že si pravdivé tvrdenia budú pamätať ako nepravdivé.

Gilbert, Krull, a Malone⁵ (1990) potvrdili tento výsledok, ukazujúc, že keď pokusným osobám predkladali nové tvrdenia označené PRAVDA a NEPRAVDA, rozptyľovanie nemalo vplyv na identifikovanie pravdivých tvrdení (55% úspech bez vyrušovania, 58% s vyrušovaním), ale ovplyvňovalo identifikovanie nepravdivých tvrdení (55% bez vyrušovania, 35% s vyrušovaním).

Ešte dramatickejšiu ilustráciu vyprodukoval v nasledujúcom experimente Gilbert, Tafarodi a Malone⁶ (1993). Pokusné osoby nahlas čítali správy o zločine rolujúce sa na obrazovke, v ktorých farba textu označovala, či je konkrétne tvrdenie pravdivé alebo nepravdivé. Niektoré správy obsahovali nepravdivé tvrdenia, ktoré zvyšovali vážnosť zločinu, iné správy obsahovali nepravdivé tvrdenia, ktoré zmierňovali (ospravedlňovali) zločin. Pokusné osoby mali počas čítania správ o zločine zároveň sledovať postupnosti písmen a hľadať číslicu “5” – to bola rozptyľujúca úloha vytvárajúca kognitívne vyťaženie. Nakoniec mali pokusné osoby odporučiť dĺžku väzenia pre daného zločincu, od 0 do 20 rokov.

Pokusné osoby v kognitívne vyťaženom stave odporučili v priemere 11,15 rokov väzenia pre zločincov s “ážnejším” zločinom, čiže zločincov, ktorých správy obsahovali výroky zvyšujúce vážnosť zločinu označené ako nepravdivé. Vyťažené

¹ pozri: 79 Priming a kontaminácia

² pozri: 78 Ukotvenie a prispôsobenie

³ pozri: 79 Priming a kontaminácia

⁴ <http://www.overcomingbias.com/2007/10/what-evidence-i.html>

⁵ <http://www.wjh.harvard.edu/~dtg/Gilbert%20et%20al%20%28UNBELIEVING%29.pdf>

⁶ <http://www.wjh.harvard.edu/~dtg/Gilbert%20et%20al%20%28EVERYTHING%20YOU%20READ%29.pdf>

osoby odporučili v priemere 5,83 rokov väzenia pre zločincov, ktorých správy obsahovali výroky ospravedlňujúce zločin označené ako nepravdivé. Tento takmer dvojnásobný rozdiel bol, ako asi očakávate, štatisticky významný.

Nevyťažení účastníci čítali presne tie isté správy, s rovnakými označeniami, a okolo nich občas prešli rovnaké postupnosti číslíc, ale oni nemali za úlohu sledovať číslicu "5". Preto mohli venovať viac pozornosti "nevereniu" výrokov označených za nepravdivé. Títo nevyťažení účastníci odporúčali 7,03 rokov verzus 6,03 rokov pre zločincov, ktorých správy nepravdivo zvyšovali vážnosť alebo nepravdivo ospravedlňovali.

Článok Gilberta, Tafarodiho a Maloneho mal názov: "Nedokážete neveriť všetkému, čo čítate."

To naznačuje – prinajmenšom – že by sme mali byť omnoho opatrnejší, keď sa vystavujeme nespoľahlivým informáciám, najmä ak popri tom robíme ešte niečo iné. Budte opatrní, keď zazriete noviny v supermarkete.

Gilbert, D. 2002. Inferenčná oprava. In *Heuristics and biases: The psychology of intuitive judgment*. Táto citácia už vám je povedomá, však?

Gilbert, D., Krull, D. and Malone, P. 1990. Prestať veriť neuveriteľnému: Niektoré problémy s odmietaním nepravdivej informácie.⁷ *Journal of Personality and Social Psychology*, 59(4), 601-613.

Gilbert, D., Tafarodi, R. and Malone, P. 1993. Nedokážete neveriť všetkému, čo čítate.⁸ *Journal of Personality and Social Psychology*, 65(2), 221-233.

Pôvodný článok s komentármi⁹

⁷ <http://www.wjh.harvard.edu/~dtg/Gilbert%20et%20al%20%28UNBELIEVING%29.pdf>

⁸ <http://www.wjh.harvard.edu/~dtg/Gilbert%20et%20al%20%28EVERYTHING%20YOU%20READ%29.pdf>

⁹ http://lesswrong.com/lw/k4/do_we_believe_everything_were_told/

Kapitola 81. Uložené myšlienky

Jednou z najväčších záhad ľudského mozgu je, ako tá prekliata vec *vôbec dokáže* fungovať, keď väčšina neurónov vysiela signál 10-20-krát za sekundu, maximálne 200 Hz. V neurológii existuje "pravidlo sto krokov", že každá predpokladaná operácia musí skončiť po *nanajvýš* 100 postupných krokoch – môžete byť paralelní, ako len chcete, ale nemôžete predpokladať viac ako 100 (a radšej menej) neurónových signálov po sebe.

Viete si predstaviť musieť programovať pomocou 100Hz procesorov, bez ohľadu na to, ako veľa by ste ich mali? Potrebovali by ste stovky miliárd procesorov, aby ste *vôbec niečo* urobili v reálnom čase.

Keby ste museli písať programy pracujúce v reálnom čase pre sto miliárd 100Hz procesorov, jeden trik, ktorý by ste používali, kde by sa len dalo, by bolo skladovanie výsledkov. To znamená, že by ste si odložili výsledky predchádzajúcich operácií a nabitúce by ste ich vyhľadali namiesto prepočítavania od nuly. A to je typický *nervový* prístup – rozoznávanie, asociácie, dopĺňanie vzoru.

Je rozumné predpokladať, že *väčšina* ľudského poznávania pozostáva z hľadania uložených údajov.

Táto myšlienka mi z času na čas prichádza na rozum.

Existuje krásny ilustrujúci príbeh, ktorý som si myslel, že som si odložil medzi záložky, ale neviem ho opäť nájsť: bol to príbeh o človeku, ktorého sused "vševed" raz mimochodom tvrdil, že najlepší spôsob, ako odstrániť z domu komín je vybúrať krb, počkať, kým tehly spadnú o úroveň nižšie, vybúrať tieto tehly, a opakovať, kým nezmizne celý komín. O niekoľko rokov neskôr, keď tento muž chcel odstrániť svoj komín, táto uložená myšlienka číhala, pripravená zaútočiť...

A ako si ten muž dodatočne uvedomil – asi uhádnete, že to nedopadlo dobre – jeho sused nebol v týchto veciach nejako zvlášť znalý, ani dôveryhodný zdroj. Keby nad touto myšlienkou *zapochyboval*, pravdepodobne by si bol uvedomil, že je hlúpa. Niektoré uložené údaje je lepšie opäť prepočítať. Ale mozog dopĺňa vzor automaticky – a ak si vedome neuvedomíte, že daný vzor treba upraviť, zostane vám doplnený vzor.

Predpokladám, že keby táto myšlienka napadla tomu mužovi samotnému – keby sám *osobne* dostal tento skvostný nápad ako odstrániť komín – bol by sa nad touto myšlienkou zamyslel kritickejšie. Ale keď už niekto *iný* túto myšlienku celú premyslel, môžete ušetriť výpočtovú silu tým, že si uložíte ich *záver*, nie?

Špeciálne v modernej civilizácii nikto nedokáže myslieť dosť rýchlo na to, aby si premyslel svoje vlastné myšlienky. Keby ma ako dieťa pohodili v lese, kde by ma vychovali vlci alebo tichí roboti, sotva by ste ma poznali ako človeka. Nikto nedokáže myslieť dosť rýchlo ani na to, aby za jeden život zrekapituloval múdrosť kmeňa lovcov a zberačov, začínajúc od nuly. A čo sa týka múdrosti gramotnej civilizácie, zabudnite na to.

Avšak odvrátenou stranou tohto celého je, že pravidelne vidím ľudí, ktorí sa pokúšajú o kritické myslenie, ako papagájujú uložené myšlienky, ktoré neboli vymyslené kritickými mysliteľmi.

Dobrým príkladom je skeptik, ktorý pripúšťa: "No, faktickým indíciami nemožno náboženstvo ani dokázať ani vyvrátiť." Ako som inde zdôraznil¹, toto je

¹ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

jednoducho nepravdivá teória pravdepodobnosti. A takisto je to jednoducho nepravda z pohľadu reálnej psychológie náboženstva – pred pár stáročiami by vás za takýto výrok upálili na hranici. Matka, ktorej dcéra má rakovinu, sa modlí: “Bože, prosím uzdrav moju dcéru,” a nie: “Milý Bože, viem, že náboženstvo nesmie mať žiadne falzifikovateľné dôsledky, čo znamená, že nie je možné, aby si uzdravil moju dcéru, takže... no, v podstate sa modlím, aby som sa cítila lepšie, namiesto robenia niečoho, čo by mojej dcére mohlo naozaj pomôcť.”

Ale ľudia si prečítajú: “Faktickým indíciami nemožno náboženstvo ani dokázať ani vyvrátiť,” a potom, keď nabudúcevidia nejaký kus indicie vyvracajúcej náboženstvo, ich mozog si doplní tento vzor. Dokonca aj niektorí ateisti opakujú túto absurditu bez zaváhania. Keby sa nad touto myšlienkou zamysleli sami, namiesto počutia od niekoho iného, boli by skeptickejší.

Smrť: doplňte vzor: “Smrť dáva životu zmysel.”

Je frustrujúce hovoriť s dobrými a slušnými ľuďmi – ľuďmi, ktorí by ani za tisíc rokov nikdy *spontánne* nepomysleli na vyhladenie ľudskej rasy – spomenúť tému existenciálneho rizika, a počuť, ako hovoria: “No, možno si ľudská rasa nezaslúži prežiť.” Za tisíc rokov by ich nikdy napadlo zastreliť svoje vlastné dieťa, ktoré je súčasťou ľudskej rasy, ale mozog doplní tento vzor.

Aké vzory, ktoré ste si nikdy nevybrali, sa dopĺňajú vnútri vašej mysle?

Rozumnosť: doplňte vzor: “Láska nie je rozumná.”

Keby táto myšlienka niekedy napadla vám osobne, ako celkom nová myšlienka, ako by ste ju kriticky skúmali? Viem, čo by som povedal² *ja*, ale čo by ste povedali vy? Môže byť ťažké to vidieť čerstvými očami. Pokúste sa zabrániť svojej mysli v doplnení vzoru štandardným, neprekvapujúcim, už známym spôsobom. Možno že neexistuje lepšia odpoveď než tá štandardná, ale nemôžete o tej odpovedi *rozmyšľať*, dokiaľ nezabránite svojmu mozgu v automatickom dopĺňaní odpovede.

Teraz, keď ste si prečítali tento článok, až bude nabudúce niekoho počuť, ako bez zaváhania opakuje mém, ktorý považujete za hlúpy alebo nepravdivý, pomyslite si: “Uložené myšlienky.” Môj názor je teraz vo vašej mysli, čaká na doplnenie vzoru. Ale je to pravda? Nedovoľte svojej mysli doplniť tento vzor! *Myslíte!*

Pôvodný článok s komentármi³

² pozri: 2 Rozumné čítanie

³ http://lesswrong.com/lw/k5/cached_thoughts/

Kapitola 82. Koľaje “mimo vychodených koľají”

Kedykoľvek vás niekto nabáda, aby ste “mysleli mimo vychodených koľají”, zvyčajne vám, *pre vaše pohodlie*, ukáže, kde presne sa “mimo vychodených koľají” nachádza. Sranda, ako sa všetci nezávisláci obliekajú rovnako...

V Umelej Inteligencii má každý mimo odboru uložený výsledok¹ *skvelej novej revolučnej myšlienky UI* – neurónové siete, ktoré fungujú rovnako ako ľudský mozog! Nová myšlienka UI: doplňte vzor: “Logickým UI sa napriek všetkých veľkých sľubom nepodarilo poskytnúť skutočnú inteligenciu celé desaťročia – potrebujeme neurónové siete!”

Táto uložená myšlienka tu je už asi tri desaťročia. Stále žiadna umelá inteligencia. Napriek tomu si akosi každý mimo odboru myslí, že neurónové siete sú Nová Myšlienka Ktorá Rúca Dominantnú Paradigmú, odkedy v 70-tych rokoch vynasli spätnú propagáciu. Hovorte mi o vašich starnúcich hipíkoch.

Imidž nezávisláka svojou podstatou nedovoľuje žiadne odchýlky od normy. Ak nechodíte v čiernom, ako majú ľudia vedieť, že ste trpiaci umelec? Ako majú ľudia rozoznať jedinečnosť, ak nezapadáte do štandardného vzoru ako má jedinečnosť vyzeráť? Ako má hocikto rozoznať, že máte novú revolučnú predstavu o UI, ak sa netýka neurónových sietí?

Ďalším príkladom rovnakého žánru je “podvratná” literatúra, ktorá všetka znie rovnako, a stojí za ňou malá vzdorovitá liga rebelov, ktorí ovládajú celú katedru angličtiny. Ako sa Anonymous opýtal² na blogu Scotta Aaronsona:

Už niekedy nejaká podvratná literatúra, ktorú ste čítali, spôsobila, že ste zmenili niektorý zo svojich politických názorov?

Alebo ako si všimol³ Lizard:

Revolúcia už bola v televízii. Revolúcia už má svoj *merchandise*. Revolúcia je spotrebný tovar, životný štýl v balíčku dostupný vo vašom miestnom hypermarkete. Za 19,95 dolára si kúpite čiernu masku, konzervu so sprejom, transparent “Smrť fašizmu” a konto na vlastný blog, na ktorom môžete písať o policajnej brutalite, ktorú ste utrpeli, keď ste sa priviazali reťazou k požiarnemu hydrantu. Kapitalizmus sa naučil predávať antikapitalizmus.

Mnohí v Silicon Valley si všimli, že drvivá väčšina špekulatívnych kapitalistov v ľubovolnej dobe naháňa rovnaké Revolučné Inovácie, a sú to tie Revolučné Inovácie, ktoré mali verejnú ponuku účastí pred šiestimi mesiacmi. To je *zvlášť* zdrvivé pozorovanie špekulatívneho kapitálu, pretože tam je priamy ekonomický motív nenasledovať dav – buď niekto iný vyvíja rovnaký produkt, alebo niekto iný ponúka príliš veľa za startup. Steve Jurvetson mi raz povedal, že vo firme Draper Fisher Jurvetson súhlas dvoch partnerov stačí na nákup ľubovolného startupu do 1,5 milióna dolárov. Ale ak sa *všetci* partneri zhodnú, že niečo vyzerá ako dobrý nápad, tak to neurobia. Kiež by grantové komisie mali toľko rozumu.

Problémom originality je, že aby ste ju dosiahli, musíte naozaj *myslieť*, a nenechať svoj mozog dopĺňať vzory⁴. Neexistuje žiadna pohodlná značka “mimo

¹ pozri: 81 Uložené myšlienky

² <http://www.scottaaronson.com/blog/?p=87>

³ <http://journalism.berkeley.edu/projects/biplog/archive/000748.html>

⁴ pozri: 81 Uložené myšlienky

vychodených koľají", ku ktorej sa môžete ihneď rozbehnúť. Funguje to takmer ako zen – nemôžete naučiť *satori* pomocou slov, pretože *satori* je zážitok, keď slová zlyhávajú. Čím viac sa snažíte nasledovať slovné pokyny zenového majstra, tým ďalej ste od dosiahnutia prázdnej mysle.

Myslím, že existuje dôvod, prečo ľudia nedosiahnu novosť tým, že sa o ňu snažia. Vlastnosti ako pravdivosť alebo dobrý dizajn nezávisia od novosti: $2 + 2 = 4$, naozaj, fakt, aj keď si to myslí každý. Ľudia, ktorí sa snažia objaviť pravdu alebo vymyslieť dobrý dizajn, môžu postupom času dosiahnuť tvorivosť. Nie každá zmena je vylepšenie, ale každé vylepšenie je zmena.

Každé vylepšenie je zmena, ale nie každá zmena je vylepšenie. Ten, kto hovorí: "Chcem postaviť originálnu pascu na myši!" a nie: "Chcem postaviť optimálnu pascu na myši!" si takmer vždy želá byť *vnímaný* ako originálny. "Originalita" v tomto zmysle je svojou podstatou sociálna, pretože sa dá určiť iba porovnaním s druhými ľuďmi. Takže ich mozog jednoducho doplní štandardný vzor toho, čo sa vníma ako "originálne" a ich priatelia súhlasne prikývnu a povedia, že je to podvrátne.

Knihy o businessse vám vždy povedia, pre vaše pohodlie, kam presne sa podel váš syr. V opačnom prípade by bol čitateľ zanechaný opakujúc: "Kde je to 'mimo vychodených koľají', kam mám ísť?"

Skutočné myslenie, podobne ako *satori*, je úkon mysle bez slov.

Pôvodný článok s komentármi⁵

⁵ http://lesswrong.com/lw/k6/the_outside_the_box_box/

Kapitola 83. Originálny pohľad

Súvisí s kapitolami: Uložené myšlienky¹, Cnosť presnosti²

Keď to Robert Pirsig vyjadril takto dobre, ja už len skopírujem, čo povedal. Nevieť, či sa tento príbeh zakladá na skutočnosti alebo nie, ale v oboch prípadoch je pravdivý.

Mal problém so žiakmi, ktorí nemali čo povedať. Najprv si myslel, že je to lenivosť, ale neskôr sa ukázalo, že nie. Jednoducho nedokázali vymyslieť, čo povedať.

Jedna z nich, dievča s okuliarmi s hrubými sklami, chcela napísať esej dĺžky päťsto slov o Spojených štátoch. Bol zvyknutý na sklamania, ktoré nasledujú po takýchto výrokoch, tak jej navrhol, bez podceňovania, aby to zúžila iba na Bozeman.

Keď prišiel termín práce, nemala ju a bola veľmi rozrušená. Skúšala a skúšala, ale nenapadlo jej nič, čo by mohla povedať.

To ho zarazilo. Teraz *on* nevedel vymyslieť, čo povedať. Nastalo ticho a potom nečakaná odpoveď: "Zúž to iba na *hlavnú ulicu* v Bozemane." Bol to záblesk osvietenia.

Usilovne prikývla a odišla. Ale hneď pred ďalšou hodinou sa vrátila so *skutočnou* úzkosťou, tentokrát so slzami, tá úzkosť tam asi bola už dlho. Stále jej nenapadalo nič, čo by sa dalo povedať, a nechápala prečo, ak nedokázala vymyslieť nič o *celom* Bozemane, by mala byť schopná vymyslieť niečo o jednej jeho ulici.

Rozčúlil sa: "Ty sa *nepozeralaš!*" Spomenul si na svoje vlastné vyhodenie z univerzity, pretože toho povedal *príliš* veľa. Ku každému faktu existuje *nekonečne* veľa hypotéz. Čím viac sa *pozeráš*, tým viac ich *vidíš*. Ona sa naozaj nepozerala, a ešte stále to nejako nepochopila.

Povedal jej nahnevane: "Zúž to na *prednú časť jednej* budovy na hlavnej ulici v Bozemane. Dom opery. Začni od ľavej hornej tehly."

Jej oči za hrubými sklami okuliarov sa naširoko otvorili.

Na ďalšiu hodinu prišla so začudovaným pohľadom a odovzdala mu esej dĺžky päťsto slov o prednej časti Domu opery na hlavnej ulici v Bozemane, v štáte Montana. "Sedela som v stánku s hamburgermi na opačnej strane cesty," povedala, "a začala som písať o prvej tehle, potom o druhej tehle, a pri tretej tehle to zrazu všetko na mňa prišlo a nevedela som sa zastaviť. Mysleli si, že mi šibe, robili si zo mňa srandu, ale tu to je. Nerozumiem tomu."

To ani *on*, ale počas dlhých prechádzok po uliciach mesta o tom rozmyšľal a došiel k záveru, že ju očividne blokovala podobná prekážka aká paralyzovala jeho počas jeho prvého dňa učenia. Bola zablokovaná, pretože sa pokúšala písomne zopakovať veci, ktoré už počula, tak ako sa *on* počas prvého dňa pokúšal zopakovať veci, ktoré sa už rozhodol povedať. Nedokázala vymyslieť nič, čo by sa dalo napísať o Bozemane, pretože si nedokázala spomenúť na

¹ pozri: 81 Uložené myšlienky

² pozri: 26 Cnosť presnosti

nič počuté, čo by stálo za zopakovanie. Zvláštne si neuvedomovala, že by sa mohla pozrieť a vidieť čerstvo sama za seba, a potom písať, bez sledovania v prvom rade, čo už bolo povedané. Zúženie na jednu tehlu zničilo tento blok, pretože bolo príliš jasné, že *musí* urobiť originálny a priamy pohľad.

—Robert M. Pirsig, Zen a údržba motocykla

Pôvodný článok s komentármi³

³ http://lesswrong.com/lw/k7/original_seeing/

Kapitola 84. Čudnejšie než história

Predstavte si, že by som vám povedal, že viem *naisto*, že nasledujúce tvrdenia sú pravdivé:

- Keď sa natriete istým *presne* daným odtieňom farby medzi modrou a zelenou, otočí sa sila gravitácie a budete padať smerom nahor.
- V budúcnosti bude obloha zaplnená miliardami vznášajúcich sa čiernych gúl. Každá guľa bude väčšia než všetky vzducholode, ktoré kedy existovali, dokopy. Keď guľi ponúknete peniaze, spustí sa z oblohy prostitút na bungee lane.
- Vaše vnúčatá si budú myslieť, že by bola nielen hlúposť, ale *zlo*, zatvárať zlodějov za mreže namiesto toho, aby dostali po zadku.

Pomysleli by ste si, že mi šibe, však?

Teraz si predstavte, že ste v roku 1901 a máte si vybrať, či budete viac veriť horeuvedeným alebo nasledujúcim tvrdeniam:

- Existuje absolútna hranica rýchlosti akou sa dva predmety môžu voči sebe pohybovať, a je to presne 670 616 629,2 míle za hodinu. Ak vyskočíte na vlak, ktorý ide takmer takto rýchlo a vystrelíte z okna, základné jednotky dĺžky sa zmenia, takže *vám* sa bude zdať, že guľka letí rýchlo pred vami, ale iní ľudia uvidia niečo iné. Aha, čas okolo vás sa tiež zmení.
- V budúcnosti bude existovať superprepojená globálna sieť miliónov sčítacích zariadení, každé z nich bude silnejšie než všetky sčítacie stroje pred rokom 1901 dohromady. Jedným z hlavných využití tejto siete bude prenos pohyblivých obrázkov lesbického sexu, predstierajúc, že sa skladajú z čísel.
- Vaše vnúčatá si budú myslieť, že by bola nielen hlúposť, ale *zlo*, povedať, že niekto nemôže byť prezidentom Spojených Štátov, pretože je čierny.

Inšpirované Robinovým komentárom¹: “Zaujímalo by ma, či by niekto vedel dostatočne podrobne opísať fiktívny príbeh o alternatívnej skutočnosti, ktorú by naši predkovia nedokázali odlíšiť od tej skutočnej, aby sa ozrejmiло, aká prekvapujúca sa ukázala skutočnosť.”

Pôvodný článok s komentármi²

¹ http://lesswrong.com/lw/j0/making_history_available/ewg

² http://lesswrong.com/lw/j1/stranger_than_history/

Kapitola 85. Logická chyba zovšeobecňovania fiktívnej indície

Keď sa pokúšam o úvod do problematiky vyspelej UI, čo je prvá vec, ktorú počujem vo viac než polovici prípadov?

“Aha, myslíš ako vo filme Terminátor / Matrix / Asimovovi roboti!”

Odpovedám: “No, nie presne tak. Snažím sa vyhnúť logickej chybe zovšeobecňovania fiktívnej indície.”

Niektorí to pochopia hneď a zasmejú sa. Iní bránia svoje použitie daného príkladu a nesúhlasia, že je to chyba.

Čo je zlé na používaní filmov alebo románov ako východiskových bodov diskusie? Nikto predsa netvrdí, že je to *pravda*. Kde je tam lož, kde je racionalistický hriech? Science fiction predstavuje autorov pokus o vizualizáciu budúcnosti; prečo nevyužiť myslenie, ktoré už za nás urobil niekto iný, namiesto začínania od nuly?

Nie každý chybný krok v presnom tanci rozumnosti spočíva vo vyslovenej viere v nepravdu; existujú aj jemnejšie spôsoby, ako sa mýliť.

Po prvé, zbavme sa predstavy, že science fiction predstavuje plnohodnotný rozumný pokus o predpovedanie budúcnosti. Ešte aj tí najusilovnejší spisovatelia science fiction sú v prvom rade rozprávačmi; požiadavky na príbeh nie sú rovnaké ako požiadavky na predpoveď. Ako poukázal¹ Nick Bostrom:

Kedy ste naposledy videli film o tom, ako ľudstvo zrazu vyhynulo (bez varovania, a bez nahradenia nejakou inou civilizáciou)? Aj keby takýto scenár bol omnoho pravdepodobnejší než scenár, kde ľudskí hrdinovia úspešne odrazia inváziu príšer alebo robotických vojakov, nebola by to veľká zábava sledovať.

Vo fikcii teda existujú konkrétne skreslenia². Ale pokúšať sa opraviť tieto konkrétne skreslenia nestačí. Príbeh *nikdy* nie je rozumným pokusom o analýzu, ani u tých najusilovnejších spisovateľov science fiction, pretože príbehy nepoužívajú pravdepodobnostné distribúcie. Ako ilustráciu uvediem:

Bob Merkelthud sa obozretne prešmykol cez dvere mimozemskej kozmickej lode, obzerajúc sa doprava a potom doľava (alebo doľava a potom doprava), či tam zostala niektorá z obávaných Vesmírnych Príšer. Po jeho boku bola jediná zbraň, ktorá proti Vesmírnym Príšerám účinkovala: s pravdepodobnosťou 30% Vesmírny Meč ukutý z čistého titanu, s pravdepodobnosťou 20% obyčajný železný sochor, s pravdepodobnosťou 45% lesklý čierny disk nájdený v dymiacich ruinách Stonehenge, pravdepodobnosť 5% je rozdelená na priveľa drobných možností, ktoré tu nebudeme menovať.

Merkelthud (aj keď je značná šanca, že tam namiesto neho bola Susan Wifflefoofer) urobil dva kroky dopredu alebo jeden krok

¹ <http://www.nickbostrom.com/existential/risks.html>

² http://www.overcomingbias.com/2006/12/biases_of_scien.html

dozadu, keď ohromný rev preťal ticho čiernej prechodovej komory!
Alebo tiché bzučanie v bielej prechodovej komore. Hoci Amfer
a Woofi (1997) tvrdia, že v tomto bode bol Merkelthud zožratý,
Spacklebackle (2003) upozorňuje na to, že...

Postavy môžu byť nevedomé, ale samotný *autor* nemôže povedať čarovné slovo: "neviem". Hlavný hrdina musí prejsť budúcnosťou po jednej čiare, plnej podrobností³, ktoré dávajú príbehu telo, od Wifflefooferovej primerane futuristických postojov voči feminizmu, až po farbu jej náušník.

Potom sú všetky tieto prirážajúce podrobnosti⁴ a pochybné predpoklady zabalené a označené krátkym názvom⁵, vytvárajúc ilúziu jedného balíka⁶.

Pri problémoch s veľkým priestorom odpovedí, najťažšie nie je *overiť* správnosť odpovede, ale jednoducho ju vôbec v priestore odpovedí nájsť⁷. Ak niekto začne otázkou, či nás UI uloží alebo neuloží do puzdier ako vo filme "Matrix", skáče tým na 100-bitový predpoklad, bez zodpovedajúcich 98 bitov indicie potrebných na nájdenie ho v priestore odpovedí s možnosťou hodnou explicitného zváženia. Po získaní prvých 98 bitov indicie by už stačilo len o pár bitov viac, aby sme túto možnosť povýšili na takmer istotu, čo vám hovorí niečo o čom, kde sa robí väčšina danej práce.

Tento "úvodný" krok nájdenia možností hodných explicitného zváženia zahŕňa kroky ako: Brať do úvahy čo vieš a čo nevieš, čo vieš a čo nevieš predpovedať, vynakladať vedomé úsilie vyhnúť sa skresleniu absurdity⁸ a rozšíriť intervaly istoty⁹, zvážiť ktoré otázky sú tie dôležité, pokúšať sa zohľadniť možné Čierne Labute a myslieť na (dovtedajšie) neznáme neznáme veci. Skočiť na "Matrix: áno alebo nie?" toto všetko preskakuje.

Každý profesionálny vyjednávač vie, že ovládať pojmy diskusie je veľmi blízko k ovládaniu výsledku diskusie. Ak začnete myslením na *Matrix*, budete myslieť na pochodujúce armády robotov, ktoré porazia ľudí po dlhom boji... a nie na superinteligenciu, ktorá luskne nanotechnologickými prstami. Sústredíte sa na súboj "my proti nim", nasmerujete pozornosť na otázky ako: "kto vyhrá?" a "kto by mal vyhrať?" a "bude UI naozaj taká?". Vznikne všeobecná atmosféra zábavy: "Aká je tvoja úžasná predstava budúcnosti?"

A v prázdnote sa stratia: úvahy o viacerých možných dizajnoch mysle, ktoré by "umelá inteligencia" mohla implementovať; závislosť budúcnosti od jej počiatočných podmienok; sila¹⁰ nadľudskej inteligencie a argument pre jej nepredvídateľnosť¹¹; ľudia, ktorí majú seriózny pohľad na celú vec a snažia sa s tým niečo urobiť.

Keby sa nejaký zákerný kazič diskusií rozhodol, že je *preňho* najlepšie, keď bude nútiť diskutérov, aby začali vyvracaním *Terminátora*, bol by v takomto posunutí debaty úspešný. Pri debate o kontrole zbraní rečník NRA nechce byť uvedený ako "fanatický strieľač", ani oponent zbraní nechce byť uvedený ako "advokát ozbrojenia obetí". Prečo by ste vy mali dovoliť rovnaké skreslenie rámca od hollywoodskych scenáristov, hoci len náhodne?

Novinári mi nehovoria: "Budúcnosť bude ako film *2001*." Ale pýtajú sa: "Bude budúcnosť ako film *2001*, alebo ako film *U.I.*?" To je asi rovnaký problém rámca

³ pozri: 6 Prirážajúce podrobnosti

⁴ pozri: 6 Prirážajúce podrobnosti

⁵ pozri: 23 Occamova britva

⁶ http://en.wikipedia.org/wiki/Package-deal_fallacy

⁷ http://wiki.lesswrong.com/wiki/Locate_the_hypothesis

⁸ pozri: 84 Čudnejšie než história

⁹ http://lesswrong.com/lw/j6/why_is_the_future_so_absurd/

¹⁰ <http://intelligence.org/2007/07/10/the-power-of-intelligence/>

¹¹ <http://intelligence.org/2007/09/30/three-major-singularity-schools/>

ako keby sa pýtali: "Mali by sme okresať dávky pre postihnutých veteránov, alebo zvýšiť dane bohatým?"

V pravekom prostredí neexistovali kiná; čo ste videli na vlastné oči, to bola pravda. Krátky pohľad na jedno slovo nás dokáže primovať¹² a urobiť kompatibilné myšlienky dostupnejšie¹³, s dokázateľným silným vplyvom na odhady pravdepodobnosti. Koľko zmätku si myslíte, že vo vašom úsudku napácha dvojhodinový film? Bude dosť ťažké odstrániť tieto škody pomocou úmyselného sústredenia – načo si teda pozývať upíra do domu? V šachu alebo v go, každý premárnený ťah je strata; v rozumnosti, každý vplyv, ktorý nie je indicia, je (v priemere) entropia.

Darí sa filmovým divákom neveriť¹⁴ tomu, čo vidia? Pokiaľ viem, málo filmových divákov sa správa ako keby *priamo* pozorovali budúcnosť Zeme. Ľudia, ktorí videli filmy *Terminátor*, sa neschovávali v jadrových úkrytoch 29. augusta 1997. Ale tí, čo sa dopúšťajú tohto omylu, napohľad konajú ako keby videli udalosti filmu odohrať sa na *nejakej inej* planéte; nie na Zemi, ale na nejakom mieste podobnom Zemi.

Poviete: "Predstavme si, že postavíme veľmi múdru UI," a oni povedia: "Ale neviedlo toto k jadrovej vojne vo filme *Terminátor*?" Pokiaľ viem, je to rovnaké uvažovanie, vrátane tónu hlasu, ako keby niekto povedal: "Ale neviedlo toto k jadrovej vojne na Alpha Centauri?" alebo "Neviedlo toto k pádu talianskeho mestského štátu Piccolo v štrnástom storočí?" Filmu neveríme, ale je dostupný¹⁵. Neberieme ho ako proroctvo, ale ako ilustračný historický príklad. Bude sa história opakovať? Kto vie?

V nedávnej diskusii o Singularite niekto spomenul, že Vinge si zrejme nemyslí, že rozhrania medzi mozgom a počítačom môžu výrazne zvýšiť inteligenciu, a spomenul *Marooned in Realtime* a Tunča Blumenthala, ktorý bol najskúsenejší cestovateľ, ale nevyzeral až taký mocný. Rozhorčene som odpovedal: "Ale Tunč stratil väčšinu svojho hardvéru! Bol zmrzačený!" A potom urobil myšlienkovú skúšku správnosti a pomyslel som si: Čo do čerta to trepem.

Nemali by sme o tejto téme argumentovať podľa nej samotnej, bez ohľadu na to, ako Vinge vykreslil svoje postavy? Tunč Blumenthal nie je "zmrzačený", on je *vymyslený*. Mohol som povedať: "Vinge sa rozhodol vykresliť Tunča ako zmrzačeného z dôvodov, ktoré môžu ale nemusia súvisieť s jeho osobnými predpovedami" a to by dalo jeho autorskému hlasu primeranú váhu indicie. Nemôže povedať: "Tunč bol zmrzačený." Ohľadom Tunča Blumenthala neexistuje žiadne *bol*.

Úmyselne som navrchu tohto článku ponechal chybu, ktorú som urobil pri svojom prvom náčrte: "Iní bránia svoje použitie daného *príkladu* a nesúhlasia, že je to omyl." Ale Matrix *nie je* príklad!

Blízky omyl je logický klam argumentovania imaginárnou indíciou: "Pozri, keby si *šiel* na koniec dúhy, *našiel* by si tam hrniec zlata – čo dokazuje, že mám pravdu!" (Aktualizovanie na základe indicie, ktorá bola predpovedaná ale nie pozorovaná je matematický zrkadlový obraz skreslenia spätného pohľadu¹⁶.)

Mozog má veľa mechanizmov na zovšeobecňovanie na základe pozorovania, nielen heuristiku dostupnosti. Vidíte tri zebry, vytvoríte si kategóriu "zebra" a táto kategória predstavuje automatický vnemový úsudok. Tvory konského tvaru

¹² pozri: 79 Priming a kontaminácia

¹³ pozri: 5 Dostupnosť

¹⁴ pozri: 80 Veríme všetkému, čo nám povedia?

¹⁵ pozri: 5 Dostupnosť

¹⁶ pozri: 34 Falošná kauzalita

s bielymi a čiernymi pruhmi sa klasifikujú ako "zebry", preto sú rýchle a dobré na jedenie; očakávame, že budú podobné zvyšným pozorovaným zebrom.

Ľudia vidia (na obrazovke) troch Borgov, ich mozgy si automaticky vytvoria kategóriu "Borgovia" a automaticky usudzujú, že ľudia s rozhraním medzi mozgom a počítačom budú z množiny "Borgovia" a budú podobní pozorovaným Borgom: chladní, bezcitní, oblečení v čiernej koži, kráčajúci ťažkými mechanickými krokmi. Novinári neveria, že budúcnosť *bude* obsahovať Borgov – neveria, že *Star Trek* je proroctvo. Ale keď niekto hovorí o rozhraniach medzi mozgom a počítačom, myslia si: "Bude budúcnosť obsahovať Borgov?" Nie: "Ako viem, že počítačom sprostredkovaná telepatia urobí ľudí menej milými?" Nie: "Nikdy som nevidel žiadneho Borga a takisto ani nikto iný." Nie: "Vytváram si rasový predsudok na základe *doslova* nulovej skúsenosti."

Ako povedal¹⁷ George Orwell o kliše:

Predovšetkým treba nechať význam, nech si vyberie slovo, a nie naopak... Keď myslíte na niečo abstraktné, máte väčší sklon začať od slov, a pokiaľ nevynaložíte vedomé úsilie, aby ste tomu zabránili, existujúci dialekt vtrhne a vykoná svoje dielo za vás, na úkor zahmlenia alebo priam zmenenia toho, čo ste mysleli.

Odhadujem, že *najškodlivejšia* stránka používania predstavivosti iných autorov je to, že odučí ľudí používať svoju vlastnú. Ako povedal¹⁸ Robert Pirsig:

Bola zablokovaná, pretože sa pokúšala písomne zopakovať veci, ktoré už počula, tak ako sa on počas prvého dňa pokúšal zopakovať veci, ktoré sa už rozhodol povedať. Nedokázala vymyslieť nič, čo by sa dalo napísať o Bozemane, pretože si nedokázala spomenúť na nič počuté, čo by stálo za zopakovanie. Zvlášť si neuvedomovala, že by sa mohla pozrieť a vidieť čerstvo sama za seba, a potom písať, bez sledovania v prvom rade, čo už bolo povedané.

Zapamätané fikcie vtrhnú a urobia vaše myslenie za vás; sú náhradou *videnia* – najnebezpečnejšou zo všetkých skratiek.

Tieto názory sú ďalej rozvinuté v článkoch: Ukotvenie¹⁹, Kontaminácia²⁰, Dostupnosť²¹, Uložené myšlienky²², Veríme všetkému, čo nám povedia?²³, Einsteinova drzosť²⁴, Priťažujúce podrobnosti²⁵

Pôvodný článok s komentármi²⁶

¹⁷ pozri: 55 Rozumnosť a anglický jazyk

¹⁸ pozri: 83 Originálny pohľad

¹⁹ pozri: 78 Ukotvenie a prispôsobenie

²⁰ pozri: 79 Priming a kontaminácia

²¹ pozri: 5 Dostupnosť

²² pozri: 81 Uložené myšlienky

²³ pozri: 80 Veríme všetkému, čo nám povedia?

²⁴ pozri: 24 Einsteinova drzosť

²⁵ pozri: 6 Priťažujúce podrobnosti

²⁶ http://lesswrong.com/lw/k9/the_logical_fallacy_of_generalization_from/

Kapitola 86. Ako vyzerat' (a byt') hlboký

Nedávno som sa zúčastnil diskusnej skupiny, ktorej téma pri danom stretnutí bola Smrť. Vyvolalo to hlboké emócie. Myslím si, že zo všetkých obedov v Silicon Valley, ktorých som sa zúčastnil, tento bol najúprimnejší; ľudia hovorili o smrti v rodine, o smrti priateľov, čo si mysleli o svojej vlastnej smrti. Ľudia sa navzájom naozaj počúvali. Kiež by som vedel tieto okolnosti spoľahlivo reprodukovať.

Bol som jediný prítomný transhumanista, a dával som si obrovský pozor, aby som to každému nestrkal pred nos. ("Fanatik je človek, ktorý nedokáže zmeniť názor a nechce zmeniť tému." Ja sa snažím sa byť schopný aspoň zmeniť tému.) Nie je prekvapivé¹, že ľudia hovorili o význame, ktorý smrť dáva životu, alebo ako je smrť v skutočnosti maskovaným požehnaním. Ja som však, veľmi opatrne, vysvetlil, že transhumanisti sú vo všeobecnosti pozitívni voči životu, ale nemajú radi smrť².

Dodatočne za mnou prišlo pár ľudí a povedalo mi, že som veľmi "hlboký". No áno, som, ale vďaka tomuto som začal rozmýšľať, čo spôsobuje, že ľudia vyzerajú hlbokí.

V jednom bode rozhovoru jedna žena povedala, že myšlienka na smrť ju vedie k tomu, aby bola k ľuďom milá, pretože nikdy nevie, či ich nevidí poslednýkrát. "Keď môžem o niekom povedať niečo pekné," povedala, "poviem mu to hneď a nečakám."

"To je krásna myšlienka," povedal som, "a aj keby z vás jedného dňa spada hrozba smrti, dúfam, že to budete robiť naďalej..."

Dodatočne, táto žena bola jedna z tých, čo mi povedali, že som hlboký.

V inom bode diskusie jeden muž hovoril o tom, že smrť má nejakú výhodu X, už si presne nespomínam čo. A ja som povedal: "Viete, pri ľudskej povahe, keby ľudia dostávali každý týždeň po hlave baseballovou pálkou, čoskoro by si vymysleli dôvody, prečo dostávať po hlave baseballovou pálkou je dobrá vec. Ale keby ste vzali niekoho, kto nedostával po hlave baseballovou pálkou, a opýtali by ste sa ho, či by chcel, povedal by, že nie. Myslím si, že keby ste vzali niekoho, kto bol nesmrteľný, a opýtali by ste sa ho, či by chcel zomrieť kvôli výhode X, povedal by, že nie."

Dodatočne mi tento muž povedal, že som hlboký.

Korelácia nie je kauzalita. Možno som v ten deň jednoducho hovoril hlbším hlasom, takže to znelo múdrejšie.

Ale mám podozrenie, že som im pripadal "hlboký", pretože som sústavne porušoval ich uložené vzory³ "hlbokej múdrosti" spôsobom, ktorý hneď dával zmysel.

Existuje stereotyp Hlbokej Múdrosti. Smrť: doplňte vzor: "Smrť dáva životu zmysel." Každý pozná túto štandardnú Hlboko Múdru odpoveď. A preto získava určité vlastnosti signálu na potlesk. Keď ju povie, ľudia možno prikývnu, pretože ich mozog si doplní vzor a oni vedia, že majú prikývnuť. Možno dokonca povedia: "Aká hlboká múdrosť!", azda v nádeji, že budú sami považovaní za hlbokých. Ale

¹ pozri: 83 Originálny pohľad

² <http://yudkowsky.net/singularity/simplified>

³ pozri: 81 Uložené myšlienky

nebudú *prekvapení*; nepočuli nič mimo vychodených kolají⁴; nepočuli nič, čo by si nemohli pomyslieť sami. Dalo by sa to nazvať vierou v múdrosť⁵ – myšlienka je označená ako “hlboko múdra”, a je to doplnený štandardný vzor pre “hlbokú múdrosť”, ale neprináša žiaden zážitok vhladu.

Ľudia, ktorí sa *snažia vyzeráť* Hlboko Múdri často nakoniec vyzerajú prázdni, ako ozvena⁶, pretože sa snažia vyzeráť Hlboko Múdri namiesto optimalizovania⁷.

Kolko rozmyšľania som potreboval ja v tomto procese vyzerania hlboký? Ľudský mozog beží iba rýchlosťou 100 Hz a ja som odpovedal v reálnom čase, takže väčšina práce musela byť vypočítaná vopred. Časť, ktorú som prežíval ako náročnú, bola vyberanie odpovede pochopiteľnej v jednom inferenčnom kroku a potom jej formulovanie pre maximálny dopad.

Z filozofického hľadiska bola takmer všetka moja práca už urobená. Doplňte vzor: Existujúca situácia X je naozaj odôvodnená, pretože má výhodu Y: “Klam naturalizmu?” / “Skreslenie status quo?” / “Mohli by sme mať Y bez X?” / “Keby sme nikdy predtým nepočuli o X, vybrali by sme si to dobrovoľne, aby sme dostali Y?” Myslím si, že je férové povedať, že vykonávam tieto myšlienkové vzory asi na rovnakej úrovni automatiky ako dýcham. Napokon, väčšina ľudského myslenia musí byť vyhľadávanie uložených údajov, ak má mozog vôbec fungovať.

A tiež som už mal rozvinutú filozofiu transhumanizmu⁸. Transhumanizmus má tiež uložené myšlienky o smrti. Smrť: doplňte vzor: “Smrť je nezmyselná tragédia, ktorú si ľudia racionalizujú.” Toto bola neštandardná uložená informácia; taká, ktorú moji poslucháči nepoznali. Mal som niekoľko príležitostí použiť neštandardnú uloženú informáciu, a pretože to všetko boli časti rozvinutej filozofie transhumanizmu, všetky viditeľne patrili k rovnakej téme. To spôsobilo, že som vyzeral *súvisle*, nielen originálne.

Predpokladám, že toto je dôvod, prečo východná filozofia pripadá hlboká západným ľuďom – má neštandardnú ale súvislú uloženú Hlbokú Múdrosť. Zrkadlovo, v dielach japonskej fikcie človek občas nájde kresťanov⁹ vykreslených ako studnice hlbokéj múdrosti a mystických tajomstiev. (A občas nie¹⁰.)

Ak si správne spomínam, jeden ekonóm raz poznamenal, že laická verejnosť má tak málo skúseností so štandardnou ekonomikou, že keď ho raz pozvali vystupovať v televízii, potreboval im iba zopakovať ekonómiu pre prvákov, aby vyznel ako brilantný originálny mysliteľ.

Kritické bolo aj to, že moji poslucháči mohli *ihneď* vidieť, že moje odpovede dávajú zmysel. Mohli ale nemuseli súhlasiť s myšlienkou, ale nevznela im ako úplne od veci. Poznáam transhumanistov, ktorí nedokážu vyzeráť hlbokí, pretože nedokážu zohľadniť, čo ich poslucháč ešte nevie. Ak chcete znieť hlboko, nikdy nehovorte nič, čo je vzdialené viac než jeden inferenčný krok od súčasného myšlienkového stavu vášho poslucháča. Tak to skrátka je.

Ak chcete *vyzeráť* hlboký, študujte neštandardné filozofie. Vyhľadávajte diskusie na témy, ktoré vám dajú príležitosť vyzeráť hlboký. Urobte svoje filozofické myslenie v predstihu, aby ste sa mohli sústrediť na dobré vysvetľovanie. Ale najmä trénujte držanie sa v hraniciach jedného inferenčného kroku.

Aby ste *boli* hlbokí, myslite sám za seba o “múdrych” aleb dôležitých alebo emocionálne nabitých témach. Myslieť sám za seba nie je to isté ako prísť

⁴ pozri: 82 Kolaje “mimo vychodených kolají”

⁵ pozri: 13 Viera vo vieru

⁶ pozri: 83 Originálny pohľad

⁷ pozri: 82 Kolaje “mimo vychodených kolají”

⁸ <http://yudkowsky.net/singularity/simplified>

⁹ <http://tvtropes.org/pmwiki/pmwiki.php/Main/JesusTaboo>

¹⁰ <http://tvtropes.org/pmwiki/pmwiki.php/Main/CreepyCoolCrosses>

s nezvyčajnou odpoveďou¹¹. Znamená to vidieť veci sám¹², nie iba nechať svoj mozog doplniť vzor¹³. Ak sa nezastavíte pri prvej odpovedi¹⁴ a vyhodíte odpovede, ktoré vyzerajú nejasne neuspokojivo¹⁵, časom vaše myšlienky vytvoria koherentný celok, pochádzajúci z jedného zdroja, z vás, namiesto úlomkov opakovania¹⁶ záverov druhých ľudí¹⁷.

Pôvodný článok s komentármi¹⁸

¹¹ pozri: 82 Koláže “mimo vychodených kolají”

¹² pozri: 83 Originálny pohľad

¹³ pozri: 81 Uložené myšlienky

¹⁴ pozri: 111 Tretia možnosť

¹⁵ pozri: 27 Tvoja sila racionalistu

¹⁶ pozri: 83 Originálny pohľad

¹⁷ pozri: 81 Uložené myšlienky

¹⁸ http://lesswrong.com/lw/k8/how_to_seem_and_be_deep/

Kapitola 87. Svoje názory meníme zriedkavejšie, než si myslíme

Počas posledných pár rokov sme diskkrétne oslovovali kolegov, ktorí stáli pred voľbou medzi ponukami práce, a žiadali sme ich o odhad pravdepodobnosti, že si vyberú tú alebo onú prácu. Priemerná istota predpovedanej voľby bola skromných 66%, ale iba 1 z 24 respondentov si vybral možnosť, ktorej pôvodne priradil nižšiu pravdepodobnosť, z čoho vychádza celková miera presnosti 96%.

—Dale Griffin a Amos Tversky, *Zvažovanie indície a determinanty istoty*. (Kognitívna psychológia, 24, str. 411-435.)

Keď som prvýkrát čítal uvedené slová – 1. augusta 2003, asi o tretej popoludní – zmenili môj spôsob myslenia. Uvedomil som si, že *akonáhle viem uhádnuť, aká bude moja odpoveď* – akonáhle viem priradiť vyššiu pravdepodobnosť rozhodnutiu sa jedným smerom oproti druhému – potom som sa, s najväčšou pravdepodobnosťou, už rozhodol. Svoje názory meníme zriedkavejšie než si myslíme. Väčšinu času by sme dokázali uhádnuť svoju vlastnú odpoveď do pol sekundy po počutí otázky.

Ako rýchlo uplynie tento nepovšimnutý okamih, keď ešte nevieme uhádnuť, aká bude naša odpoveď; to malé okno príležitosti, aby inteligencia konala. V otázkach výberu, ako aj v otázkach faktu.

Princíp spodného riadku¹ je, že iba skutočné príčiny vašich názorov určujú vašu efektivitu ako racionalistu. Akonáhle je váš názor pevne daný, žiadne množstvo argumentov nezmení jeho pravdivostnú hodnotu; akonáhle je vaše rozhodnutie pevne dané, žiadne množstvo argumentov nezmení jeho dôsledky.

Možno si myslíte, že môžete dôjsť k názoru alebo rozhodnutiu pomocou nerozumných prostriedkov, potom ho skúsate zdôvodniť, a keď zistíte, že ho nedokázate zdôvodniť, odmietnete ho.

Lenže my svoje názory meníme zriedkavejšie – *omnoho* zriedkavejšie – než si myslíme.

Som si istý, že si dokázate spomenúť aspoň na jednu okolnosť vo svojom živote, keď ste zmenili názor. To dokáže každý. Ale čo všetky tie okolnosti vo vašom živote, keď ste svoj názor nezmenili? Sú aj tie dostupné² vo vašom heuristickom odhade vlastnej kompetencie³?

Uprostred skreslenia spätného pohľadu⁴, falošnej kauzality⁵, pozitívneho skreslenia⁶, ukotvovania⁷, atď atď, a najmä obávaného sklonu potvrdzovať⁸, akonáhle sa raz nejaká myšlienka dostane do vašej hlavy, pravdepodobne tam aj zostane.

¹ pozri: 93 Spodný riadok

² pozri: 5 Dostupnosť

³ http://en.wikipedia.org/wiki/Lake_Wobegon_effect

⁴ http://lesswrong.com/lw/il/hindsight_bias/

⁵ pozri: 34 Falošná kauzalita

⁶ pozri: 39 Pozitívne skreslenie: Pozeraj do tmy

⁷ pozri: 78 Ukotvenie a prispôsobenie

⁸ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

Pôvodný článok s komentármi⁹

⁹ http://lesswrong.com/lw/jx/we_change_our_minds_less_often_than_we_think/

Kapitola 88. Odložte navrhovanie riešení

Z knihy Robyna Dawesa *Rozumná voľba v neistom svete*, strany 55-56. Pridané zvýraznenie.

Norman R. F. Maier si všimol, že keď skupina čelí problému, jej členovia majú prirodzený sklon na začiatku diskusie o probléme navrhovať možné riešenia. V dôsledku toho sa interakcia skupiny zameriava na výhody a nedostatky navrhovaných riešení, ľudia emocionálne prilnú k riešeniam, ktoré sami navrhli, a lepšie riešenia už sa nenavrhujú. Maier vydal nariadenie ako zlepšiť skupinové riešenie problémov: **“Nenavrhujte riešenia, dokiaľ nebol problém nebol prediskutovaný tak dôkladne, ako sa len dá bez ich navrhovania.”** Je jednoduché ukázať, že toto nariadenie funguje v kontextoch, kde existujú objektívne definované dobré riešenia problémov.

Maier vymyslel nasledujúci pokus “hrania rolí”, aby demonštroval tento bod. Pri bežiacom páse pracujú traja zamestnanci s rôznymi úrovňami zručnosti. Striedajú sa medzi tromi činnosťami, ktoré vyžadujú rôznu úroveň zručnosti, pretože ten najšikovnejší – ktorý je zároveň najdominantnejší – má silnú motiváciu vyhnúť sa núde. V kontraste s ním si ten najmenej šikovný pracovník uvedomuje, že tie zložitejšie úkony nerobí rovnako dobre ako zvyšní dvaja, ale súhlasil so striedaním kvôli dominancii jeho šikovného kolegu. “Odborník na efektivitu” si všimol, že keby najšikovnejší zamestnanec dostal tú najnáročnejšiu prácu a ten najmenej šikovný tú najmenej náročnú, produktivita by sa mohla zvýšiť o 20 % a tak odborník navrhuje, aby sa prestali striedať. Traja robotníci a štvrtá osoba, ktorá má pridelenú rolu predáka, majú za úlohu diskutovať o odporúčaní odborníka. Niektoré skupiny dostali Maierovo nariadenie, aby nediskutovali o riešeníach, kým dôkladne neprediskutujú samotný problém, iné skupiny toto nariadenie nedostali. Tí, ktorí tento pokyn nedostali, okamžite začali argumentovať o dôležitosti produktivity verzus autonómie robotníka a o vyhýbaní sa núde. Skupiny s daným pokynom s omnoho väčšou pravdepodobnosťou objavili riešenie, že dvaja šikovnejší robotníci sa môžu striedať, zatiaľ čo ten najmenej šikovný zostane pri najmenej náročnej práci – riešenie, ktoré viedlo k zvýšeniu produktivity o 19%.

Často som používal tento pokyn, keď som viedol skupiny – **najmä keď čelili veľmi ťažkému problému, čo je situácia, kde majú členovia skupiny najväčší sklon ihneď navrhovať riešenia.** Aj keď som nemal objektívne kritériá na posúdenie ako kvalitne skupina vyriešila problém, zdalo sa mi, že Maierov pokyn podporuje lepšie riešenia problémov.

To je také pravdivé, že to ani nie je smiešne. A je to horšie a horšie čím ťažší je daný problém. Vezmite si napríklad umelú inteligenciu. Prekvapujúce množstvo ľudí, s ktorými sa stretnem, vie napohľad úplne presne, ako sa má postaviť všeobecná umelá inteligencia, hoci nevedia napríklad povedať, ako postaviť optický rozpoznávač znakov alebo kolaboratívny filtrujúci systém (omnoho ľahšie problémy). Pokiaľ ide o zostrojenie umelej inteligencie s pozitívnym dopadom

na svet – Priateľskej UI, voľne povedané – nuž *ten* problém je taký neuveriteľne zložitý, že *väčšina* vyrieši celú túto tému za 15 sekúnd. *Dajte mi pokoj.*

(Doplnené: Tento problém zďaleka nie je jedinečný pre UI. Fyzici stretávajú mnoho nefyzikov, ktorí majú svoje vlastné fyzikálne teórie; ekonómovia počujú mnoho úžasných nových ekonomických teórií. Ak ste evolučný biológ, každý, koho stretnete, dokáže okamžite vyriešiť ľubovoľný otvorený problém vo vašej oblasti, zvyčajne predpokladajúc skupinovú selekciu. A tak ďalej.)

Maierova rada zdôrazňuje princíp spodného riadku¹, že efektivitu nášho rozhodovania určuje iba to, ktoré indície a postupy sme použili, keď sme prvýkrát došli k nášmu riešeniu – keď už napíšete spodný riadok, je neskoro písať nad neho ďalšie dôvody². Ak urobíte svoje rozhodnutie príliš zavčasu, naozaj bude založené na veľmi malom množstve rozmýšľania, bez ohľadu na to, koľko úžasných argumentov vymyslíte dodatočne.

Ďalej uvážte, že svoje názory meníme zriedkavejšie než si myslíme³: 24 ľudí priradilo v priemere pravdepodobnosť 66 %, že si v budúcnosti vyberú možnosť, ktorá im vtedy pripadala pravdepodobnejšia, ale iba 1 z 24 si naozaj vybral možnosť, ktorá mu vtedy pripadala menej pravdepodobná. **Akonáhle viete uhádnuť, aká bude vaša odpoveď, už ste sa pravdepodobne rozhodli.** Ak dokážete uhádnuť svoju odpoveď pol sekundy po tom, čo počujete otázku, mali ste iba pol sekundy na to, aby ste boli inteligentní. To nie je veľa času.

Tradičná rozumnosť⁴ zdôrazňuje *falzifikáciu* – schopnosť *vzdať sa* pôvodného názoru, keď sa stretnete s jasnou indíciou proti nemu. Ale akonáhle sa myšlienka dostane do vašej hlavy, pravdepodobne bude vyžadovať príliš veľa indícií opäť ju odtiaľ dostať. Čo je horšie, niekedy nemáme luxus prevahy indícií.

Domnievam sa, že omnoho mocnejšia (a náročnejšia) metóda je *odložiť myslenie na odpoveď*. Zastaviť sa, predĺžiť tú chvíľku, keď ešte nevieme uhádnuť, aká bude naša odpoveď; čím svojej inteligencii doprajeme viac času.

Aj pol minúty by bolo zlepšenie oproti polsekunde.

Pôvodný článok s komentármi⁵

¹ pozri: 93 Spodný riadok

² pozri: 95 Racionalizovanie

³ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁴ pozri: 124 Nikto vám nemôže udeliť výnimku zo zákonov rozumnosti

⁵ http://lesswrong.com/lw/ka/hold_off_on_proposing_solutions/

Kapitola 89. Klam pôvodu

V¹ zoznamoch² logických³ chýb⁴ nájdete “klam pôvodu” - chybné napádanie názoru na základe príčin, prečo tomu niekto verí.

To je na prvý pohľad veľmi zvláštna myšlienka – ak príčiny názoru neovplyvňujú jeho systematickú spoľahlivosť, tak čo teda? Ak nám Deep Blue odporučí ťah v šachu, budeme mu dôverovať na základe nášho porozumenia *kódu*, ktorý prehľadáva herný strom, keďže tento herný strom nedokážeme vyhodnotiť sami. Čo by označovalo nejaký pravdepodobnostný argument ako “rozumný”, ak nie to, že bol vytvorený nejakým systematicky spoľahlivým procesom?

Články o klame pôvodu vám povedia, že uvažovanie o pôvode nie je vždy chyba – že pôvod indície *môže* byť relevantný pre jej vyhodnotenie, napríklad v prípade dôveryhodného odborníka. Ale inokedy, hovoria⁵ tieto články, *to je omyl*; chemik Kekulé prvýkrát uvidel prstencovú štruktúru benzénu vo sne, ale to neznamená, že tomto názoru nesmieme nikdy dôverovať.

Takže niekedy je klam pôvodu omylom a niekedy nie?

Klam pôvodu je formálne omylom, pretože *pôvodná príčina* názoru nie je to isté ako jeho *súčasný stav zdôvodniteľnosti*, súhrn všetkých *dnes* známych argumentov za a proti.

Svoje názory však meníme zriedkavejšie než si myslíme.⁶ Obvinenia z pôvodu majú medzi ľuďmi silu, ktorú by nemali medzi ideálnymi bayesiánmi.

Vyčistiť si myseľ je *silná heuristika*, ak nadobudnete nové podozrenie, že mnohé z vašich názorov môžu pochádzať z chybného zdroja.

Akonáhle sa nejaká myšlienka dostane do našich hláv, nie je vždy ľahké vykoreniť ju pomocou indícií. Vezmite si všetkých tých ľudí, ktorí vyrástli veriac Biblii; neskôr odmietli (na vedomej úrovni) predstavu, že Bibliu napísala ruka Boha; a ktorí si predsa myslia, že Biblia obsahuje nevyhnutnú etickú múdrosť⁷. Zabudli si vyčistiť myseľ; robili by omnoho lepšie, keby pochybovali o všetkom, čo Biblia hovorí, *pretože to Biblia hovorí*.

Zároveň by mali v hlavne pevne držať princíp, že obrátená hlúposť nie je inteligencia⁸; cieľom je skutočne si uvoľniť hlavu a myslieť nezávisle, nie negovať Bibliu a mať toto ako svoj algoritmus.

Akonáhle sa nejaká myšlienka dostane do vašej hlavy, budete mať sklon vidieť pre ňu podporu kamkoľvek sa pozriete – a preto, keď je pôvodný zdroj náhle v podozrení, bolo by veru veľmi múdre pochybovať o listoch, ktoré pôvodne vyrástli na tejto vetve.

Ak to dokážete! Nie je ľahké vyčistiť si myseľ. Vyžaduje to krčovité úsilie *naozaj znovu uvažovať*, namiesto nechania svojej mysle, aby zapadla do vzorca opakovania⁹ uložených¹⁰ argumentov. “Nie je to skutočná kríza viery, dokiaľ by veci nemohli rovnako ľahko ísť ľubovoľným smerom,” povedal Thor Shenkel.

¹ <http://www.nizkor.org/features/fallacies/genetic-fallacy.html>

² http://en.wikipedia.org/wiki/List_of_fallacies

³ http://atheism.about.com/library/FAQs/skepticism/blfaq_fall_genetic.htm

⁴ <http://www.fallacyfiles.org/genefall.html>

⁵ <http://www.fallacyfiles.org/genefall.html>

⁶ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁷ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

⁸ pozri: 52 Obrátená hlúposť nie je inteligencia

⁹ pozri: 92 Jeden argument proti armáde

Mali by ste byť *mimoriadne podozrievaví*, ak máte mnoho myšlienok, ktoré vám odporučil zdroj, o ktorom teraz viete, že je nedôveryhodný, ale zhodou okolností sa zdá, že všetky tieto myšlienky boli správne – Biblia je tu samozrejým archetypálnym príkladom.

Na druhej strane... existuje niečo také ako indícia dostatočne jasná, že už významne nezáleží na tom, odkiaľ táto myšlienka pôvodne prišla. Zhromažďovanie tohto druhu jasných indícií, o tom je celá veda¹¹. Nezáleží už na tom, že Kekulé prvýkrát uvidel prstencovú štruktúru benzénu v sne – nezáležalo by na tom, keby sme túto hypotézu na testovanie¹² našli pomocou generovania náhodných obrázkov na počítači, od spiritualistu, ktorý bol odhalený ako podvodník, alebo dokonca z Biblie. Prstencová štruktúra benzénu je podporená toľkými pokusnými indíciami, že zdroj tohto námetu je nepodstatný.

V neprítomnosti takejto jasnej indície musíte venovať pozornosť pôvodným zdrojom myšlienok – dôverovať odborníkovi viac než laikovi, ak si ich oblasť získala rešpekt – podozrievať myšlienky, ktoré ste pôvodne získali z podozrivých zdrojov – nedôverovať tým, ktorých motívy sú nedôveryhodné, *ak* nedokážu predložiť argumenty nezávislé na ich vlastnej autorite.

Klam pôvodu je *omylom* vtedy, keď existujú zdôvodnenia *okrem* pôvodného tvrdenia faktu, ale keď sa obvinenie z pôvodu prekladá, akoby to uzavrelo tému.

Niektoré dobré heuristiky (pre ľudí):

- Dávajte si pozor na všeobecné obvinenia voči názorom, ktoré sa vám nepáčia, najmä ak váš oponent tvrdí, že má aj iné dôvody okrem jednoduchej autority hovoriaceho. "Lietanie je náboženská predstava, preto bratia Wrightovci museli byť klamári" je jeden z klasicky uvádzaných príkladov.
- Rovnako si nemyslite, že dostanete dobrú informáciu o technickej téme tým, že budete múdro psychoanalyzovať zúčastnené osobnosti a ich pomýlené motívy. Ak existujú technické argumenty, majú prednosť.
- Keď vznikne nové podozrenie ohľadom niektorého z vašich základných zdrojov, naozaj by ste *mali* pochybovať o všetkých vetvách a listoch, ktoré z tohto koreňa vyrástli. Neopravňuje vás ho priamo tieto závery vyhodiť, lebo obrátená hlúposť nie je inteligencia, ale...
- Budte extrémne podozrievaví, ak zistíte, že stále veríte dávnyim odporúčeniam zdroja, ktorý ste neskôr odmietli.

Doplnené: Hal Finney navrhuje, aby sme to volali: „heuristika pôvodu“.

Pôvodný článok s komentármi¹³

¹⁰ pozri: 81 Uložené myšlienky

¹¹ pozri: 21 Vedecká indícia, právna indícia, rozumná indícia

¹² pozri: 24 Einsteinova drzosť

¹³ http://lesswrong.com/lw/s3/the_genetic_fallacy/

Časť IX. Proti racionalizácii

Obsah

90 Vedieť o skresleniach môže ľuďom ublížiť	110
91 Aktualizujte postupne	112
92 Jeden argument proti armáde	114
93 Spodný riadok	116
94 Filtrovaná indícia	118
95 Racionalizovanie	120
96 Rozumný argument	122
97 Vyhýbanie sa naozaj slabým miestam vašich názorov	124
98 Motivované zastavenie a motivované pokračovanie	127
99 Falošné zdôvodnenie	129
100 Je toto vaše skutočné odmietnutie?	131
101 Previazané pravdy, nákazlivé lži	134
102 O klamstvách a výbuchoch čiernych labutí	137
103 Epistemológia temnej strany	138

Kapitola 90. Vedieť o skresleniach môže ľuďom ublížiť

Jedného dňa som skúšal povedať mojej mame o probléme kalibrácie odborníkov a povedal som: "Takže keď odborník povie, že si niečím na 99 % istý, v skutočnosti sa to stáva iba v 70 % prípadov." Po krátkej pauze, keď som náhle uvedomil, že hovorím so svojou mamou, som rýchlo dodal: "Samozrejme, takýto skepticizmus musíš uplatňovať nestranne, aj na seba samého, nielen používať ho ako argument proti všetkému, s čím nesúhlasíš..."

A moja mama na to: "Robíš si srandu? Toto je skvelé! To budem používať vždy!"

Taberov a Lodgeov Motivovaný skepticizmus vo vyhodnocovaní politických názorov¹ popisuje potvrdenie šiestich predpovedí:

1. Efekt pôvodného postoja. Osoby, ktoré majú na danú tému silný názor – aj keď sú vyzvané, aby boli objektívne – hodnotia argumenty súhlasiace s ich názorom priaznivejšie než argumenty proti.
2. Sklon nesúhlasiť. Osoby strávia viac času a kognitívnych zdrojov ohováraním argumentov proti než súhlasiacich argumentov.
3. Sklon súhlasiť. Osoby s možnosťou slobodne si vybrať svoje informačné zdroje, budú hľadať skôr súhlasiace než odporujúce zdroje.
4. **Polarizácia postoja. Vystaviť osoby napohľad vyváženej množine argumentov za a proti posilní ich pôvodný postoj.**
5. Efekt sily postoja. Osoby vyjadrujúce silnejšie postoje budú viac náchylné k horeuvedeným sklonom.
6. **Efekt sofistikovanejosti. Osoby, ktoré sa vyznajú v politike, keďže majú viac zbraní na odrážanie nesúhlasiacich faktov a argumentov, budú viac náchylné k horeuvedeným sklonom.**

Ak ste na začiatku iracionálni, viac vedomostí vám môže ublížiť. Pre skutočného bayesiánca informácia nikdy nemôže mať negatívny očakávaný úžitok. Ale ľudia nie sú dokonalí bayesovskí majstri; keď si nedávame pozor, môžeme sa porezať.

Videl som ľudí vážne pošahaných vlastnou znalosťou skreslení. Mali viac zbraní na argumentovanie proti všetkému, čo sa im nepáčilo. A tento problém – príliš veľa pripravených zbraní – je jedným z hlavných dôvodov, prečo ľudia s vysokou mentálnou obratnosťou často skončia ako hlupáci, v tom zmysle hlúposti, ktorý Stanovich nazýva "dysracionalita".

Napadajú vám ľudia, na ktorých tento popis sedí, však? Ľudia s vysokým g-faktorom, ktorí skončia ako menej efektívni, pretože sú príliš sofistickovanými diskutérmi? Myslíte si, že by ste im pomohli – urobili ich efektívnejšími racionalistami – keby ste im jednoducho povedali o zozname klasických sklonov?

Spomínam si na človeka, ktorý sa dozvedel o probléme kalibrácie / prehnanej sebadôvery. Čoskoro na to povedal: "Pozri, nemôžeš dôverovať odborníkovi; ako ukazujú experimenty, veľmi často sa mýlia. Preto, keď predpovedám budúcnosť,

¹ <http://www.stonybrook.edu/polsci/mlodge/lodgemotivated.pdf>

radšej vychádzam z toho, že veci budú pokračovať tak, ako historicky..." a pustil sa do veľmi zložitej, náchylnej k chybám, vysoko otáznej extrapolácie. Akosi, keď došlo na dôveru k jeho vlastným obľúbeným záverom, všetky tieto sklony a omyly sa zdali omnoho menej *výrazné* – napadli mu omnoho zriedkavejšie – než keď potreboval protiargument proti niekomu inému.

Povedal som mu teda o probléme so sklonom nesúhlasiť a sofistikovane argumentovať, a hľa, keď som nabudúce povedal niečo, čo sa mu nepáčilo, označil ma za sofistikovaného argumentátora. Nepokúšal sa ukázať na žiaden konkrétny sofistikovaný argument, žiadnu konkrétnu chybu – iba potriasol hlavou a smutne si vzdychol, ako očividne používam svoju inteligenciu, aby som porazil sám seba. Získal ďalší úplne všeobecný protiargument.

Ešte aj samotný pojem "sofistikovaného argumentátora" môže byť zhubný, ak vám príliš pohotovo prichádza na rozum, keď sa stretnete s napohľad inteligentným človekom, ktorý povie niečo, čo sa vám nepáči.

Snažím sa poučiť zo svojich chýb. Keď som mal naposledy prednášku o heuristikách a sklonoch, na začiatku som tejto všeobecný pojem ilustroval pomocou klamu konjunkcie a heuristiky reprezentatívnosti. Až potom som prešiel k sklonu súhlasiť, sklonu nesúhlasiť, sofistikovanej argumentácii, motivovanému skepticizmu, a ďalším efektom postoja. Ďalších tridsať minút som strávil *zdôrazňovaním* tejto témy, vracal som sa k nej z toľkých rôznych uhlov, ako som len vedel.

Chcel som, aby sa moje obecenstvo o túto tému začalo zaujímať. Nuž, na to by stačil jednoduchý popis klamu konjunkcie a heuristiky reprezentatívnosti. Ale predpokladajme, že sa zaujímať začali. Čo ďalej? Literatúra o sklonoch je väčšinou kognitívna psychológia pre samotnú kognitívnu psychológiu. Musel som svojmu obecenstvu dať veľké varovanie počas tejto jednej lekcie, inak by si pravdepodobne už nikdy nevypočuli.

Či v článku alebo na prednáške, odteraz sa vždy snažím nespomínať kalibráciu a prehnanú sebadôveru, dokiaľ najprv neobjasním sklon nesúhlasiť, motivovaný skepticizmus, sofistikovaných argumentátorov a dysracionalitu mentálne obratných. V prvom rade, neublížiť!

Pôvodný článok s komentármi²

² http://lesswrong.com/lw/he/knowning_about_biases_can_hurt_people/

Kapitola 91. Aktualizujte postupne

Politika zabíja myslenie.¹ Debata je vojna, argumenty sú vojaci². Existuje pokúšenie hľadať spôsoby, ako interpretovať všetky možné experimentálne výsledky³ ako potvrdzujúce vašu teóriu, ako keď opevňujete citadelu pred každým možným smerom útoku. To nemôžete. Je to matematicky nemožné. Pre každé očakávanie indície existuje rovnaké očakávanie protiindície v opačnom smere.⁴

Ale je v poriadku, ak vaša obľúbená teória nie je *dokonale* chránená. Ak hypotéza znie, že na minci padne hlava v 95% prípadov, potom v jednom prípade z dvadsiatich uvidíte niečo, čo vyzerá ako indícia proti. To je v poriadku. To je normálne. Dokonca sa to očakáva, pokiaľ máte devätnásť potvrdzujúcich pozorovaní na každé odporujúce. Pravdepodobnostný model môže dostať jeden či dva zásahy⁵ a stále prežiť, pokiaľ tieto zásahy *nepokračujú*.

Napriek tomu sa všeobecne verí, najmä na súde verejnej mienky, že pravdivá teória nemôže mať *žiadne* zlyhania, a nepravdivá teória *žiadne* úspechy.

Nájdete ľudí, ktorí sa držia jediného kúska toho, čo považujú za indíciu, a tvrdia, že ich teória to dokáže "vysvetliť", akoby toto bola celá podpora, akú len môže teória potrebovať. Podľa všetkého nepravdivá teória nemôže mať *žiadnu* indíciu vo svoj prospech; je nemožné, aby nepravdivá teória dokázala vysvetliť čo len jedinou udalosť. Tým pádom je jeden kus potvrdzujúcej indície všetkým, čo teória potrebuje.

Je iba o čosi menej hlúpe držať sa jediného kúska *pravdepodobnostnej* protiindície ako vyvrátenia, akoby sa proti správnej teórii nemohlo ani *trošku* argumentovať. Ale presne takto ľudia argumentovali celé veky, snažili sa poraziť všetky nepriateľské argumenty, zatiaľ čo nepriateľovi nedopriali ani jediný kúsok podpory. Ľudia chcú, aby ich debaty boli jednostranné; sú zvyknutí na svet, v ktorom proti ich obľúbeným teóriám neexistuje jediný kúsok protipodpory. Preto, pripustiť jediný kus pravdepodobnostnej protiindície by znamenalo koniec sveta.

Ja viem, že niekto v obecenstve ide povedať: "Ale v skutočnom svete *nemôžeš* pripustiť ani jediný bod, ak chceš vyhrať debatu! Ak pripustiš, že existuje nejaký protiargument, nepriateľ ti ho bude omieľať dokola a dokola – to *nemôžeš* nepriateľovi dovoliť! *Prehráš!* Čo by mohlo byť fyzicky desivejšie než *toto*?"

No a čo. Rozumnosť nie je na vyhrávanie debát, je na rozhodovanie, ku ktorej strane sa pridať. Ak ste sa už rozhodli, za ktorú stranu idete argumentovať, rozumnú časť ste tým *ukončili*, či už dobre alebo úboho. Ale ako sa môžete, sami od seba, rozhodnúť, za ktorú stranu budete argumentovať? Ak *vybrať si nesprávnu stranu* je fyzicky desivé, hoci len trošku fyzicky desivé, mali by ste spracovať *všetky* indície.

Rozumnosť nie je chôdza, ale tanec. Pri každom kroku tohto tanca by vaša noha mala dopadnúť presne na správne miesto, nie viac doľava ani doprava. Posuňte názor vyššie s každým kuskom potvrdzujúcej indície. Posuňte názor nižšie s každým kuskom opačnej indície. Áno, *nižšie*. Aj keď je model správny, ak to nie je presný model, niekedy budete musieť prehodnotiť svoj názor smerom *nadol*.

¹ pozri: 47 Politika zabíja myslenie

² pozri: 48 Debaty o pravidlách by nemali vyzeráť jednostranne

³ pozri: 13 Viera vo vieru

⁴ pozri: 29 Zákon zachovania očakávanej indície

⁵ http://lesswrong.com/lw/ig/i_defy_the_data/

Ak jeden či dva kúsky indície sú náhodou protipodporou vášho názoru, to je v poriadku. To sa pravdepodobnostným indíciám pre nepresné teórie občas stáva. (Keď zlyhá presná teória, vtedy *máte* problém!) Len posuňte svoj názor o čosi nadol – pravdepodobnosť, pomer šancí, alebo hoci aj neverbálnu silu dôvery vo vašej mysli. Len ho posuňte o čosi nadol a čakajte na ďalšie indície⁶. Ak je teória pravdivá, čoskoro príde podporujúca indícia a pravdepodobnosť opäť začne šplhať nahor. Ak je teória nepravdivá, tak ju predsa naozaj nechcete.

Problém používania čierno-bieleho, binárneho, kvalitatívneho rozmyšľania je, že ľubovoľné jedno pozorovanie buď zničí teóriu alebo ju nezničí. Keď nie je dovolené jediné odporujúce pozorovanie, vytvára to kognitívnu disonanciu a treba to odargumentovať⁷. A toto vylučuje postupný pokrok; vylučuje to správne zapojenie všetkých indícií. Pri pravdepodobnostnom uvažovaní si uvedomujeme, že správna teória bude v priemere vytvárať väčšiu váhu podpory než protipodpory. A tak si môžete *bez strachu* povedať: “Toto je mierna indícia proti, posuniem svoj názor nadol.” Áno, *nadol*. To neničí vašu obľúbenú teóriu. To by bolo kvalitatívne uvažovanie; myslíte kvantitatívne.

Pre každé očakávanie indície existuje očakávanie rovnakej protiindície v opačnom smere.⁸ V každej situácii musíte v priemere očakávať upravenie svojho názoru nadol rovnako ako musíte očakávať upravenie svojho názoru nahor. Ak si myslíte, že už viete, aká indícia príde, potom si musíte byť svojou teóriou pomerne istí – pravdepodobnosť blízka k 1 – čo vám nenecháva veľa miesta na ďalšie zvyšovanie pravdepodobnosti. A akokoľvek nepravdepodobné sa zdá, že by ste natrafili na vyvracajúcu indíciu, výsledný posun nadol musí byť dosť veľký na to, aby presne vyvážil očakávaný zisk na opačnej strane. Vážený priemer vašej očakávanej výslednej pravdepodobnosti sa musí rovnať vašej pôvodnej pravdepodobnosti.

Aké bláznivé je potom báť sa⁹ upravovania svojej pravdepodobnosti nadol, ak sa vôbec unívate danú záležitosť vyšetrovať? V priemere musíte očakávať rovnako veľa posunu nadol ako nahor z každého jednotlivého pozorovania.

Možno sa stane, že kúsok protipodpory príde zase, a zase, a zase, zatiaľ čo nová podpora priteká pomaly. Možno zistíte, že váš názor sa posúva stále nižšie a nižšie. Až si nakoniec uvedomíte, z ktorého smeru proti vám fúka vietor indícií. V tej chvíli uvedomenia nemá zmysel zostavovať výhovorky. V tej chvíli uvedomenia ste sa svojho obľúbeného názoru *už vzdali*. Hurá! Čas na oslavu! Otvorte fľašu šampanského a pošlite pre pizzu! Držaním sa svojich pôvodných názorov sa predsa nemôžete stať silnejší¹⁰.

Pôvodný článok s komentármi¹¹

⁶ pozri: 29 Zákon zachovania očakávanej indície

⁷ http://lesswrong.com/lw/ig/i_defy_the_data/

⁸ pozri: 29 Zákon zachovania očakávanej indície

⁹ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

¹⁰ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

¹¹ http://lesswrong.com/lw/ij/update_yourself_incrementally/

Kapitola 92. Jeden argument proti armáde

Včera som hovoril o spôsobe uvažovania, pri ktorom nie je dovolený jediný argument proti, s tým výsledkom, že každú nepodporujúcu indíciu treba odargumentovať.¹ Dnes naznačujem, že keď sa ľudia stretnú s argumentom proti, zabráni zníženiu svojej istoty tak, že si *zopakujú* už známu podporu.

Predstavte si, že krajina Freedonia debatuje, či ich sused, Sylvania, môže za nedávnu hromadu meteoritov padajúcich na ich mestá. Sú niektoré indície, ktoré to naznačujú: meteority zasiahli mestá blízko sylvaniánskej hranice; *pred* dopadmi bola nezvyčajná aktivita na sylvaniánskej burze cenných papierov; a sylvaniánskeho veľvyslanca Trentina niekto počul šomrať o "nebeskej pomste".

Niektori prídu a povie vám: "Nemyslím si, že Sylvania je zodpovedná za tieto padajúce meteority. Každoročne s nami obchodujú v miliardách dinárov." "Pozri," odpoviete mu, "meteority zasiahli mestá blízko Sylvanie, na ich burze cenných papierov bola podozrivá aktivita, a ich veľvyslanec potom hovoril o nebeskej pomste." Keďže tieto tri argumenty prevážia onen jeden, *zachováte* si názor, že za to môže Sylvania – veríte namiesto neverenia, kvalitatívne. Rovnováha indícií je zrejme proti Sylvanii.

Potom za vami prídu ďalší a povie: "Nemyslím si, že Sylvania je zodpovedná za tieto padajúce meteority. Zmeniť smer padajúcich meteoritov je veľmi ťažké. Sylvania dokonca ani nemá vesmírny program." Odpoviete: "Ale meteority zasiahli mestá blízko Sylvanie, a ich investori to vedeli, a veľvyslanec to priamo priznal!" Opäť, tieto tri argumenty prevážia onen jeden (prevahou troch proti jednému), takže si zachováte názor, že za to môže Sylvania.

Ba čo viac, vaše presvedčenie sa *posilnilo*. Teraz ste už v dvoch oddelených situáciách vyhodnotili rovnováhu dôkazov, a oba razy bola rovnováha naklonená proti Sylvanii s pomerom 3 ku 1.

Stretávate sa s ďalšími argumentmi prosylvanských zradcov – opäť, a opäť, a ešte stokrát – ale zakaždým je nový argument hravo porazený 3 ku 1. A v každej situácii sa cítite viac presvedčení, že Sylvania za to naozaj môže; posúvate svoje pôvodné presvedčenie podľa pocítovanej rovnováhy indícií.

Problém je samozrejme v tom, že *opakovaním* argumentov, ktoré *ste už vedeli*, počítate danú indíciu dvojmo. To by bol vážny hriech ešte aj v prípade, že by ste dvojmo počítali *všetky* indície. (Predstavte si vedca, ktorý urobí pokus s 50 jedincami a nezíska štatisticky významné výsledky, a tak všetky údaje započíta dvakrát.)

Ale selektívne počítať dvojmo *iba niektoré indície*, to je číra fraška. Pamätám si, že ako dieťa som videl kreslený film, v ktorom záporný hrdina rozdeľoval korisť pomocou nasledujúceho algoritmu: "Jedna zlatka tebe, jedna mne. Druhá tebe, jedna-dve mne. Tretia tebe, jedna-dve-tri mne."

Ako som včera² zdôraznil, aj keď je obľúbený názor *pravdivý*, racionalista môže občas potrebovať znížiť jeho pravdepodobnosť pri zapájaní *všetkých* indícií. Áno, rovnováha podpory môže stále uprednostniť váš obľúbený názor. Ale aj tak musíte posunúť pravdepodobnosť *nadol* – áno, *nadol* – z akejkoľvek hodnoty, kde bola predtým než ste počuli indíciu proti. Nepomôže *zopakovať* si podporujúce argumenty, pretože tie ste už zohľadnili.

¹ pozri: 91 Aktualizujte postupne

² pozri: 91 Aktualizujte postupne

A predsa sa mi zdá, že keď sa ľudia stretnú s *novým* protiargumentom, hľadajú ospravedlnenie, prečo neposunúť svoje presvedčenie nadol, a samozrejme nájdú podporujúce argumenty, ktoré *už poznali*. Musím si neustále dávať pozor, aby som to nerobil ja sám! Pripadá mi to rovnako prirodzené ako zablokovať úder meča pripraveným štítom.

Pri správnom druhu nesprávneho rozmýšľania, hrstka podpory – dokonca aj jediný argument – dokáže odraziť celú armádu protirečení.

Pôvodný článok s komentármi³

³ http://lesswrong.com/lw/ik/one_argument_against_an_army/

Kapitola 93. Spodný riadok

V aukcii sú dve zapečatené krabice, krabica A a krabica B. Práve jedna z týchto krabíc obsahuje vzácny diamant. Existujú všelijaké príznaky a znamenia naznačujúce, či daná krabica obsahuje diamant; ja však o žiadnom z nich *neviem*, že by bolo celkom spoľahlivé. Napríklad na jednej krabici je modrá pečiatka, a ja viem, že krabice, ktoré obsahujú diamant, majú väčšiu šancu mať modrú pečiatku než prázdne krabice. Alebo jedna krabica má lesklý povrch a ja mám podozrenie – ale nie som si istý – že krabice s diamantom nebývajú lesklé.

Predpokladajme, že je tam chytrý rečník, ktorý drží hárok papiera a hovorí majiteľom krabice A a krabice B: “Vydražte si moje služby, a ktorý vás získa moje služby, za toho budem argumentovať, že jeho krabica obsahuje diamant, takže za tú krabicu dostane vyššiu cenu.” Takže majitelia krabíc dražia, a majiteľ krabice B ponúkne viac, čím získa služby chytrého rečníka.

Chytrý rečník si začne organizovať myšlienky. Najprv napíše: “A *preto*, krabica B obsahuje diamant!” na spodok svojho hárku papiera. Potom na vrch papiera napíše: “Krabica B má na sebe modrú pečiatku,” a pod to: “Krabica A je lesklá,” a ešte: “Krabica B je ľahšia než krabica A,” a tak ďalej, pre mnohé ďalšie príznaky a znamenia; chytrý rečník akurát ignoruje všetky znamenia, ktoré by mohli svedčiť v prospech krabice A. Nakoniec chytrý rečník príde za mnou a prečíta mi zo svojho hárku papiera: “Krabica B má na sebe modrú pečiatku, a krabica A je lesklá,” a tak ďalej, až kým sa dostane k: “A *preto*, krabica B obsahuje diamant!”

Zvážte však: Vo chvíli, keď chytrý diskutér napísal svoj záver, vo chvíli, keď atrament vsiakol do jeho hárku papiera, previazanosť indíciami¹ medzi týmto fyzickým atramentom a fyzickými krabicami bola pevne daná.

Možno vám pomôže predstaviť si množinu svetov – Everettove vetvy alebo Tegmarkove duplikáty² – v ktorých je nejaká objektívna frekvencia, s ktorou krabica A alebo krabica B obsahuje diamant. Podobne je nejaká objektívna frekvencia v rámci podmnožiny “svety s lesklou krabicou A”, kde krabica B obsahuje diamant; a nejaká objektívna frekvencia v “svetoch s lesklou krabicou A a krabicou B s modrou pečiatkou,” kde krabica B obsahuje diamant.

Atrament na papieri je sformovaný do čudných tvarov a kriviek, ktoré vyzerajú ako text: “A *preto*, krabica B obsahuje diamant.” Keby ste boli gramotný anglicky hovoriaci človek, mohlo by vás to popliesť a mohli by ste si myslieť, že takto pokrútený atrament nejakú *znamená*, že krabica B obsahuje diamant. Pokusné osoby, ktoré majú za úlohu povedať farbu na vytlačennom obrázku a dostanú obrázok “zelená” (napísané červenou farbou), často povedia “zelená” namiesto “červená”. Pomáha byť negramotný, vtedy vás nemýli tvar atramentu.

Pre nás je dôležitou stránkou veci jej previazanosť s inými vecami. Vezmime si opäť tú množinu svetov, Everettových vetiev alebo Tegmarkových duplikátov. Vo chvíli, keď všetci chytrí rečníci vo všetkých svetoch umiestnia atrament na spodok svojho papiera – predpokladajme, že je to v jednej chvíli – pevne tým určili koreláciu atramentu s krabicami. Chytrý rečník píše nezmizikovateľným perom; atrament sa už nezmení. Krabice sa už nezmenia. V podmnožine svetov, kde atrament hovorí “A *preto*, krabica B obsahuje diamant,” je už pevne dané nejaké percento svetov, kde krabica A obsahuje diamant. A to sa nezmení bez ohľadu na to, čo sa dopíše na prázdne riadky nad tým.

Takže indíciová previazanosť atramentu je pevne daná a ponechávam na vás, aby ste sa rozhodli, aká môže byť. Možno majitelia krabíc, ktorí veria, že sa

¹ pozri: 20 Čo je indícia?

² <http://www.arxiv.org/abs/astro-ph/0302131>

v ich prospech dá zostaviť lepší argument, sú ochotnejší prenajať si reklamu; možno majitelia krabíc, ktorí sa boja svojej nedostatočnosti, dražia vyššie. Ak majitelia krabíc sami nerozumejú príznakom a znameniam, potom bude atrament naprosto nepreviazaný s obsahmi krabíc, hoci vám môže povedať niečo o majiteľových financiách a dražiacich zvykoch.

Teraz si predstavte, že iný človek je úprimne zvedavý, a *najprv* si na papier napíše všetky odlišné príznaky *oboch* krabíc, potom aplikuje svoje poznanie a zákony pravdepodobnosti, a napíše naspodok: "*Preto* odhadujem s pravdepodobnosťou 85 %, že krabica B obsahuje diamant." Čoho indíciou je tento nápis? Keď skúmam reťaz príčiny a následku vedúcu k tomuto fyzickému atramentu na fyzickom papieri, zisťujem, že reťaz kauzality sa vinie cez všetky príznaky a znamenia krabíc, a závisí od týchto príznakov; vo svetoch s odlišnými príznakmi je naspodku papiera napísaná iná pravdepodobnosť.

Preto je písmo zvedavého výskumníka previazané s príznakmi a znameniami a s obsahom krabíc, zatiaľ čo písmo chytrého rečníka je indíciou iba toho, ktorý majiteľ mu v dražbe ponúkol viac. Toto je veľký rozdiel vo význame atramentu, hoci ten, kto hlúpo nahlas prečíta tvar atramentu, si môže myslieť, že napísané slová znejú podobne.

Vaša efektivita ako racionalistu je daný tým, ktorý algoritmus naozaj napíše spodný riadok vašich myšlienok. Ak vaše auto vydáva pri brzdení kovové škripajúce zvuky, ale vy nie ste ochotný čeliť finančným nákladom na opravu brzd, môžete sa rozhodnúť hľadať dôvody, prečo vaše auto netreba opraviť. Ale skutočné percento vás, ktoré prežije v Everettových vetvách alebo Tegmarkových svetoch – ktoré použijeme na opis vaše efektivity ako racionalistu – je dané algoritmom, ktorý rozhodne, pre *ktorý* záver budete hľadať argumenty. V tomto prípade je skutočný algoritmus: "Nikdy neopravuj nič drahé." Ak je to dobrý algoritmus, fajn; ak je to zlý algoritmus, smola. Argumenty, ktoré dopíšete potom, nad spodný riadok, nezmenia nič žiadnym smerom.

Dodatok: Toto je myslené ako varovanie proti vášmu vlastnému uvažovaniu, nie ako Úplne Všeobecný Protiargument proti záverom, ktoré sa vám nepáčia. Pretože povedať "Môj oponent je chytrý rečník" je chytrý argument, ak je vaším cieľom zachovať si všetky názory, ktoré ste mali na začiatku. Najchytrejší rečník na svete môže poukázať na to, že slnko svieti, a predsa bude naďalej deň. Viac na túto tému v článku Filtrovaná indícia³.

Pôvodný článok s komentármi⁴

³ pozri: 94 Filtrovaná indícia

⁴ http://www.lesswrong.com/lw/js/the_bottom_line/

Kapitola 94. Filtrovaná indícia

Včera som rozoberal dilemu chytrého rečníka¹, najatého, aby vám predal krabicu, ktorá možno obsahuje diamant a možno nie. Chytrý rečník vám zdôrazní, že na krabici je modrá pečiatka, a že je známy fakt, že krabice obsahujúce diamant majú väčšiu šancu mať modrú pečiatku než prázdne krabice. Čo sa v tomto bode deje z bayesiánskeho pohľadu? Musíte bezmocne aktualizovať svoje pravdepodobnosti tak, ako si to chytrý rečník želá?

Ak sa môžete pozrieť na danú krabicu osobne, môžete si všetky príznaky zrátať sami. Čo ak sa nemôžete pozrieť? Čo ak jediná indícia, ktorú máte, je slovo chytrého rečníka, ktorý je zákonom obmedzený na hovorenie iba pravdivých výrokov, ale nehovorí vám všetko, čo vie? Každý výrok, ktorý povie, je platná indícia – ako by ste mohli *neaktualizovať* svoje pravdepodobnosti? Prestalo azda byť pravdou, že v takom a takom percente Everettových vetiev alebo Tegmarkových duplikátov, v ktorých krabica B má modrú pečiatku, krabica B obsahuje diamant? Podľa Jaynesa, bayesiánec musí vždy zohľadniť všetky známe indicie, pod hrozbou paradoxu. Lenže potom vás chytrý rečník dokáže presvedčiť o hocičom, čo bude chcieť, pokiaľ je dostatočná paleta príznakov, o ktorých vás môže selektívne informovať. To neznie správne.

Vezmite si jednoduchý prípad, nevyváženú mincu, na ktorej buď padá 2/3 hlava a 1/3 znak, alebo 1/3 hlava a 2/3 znak; obe tieto možnosti sú a priori rovnako pravdepodobné. Každé pozorované H je 1 bit² indicie pre H-nevyváženú mincu; každé pozorované Z je jeden bit indicie pre Z-nevyváženú mincu. Hodím mincou desaťkrát a potom vám poviem: “Pri 4. hode, 6. hode, a 9. hode padla hlava.” Aká je vaša výsledná pravdepodobnosť, že minca je H-nevyvážená.

A odpoveď je, že to môže byť takmer čokoľvek, podľa toho, aká reťaz príčin a následkov leží za mojím vyslovením týchto slov – mojím výberom, ktoré hody vám oznámiť.

- Možno sa riadim algoritmom, že poviem výsledok 4., 6. a 9. hodu bez ohľadu na výsledky, ktoré padli v týchto alebo ostatných hodoch. Ak viete, že som použil tento algoritmus, výsledné šance sú 8:1 v prospech H-nevyvázenej mince.
- Možno oznamujem všetky hody, na ktorých padla hlava, a nič iné. V tom prípade viete, že vo všetkých zvyšných 7 hodoch padol znak, a výsledné šance sú 1:16 proti H-nevyvázenej minci.
- Možno som sa dopredu rozhodol, že vám poviem výsledok 4., 6. a 9. hodu iba vtedy, ak pravdepodobnosť, že minca je H-nevyvážená prekročí 98 %. A tak ďalej.

Alebo si vezmite problém Montyho Halla:

V hernej show dostanete na výber tri dvere vedúce do troch miestností. Viete, že v jednej miestnosti je 100 000 dolárov, a zvyšné dve sú prázdne. Moderátor vás požiada, aby ste si vybrali dvere, a vy si vyberiete dvere číslo 1. Potom moderátor otvorí dvere číslo 2 a ukáže prázdnu izbu. Chcete zameniť svoje rozhodnutie za číslo 3 alebo zostávate s dverami číslo 1?

Odpoveď závisí od moderátorovho algoritmu. Ak moderátor vždy otvorí niektoré dvere a vždy si vyberie dvere vedúce do prázdnej miestnosti, potom by ste mali zameniť rozhodnutie za dvere číslo 3. Ak moderátor vždy otvára dvere číslo 2

¹ pozri: 93 Spodný riadok

² pozri: 22 Koľko indicie treba?

bez ohľadu na to, čo je za nimi, dvere 1 aj 3 majú obe pravdepodobnosť 50 %, že obsahujú peniaze. Ak moderátor otvára nejaké dvere iba v tých prípadoch, keď ste si na prvýkrát vybrali dvere s peniazmi, potom by ste mali jednoznačne zostať s číslom 1.

Nemali by ste zohľadniť iba to, že dvere číslo 2 sú prázdne, ale tento fakt plus fakt, že sa moderátor *rozhodol* otvoriť dvere číslo 2. Mnohí ľudia sú zo štandardného problému Montyho Halla zmätení, pretože aktualizujú len o informáciu, že číslo 2 je prázdne, a v tom prípade majú čísla 1 a 3 rovnakú pravdepodobnosť obsahovať peniaze. To je dôvod, prečo majú bayesiánci prikázané zohľadňovať všetky svoje vedomosti, pod hrozbou paradoxu.

Keď niekto povie: "Pri 4. hode mincou padla hlava," nezohľadňujeme iba to, že pri 4. hode mincou padla hlava - neberieme si podmnožinu všetkých možných svetov, kde v 4. hode padla hlava - ale zohľadňujeme podmnožinu všetkých možných svetov, kde rečník riadiaci sa nejakým konkrétnym algoritmom *povedal*: "Pri 4. hode mincou padla hlava." Vyslovená veta nie je sama osebe faktom; nenechajte sa zviazať z cesty púhym významom slov.

Väčšina súdnych procesov funguje podľa teórie, že každý prípad má práve dve odporujúce si strany³, a že je jednoduchšie nájsť dvoch zaujatých ľudí než jedného nezaujatého. Či už žalobca alebo obhajca, *niekto* z nich má motív predložiť každý kúsok indície, takže súd uvidí všetky indície; to je teória. Ak sú v dileme s dvoma krabicami dvaja chytrí rečníci, nie je to až také dobré ako jeden zvedavý výskumník, ale je to skoro také dobré. Ale to platí, keď sú dve krabice. Skutočnosť často obsahuje mnohostranné problémy, a hlboké problémy, a nesamozrejmé odpovede, ktoré sa nenájdu ľahko tým, že Modrý a Zelený kričia jeden na druhého⁴.

Dajte si pozor, aby ste nezneužívali pojem filtrovania indície ako Úplne Všeobecný Protiargument na vylúčenie všetkých indícií, ktoré sa vám nepáčia: "Tento argument bol filtrovaný, preto ho môžem ignorovať." Ak vás podráždil argument opačnej strany, potom ste už zoznámení s prípadom, a dosť zaujatí na to, aby ste si vybrali jednu stranu. Pravdepodobne už poznáte najsilnejšie argumenty vašej strany. Nemáte žiaden dôvod dedukovať na základe argumentu proti, že existujú nové⁵ príznaky a znamenia vo váš prospech, ktoré ste ešte nevideli⁶. Zostávajú vám teda samotné nepohodlné fakty; modrá pečiatka na krabici B je stále indíciou.

Ale ak počujete daný argument po prvýkrát, a ak počujete iba jednu stranu argumentu, potom by ste si veru mali dať pozor! V istom zmysle, nikto nemôže *naozaj* veriť teórii prirodzeného výberu, dokiaľ nepočúval kreacionistov aspoň päť minút; až *potom* vie, že je to nabetón.

Pôvodný článok s komentármi⁷

³ pozri: 49 Váhy spravodlivosti, zápisník rozumnosti

⁴ pozri: 47 Politika zabíja myslenie

⁵ pozri: 92 Jeden argument proti armáde

⁶ pozri: 92 Jeden argument proti armáde

⁷ http://www.lesswrong.com/lw/jt/what_evidence_filtered_evidence/

Kapitola 95. Racionalizovanie

V článku “Spodný riadok” som predložil dilemu dvoch krabíc, z ktorých iba jedna obsahuje diamant, s rôznymi príznakmi a znameniami ako indíciami. Definoval som rozdiel medzi zvedavým výskumníkom a chytrým rečníkom. Zvedavý výskumník si zapíše všetky príznaky a znamenania, spracuje ich, a nakoniec napíše: “*Preto* odhadujem, že s pravdepodobnosťou 85% krabica B obsahuje diamant.” Chytrý rečník pracuje pre toho, kto si ho vydražil, a začne tým, že napíše: “*Preto*, krabica B obsahuje diamant,” a potom vyberá priaznivé príznaky a znamenania do riadkov nad tým.

Ten prvý postup je rozumnosť. Ten druhý postup je všeobecne známy ako “racionalizovanie”.

“Racionalizovanie.” Aký čudný názov. Podľa mňa je to *nesprávne slovo*. Nemôžete “racionalizovať” niečo, čo nie je už racionálne. To je ako keby sme “klamanie” nazývali “pravdizovanie”.

Na čisto výpočtovej úrovni je obrovský rozdiel medzi:

1. Začatím od indícií, a potom spracovaním tokov pravdepodobnosti, s cieľom vypísať pravdepodobný záver. (Zapísať si všetky príznaky a znamenania, a potom sa nechať unášať k pravdepodobnosti na spodnom riadku¹, ktorá závisí od týchto príznakov a znamení.)
2. Začatím od záveru, a potom spracovaním tokov pravdepodobnosti s cieľom vypísať indície, ktoré napohľad podporujú tento záver. (Zapísať si spodný riadok, a potom plávať proti prúdu a vyberať² príznaky a znamenania na prezentáciu³ v horných riadkoch.)

Ktorý blázon vymyslel také mätúco podobné slová “racionalita” a “racionalizovanie”, aby opísal takéto mimoriadne odlišné myšlienkové procesy? Dal by som prednosť pojmom, ktoré by urobili tento algoritmickej rozdiel zrejším, povedzme “rozumnosť” verzus “obrovská vsávajúca kognitívna čierna diera”.

Nie každá zmena je zlepšenie, ale každé zlepšenie je nevyhnutne zmena. Nemôžete pre pevne daný výrok získať viac pravdy tým, že preň budete argumentovať; môžete o ňom presvedčiť viac ľudí, ale nemôžete ho urobiť *pravdivejším*. Aby sme zlepšili svoje názory, nevyhnutne musíme meniť svoje názory. Rozumnosť je operácia, ktorú používame na získanie väčšej pravdivostnej hodnoty pre naše názory tým, že ich meníme. Racionalizovanie funguje, aby upevnilo názory; lepšie by ho bolo nazvať “anti-rozumnosť”, aj kvôli praktickým výsledkom, aj kvôli otočenému algoritmu.

“Rozumnosť” je plávanie *po prúde*, ktoré zbiera indície, zvažuje ich, a vypisuje záver. Zvedavý výskumník používa algoritmus plávania po prúde: *najprv* zbiera indície, píše si zoznam všetkých viditeľných príznakov a znamení, ktoré potom *po prúde* spracuje, aby získal predtým neznámu pravdepodobnosť, že krabica obsahuje diamant. Po celý ten čas proces rozumnosti šiel dopredu, zvedavý výskumník ešte nepoznal svoj cieľ, a práve preto bol *zvedavý*. Pri Bayesovej Ceste sa pôvodná pravdepodobnosť rovná očakávanej výslednej pravdepodobnosti⁴. Ak poznáte svoj cieľ, už ste tam.

¹ pozri: 93 Spodný riadok

² pozri: 94 Filtrovaná indícia

³ pozri: 94 Filtrovaná indícia

⁴ pozri: 29 Zákon zachovania očakávanej indície

“Racionalizácia” je plávanie *proti prúdu*, od záveru k vybraným indíciám. Najprv si napíšete spodný riadok, ktorá je známy a pevne daný; cieľom vášho spracovávania je zistiť, ktoré argumenty máte napísať na predchádzajúce riadky. Toto, a nie záver, je neznámou premennou daného procesu.

Obávam sa, že Tradičná Rozumnosť nerobí svojich používateľov primerane citlivými na rozdiely medzi plávaním po prúde a proti prúdu. V Tradičnej Rozumnosti nie je nič zlé na tom, keď si vedec nájde svoju oblúbenú hypotézu a potom začne hľadať pokus, ktorý ju dokáže. Tradičný Racionalista by sa na toto pozrel súhlasne a povedal: “Pýcha je motorom, ktorý poháňa Vedu vpred.” Áno, *je* to motor, ktorý poháňa Vedu vpred. Lahšie nájdete žalobcu a obhajcu zaujatých opačnými smermi, než jedného nezaujatého človeka.

Ale že niečo robí každý, to ešte danú vec nerobí správnou. Bolo by ešte lepšie, keby ten vedec, keď nájde svoju oblúbenú hypotézu, ju začal *testovať* z čírej *zvedavosti* – vytvárať pokusy, ktoré budú posúvať jeho vlastné názory neznámym smerom⁵.

Ak naozaj nevíete, kam idete, pravdepodobne ste celkom zvedaví, kam to bude. Zvedavosť je prvou cnosťou⁶, bez ktorej vaše otázky budú bezcieľné a vaše schopnosti nesústreďené.

Pocíťujte prúd Sily, a dajte si pozor, aby neprúdila naopak.

Pôvodný článok s komentármi⁷

⁵ pozri: 29 Zákon zachovania očakávanej indície

⁶ pozri: 258 Dvanásť cností rozumnosti

⁷ <http://www.lesswrong.com/lw/ju/rationalization/>

Kapitola 96. Rozumný argument

Vaše povolanie je menežer kampaní a práve vás najal Mortimer Q. Snodgrass, kandidát Zelených na starostu Hadleyburgu. Ako menežer kampane si čítate blog o rozumnosti a po rozume vám behá najmä jedna myšlienka: “Ako môžem zostaviť bezchybne rozumný argument, že Mortimer Q. Snodgrass je najlepší kandidát na starostu Hadleyburgu?”

Prepáčte. To sa nedá.

“Čože?” skríknete. “Ale čo ak použijem iba platnú podporu na konštrukciu svojej štruktúry rozmyšľania? Čo ak každý citovaný fakt bude pravdivý podľa môjho najlepšieho vedomia, a bude to relevantná indícia¹ podľa Bayesovho pravidla²?”

Prepáčte. Ani tak sa to nedá. Porazili ste sám seba v tom okamihu, keď ste si dopredu určili záver vášho argumentu.

Tento rok *Hadleyburský Hlásnik* poslal dotazník so 16 položkami všetkým kandidátom na starostu, s otázkami ako: “Dokážete namaľovať všetky farby vetra?” a “Nadychuteje sa?” Žiaľ, budovu *Hlásnika* zničil meteorit pred uverejnením výsledkov. Čo je škoda, pretože váš kandidát, Mortimer Q. Snodgrass, vychádza dobre v porovnaní so svojimi súpermi v 15 zo 16 otázok. Jediná problematická časť bola otázka číslo 11: “Ste teraz, alebo ste niekedy v minulosti boli, superzloduch?”

Ste teda v pokušení uverejniť tento dotazník ako súčasť svojej kampaňovej literatúry... samozrejme s vynechaním otázky číslo 11.

Čím ste prekročili hranicu medzi *rozumnosťou* a *racionalizáciou*. Už nie je možné, aby sa voliči rozhodovali iba na základe faktov; musia sa rozhodovať aj na základe doplnujúceho faktu³ ich prezentácie, a odvodzovať existenciu skrytých indícií.

Tú hranicu ste vlastne prekročili v bode, keď ste zvažovali, či tento dotazník je pre vášho kandidáta priaznivý alebo nepriaznivý pred tým, ako ste sa rozhodli, či ho uverejníte. “Čože!” vykriknete. “Kampaň by mala uverejňovať fakty nepriaznivé pre svojho kandidáta?” Ale postavte sa do role voliča, ktorý si stále skúša vybrať kandidáta – prečo by ste cenzurovali užitočnú informáciu? Keby ste boli úprimne zvedaví, nerobili by ste to. Keby ste plávali *po prúde* od indícií k neznámej voľbe kandidáta, namiesto plávania *proti prúdu* od vopred určeného kandidáta smerom k výberu argumentov.

“Logický” argument je taký, ktorý vyplýva zo svojich predpokladov. Napríklad nasledujúci argument je *nelogický*:

- Všetky obdĺžniky majú štyri strany.
- Všetky štvorce majú štyri strany.
- Preto všetky štvorce sú obdĺžniky.

Tento sylogizmus nezachráni pred nelogickosťou pravdivosť jeho predpokladov, dokonca ani pravdivosť jeho záveru. Oplatí sa odlišovať logické úsudky od

¹ pozri: 20 Čo je indícia?

² pozri: 178 Intuitívne vysvetlenie Bayesovej vety

³ pozri: 94 Filtrovaná indícia

nelogických, a odmietajú ich ospravedlňovanie dokonca aj v prípade, že je záver zhodou okolností pravdivý. Po prvé, toto rozlišovanie môže ovplyvniť, ako budeme prehodnocovať svoje názory vo svetle budúcich indícií. Po druhé, lajdáctvo je návykové.

Hlavne ide o to, že tento sylogizmus nepredkladá skutočné vysvetlenie. Možno všetky štvorce sú obdĺžniky, ale ak áno, nie je to *preto*, že jedny aj druhé majú štyri strany. Mohli by sme to označiť ako pokrytecký sylogizmus – taký, v ktorom je nesúlad medzi jeho uvedenými dôvodmi a skutočnými dôvodmi.

Ak naozaj chcete predložiť čestný, rozumný argument *za vášho kandidáta* v politickej kampani, existuje iba jeden spôsob, ako to urobiť:

- *Skôr než vás niekto najme*, zhromažďte si všetky dostupné indície o rôznych kandidátoch.
- Urobte si zoznam bodov, ktoré použijete pri rozhodovaní, ktorý kandidát vyzerá najlepšie.
- Spracujte tento zoznam.
- Chodte za vyhrávacím kandidátom.
- Ponúknite mu, že sa stanete menežerom jeho kampane.
- Keď sa vás opýtajú na kampaňovú literatúru, vytlačte svoj zoznam.

Iba takto dokážete ponúknuť *rozumnú* reťaz argumentov, ktorej spodný riadok⁴ bol napísaný smerom *po prúde* na základe riadkov napísaných nad ním. To, čo *naozaj* rozhoduje o vašom spodnom riadku, je jediná vec, ktorú môžete *poctivo* napísať do riadkov nad ním.

Pôvodný článok s komentármi⁵

⁴ pozri: 93 Spodný riadok

⁵ http://lesswrong.com/lw/jw/a_rational_argument/

Kapitola 97. Vyhýbanie sa naozaj slabým miestam vašich názorov

Pred pár rokmi zomrela moja prababička, vyše deväťdesiatročná, po dlhom, pomalom a krutom rozpade. Nikdy som ju nepoznal ako osobu, ale v mojom vzdialenom detstve varievala pre svoju rodinu; pamätám si jej faširované ryby, a jej tvár, a že bola ku mne milá. Na jej pohrebe rečnil môj prastrýc, ktorý sa o ňu staral počas jej posledných rokov: povedal, prehltajúc slzy, že Boh si jeho matku vzal naspäť po kúskoch: jej pamäť, jej reč, a nakoniec jej úsmev; a že keď si Boh nakoniec vzal jej úsmev, vedel, že už nebude dlho trvať kým zomrie, pretože to znamenalo, že už je takmer celá preč.

Počul som to a bol som v rozpakoch, pretože bola nemysliteľne krutá vec urobiť toto *komukoľvek*, a preto som nečakal, že to môj prastrýc pripíše Bohu. Zvyčajne, žid akosi skrátka-nemyslí-na logické dôsledky toho, že Boh dopustil tragédiu. Podľa židovskej teológie, Boh nepretržite udržiava celý vesmír a vyberá každú jednu udalosť v ňom; ale bežne je odvodzovanie logických dôsledkov z tohto presvedčenia vyhradené na šťastné udalosti. Keď povieť "to urobil Boh!" iba keď ste požehnaní bábätkom, ale skrátka-nemyslíte-na "to urobil Boh!" pri potratoch a úmrtiach v kolíske, dokážete si vybudovať dosť pokrivený¹ obraz o láskavej povahe vášho Boha.

Preto som bol prekvapený, keď som počul môjho prastrýca prisudzovať pomalý rozpad svojej matky vedomému, strategicky naplánovanému Božiemu činu. Porušovalo to pravidlá náboženského sebaklamu, ako som ich ja chápal.

Keby som si bol všimol svoj vlastný zmätok², mohol som urobiť úspešnú prekvapujúcu predpoveď. Onedlho na to sa môj prastrýc vzdal židovského náboženstva. (Ako jediný člen mojej širokej rodiny okrem mňa, pokiaľ viem.)

Moderný ortodoxný judaizmus³ sa nepodobá na žiadne iné náboženstvo, o ktorom som počul, a neviem ako ho opísať niekomu, kto nebol nútený študovať Mišnu a Gemaru. Je v ňom tradícia pochybovania, ale ten *druh* pochybovania... Nebolo by vôbec prekvapujúce počuť rabína pri týždennej kázni, ktorý by poukázal na konflikt medzi siedmimi dňami stvorenia a 13,7 miliardami rokov od Veľkého Tresku – pretože by si myslel, že pre to našiel naozaj chytré vysvetlenie, zahrňajúce tri ďalšie biblické odkazy, Midraš, a napoly pochopený článok zo *Scientific American*. V ortodoxnom judaizme máte povolené všimnúť si nezrovnalosti a protirečenia, ale iba za účelom ich odkecania, a kto príde s najkomplikovanejším vysvetlením, dostane odmenu.

Je tam tradícia skúmania. Ale útočíte na ciele iba so zámerom obrániť ich. Útočíte iba na tie ciele, o ktorých viete, že ich obránite.

V modernom ortodoxnom judaizme som nepočul veľké zdôrazňovanie cnosti slepej viery. Máte dovolené pochybovať. Akurát nemáte dovolené *úspešne* pochybovať⁴.

Predpokladám, že veľká väčšina vzdelaných ortodoxných židov v niektorom bode svojho života pochybovala o svojej viere. Ale to pochybovanie bolo

¹ pozri: 95 Racionalizovanie

² pozri: 27 Tvoja sila racionalistu

³ http://en.wikipedia.org/wiki/Modern_Orthodox_Judaism

⁴ pozri: 121 Správne použitie pochybnosti

pravdepodobne niečo ako toto: "Podľa skeptikov, Tóra hovorí, že vesmír bol stvorený za sedem dní, čo nie je vedecky presné. Ale boli by pôvodní príslušníci Izraela, zhromaždení pri hore Sinaj, schopní pochopiť vedeckú pravdu, aj keby im bola predložená? Mali vôbec slovo 'miliarda'? Je jednoduchšie vidieť sedemdnňový príbeh ako metaforu – najprv Boh stvoril svetlo, čo predstavuje Veľký Tresk..."

Je toto ten najslabší bod, v ktorom má človek napadnúť svoj vlastný judaizmus? Čítajte v Tóre o kúsok ďalej a nájdete, ako Boh zabíja deti, prvorodených synov v Egypte, aby presvedčil nevoleného faraóna, aby prepustil otrokov, ktorý logicky mohli byť z krajiny odteleportovaní. Ortodoxný žid túto pasáž celkom iste pozná, pretože by si mal raz za rok prečítať celú Tóru v synagóge, a s touto udalosťou sa spája veľký sviatok. Meno "Pascha" (Veľká Noc; po anglicky Passover) pochádza z toho, že Boh *obišiel* (po anglicky: pass over) židovské domácnosti, keď zabíjal každého prvorodeného syna v Egypte.

Moderní ortodoxní židia sú, prevažne, milí a civilizovaní ľudia; omnoho civilizovanejší než tých niekoľko redaktorov Starého Zákona. Ešte aj starí rabíni boli civilizovanejší. V Sederi je rituál, kde vezmete desať kvapiek vína zo svojho pohára, jednu kvapku za každú z desiatich pohrôm, aby ste zdôraznili utrpenie Egyptanov. (Samozrejme, že by ste mali súcitiť s utrpením Egyptanov, ale nie *natolko* súcitiť, aby by ste vstali a povedali: "Toto nie je správne! Je zlé urobiť niečo také!") Ukazuje to zaujímavý kontrast – rabíni boli podstatne láskavejší než redaktori Starého Zákona, keďže chápali krutosť pohrôm. Lenže Veda bola za oných dní slabšia, takže rabíni mohli uvažovať nad menej príjemnými stránkami Písma bez obáv, že by to celkom zlomilo ich vieru.

Keď sa ani *neopýtate*, či táto udalosť nevrhá na Boha zlé svetlo, nie ste nútení rýchlo vyhrnúť: "Cesty Božie sú tajomné!" alebo "Nie sme dosť múdri na to, aby sme spochybňovali Božie rozhodnutia!" alebo "Zabíjať bábätká je okej, keď to robí Boh!" Na túto časť otázky skrátka-nemyslite.

Dôvod, prečo vzdelaní ľudia zostanú veriacimi, je podľa mňa to, že keď pochybujú, podvedome sú veľmi opatrní, aby napádali svoje vlastné presvedčenie iba na jeho najsilnejších miestach – miestach, o ktorých vedia, že ich obránia. Navyše, na miestach, kde precvičenie⁵ si štandardnej obrany pôsobí posilňujúco.

Pravdepodobne pôsobí naozaj dobre precvičiť si, napríklad, svoju predpísanú obranu: "Nehovorí azda Veda, že vesmír sú atómy, ktoré sa iba nezmyselne od seba odrážajú?", pretože potvrdzujú zmysel vesmíru a ako pochádza od Boha, atď. Omnoho príjemnejšia téma na rozmyšľanie, než negramotná egyptská matka bedákajúca nad kolískou svojho zamordovaného syna. Každý, kto *spontánne* pomyslí na to druhé, zatiaľ čo pochybuje o svojej viere v judaizmus, ten o nej *naozaj* pochybuje, a pravdepodobne už dlho nezostane židom.

Mojou pointou teraz nie je mlátiť do ortodoxného judaizmu. Som si istý, že na zabíjanie prvorodených existuje taká či onaká odpoveď, pravdepodobne celý tucet. Mojou pointou je, že keď ide o spontánne sebaspochybňovanie, človek má sklon spontánne zaútočiť na svoje silné stránky, kde si môže precvičiť upokojujúcu odpoveď, než spontánne zaútočiť na tie najslabšie, najzraniteľnejšie body. Podobne, človek má sklon zastaviť sa pri prvej odpovedi a upokojiť sa, namiesto ďalšieho kritizovania odpovede. Lepší nadpis než "Vyhýbanie sa naozaj slabým miestam vašich názorov" by bol "Spontánne nemyslenie na najboľavejšie slabé miesta vašich názorov".

Viac než čokoľvek iné, vplyv náboženstva udržiavajú ľudia, ktorí skrátka-nemyslia-na skutočné slabé miesta svojho náboženstva. Nemyslím si, že je to vecou tréningu, ale vecou inštinktu. Ľudia nemyslia na skutočné slabé miesta

⁵ pozri: 92 Jeden argument proti armáde

svojich vlastných názorov z rovnakého dôvodu, ako sa nedotýkajú do červena rozpálenej pece; pretože to *bolí*.

Aby ste to robili lepšie⁶: Keď pochybujete o niektorom zo svojich najobľúbenejších názorov, zavrite oči, vyprázdňte si myseľ, zatnite zuby, a úmyselne myslite na to, čo najviac bolí. Neprecvičujte si štandardné námietky, ktorých štandardné obrany vám prinášajú dobrý pocit. Opýtajte sa sami seba, čo by *múdri* ľudia, ktorí nesúhlasia, povedali na vašu prvú odpoveď, aj na vašu druhú odpoveď. Kedykoľvek sa pristihnete, že cítate pred námietkou, na ktorú ste letmo pomysleli, vytiahnite ju do popredia svojej mysle. Udríte sami seba na solar plexus. Pichnite si nôž do srdca a pokývajte ním, aby ste rozšírili dieru. Zoči-voči bolesti si precvičujte iba toto:

Čo je pravda, to už je pravda.

Priznať si to, to nerobí o nič horším.

Nebyť voči tomu otvorený nespôsobí, že to odíde.

A pretože je to pravda, je to to, s čím treba interagovať.

Čokoľvek nepravdivé nie je tu na prežívanie.

Ľudia dokážu zvládnuť to, čo je pravda,

pretože to už zvládajú.

—Eugene Gendlin

Pôvodný článok s komentármi⁷

⁶ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

⁷ http://lesswrong.com/lw/jy/avoiding_your_beliefs_real_weak_points/

Kapitola 98. Motivované zastavenie a motivované pokračovanie

Súvisí s kapitolami: Tretia možnosť¹, Meditácia o zvedavosti²

Aj keď nesúhlasím s niektorými názormi skupiny Rýchlo a šetrne³ – myslím si, že sa snažia premeniť na limonádu *príliš* veľa citrónov – zároveň sa mi zdá, že majú sklon vyvinúť si *psychologicky najrealistickejšie* modely zo všetkých škôl rozhodovania. Väčšina pokusov predkladá pokusným osobám možnosti, pokusná osoba si vyberie možnosť, a to je výsledok pokusu. Zástancovia šetrnosti si uvedomujú, že v skutočnom živote si musíte *vytvárať* vlastné možnosti, a študovali, ako pokusné osoby robia *toto*.

Podobne mnohé pokusy predkladajú indície na striebornom podnose, ale v skutočnom živote musíte indície zbierať, čo môže byť drahé, a v niektorom bude sa musíte rozhodnúť, že máte dosť indícií, aby ste sa zastavili a vybrali. Keď si idete kúpiť dom, nemáte na výber presne 10 domov a nikto vás nevedie na prehliadku všetkých z nich skôr než sa môžete rozhodnúť. Pozriete sa na jeden dom, na druhý, porovnáte ich navzájom; prispôsobíte svoje očakávania – prehodnotíte nakoľko naozaj chcete byť blízko svojho pracoviska a koľko ste naozaj ochotní zaplatiť; rozhodnete sa, ktorý dom si pozriete potom, a v niektorom bode sa rozhodnete, že už ste videli domov dosť a vyberiete si.

Gilovichovo rozlišovanie medzi *motivovaným skepticizmom* a *motivovanou dôverčivosťou* zdôrazňuje, ako závery, ktorým daná osoba nechce veriť, posudzuje podľa prísnejšieho štandardu ako závery, ktorým chce veriť. Motivovaný skeptik sa pýta, či ho indície *núti* prijať daný záver; motivovaný dôverčivec sa pýta, či mu indície *dovoľujú* prijať daný záver.

Tvrdím, že analogické skreslenie v psychologicky realistickom hľadaní je *motivované zastavenie* a *motivované pokračovanie*: keď máme skrytý motív vybrať si momentálne “najlepšiu” možnosť, máme skrytý motív zastaviť sa a vybrať si, a odmietnuť zvažovanie prípadných ďalších možností. Ak máme skrytý motív odmietnuť momentálne najlepšiu možnosť, máme skrytý motív počkať s rozhodnutím až do ďalších indícií, aby sme vytvorili viac možností – aby sme našli niečo iné, hocičo, čo by sme mohli urobiť *namiesto* daného záveru.

Veľkým historickým škandálom v štatistike bol R. A. Fisher, významný zakladateľ odvetvia, ktorý trval na tom, že sa nepodarilo dokázať *kauzálne* spojenie medzi fajčením a rakovinou pľúc. “Korelácia nie je kauzalita,” vypovedal pred kongresom. Možno fajčiari majú gén, ktorý zároveň spôsobuje, že fajčia, aj že dostanú rakovinu pľúc.

Alebo možno fakt, že Fisher bol zamestnaný ako konzultant pre tabakové firmy, mu dal skrytý motív rozhodnúť sa, že doteraz zhromaždené indície sú nedostatočné na vyvodenie záveru, a že je lepšie hľadať ďalej. Fisher bol sám fajčiar a zomrel na rakovinu hrubého čreva v roku 1962.

(Poznámka ad hominem: Fisher bol frekventista. Bayesovci⁴ sú rozumnejší pri usudzovaní o pravdepodobnej kauzalite.)

¹ pozri: 111 Tretia možnosť

² pozri: 123 Meditácia o zvedavosti

³ <http://fastandfrugal.com/>

⁴ <http://bayes.cs.ucla.edu/BOOK-2K/>

Tak ako mnohé iné formy motivovaného skepticizmu, motivované pokračovanie sa môže maskovať za cnostnú rozumnosť. Kto už môže argumentovať proti zhromažďovaniu ďalších indícií? Ja. Indície sú často drahé, a ešte horšie, pomalé, a určite nie je nič cnostné na odmietaní spracovať tie indície, ktoré už máte. Vždy si to môžete neskôr rozmyslieť.⁵ (Zdanlivý rozpor sa rieši takto: Stráviť *jednu hodinu* diskutovaním problému, keď máte myseľ starostlivo zbavenú riešení, je niečo iné ako čakať na ďalšiu štúdiu za 20 miliónov dolárov.)

Motivované zastavenie sa vyskytuje všade, kde sa bojíme tretej možnosti⁶, a kde máme argument, ktorého zrejmy protiargument⁷ by sme radšej nevideli, a na ďalších miestach. Vyskytuje sa, keď hľadáte plán činnosti, ktorý vám dá dobrý pocit zo samotnej činnosti⁸ a tak by ste radšej neriešili, ako dobre váš plán *naozaj* funguje, zo strachu, že zničíte hrejivú žiaru morálneho uspokojenia⁹, za ktorú ste zaplatili kopec peňazí. Vyskytuje sa, keď sa vaše názory a očakávania rozsynchronizujú¹⁰, takže máte dôvod báť sa zhromažďovania nových indícií.

Ponaučenie je, že samotné rozhodnutie ukončiť proces hľadania (dočasne alebo natrvalo) je rovnako ako samotný proces hľadania predmetom skreslení a skrytých motívov. Mali by ste mať podozrenie na motivované zastavenie, keď ukončíte hľadanie po tom, čo nájdete pohodlný záver, a predsa je veľa rýchlych a lacných indícií, ktoré ste ešte nezhrmáždili - webové stránky, ktoré ste mohli navštíviť, protiargumenty na protiargumenty, ktoré ste mohli zvážiť, alebo ste nezavreli oči na päť minút podľa hodín, aby ste skúsili vymyslieť lepšiu možnosť. Mali by ste mať podozrenie na motivované pokračovanie, keď nejaká indícia ukazuje smerom, ktorý sa vám nepáči, ale rozhodnete sa, že potrebujete viac indícií - *drahých* indícií, o ktorých viete, že ich tak rýchlo nezhrmáždíte, v porovnaní s tým, čo si nájdete na Googli za 30 minút - skôr než urobíte niečo nepohodlné.

Pôvodný článok s komentármi¹¹

⁵ pozri: 88 Odložte navrhovanie riešení

⁶ pozri: 111 Tretia možnosť

⁷ pozri: 123 Meditácia o zvedavosti

⁸ http://lesswrong.com/lw/kb/cant_say_no_spending/

⁹ pozri: 283 Necitlivosť k rozsahu

¹⁰ pozri: 13 Viera vo vieru

¹¹ http://lesswrong.com/lw/km/motivated_stopping_and_motivated_continuation/

Kapitola 99. Falošné zdôvodnenie

Mnohí kresťania, ktorí prestali naozaj veriť¹, teraz tvrdia, že si vážia Bibliu ako zdroj etických rád. Štandardnú ateistickú odpoveď dáva Sam Harris: “Ty aj ja obaja vieme, že by nám trvalo päť minút vyrobiť knihu, ktorá ponúka súvislejšiu a súcitnejšiu morálku než Biblia.” Podobne, človek môže skúsiť tvrdiť, že si vážia Bibliu ako literárne dielo. Prečo si potom nevážia *Pána Prsteňov*, čo je omnoho kvalitnejšie literárne dielo? A napriek štandardnej kritike Tolkienovej morálky je *Pán Prsteňov* oproti Biblii lepší prinajlepšom ako zdroj etiky. Prečo teda ľudia nenosia okolo krku prstienky namiesto krížikov? Dokonca aj *Harry Potter* je lepší ako Biblia, aj ako literárne dielo, aj ako morálna filozofia. Keby som chcel byť naozaj krutý, porovnal by som Bibliu so sériou *Kushiel* od Jacqueline Carey.

“Ako môžeš zdôvodniť, že si kupuješ laptop vykladaný drahokamami za 1 milión dolárov²,” opýtate sa priateľa, “keď mnoho ľudí nemá vôbec žiaden laptop?” A váš priateľ povie: “Ale pomysli na pracovné miesta, ktoré to poskytne – výrobcovi laptopov, reklamnej agentúre výrobcu laptopov – a oni si potom kúpia jedlo a účesy – to bude stimulovať ekonomiku a nakoniec mnohí ľudia budú mať svoje vlastné laptopy.” Bolo by však *efektívnejšie* kúpiť 5000 laptopov OLPC, čím by ste poskytli zamestnanie výrobcom OLPC a zároveň by ste priamo rozdelili laptopy.

Už som sa dotkol témy zlyhania v hľadaní tretej možnosti³. Ale toto nie je ozajstné motivované zastavenie⁴. Nazvať to “motivované zastavenie” by naznačovalo, že sa vôbec niekedy niečo hľadalo.

V kapitole Spodný riadok⁵ som ukázal, že iba skutočné príčiny našich rozhodnutí dokážu ovplyvniť našu efektivitu pri dosahovaní svojich cieľov. Nieкто, kto si kupuje laptop za milión dolárov, si v skutočnosti myslel: “Aha, ligoce sa” a to bola jediná kauzálna história jeho rozhodnutia kúpiť si laptop. Žiadne množstvo “zdôvodnenia” to nezmení, pokiaľ to zdôvodnenie nie je skutočný proces nového vyhľadávania, ktoré by mohlo zmeniť záver. *Naozaj zmeniť záver*. Väčšina kritiky, ktorú robíme z pocitu povinnosti⁶, je skôr symbolickou kontrolou než niečím iným. Ako slobodné voľby v krajine s jedinou politickou stranou.

Aby ste naozaj zdôvodnili úctyhodnosť Biblie jej literárnou kvalitou, museli by ste najprv nejako neutrálne prečítať kandidujúce knihy, dokiaľ nenájdete knihu s najvyššou literárnou kvalitou. Používajúc preslávenosť ako jedno možné rozumné kritérium na generovanie kandidátov navrhujem, aby ste si legitímne prečítali Shakespeara, Bibliu, a *Gödel, Escher, Bach*. (Inak by bola príliš veľká náhoda, aby sa Biblia vyskytla ako kandidát pri výbere z miliónov kníh.) Skutočne zložitá je tá časť “neutrálne prečítať”. Pomerne ľahké, ak nie ste kresťan, lenže ak ste...

Lenže nič takéto sa samozrejme nestalo. Nikto nič nehľadal. Napísať zdôvodnenie “literárna kvalita” nad spodný riadok⁷ “milujem Bibliu” je historické prekrútenie skutočnosti, ako sa tam ten spodný riadok⁸ naozaj dostal, je to ako predávať mačacie mlieko ako kravské mlieko. Toto nie je to, odkiaľ ten spodný riadok⁹ naozaj prišiel. Toto nie je to, čo sa pôvodne stalo a vytvorilo tento záver.

¹ pozri: 13 Viera vo vieru

² <http://entertainment.slashdot.org/story/07/03/26/197253/a-million-dollar-laptop-created>

³ pozri: 111 Tretia možnosť

⁴ pozri: 98 Motivované zastavenie a motivované pokračovanie

⁵ pozri: 93 Spodný riadok

⁶ pozri: 123 Meditácia o zvedavosti

⁷ pozri: 93 Spodný riadok

⁸ pozri: 93 Spodný riadok

⁹ pozri: 93 Spodný riadok

Ak naozaj podrobíte svoj záver kritike, ktorá ho môže potenciálne od-záverovať – ak tá kritika *naozaj* má túto moc – to potom mení “skutočný algoritmus” vášho záveru. To mení previazanosť vášho záveru v možných svetoch. Ale ľudia veľmi preceňujú *skutočnú* pravdepodobnosť, že zmenia svoj názor¹⁰.

Pri toľkých otvorených myšliach by si človek myslel, že tu bude viac aktualizovania názorov.

Skúsím hádať: Áno, pripúšťate, že ste sa pôvodne rozhodli kúpiť si milióndolárový laptop na základe myšlienky: “Aha, ligoce sa.” Áno, pripúšťate, že toto nie je rozhodovací proces v súlade s vašimi vyjadrenými cieľmi. Lenže odvtedy ste sa rozhodli, že by ste v skutočnosti mali minúť svoje peniaze takým spôsobom, ktorý poskytne laptopy čo najväčšiemu množstvu chudákov bez laptopov. A predsa ste *nedokázali* nájsť efektívnejší spôsob, ako to dosiahnuť, než kúpiť si milióndolárový laptop vykladaný diamantmi – pretože tak predsa dáte peniaze obchodu s laptopmi a stimulujete tak ekonomiku! To sa nedá prekonať!

Drahý priateľ, čertovsky podozrievam túto úžasnú zhodu okolností. Čertovsky podozrievam to, že najlepšia odpoveď podľa tohto milého, rozumného, altruistického kritéria X je zároveň tá istá myšlienka, aká vám pôvodne napadla na základe nesúvisiaceho, neobhájiteľného procesu Y. Ak si nemyslíte, že ste správnu otázku pravdepodobne získali hodom kockou, ako pravdepodobne by ste ju získali na základe iného nerozumného myslenia?

Je nepravdepodobné, že používate pomýlené uvažovanie, a predsa nerobíte žiadne chyby.

Pôvodný článok s komentármi¹¹

¹⁰ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

¹¹ http://lesswrong.com/lw/kq/fake_justification/

Kapitola 100. Je toto vaše skutočné odmietnutie?

Z času na čas sa stane, že niekto natrafí na niektorý z mojich názorov ohľadom transhumanizmu – na rozdiel od mojich názorov týkajúcich sa ľudskej rozumnosti – čudnú, exoticky znejúcu myšlienku ako superinteligencia a Priateľská UI. A dotyčný ich odmietne.

Keď je potom vyzvaný, aby toto odmietnutie vysvetlil, neraz povie:

“Prečo by som mal veriť niečomu z toho, čo Yudkowsky hovorí? Ved’ nemá PhD!”

A občas niekto iný, kto to počuje, povie: “Ach, mal by si si urobiť PhD, aby ťa ľudia počúvali.” Alebo túto radu poskytne priamo ten, kto mi neverí, slovami: “Vráť sa až budeš mať PhD.”

Nuž, existujú dobré i zlé dôvody urobiť si PhD, ale toto je jeden z tých zlých.

Existuje veľa dôvodov, prečo niekto *naozaj* nepriateľsky reaguje na transhumanistické myšlienky. Väčšinou je to otázka rozoznávania vzorov, nie slovného uvažovania: myšlienka sa podobá¹ na vzory “divná myšlienka” alebo “vedecká fantastika” alebo “seкта ohlasujúca koniec sveta” alebo “prehnané mladícke nadšenie”.

Preto je okamžite, rýchlosťou vnemu, táto myšlienka odmietnutá. Ak sa potom niekto opýta: “Prečo nie?”, spustí sa tým hľadanie zdôvodnenia. Ale toto hľadanie nemusí nevyhnutne natrafiť na skutočný dôvod – pričom slovami “skutočný dôvod” nemyslím *najlepší* dôvod, ktorý sa dá poskytnúť, ale skôr tú príčinu, ktorá ako historický fakt rozhodla² v prvom momente, keď nastalo odmietnutie³.

Namiesto toho hľadanie zdôvodnenia natrafí na fakt znejúci ako zdôvodnenie: “Tento rečník nemá PhD.”

Lenže ja nemám PhD ani vtedy, keď rozprávam o ľudskej rozumnosti, prečo teda aj tam neprichádza tá istá námietka⁴?

A ešte viac k veci, *keby* som mal PhD, ľudia by to nepovažovali za rozhodujúci faktor naznačujúci, že by mali veriť všetkému, čo poviem. Namiesto toho by nastalo rovnaké počiatočné odmietnutie, z rovnakého dôvodu; a dodatočné hľadanie zdôvodnenia by sa potom zastavilo v odlišnom bode.

Povedali by: “Prečo by sme *tebe* mali veriť? Ty si len nejaký chlap s PhD! Takých je na svete kopa. Vráť sa, až sa presláviš vo svojej oblasti a získaš doživotnú zmluvu na významnej univerzite.”

Ale veria ľudia *naozaj* náhodným profesorom z Harvardu, keď hovoria čudné veci? Samozrejme nie. (Keby som však bol profesorom na Harvarde, bolo by *naozaj* jednoduchšie získať *pozornosť médií*. Reportéri, ktorí by mali od začiatku sklon neveriť mi – ktorí by pravdepodobne mali rovnaký sklon neveriť náhodnému človeku s PhD – by aj tak o mne napísali článok, pretože aj to by bola novinka, že profesor z Harvardu verí takú divnú vec.)

Ak hovoríte veci, ktoré začiatočníkovi znejú *nesprávne*, na rozdiel od iba vychrlenia magicky znejúcich technoblábolov o splietaní leptických kvarkov v N

¹ pozri: 33 Veda ako uniforma

² pozri: 93 Spodný riadok

³ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁴ pozri: 77 Sektárske antisektárstvo

+2 rozmeroch; a ak je poslucháč cudzinec, ktorý osobne nepozná *ani* vás *ani* hlavnú tému vášho odboru; potom si myslím, že bod, v ktorom náhodný človek *naozaj* začne prikladať dôveryhodnosť prekračujúcu jeho pôvodný dojem, iba *kvôli* akademickým dokladom, je niekde na úrovni nositeľa Nobelovej ceny. Ak vôbec. Približne povedané, potrebujete takú úroveň akademických dokladov, ktorá vás posúva "mimo dosah smrteľníkov".

Pokiaľ viem, približne toto sa stalo Ericovi Drexlerovi. Predložil svoju predstavu nanotechnológie a ľudia povedali: "Kde sú technické detaily?" a "Vráť sa, až budeš mať PhD!" A Eric Drexler strávil šesť rokov spisovaním technických detailov a dostal za to PhD pod vedením Marvina Minskeho. A *Nanosystémy* je skvelá kniha. Ale zmenili tí ľudia, ktorí hovorili: "Vráť sa, až budeš mať PhD", naozaj svoj názor na molekulárnu nanotechnológiu? Pokiaľ viem, nie.

Podobne bolo všeobecným pravidlom v Inštitúte Singularity [pozn. prekl: dnešný Inštitút Výskumu Strojovej Inteligencie], že čokoľvek sa od nás žiadalo, aby sme boli dôveryhodnejší, keď sme to naozaj urobili, takmer nič sa nezmenilo. "Vyvíjate nejaký kód? Nemám záujem podporovať organizáciu, ktorá nevyvíja kód" -> OpenCog -> nič sa nezmenilo. "Eliezer Yudkowsky nemá akademické tituly" -> ako riaditeľ výskumu bol dosadený profesor Ben Goertzel -> nič sa nezmenilo. Jediná vec, ktorá naozaj vyzerá, že zvýšila dôveryhodnosť, sú známi ľudia spájaní s organizáciou, napríklad náš sponzor Peter Thiel, alebo člen rady Ray Kurzweil.

Toto môže byť dôležitá vec, ktorú sa oplatí zapamätať začínajúcim podnikateľom a čerstvým konzultantom – čo vám tí, ktorých ste si nezískali ako zákazníkov, *povedia* ako dôvod odmietnutia, nemusí *v skutočnosti* robiť žiaden rozdiel; a mali by ste sa nad tým starostlivo zamyslieť skôr než vynaložíte obrovské úsilie. Ak vám rizikový kapitalista povie: "Keby len váš predaj rástol o čosi rýchlejšie!", alebo ak vám potenciálny zákazník povie: "Vyzerá to dobre, ale nemá to vlastnosť X," nemusí to byť ich *skutočné* odmietnutie. Napraviť túto vec môže niečo zmeniť, ale nemusí.

A bolo by dobré pamätať na to počas hádok. Robin a ja máme spoločný názor, že dvaja racionalisti by sa nemali zhodnúť na tom, že sa nezhodnú⁵; nemali by mať spoločné poznanie epistenického nesúhlasu, pokiaľ niečo nie je veľmi zle.

Predpokladám, že ak sa vo všeobecnosti dvaja racionalisti rozhodnú vyriešiť nezhodu, ktorá pretrvala po prvej výmene, mali by očakávať, že skutočné dôvody nesúhlasu sa buď ťažko komunikujú, alebo sa ťažko odhaľujú. Napríklad:

- Nezvyčajné, ale dobre podložené vedecké poznanie alebo matematika;
- Dlhé inferenčné vzdialenosti⁶;
- Intuícia ťažko vyjadriteľná slovami, možno vychádzajúca z konkrétnych vizualizácií;
- Duch doby získaný zo svojej profesie (ktorá na to môže mať dobrý dôvod);
- Vnemovo rozoznané vzory na základe skúsenosti;
- Číre myšlienkové zvyky;
- Emocionálna pripútanosť veriť konkrétnemu výsledku;
- Strach z odhalenia minulých chýb;

⁵ http://www.overcomingbias.com/2006/12/agreeing_to_agr.html

⁶ pozri: 9 Očakávame krátke inferenčné vzdialenosti

- Sebaklam za účelom pýchy alebo iných osobných výhod.

Keby išlo o záležitosť, v ktorej sa *všetky* skutočné odmietnutia dajú *ľahko* vyložiť na stôl, nehoda by sa pravdepodobne vyriešila tak priamočiaro, že by nezostala po prvom stretnutí.

“Je toto moje skutočné odmietnutie?” je otázka, ktorú by obaja nesúhlasiaci mali určite položiť *sami sebe*, aby to uľahčili tomu druhému. Avšak snahy priamo, verejne psychoanalyzovať toho druhého podľa môjho pozorovania zvyknú konverzáciu *veľmi* rýchlo zvrhnúť.

Aj tak by však malo by povolené, aby sa nesúhlasiaci pokorne opýtal: “Je toto tvoje skutočné odmietnutie?”, ak existuje nejaký produktívny spôsob, ako sa venovať tejto podtému. Možno by pravidlo malo znieť tak, že sa môžete otvorene spýtať: “Je tento jednoduchý priamočiaro znejúci dôvod tvoje *skutočné* odmietnutie, alebo je to skôr intuícia X, alebo profesionálny duch doby Y?” Zatiaľ čo tie zahanbujúcejšie možnosti uvedené naspodku tabuľky sú ponechané na svedomie toho druhého, a je jeho zodpovednosť, ako sa s tým vyrovná.

Post scriptum:

Tento článok nie je *naozaj* o PhD vo všeobecnosti, ani o jeho konkrétnom vplyve na dôveryhodnosť. Ale vždy som si myslel, že do tej miery, do akej je toto strategicky dôležitá vec, by dávalo viac zmyslu naverbovať akademika s už existujúcim vysokým postavením, než vynaložiť hromadu času v snahe dosiahnuť nízke alebo stredné akademické postavenie.

Ak by však nejaký profesor tam vonku chcel, aby som prišiel a *iba* si urobil PhD v analytickej filozofii – *iba* napísal dizertačku a obhájil ju – v tom prípade som si, na svoje vlastné použitie, vypracoval všeobecnú a matematicky elegantnú teóriu na rozhodovacie problémy ako je Newcombov problém⁷. Myslím si, že by z toho bola pekná dizertačná práca, a je to pripravené na napísanie – ak niekto má tú moc nechať ma robiť veci klasickým spôsobom.

Pôvodný článok s komentármi⁸

⁷ pozri: 291 Newcombov problém a ľutovanie rozumnosti

⁸ http://lesswrong.com/lw/wj/is_that_your_true_rejection/

Kapitola 101. Previazané pravdy, nákazlivé lži

Jeden z našich veľmi raných filozofov došiel k záveru, že plne kompetentná myseľ by zo štúdia jedného faktu alebo predmetu patriaceho do nejakého vesmíru dokázala zostaviť alebo si predstaviť tento vesmír, od okamihu jeho stvorenia až po jeho úplný koniec...

—First Lensman

Ak sa niekto z vás sústreďí na jediný fakt alebo malý predmet, ako napríklad kamienok alebo semienko rastliny alebo iného tvora, na celkom krátke obdobie sto vašich rokov, začnete vnímať jeho pravdu.

—Gray Lensman

Som si pomerne istý, že jediný kamienok vzatý z pláže na našej vlastnej Zemi by neurčoval svetadiely a krajiny, politiku a ľudí na tento Zemi. Iné planéty v priestore a čase, iné Everettové vetvy¹, by vytvorili rovnaký kamienok. Na druhej strane, totožnosť jedného kamienka by zrejme zahŕňala naše fyzikálne zákony. V tom zmysle by celý náš vesmír – všetky Everettové vetvy – vyplývali z tohto kamienka. (Ak, ako sa zdá pravdepodobné, neexistujú žiadne naozaj voľné premenné.)

Jediný kamienok teda pravdepodobne neurčuje našu celú Zem. Ale z jedného kamienka vyplýva veľmi veľa. Zo štúdia toho jedného kamienka by ste mohli uvidieť zákony fyziky a všetko, čo vyplýva z nich. Rozmýšľajúc o týchto zákonoch fyziky by ste mohli uvidieť, ako sa tvoria planéty, a mohli by ste uhádnuť, že kamienok pochádza z niektorej takejto planéty. Vnútorne kryštály a molekulárne útvary kamienka vytvorené pod vplyvom gravitácie by vám povedali niečo o hmotnosti planéty; zmes prvkov v kamienku by vám povedala niečo o zložení planéty.

Nie som geológ, takže neviem, aké tajomstvá sú geológom k dispozícii. Ale viem si veľmi ľahko predstaviť, že ukážem geológovi kamienok a poviem: “Tento kamienok je z pláže na Half Moon Bay”, na čo mi geológ ihneď povie: “Som zmätený²” alebo priam “Klameš”. Možno je to nesprávny druh kameňa, alebo ten kamienok nie je dostatočne zodraný na to, aby mohol byť z pláže – ja nepoznám kamienky dosť dobre na to, aby som uhádol súvislosti a podpisy, podľa ktorých by ma mohli prichytiť, a práve to je pointa.

“Iba Boh dokáže povedať naozaj dôveryhodnú lož.” Zaujímalo by ma, či niekedy existovalo náboženstvo, ktoré by si vytvorilo takéto príslovie. Tipol by som si (falzifikovateľne), že nie: je to racionalistický postoj, aj keď ho vyjadríte ako teologickú metaforu. Povedať: “všetko súvisí so všetkým, pretože Boh stvoril celý svet a udržiava ho” môže vyvolať pekné hrejivé pocity počas kázne, ale nedostane vás to veľmi ďaleko, keď dôjde na priradovanie kamienkov k plážam.

Minca na Zemi vytvorí gravitačné zrýchlenie na Mesiaci okolo $4,5 \cdot 10^{-31} \text{ m/s}^2$, takže v istom zmysle nie je celkom nesprávne povedať, že každá udalosť je previazaná s celým svojím svetelným kužeľom minulosti. A keďže odvodenia sa môžu prenášať kauzálnymi sieťami dopredu aj dozadu, *epistemické* previazania môžu ľahko prekročiť hranice svetelných kužeľov. Ale nechcem by som byť

¹ http://lesswrong.com/lw/r8/and_the_winner_is_manyworlds/

² pozri: 27 Tvoja sila racionalistu

forenzným astronómom, ktorý sa musí pozrieť na Mesiaca a zistiť, či na tej minci padla hlava alebo znak – ten vzťah je menší než kvantová neistota a tepelný šum.

Ak povie: “Všetko je previazané s niečím iným” alebo “Všetko je inferenčne previazané a niektoré previazania sú omnoho silnejšie než iné,” môžete byť naozaj múdry a nie iba Hlboko Múdry³.

Fyzicky je každá udalosť v istom zmysle súčtom celého svojho svetelného kužela minulosti, bez hraníc. Ale zoznam *pozorovateľných* previazaní je omnoho kratší a dáva vám niečo ako sieť. Táto pravidelnosť na vysokej úrovni⁴ je to, o čom hovorím, keď poviem Veľká Sieť Kauzality.

Používam tieto Veľké Písmená tak trochu ironicky; ale ak si niečo zaslúži Veľké Písmená, jednoznačne je to Veľká Sieť Kauzality.

“Ach, akú zamotanú sieť to tkáme, keď sa prvýkrát klamať pokúšame,” povedal Sir Walter Scott. Nie všetky lži sa nám vymykajú spod kontroly – nežijeme v takom spravodlivom vesmíre⁵. Ale občas sa stane, že niekto klame o nejakom fakte, a potom musí klamať o previazanom fakte, a potom o ďalšom fakte, ktorý je previazaný s týmto:

“Kde si bol?”

“Ehm, bol som na služobnej ceste.”

“O čom bola tá služobná cesta?”

“To ti nemôžem povedať; to bolo dôverné jednanie s významným zákazníkom.”

“Och – teba na také pozývajú? Dobrá správa! Mal by som zavolať tvojemu šéfovi a poďakovať mu za to, že ťa tam vzal.”

“Prepáč – teraz práve nie je vo svojej kancelári...”

Ľuďom, ktorí nie sú bohovia, sa často nepodariť *predstaviť* si všetky fakty, ktoré by museli prekrútiť, aby povedali naozaj uveriteľnú lož. “Boh spôsobil, že som tehotná⁶” znelo o trochu dôveryhodnejšie za starých čias, predtým než náš model sveta začal obsahovať (pojem) chromozóm Y. Mnohé podobné lži môžu dnes prasknúť, keď sa genetické testovanie stáva čoraz častejším. Ľudia boli odsúdení za znásilnenie, a falošné obvinenia boli odhalené, o niekoľko rokov neskôr, na základe indícií, o ktorých si neuvedomovali, že ich nechávajú. Študent evolučnej biológie môže vidieť rukopis prirodzeného výberu⁷ na každom vlkovi, ktorý naháňa zajaca; a na každom zajacovi, ktorý uteká preč; a na každej včele, ktorá uštipne namiesto toho, aby slušne upozornila – ale samotným kreacionistom *ich vlastné* lži znejú dôveryhodne, to som si istý.

Nie všetky lži sú odhalené, nie všetci klamári sú potrestaní; nežijeme v takom spravodlivom vesmíre. Ale nie všetky lži sú také bezpečné, ako klamári veria. Ktovie, koľko hriechov by odhalila bayesovská superinteligencia, keby urobila (nedeštruktívny?) nanotechnologický sken Zeme? Prínajmenšom všetky tie lži, o ktorých stále existujú indície v nejakom mozgu. Niektoré také lži sa môžu stať známymi ešte skôr, ak sa neurovedcom niekedy podarí zostaviť naozaj dobrý detektor lži pomocou neurozobrazovania. Paul Ekman (priekopník v štúdiu drobných pohybov tvárových svalov) by pravdepodobne teraz dokázal prečítať veľký zlomok svetových lží, keby dostal príležitosť.

³ pozri: 86 Ako vyzerá (a byť) hlboký

⁴ pozri: 197 Redukcionizmus

⁵ pozri: 305 Mimo dosahu boha

⁶ http://lesswrong.com/lw/m8/the_amazing_virgin_pregnancy/

⁷ pozri: 129 Mimozemský boh

Nie všetky lži sú odhalené, nie všetci klamári sú potrestaní. Ale Veľká Sieť zvykne byť často podceňovaná. Naštudovanie len tých vedomostí, ktoré ľudia už *zhromaždili*, by trvalo mnoho ľudských životov⁸. Každý, kto si myslí, že nielen Boh dokáže povedať dokonalú lož, bez rizika, podceňuje zamotanosť Veľkej Sieť.

Je úprimnosť najlepšou taktikou? Neviem, či chcem zájsť až tak ďaleko: Aj v mojej etike je niekedy okej držať hubu. Ale v porovnaní s priamymi lžami, aj úprimnosť aj mlčanie zahŕňa menšie vystavenie sa rekurzívne propagovaniu rizík, o ktorých neviete, že ich podstupujete.

Pôvodný článok s komentármi⁹

⁸ http://lesswrong.com/lw/kj/no_one_knows_what_science_doesnt_know/

⁹ http://lesswrong.com/lw/uw/entangled_truths_contagious_lies/

Kapitola 102. O klamstvách a výbuchoch čiernych labutí

Sudca Marcus Einfeld, 70-ročný, kráľovský advokát od roku 1977, austrálsky živý poklad 1997, cena mieru OSN 2002, zakladajúci predseda Austrálskej komisie pre ľudské práva a rovnosť príležitostí, ktorý pred niekoľkými rokmi odišiel na dôchodok, ale pravidelne je privolávaný súdiť dôležité prípady...

...pôjde do väzenia aspoň na dva roky za sériu krivých prisah a klamstiev, ktoré začali 36-librovou pokutou za prekročenie povolenej rýchlosti o 10 km/h¹.

Celá tá *podozrivo cnostne znejúca* teória² o tom, že poctiví ľudia nevedia dobre klamať, a že niekde zostanú previazané stopy, a že celá vec môže vybuchnúť v epickom zlyhaní typu čierna labuť, naozaj *má* niekoľko príkladov v skutočnom živote, hoci tu samozrejme pracuje selektívne spravodajstvo, keď o tomto počujeme.

Pôvodný článok s komentármi³

¹ http://news.bbc.co.uk/2/hi/uk_news/magazine/7967982.stm

² pozri: 101 Previazané pravdy, nákazlivé lži

³ http://lesswrong.com/lw/9a/of_lies_and_black_swan_blowups/

Kapitola 103. Epistemológia temnej strany

Ak raz zaklamete, pravda bude už naveky vašim nepriateľom.

Už som hovoril o tom¹, že pravdy sú previazané a lži sú nákazlivé. Ak zdvihnete kamienok z cesty a povieť geológovi, že ste ho našli na pláži – nuž, viete vy, čo taký geológ vie o kameňoch? Ja nie. Ale predstavujem si, že kamienok ošúchavaný vodou nebude vyzeráť ako kvapka stuhnutej lávy z výbuchu sopky. Viete, odkiaľ ten kamienok na vašej ceste naozaj pochádza? Vo vesmíre, ktorý sa riadi zákonmi, veci nesú stopy svojho miesta; v tejto sieti lož nezapadá. [Doplnené: Geológ mi v komentároch povedal, že väčšina kamienkov na cestách pochádza z pláží, takže by nevedel určiť rozdiel medzi kamienkom z cesty a kamienkom z pláže, ale vedel by povedať rozdiel medzi kamienkom z hory a kamienkom z cesty/pláže. Ilustračný príklad...]

Čo jednej mysli znie ako nesúvisiaca pravda – ktorá by sa ľahko dala nahradiť uveriteľnou lžou – môže byť určené tuctom spojení pohľadom väčšieho poznania. Kreacionistovi môže myšlienka, že život utváral “inteligentný dizajn” namiesto “prirodzeného výberu²” znieť ako názov športového tímu, ktorému fandí. Pre biológa, hodnoverne argumentovať, že nejaký organizmus bol inteligentne navrhnutý, by vyžadovalo klamať o takmer každej stránke tohto organizmu. Aby ste hodnoverne argumentovali, že “ľudia” boli inteligentne nadizajnovaní, museli by ste klamať o dizajne ľudskej siete, o architektúre ľudského mozgu, o bielkovinách spojených slabými van der Waalsovými silami namiesto silných kovalentných väzieb...

Alebo by ste jednoducho klamali o evolučnej teórii, čo je cesta, ktorú si vyberie väčšina kreacionistov. Namiesto klamanie o spojených uzloch v sieti, klamú o všeobecných zákonoch riadiacich tieto spojenia.

A aby potom *toto* celé zakryli, klamú o pravidlách vedy – napríklad čo znamená nazývať niečo “teória”, alebo čo to znamená, keď vedec povie, že si niečím nie je absolútne istý.

Takže prechádzajú od klamanie o konkrétnych faktoch, cez klamanie o všeobecných pravidlách, ku klamaniu o pravidlách rozmyšľania. Aby ste klamali o tom, či sa ľudia vyvinuli, musíte klamať o evolúcii; a potom musíte klamať o pravidlách vedy, ktoré určujú naše chápanie evolúcie.

Ale ako inak? Rovnako ako by človek nezapadal do kolektívu *naozaj* inteligentne navrhnutých životných foriem, a museli by ste klamať o evolúcii, aby to vyzeralo inak; takisto samotné názory o kreacionizme nezapadajú do vedy – nenašli by ste ich v dobre usporiadanej mysli, rovnako ako by ste nenašli palmy rásť na ľadovci. A preto musíte prelomiť bariéry, ktoré tomu bránia.

Čo nás vedie k prípade sebaklamu.

Jedna lož, ktorú povieť *sami sebe*, môže znieť celkom prijateľne, keď nevíete nič o pravidlách, ktorými sa riadia myšlienky, prípadne ani len, že také pravidlá *existujú*; a táto voľba³ vyzerá svojvoľne, ako keď si vyberáte chuť zmrzliny, izolovane ako kamienok na pobreží...

¹ pozri: 101 Previazané pravdy, nákazlivé lži

² pozri: 129 Mimozemský boh

³ http://lesswrong.com/lw/rb/possibility_and_couldness/

...ale potom niekto spochybní váš názor, pomocou pravidiel rozmýšľania, ktoré sa on naučil. Povie: "Kde sú tvoje indície?"

A vy povie: "Čože? Prečo by som potreboval indície?"

On povie: "Vo všeobecnosti si názory vyžadujú indície."

Tento argument je jasný vojak bojujúci za opačnú stranu⁴, ktorého musíte poraziť. Preto povie: "Nesúhlasím! Nie všetky názory vyžadujú indície. Konkrétne, názory o drakoch si nevyžadujú indície. Pokiaľ ide o drakov, môžeš veriť hocičomu, čo len chceš. Takže ja nepotrebujem indície, aby som veril, že v mojej garáži je drak⁵."

A on povie: "Čo? Nemôžeš len tak vynechať drakov. Pravidlo, že názory vyžadujú indície má svoj dôvod. Aby si nakreslil správnu mapu⁶ mesta, musíš ísť cez ulice⁷ a robiť na papieri čiary zodpovedajúce tomu, čo vidíš. To nie je svojvoľné nariadenie – ak sedíš vo svojej obývačke a kreslíš na papieri čiary náhodne, mapa bude nesprávna. S extrémne vysokou pravdepodobnosťou⁸. To platí rovnako o mape draka ako hocičoho iného."

Takže teraz už aj *toto* vysvetlenie, *prečo* názory vyžadujú indície, je *d'alší* nepriateľský vojak. Preto povie: "Nesprávne s extrémne vysokou pravdepodobnosťou? Takže je tam stále šanca, však?⁹ Nemusím ti veriť, ak to nie je absolútne isté¹⁰."

Alebo možno začnete mať sami podozrenie, že "názory si vyžadujú indície". Lenže to ohrozuje lož, ktorá je vám drahá; preto odmietnete svitanie vo svojom vnútri a zatlačíte slnko späť pod horizont.

Alebo ste už predtým počuli príslovie "názory si vyžadujú indície" a znelo vám to dosť múdro, a verejne ste to schvaľovali. Ale nikdy vám celkom nedoplo, dokiaľ vás na to niekto iný neupozornil, že by sa toto príslovie mohlo *vzťahovať* aj na váš názor, že vo vašej garáži je drak. Tak sa rýchlo zamyslíte a povie: "Ten drak je oddelené magistérium¹¹."

Mať nepravdivé názory nie je dobré, ale nemusí to byť trvalo poškodzujúce – ak svoju chybu po odhalení opustíte. Nebezpečná vec je mať nepravdivý názor, o ktorom *veríte, že musí byť chránený ako viera* – viera vo vieru¹², či už je sprevádzaná skutočnou vierou alebo nie.

Jediná Lož, Ktorú Treba Chrániť dokáže zablokovať niekoho postup do pokročilej rozumnosti. Nie, nie je to neškodná zábava.

Tak ako je samotný svet omnoho zamotanejší¹³ než sa na povrchu zdá; rovnako sú aj prísnejšie pravidlá rozmýšľania, obmedzujúce názory silnejšie, než by netrénovaný človek predpokladal. Svet je pevne prepletený, riadený všeobecnými zákonmi, a takisto aj *rozumné* názory.

Predstavte si, čo by to vyžadovalo poprieť evolúciu alebo heliocentrizmus – všetky tie previazané pravdy a riadiace zákony, ktoré by ste nemali dovolené poznať. Potom si dokážete predstaviť, ako jediný akt sebaklamu dokáže

⁴ pozri: 47 Politika zabíja myslenie

⁵ pozri: 13 Viera vo vieru

⁶ pozri: 46 Jednoduchá pravda

⁷ pozri: 185 Druhý zákon termodynamiky a stroje na poznanie

⁸ pozri: 186 Viera vo večný pohyb

⁹ pozri: 112 Ale stále je tu šanca, nie?

¹⁰ pozri: 114 Absolútna autorita

¹¹ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

¹² pozri: 13 Viera vo vieru

¹³ pozri: 101 Previazané pravdy, nákazlivé lži

zablokovať celú meta-úroveň hľadania pravdy, keď sa vaša myseľ začne cítiť ohrozená videním súvislostí. Zakáže všetky stredne pokročilé a vyššie úrovne racionalistovho Umenia. Namiesto toho vytvorí rozľahlú sústavu antizákonov, pravidiel antimyslenia, všeobecné zdôvodnenia prečo veriť nepravdivému.

Steven Kaas povedal¹⁴: "Šíriť menej než maximálne presné názory je akt sabotáže. Nerobte to nikomu, pokiaľ by ste mu zároveň nechceli prezeráť pneumatiky." Dať niekomu *ochraňovať* nepravdivý názor – presvedčiť ho, že *samotný tento názor* treba chrániť pred každou myšlienkou, ktorá vyzerá, že ho ohrozuje – nuž, toto by ste nemali robiť nikomu, pokiaľ by ste mu zároveň nechceli urobiť frontálnu lobotómiu.

Keď raz zaklamete, pravda je vašim nepriateľom, aj každá pravda spojená s touto pravdou, a každý spojenec pravdy vo všeobecnosti; proti všetkému z tohto musíte bojovať, aby ste tú lož ochránili. Či už klamete druhým alebo sám sebe.

Musíte popierať, že názory si vyžadujú indície, a potom musíte popierať, že by mapy mali odrážať územie, a potom musíte popierať, že pravda je dobrá vec...

Takto vzniká Temná Strana.

Obávam sa, že ľudia si to neuvedomujú, alebo že nie sú dostatočne opatrní – že ako chodíme po našom ľudskom svete, môžeme očakávať, že sa stretneme so *systematicky* zlou epistemológiou.

Mémy "ako myslieť" sa vznášajú okolo nás, uložené myšlienky¹⁵ Hlbokej Múdrosti¹⁶ – niektoré z nich sú dobré rady vymyslené racionalistami. Ale iné predstavy boli vymyslené, aby chránili lož alebo sebaklam: výtvory Temnej Strany.

"Každý má právo na svoj vlastný názor." Keď sa nad tým zamyslíte, kde vzniklo toto príslovie? Je to niečo, čo by niekto povedal v procese chránenia pravdy, alebo v procese chránenia *pred* pravdou? Ale ľudia nevyskočia a nepovedia: "Aha! Cítim tu prítomnosť Temnej Strany!" Pokiaľ viem, väčšina si ani neuvedomuje, že Temná Strana existuje.

Ale ako inak? Či klamete druhých alebo iba seba, Lož, Ktorú Treba Chrániť sa bude rekurzívne šíriť po sieti empirickej kauzality, po sieti všeobecných empirických pravidiel, a pravidiel samotného rozmýšľania, a porozumení týchto pravidiel. Ak na svete existuje *dobrá* epistemológia, a zároveň aj lži alebo sebaklamy, ktoré sa ľudia snažia chrániť, potom vznikne aj zlá epistemológia, aby odporovala tej dobrej. Sotva môžeme v tomto svete očakávať, že nájdeme Svetlú Stranu bez Temnej Strany; existuje Slnko, aj to, čo sa od neho odťahuje a vytvára si ochranný Tieň.

Pripomína, že to nemusia nevyhnutne byť *zlí* ľudia¹⁷. Prevažná väčšina z tých, ktorí opakujú Hlbokú Múdrost sú viac oklamaní než neúprimní, viac sebaklamúci než klamúci. Aspoň si to myslím.

A určite nie je mojím zámerom poskytnúť vám Úplne Všeobecný Protiargument¹⁸, aby ste hocikedy, keď vám niekto ponúkne epistemológiu, ktorá sa vám nepáči, mohli povedať: "Ach, toto vymyslel niekto z Temnej Strany." Je jedno z pravidiel Svetlej Strany, že návrh musíte odmietnuť kvôli nemu samotnému, nie obvinením jeho autora zo zlých úmyslov¹⁹.

¹⁴ <http://www.acceleratingfuture.com/steven/?p=124>

¹⁵ pozri: 81 Uložené myšlienky

¹⁶ pozri: 86 Ako vyzerá (a byť) hlboký

¹⁷ pozri: 51 Sú vaši nepriatelia od narodenia zlí?

¹⁸ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

¹⁹ pozri: 89 Klam pôvodu

Lenže Temná strana existuje. Strach je tá cesta, ktorá k nej vedie, a jedna zrada vás môže obrátiť. Nie všetci, ktorí nosia rúcha, sú buď Jediovia alebo podvodníci; existujú aj Sithovia, majstri a nevedomí učni. Dávajte si pozor, buďte opatrní.

Pokiaľ ide o vymenovanie častých mémov, ktoré vytvorila Temná Strana – pripomínam, nie náhodné nepravdivé názory, ale zlá epistemológia, Všeobecné Obrany Zlyhania – nuž, chcelo by sa vám do toho pustiť, milí čitatelia?

Pôvodný článok s komentármi²⁰

²⁰ http://lesswrong.com/lw/uy/dark_side_epistemology/

Časť X. Proti doublethinku

Obsah

104 Singlethink	143
105 Doublethink (dobrovoľné skreslenie)	145
106 Nie, naozaj, ja som sa oklamal	148
107 Viera v sebaklam	149
108 Moorov paradox	152
109 Neverte, že sa sami oklamete	154

Kapitola 104. Singlethink

Pamätám si ten presný okamih, keď som začal svoje putovanie ako racionalista.

Nebolo to počas čítania *To nemyslíte vážne, pán Feynman*, ani žiadneho existujúceho diela o rozumnosti; tie som jednoducho prijal ako samozrejmé. Putovanie začne, keď uvidíte veľkú chybu vo svojom existujúcom umení a objavíte túžbu zlepšiť, vytvoriť *nové* zručnosti presahujúce tie užitočné ale nedostačujúce, ktoré ste našli v knihách.

V posledných okamihoch svojho prvého života som mal pätnásť rokov a opakoval som si príjemnú spomienku na vlastnú spravodlivosť z čias, keď som bol oveľa mladší. Moje spomienky na také dávne časy sú hmlisté; mám v myšlienkach predstavu, ale nepamätám si, koľko rokov som presne mal. Myslím, že to bolo šesť alebo sedem, a že pôvodná udalosť sa stala počas letného tábora.

Pôvodne sa stalo to, že táborový vedúci, teenager, povedal nám, omnoho mladším chlapcom, aby sme sa zoradili do radu, a navrhol nasledujúcu hru: chlapec na konci radu sa bude plaziť pomedzi naše nohy, a my ho budeme plieskať po zadku, keď pôjde okolo, a potom bude na rade ďalší osemročný chlapec na konci radu. (Možno je to tým, že som stratil svoju detskú nevinnosť, ale nemôžem sa ubrániť úvahám...) Odmietol som hrať túto hru, tak som bol poslaný sedieť do kúta.

Táto spomienka – odmietania biť a byť bitý – mi začala symbolizovať, že dokonca v takomto veľmi ranom veku som odmietal mať radosť z ubližovania druhým. Že som nebol ochotný kúpiť si úder po cudzom zadku za cenu úderu po mojom vlastnom; nebol som ochotný platiť bolesťou za príležitosť spôsobiť bolesť. Odmietol som hrať hry so záporným súčtom.

A potom, vo veku pätnásť rokov, som si náhle uvedomil, že to nie je pravda. *Neodmietol* som kvôli principiálnemu postoju proti hrám s nulovým súčtom. O väzenskej dileme som sa dočítal pomerne skoro vo svojom živote, ale nie ako sedemročný. Odmietol som jednoducho preto, lebo som nechcel, aby mi ubližovali, a stať v rohu bola prijateľná cena za to, že mi nebudú ubližovať.

Čo je dôležitejšie, uvedomil si, že som toto *vždy* vedel – že skutočná spomienka *vždy* číhala v kútiku mojej mysle, moje myšlienkové oko sa na ňu na zlomok sekundy zahľadelo a potom sa odvrátilo preč.

V mojom prvom kroku po Ceste som *prichytil tento pocit* – zovšeobecnený nad subjektívnu skúsenosť – a povedal som: “Takže *takýto* je to pocit, keď zatlačím neželanú pravdu do kúta svojej mysle! Odteraz si budem všímať vždy, keď to urobím, a vyčistím *všetky* svoje kúty!”

Toto cvičenie som si nazval *singlethink*, podľa Orwellovho doublethinku. V doublethinku¹ zabudnete, a potom zabudnete, že ste zabudli. V singlethinku si všimnete, že zabúdate, a potom si spomeniete. V hlave naraz držíte iba jednu vnútorne neprotirečivú myšlienku.

“Singlethink” bola prvá *nová* racionalistická zručnosť, ktorú som si vytvoril, o ktorej som predtým nečítal v knihách. Pochybujem, že je pôvodná v zmysle akademickej pôvodnosti, ale našťastie to nie je potrebné.

Och, moje pätnásťročné ja rado pomenovávalo veci.

Desivé hĺbky sklonu potvrdzovať idú ďalej a ďalej. Nie navždy, pretože váš mozog má konečnú zložitosť, ale dosť dlho na to, aby vám to pripadalo ako večnosť. Stále

¹ pozri: 105 Doublethink (dobrovoľné skreslenie)

objavujete nové mechanizmy (alebo o nich čítate), ktorými váš mozog odstrkuje veci z cesty.

Ale moje mladé ja pomocou tejto prvej metly vymietlo pekných pár kútov.

Pôvodný článok s komentármi²

² <http://lesswrong.com/lw/k0/singlethink/>

Kapitola 105. Doublethink (dobrovoľné skreslenie)

Medzi O'Brienovými prstami sa zjavil podlhovastý ústrižok novín. Asi päť sekúnd bol vo Winstonovom zornom poli. Bola to fotografia, a nebolo pochyb o jej identite. Bola to tá fotografia. Bola to ďalšia kópia fotografie Jonesa, Aaronsona a Rutherforda na zasadnutí strany v New Yorku, ktorú pred jedenástimi rokmi náhodne zazrel a okamžite zničil. Iba krátku chvíľku bola pred jeho očami, potom bola opäť v nedohľadne. Ale videl ju, nepochybne ju videl! Urobil zúfalý bolestivý pokus oslobodiť svoju hornú polovicu tela. Nedalo sa pohnúť ani o centimeter žiadnym smerom. Na chvíľu dokonca zabudol aj na stupnicu. Všetko, čo chcel, bolo držať túto fotografiu ešte raz vo svojich prstoch, alebo ju aspoň vidieť.

“Existuje!” vykrikol.

“Nie,” povedal O'Brien.

Prešiel krížom cez miestnosť.

V protihľej stene bola pamäťová diera. O'Brien nadvihol mriežku. Útly prúžok papiera sa nevidený vznášal preč na prúde teplého vzduchu; zanikal v záblesku plameňa. O'Brien sa odvrátil od steny.

“Popol,” povedal. “Neidenifikovateľný popol. Prach. Neexistuje. Nikdy neexistovala.”

“Ale existovala! Stále existuje! Existuje v pamäti. Pamätám si ju. Ty si ju pamätáš.”

“Ja si ju nepamätám,” povedal O'Brien.

Winstonovo srdce pokleslo. Toto bol doublethink. Cítil sa na smrť bezmocný. Aj keby si bol istý, že O'Brien klame, pravdepodobne by na tom nezáležalo. Bolo však dokonale možné, že O'Brien na fotografiu naozaj zabudol. A ak áno, potom už zabudol aj na svoje popieranie, že si ju pamätá, a zabudol aj na samotný akt zabudnutia. Ako si človek mohol byť istý, že je to jednoducho trik? Možno sa myseľ naozaj mohla takto šialene vyklbiť; to bola myšlienka, ktorá ho premohla.

—George Orwell¹, 1984

Čo ak nám sebaklam pomáha byť šťastnými? Čo ak sa rozbehneme a prekonáme skreslenia a urobí nás to – ach! – *nešťastnými*? Iste by pravou múdrosťou bola rozumnosť *druhého rádu*; vyberanie si, kedy byť rozumným. Potom by ste sa mohli rozhodnúť, ktoré kognitívne skreslenia vás budú ovládať, aby ste maximalizovali svoje šťastie.

Ponechajúc bokom morálku, pochybujem, že by sa myseľ naozaj mohla takto šialene vyklbiť.

Z rozumnosti druhého rádu by vyplývalo, že si v istom bode pomyslíte: “A teraz budem nerozumne veriť, že vyhrám v lotérii, aby som sa urobil šťastným.” My však nemáme takúto priamu kontrolu nad svojimi názormi. Nedokážete sa primäť

¹ pozri: 56 Ľudské zlo a zahmlené myslenie

veriť, že obloha je zelená, aktom vôle. Mohli by ste uveriť, že tomu veríte² – hoci som vám to práve skomplikoval tým, že som poukázal na rozdiel. (Nie je za čo!) Mohli by ste dokonca *uveriť*, že ste šťastný a oklamáný; ale neboli by ste *naozaj* ani šťastný ani oklamáný.

Aby rozumnosť druhého rádu bola skutočne *rozumná*, najprv by ste potrebovali dobrý model skutočnosti, aby ste vedeli odvodiť dôsledky rozumnosti a nerozumnosti. Keby ste sa potom rozhodli byť v prvom ráde nerozumný, potrebovali by ste zabudnúť na tento presný pohľad. A potom zabudnúť na akt zabudnutia. Nechcem sa dopustiť logickej chyby zovšeobecňovania fiktívnej indicie, ale myslím si, že Orwell celkom dobre extrapoloval, kam táto cesta vedie.

Nemôžete poznať dôsledky skreslenia, dokiaľ ste sa nezbavili skreslení. A potom už je príliš neskoro na sebaklam.

Druhou možnosťou je rozhodnúť sa zostať slepo skreslený, bez akejkoľvek jasnej predstavy dôsledkov. To nie je rozumnosť druhého rádu. To je zámerná hlúposť.

Buďte nerozumne optimistickí ohľadom svojich vodičských schopností, a budete veselo bezstarostní tam, kde sa ostatní budú triasť od strachu. Nebudete sa ani zaťažovať nepohodlným bezpečnostným pásom. Budete veselo bezstarostní celý deň, týždeň, rok. Potom TRESK, a zvyšok života strávite želaním si, aby ste si mohli poškrabať svrbiacu fantómovú končatinu. Alebo paralyzovaní od krku nižšie. Alebo mŕtvi. Nie je to nevyhnutné, ale je to možné. Aké je to pravdepodobné? Takýto obchod nemôžete urobiť rozumne, dokiaľ nepoznáte svoje *skutočné* vodičské schopnosti, aby ste vedeli zistiť, akému veľkému nebezpečenstvu sa vystavujete. Takýto obchod nemôžete urobiť rozumne, dokiaľ nevíete o skresleniach ako je zanedbanie pravdepodobnosti³.

Nezáleží na tom, koľko dní strávite v blaženej nevedomosti, jediná chyba stačí na zničenie ľudského života, na vyváženie každého haliera, ktorý ste zodvihli z kolajníc hlúposti.

Jedna z hlavných rád, ktoré dávam ašpirujúcim racionalistom, je: “Nepokúšaj sa byť chytrý.” A tiež: “Počúvaj tie tiché, otravujúce pochybnosti.” Ak nevíete, potom nevíete, čo nevíete; nevíete, *koľko* toho nevíete; a nevíete, ako veľmi *potrebujete* vedieť.

Neexistuje rozumnosť druhého rádu. Existuje iba skok naslepo, ktorý môže alebo nemusí viesť do jamy s blčiacou lávou. Keď toto *víete*, už je neskoro na slepotu.

Ale ľudia to ignorujú, pretože nevedia, čo nevedia. Neznáme neznáme veci im nie sú dostupné⁴. Nesústredia sa na prázdne miesto na mape, ale berú ho akoby zodpovedalo prázdnomu územiu. Keď sa rozhodnú naslepo skočiť, skontrolujú si, čo majú v pamäti nejaké nebezpečenstvo, ale na prázdnej mape nenájdu žiadnu jamu s blčiacou lávou. Prečo neskočiť?

Bol som tam. Skúšal som to. Popálil som sa. Nesnažte sa byť chytrý.

Jednej kamarátke som raz povedal, že mám podozrenie, že šťastie z hlúposti je veľmi preceňované. Ona vážne pokrútila hlavou a povedala: “Nie, nie je; naozaj nie je.”

Možno niekde existujú hlúpi šťastní ľudia. Možno sú šťastnejší než vy. Ale život nie je fér, a vy sa nestanete šťastnejšími, keď budete žiarliť na to, čo nemáte. Mám podozrenie, že prevažná väčšina čitateľov *Overcoming Bias* by nedokázala dosiahnuť “šťastie z hlúposti”, keby to skúšali. Táto cesta je vám uzavretá.

² pozri: 13 Viera vo vieru

³ http://en.wikipedia.org/wiki/Neglect_of_probability

⁴ pozri: 5 Dostupnosť

Nikdy nedosiahnete tento stupeň nevedomosti, nemôžete zabudnúť to, čo viete, nemôžete nevidieť to, čo ste videli.

Šťastná hlúposť je pre vás uzavretá. Nikdy ju nedosiahnete, okrem skutočného poškodenia mozgu, a možno ani vtedy. Mohli by ste sa azda zamýšľať, či je šťastie z hlúposti *optimálne* – či to nie je najväčšie šťastie, o aké sa človek môže snažiť – ale na tom nezáleží. Pre vás je táto cesta uzavretá, ak vôbec niekedy otvorená bola.

Všetko, čo vám teraz zostáva, je snažiť sa o takú šťastie, aké môže dosiahnuť racionalista. Myslím si, že sa v konečnom dôsledku môže ukázať ako väčšie. Sú tam ohraničené cesty a cesty s otvoreným koncom; rovinky, na ktorých možno leňošiť, a hory, na ktoré treba vyliezť; a ak to lezenie zaberá veľa úsilia, o to vyššia je tá hora.

Navyše, k životu patrí viac než len šťastie; a od vašich rozhodnutí môže závisieť šťastie nielen vás samotných.

Ale to všetko je akademická otázka. V čase, keď si uvedomíte, že ste mali na výber, už na výber nemáte. Nedokážete nevidieť to, čo ste raz videli. Tá druhá cesta je uzavretá.

Pôvodný článok s komentármi⁵

⁵ http://lesswrong.com/lw/je/doublethink_choosing_to_be_biased/

Kapitola 106. Nie, naozaj, ja som sa oklamal

Súvisí s kapitolou: Viera vo vieru¹

Nedávno som sa rozprával s osobou, ktorá... ťažko to opísať. Formálne bola ortodoxná židovka. Bola aj vysoko inteligentná, vedomá si niektorých archeologických indícií proti jej náboženstvu, a štandardných plytkých argumentov proti náboženstvu, o ktorých veriaci vedia. Napríklad vedela, že Mordecai, Ester, Haman a Vašti neboli v perzských historických záznamoch, ale že existovala zodpovedajúca stará perzská legenda o babylonských bohoch Marduk a Ištar, a súperiacich elamitských bohoch Humman a Vašti. *Vedela* to, a predsa oslavovala Purim. Jeden z tých vysoko inteligentných veriacich ľudí, ktorí sa celé roky dusia vo svojich rozporoch, rozoberajú a prekrúcajú, dokiaľ ich myseľ nevyzerá ako vnútro kresby M. C. Eschera.

Väčšina takýchto ľudí bude predstierať, že sú príliš múdri² na to, aby sa rozprávali s ateistami, ale ona bola ochotná so mnou pár hodín debatovať.

Vďaka tomu teraz rozumiem prinajmenšom jednu ďalšiu vec o sebaklame, ktorej som predtým explicitne nerozumel – konkrétne, že nie je potrebné, aby ste *naozaj* oklamali sami seba, dokiaľ *veríte*, že ste sami seba oklamali. Môžete to nazvať “viera v sebaklam”.

Keď bola táto žena na strednej škole, myslela si, že je ateistka. Ale vtedy sa rozhodla, že sa bude správať, akoby verila v Boha. A potom – povedala mi vážne – postupom času začala naozaj veriť v Boha.

Pokiaľ môžem povedať, v tomto bode sa úplne mýli. Počas nášho rozhovoru znovu a znovu hovorila: “*Verím* v Boha”, ale ani jediný raz nepovedala: “*Boh existuje*.” Keď som sa jej pýtal, prečo je veriaca, ani raz nehovorila o dôsledkoch existencie Boha, iba o dôsledkoch viery v Boha. Nikdy: “Boh mi pomôže,” ale vždy: “moja viera v Boha mi pomáha.” Keď som povedal: “Keby niekomu záležalo iba na pravde, a pozrel by sa na náš vesmír, nebol by ani vymyslel Boha ako hypotézu,” jednoznačne so mnou súhlasila.

Nepresvedčila *naozaj* samu seba, aby verila, že Boh existuje alebo že židovské náboženstvo je pravdivé. Ani zďaleka nie, pokiaľ môžem povedať.

Na druhej strane si myslím, že *naozaj* verí, že sama seba oklamala.

Takže aj keď nemá žiaden úžitok z viery v Boha – pretože neverí – úprimne *verí*, že oklamala sama seba na vieru v Boha, a preto úprimne *očakáva*, že bude mať úžitok, ktorý so spája s oklamaním seba samej na vieru v Boha; a *toto* by asi malo tvoriť zhruba rovnaké placebo ako *naozajstná* viera v Boha.

A to by mohlo vysvetliť, prečo mala motiváciu zapálené brániť tvrdenie, že *verí* v Boha, pred mojím skeptickým spochybňovaním, aj keď nikdy nepovedala: “Ach, mimochodom, Boh naozaj existuje”, ani nevyzerala, že by sa čo len trochu o takýto výrok zaujíkala.

Pôvodný článok s komentármi³

¹ pozri: 13 Viera vo vieru

² pozri: 18 Predstieranie múdrosti

³ http://lesswrong.com/lw/r/no_really_ive_deceived_myself/

Kapitola 107. Viera v sebaklam

Súvisí s kapitolou: Epistemológia temnej strany¹

Včera som písal o svojom rozhovore s formálnou ortodoxnou židovkou, ktorá energicky obhajovala tvrdenie, že verí v Boha, pričom vyzerala, že v Boha v skutočnosti vôbec neverí.

Kým som sa jej pýtal, aké výhody podľa nej má viera v Boha, povedal som jej Tarskeho litániu² – čo je vlastne nekonečná množina litánií, z ktorých konkrétny príklad je:

Ak je obloha modrá,

Chcem veriť: "Obloha je modrá."

Ak obloha nie je modrá,

Chcem veriť: "Obloha nie je modrá."

"Toto nie je moja filozofia," povedala mi.

"Nemyslel som si, že je," odpovedal som jej. "Len sa pýtam – za predpokladu, že by Boh *neexistoval*, a že by sa o tom vedelo, či by ste aj tak mali veriť v Boha?"

Zaváhala. Zdalo sa, že sa naozaj pokúša rozmýšľať o tom, čo ma prekvapilo.

"Takže je to hypotetická otázka..." povedala pomaly.

Vtedy som si myslel, že má problém dovoliť si predstaviť si svet, v ktorom Boh *neexistuje*, pretože je pripútaná k svetu, v ktorom Boh *existuje*.

Teraz si skôr myslím, že mala problém predstaviť si kontrast medzi tým, ako by vyzeral svet, keby Boh *existoval* a keby *neexistoval*, pretože všetky jej myšlienky boli o jej *viere v Boha*, ale jej kauzálna sieť modelovala svet, ktorý Boha ako uzol neobsahoval. Dokázala by teda ľahko odpovedať na: "Ako by sa tento svet odlišoval, keby som neverila v Boha?" ale nie: "Ako by sa tento svet odlišoval, keby Boh *neexistoval*?"

Vtedy mi na túto otázku neodpovedala. Ale našla *protipríklad* k Tarskeho litánii:

Povedala: "Verím, že ľudia sú lepší, než naozaj sú."

Pokúsil som sa jej vysvetliť, že keď poviete: "Ľudia sú zlí," znamená to, že veríte, že ľudia sú zlí, a keď poviete: "Verím, že ľudia sú dobrí," znamená to, že veríte že veríte, že ľudia sú dobrí. Takže povedať: "Ľudia sú zlí a ja verím, že ľudia sú dobrí," znamená, že veríte, že ľudia sú zlí, ale veríte že veríte, že ľudia sú dobrí.

Zacitoval som jej:

Keby existovalo sloveso s významom "veriť nepravdivo", nemalo by v prítomnom čase prvú osobu jednotného čísla.

—Ludwig Wittgenstein

Ona povedala, s úsmevom: "Áno, ja verím, že ľudia sú lepší než naozaj sú. Len som ti to chcela takto vysvetliť."

¹ pozri: 103 Epistemológia temnej strany

² pozri: 123 Meditácia o zvedavosti

“Myslím si, že by sa Babička mala na teba pozrieť, Walter,” povedala Nanny. “Myslím si, že tvoja myseľ je pomotaná ako kľbko nití, ktoré spadlo.”

—Terry Pratchett, Maškaráda

A ja dokážem napísať slová: “Nuž, myslím si, že neverila tomu, že by jej myslenie malo byť reflexívne konzistentné³,” ale stále mám problém sa s tým vyrovnáť⁴.

Dokážem vidieť vzorce v slovách, ktoré vychádzajú z jej úst, ale nedokážem na úrovni empatie pochopiť myseľ, ktorá je za nimi. Viem sa vžiť do kože mimozemšťanov, ktorí jedia deti⁵ a Tretej Lady Kiritsugu⁶, ale nedokážem si predstaviť, aké to musí byť na jej mieste. Alebo iba *nechcem*?

Toto je dôvod, prečo inteligentní ľudia majú iba určité množstvo času (meraného v subjektívnom čase rozmyšľania o náboženstve), aby sa stali ateistami. Po istom bode, ak ste chytrý, ak ste strávili ten čas myslením na vaše náboženstvo a jeho obhajobu, a ak ste stále neunikli z pazúrov Epistemológie temnej strany⁷, vaša myseľ znútra skončí ako Escherov obrázok.

(Jeden z ďalších pár momentov, keď sa zamyslela – spomínam to, keby ste mali príležitosť to použiť – bol, keď hovorila o tom, aké dobré je veriť, že sa niekto zaujíma o to, či robíte dobré alebo zlé veci – samozrejme *nie* o tom, že naozaj *existuje* Boh, ktorý sa zaujíma o to, či robíte dobré alebo zlé veci, takýto výrok nebol súčasťou jej náboženstva...

A ja som povedal: “Ale mne záleží na tom, či robíte dobro alebo zlo. Takže vlastne hovoríte, že toto nestačí, a že potrebujete zároveň veriť v niečo *nad* ľudstvom, čomu záleží na tom, či robíte dobro alebo zlo.” To ju na chvíľu zastavilo, pretože samozrejme týmto spôsobom o tom predtým nerozmýšľala. Iba štandardné použitie neštandardnej sady nástrojov⁸.)

Neskôr som sa jej v istom bode opýtal, či by bolo dobré robiť *čokoľvek* odlišne, keby Boh jednoznačne neexistoval, a tentokrát odpovedala: “Nie.”

“Takže,” povedal som neveriaco, “či Boh existuje alebo neexistuje, to nemá absolútne žiaden účinok na to, ako by ľudia mali rozmyšľať alebo konať? Myslím si, že aj rabín by sa na toto pozeral podozrievavo.”

Zdá sa, že jej náboženstvo sa teraz skladá *výlučne* z uctievania uctievania. Kým skutoční veriaci starých čias asi verili, že ich zachráni všetko vidiaci otec, ona dnes verí, že ju zachráni viera v Boha.

Potom, čo povedala: “Verím, že ľudia sú lepší než naozaj sú,” som sa opýtal: “Takže ste pravidelne prekvapená, keď ľudia sklamú vaše očakávania?” Nasledovalo dlhé ticho, a potom, pomaly: “No... som *prekvapená*, keď ľudia... sklamú moje očakávania?”

Vtedy som tejto pauze nerozumel. Vtedy som tým chcel naznačiť, že ak je pravidelne sklamaná skutočnosťou, potom je toto nevýhoda viery v nepravdivú vec. Ale zdá sa, že ju zaskočili dôsledky toho, že *nie je* zaskočená.

Teraz si uvedomujem, že celá podstata jej filozofie bola *jej viera*, že *sama seba oklamala*, a možnosť, že jej odhady ľudí boli *v skutočnosti presné*, ohrozila

³ pozri: 10 Objektív, ktorý vidí svoje chyby

⁴ pozri: 196 Mysli ako skutočnosť

⁵ http://lesswrong.com/lw/y5/the_babyeating_alien_18/

⁶ http://lesswrong.com/lw/y7/the_super_happy_people_38/

⁷ pozri: 103 Epistemológia temnej strany

⁸ pozri: 86 Ako vyzerá (a byť) hlboký

Epistemológiu Temnej Strany⁹, ktorú si postavila okolo názorov ako: "Mám úžitok z toho, že verím, že ľudia sú lepší než naozaj sú."

Zosadila z trónu starú modlu a nahradila ju vysloveným uctievaním Epistemológie Temnej Strany, ktorá pôvodne vznikla, aby túto modlu chránila; uctievala svoj vlastný pokus o sebaklam. Pokus zlyhal, ale ona si to úprimne neuvedomovala.

A tak musia symbolickí strážcovia príčetnosti ľudstva (motto: "kazíme vaše pomätené zábavky už od Epikurových čias") dnes bojovať proti aktívnemu uctievaniu sebaklamu – uctievaniu *domnelých výhod viery*, namiesto Boha.

Toto v skutočnosti vysvetľuje fakt o *mne*, ktorému som predtým nerozumel – dôvod, prečo ma vytáča, keď ľudia hovoria o tom, aký je sebaklam *ľahký*, a prečo píšem celé články na blogu¹⁰ o tom, že urobiť vedomé rozhodnutie veriť, že obloha je zelená, je omnoho ťažšie, než si ľudia asi myslia.

Je to preto, lebo – hoci sa *nemôžete* len tak rozhodnúť veriť, že obloha je zelená – ak si tento fakt *neuvedomíte*, v skutočnosti sa *môžete* oklamať, že veríte, že ste sa úspešne oklamali.

A keďže potom úprimne *očakávate*, že obdržíte výhody, o ktorých si myslíte, že pochádzajú zo sebaklamu, dostanete rovnaký druh placebo výhod, aké by ste boli naozaj dostali z úspešného sebaklamu.

Keď teda chodím a vysvetľujem, aký *ťažký* je sebaklam, v skutočnosti priamo mierim na tie placebo výhody, ktoré majú ľudia z viery, že sa úspešne oklamali, a útočím na nový druh náboženstva, ktoré uctieva iba uctievanie Boha.

Rozmýšľam, či táto bitka nevytvorí nový zoznam dôvodov, prečo nie viera, ale *samotná* viera vo vieru¹¹, je dobrá vec? Prečo ľudia majú veľký úžitok z uctievania svojho uctievania? Budeme to musieť robiť znovu s vierou vo vieru vo vieru a s uctievaním uctievania uctievania? Alebo sa inteligentní veriaci nakoniec skrátka vzdajú tejto línie argumentovania?

Rád by som veril, že nikto nemôže naozaj veriť, že verí, že verí, že verí, lenže filozofické argumenty o svete zombií sú už omnoho zamotanejšie než toto¹² a ich zástancovia ich stále neopustili.

Očakávam zapálenú obhajobu viery vo vieru v komentároch, ale rozmýšľam, či sa niekomu bude chcieť predbehnúť dobu a začať obhajovať vieru vo vieru vo vieru? Poďme teda na to, nech je to za nami.

Pôvodný článok s komentármi¹³

⁹ pozri: 103 Epistemológia temnej strany

¹⁰ pozri: 105 Doublethink (dobrovolné skreslenie)

¹¹ pozri: 13 Viera vo vieru

¹² pozri: 220 Zombie? Zombie!

¹³ http://lesswrong.com/lw/s/belief_in_selfdeception/

Kapitola 108. Moorov paradox

Moorov paradox¹ je štandardné označenie pre vetu: “Vonku prší, ale ja neverím, že prší.”

Myslím si, že po prečítaní niektorých komentárov na Less Wrong rozumiem Moorovmu paradoxu lepšie. Jimrandomh² hovorí:

Mnoho ľudí nedokáže rozlíšiť medzi úrovňami nepriamosti. Pre nich “verím v X” a “X” sú to isté, a preto dôvody, prečo sa oplatí veriť v X sú zároveň dôvodmi, prečo je X pravda.

Nemyslím si, že je to správne – relatívne malé deti dokážu pochopiť pojem, že majú nesprávny názor, čo si vyžaduje oddelené vedierka pre mapu a pre územie. Ale ukazuje to smerom k podobnej myšlienke:

Mnoho ľudí asi nedokáže vedome rozlíšiť medzi *vierou* v niečo a *schvaľovaním* niečoho.

Napokon... “verím v demokraciu” znamená, v bežnej reči, že schvaľujete tento systém, a nie že veríte, že demokracia existuje. Slovo “verím” má teda viac než jeden význam. Možno hľadáme zmätené slovo³, ktoré spôsobuje zmätené myslenie (alebo možno len odráža predchádzajúce zmätené myslenie).

Takže: v pôvodnom príklade⁴: “Verím že ľudia sú lepší než naozaj sú,” uviedla nejaké dôvody, prečo by bolo dobré veriť, že ľudia sú dobrí – zdravotné výhody a podobne – a keďže potom mala hrejivý pocit ohľadom “veriť, že ľudia sú dobrí”, preskúmala tento hrejivý pocit a došla k záveru: “Verím, že ľudia sú dobrí.” Čiže si pomýlila *kladný účinok* priradený citovanému názoru za signál *jej viery v daný výrok*. Zároveň samotný svet vyzerá, akoby ľudia neboli takí dobrí. Preto povedala: “Verím, že ľudia sú lepší než naozaj sú.”

A to už hraničí s poctivou chybou – v istom zmysle – keďže ľudia nie sú vyslovene učení, ako majú vedieť, že niečomu veria. Ako v podobenstve o drakovi v garáži⁵; ten, kto hovorí: “V mojej garáži je drak – ale je neviditeľný”, si neuvedomuje svoje *očakávanie*, že neuvidí žiadneho draka ako náznak, že má (správny) model bez draka.

Nie je to tak, že by ľudí *učili* rozoznať, kedy niečomu veria. Nie je to tak, že by ich niekedy učili na strednej škole: “Skutočne niečomu veriť – mať tento výrok vo svojom zozname presvedčení – vám pripadá tak, že to skrátka vyzerá, že presne takto svet *je*. Mali by ste rozoznať tento pocit, čo je skutočný (bez úvodzoviek) názor, a odlíšiť ho od toho, že máte dobrý pocit z nejakého názoru, ktorý rozoznávate ako názor (čo znamená, že je v úvodzovkách).”

Toto výrazne pomáha, aby nám príklady Moorovho paradoxu zo skutočného života pripadali menej cudzie⁶, a poskytuje nám to ďalší mechanizmus, ktorým ľudia môžu zároveň mať aj nemať pravdu⁷.

Podobne Kurige⁸, ktorý napísal:

¹ http://en.wikipedia.org/wiki/Moore%27s_paradox

² http://lesswrong.com/lw/r/no_really_ive_deceived_myself/#ga

³ pozri: 166 Klam zhustenia

⁴ pozri: 107 Viera v sebaklam

⁵ pozri: 13 Viera vo vieru

⁶ pozri: 107 Viera v sebaklam

⁷ http://lesswrong.com/lw/1d/simultaneously_right_and_wrong/

⁸ http://lesswrong.com/lw/r/no_really_ive_deceived_myself/#gk

Verím, že existuje Boh – a že nám vštepil zmysel pre dobro a zlo, ktorým dokážeme vyhodnocovať svet okolo nás. Zároveň verím, že zmysel pre morálku nám naprogramovala evolúcia – zmysel pre morálku, ktorý je s najväčšou pravdepodobnosťou výsledkom vytvárania zoskupení a meta-politických koalícií v komunitách bonobov veľmi, veľmi dávno. Tieto dva názory si neodporujú, ale je zložité ich oba zosúladiť.

Obávam sa, Kurige, že si sa rozhodol, že máš dôvody podporovať citovaný názor, že nám Boh vštepil zmysel pre dobro a zlo. A tiež, že máš dôvody schvaľovať verdikt vedy. Obe z toho vyzerajú byť dobré komunity, do ktorých sa chceš pridať, však? Obe tieto množiny názorov majú svoje výhody? Pozrieš sa do svojho vnútra a zistiš, že máš z oboch týchto názorov dobrý pocit?

Ale *nepovedal* si:

“Boh nám vštepil zmysel pre dobro a zlo, a zároveň nám evolúcia naprogramovala zmysel pre morálku. Tieto dva stavy skutočnosti si neodporujú, ale je zložité ich oba zosúladiť.”

Ak toto čítaš, Kurige, mal by si veľmi rýchlo vysloviť uvedené nahlas, aby si si všimol, že to vyzerá aspoň o trochu ťažšie na prehĺtnutie – všimni si ten *subjektívny rozdiel* – skôr než si dáš tú námahu znovu racionalizovať.

Toto je subjektívny rozdiel medzi tým, keď máte dôvody schvaľovať dva odlišné názory, a keď máte mentálny model jedného sveta, jedného stavu, ako sa veci majú.

Pôvodný článok s komentármi⁹

⁹ http://lesswrong.com/lw/1f/moores_paradox/

Kapitola 109. Neverte, že sa sami oklamete

Nechcem vyzeráť, že zapáram do Kurigeho, ale myslím si, že musíte očakávať isté množstvo spochybňovania, ak prídete na LessWrong a povieť¹:

Jedna z vecí, ktoré som si uvedomil, ktoré pomáhajú vysvetliť rozdiel, ktorý cítim, keď hovorím s väčšinou ostatných kresťanov, je fakt, že niekde po ceste môj svetonázor urobil veľkú odbočku preč od slepej viery a skončil niekde v susedstve Orwellovského double-thinku.

“Ak vieš, že je to double-think²...”

“...ako tomu ešte môžeš veriť?” chce sa mi bezmocne povedať.

Alebo³:

Rozhodol som sa veriť v existenciu Boha – úmyselne a vedome. Toto rozhodnutie však má absolútne nulový vplyv na skutočnú existenciu Boha.

Ak vieš, že tvoj názor nekoreluje so skutočnosťou, ako ho ešte stále môžeš mať?

Nemalo by *inštinktívne* uvedomenie si: “Aha, moment, obloha naozaj *nie* je zelená,” vyplývať z uvedomenia si: “Moja mapa, ktorá hovorí: ‘obloha je zelená’ nemá žiaden dôvod korelovať s územím?”

Nuž... očividne nie.

Jedna časť tohto hlavolamu môže byť moje vysvetlenie Moorovho paradoxu⁴ (“Prší, ale ja neverím, že prší”) – že si ľudia introspektívne mýlia pozitívny pocit spojený s citovaným názorom so skutočnou vierou.

Ale ďalšia časť toho môže skrátka byť, že – v rozpore s rozhorčením, ktoré som tu pôvodne chcel napísať – v skutočnosti je pomerne *lahké* neurobiť skok z “Mapa, ktorá odráža územie by povedala: ‘X’,” k skutočnej viere v “X”. Vyžaduje si to istú prácu *vysvetliť* predstavu mysle ako konštruktéra súladu medzi mapou a územím⁵, a ešte aj potom si môže vyžadovať veľa práce dostať tieto dôsledky na úroveň *inštinktu*.

Teraz si uvedomujem, že keď som napísal⁶: “Nemôžete sa prinútiť veriť, že obloha je zelená, aktom vôle,” nebol som nezaujatým spravodajcom o existujúcich faktoch. Snažil som sa vytvoriť sebanapíňajúce proroctvo.

Môže byť múdre chodiť okolo a úmyselne opakovať: “Mne by double-think nefungoval! Hlboko vnútri by som vedel, že to nie je pravda! Ak viem, že moja mapa nemá dôvod byť korelovaná s územím, to znamená, že tomu neverím!”

Pretože vtedy – ak by vás to vôbec niekedy pokúšalo – vám rýchlo napadnú myšlienky: “Ale ja viem, že to nie je naozaj pravda!” a “Nedokážem oklamať sám seba!”; a takto bude naozaj menšia pravdepodobnosť, že úspešne oklamete sami

¹ http://lesswrong.com/lw/r/no_really_ive_deceived_myself/#gk

² pozri: 105 Doublethink (dobrovoľné skreslenie)

³ http://lesswrong.com/lw/1f/moores_paradox/#u3

⁴ pozri: 108 Moorov paradox

⁵ pozri: 20 Čo je indícia?

⁶ pozri: 108 Moorov paradox

seba. Máte väčšiu šancu pochopiť, na inštinktívnej úrovni, že povedať sám sebe X ešte z X neurobí pravdu: a preto v skutočnosti nie-X.

Ak si budete stále hovoriť, že si *nemôžete* len tak zvoliť, že budete veriť, obloha je zelená – potom je menšia pravdepodobnosť, že sa úspešne oklamete na nejakej úrovni, či už v zmysle, že by ste tomu naozaj uverili, alebo že by ste upadli do Moorovho paradoxu⁷, viery vo vieru⁸, alebo viery v sebaklam⁹.

Ak si budete stále hovoriť, že hlboko vnútri viete...

Ak si budete stále hovoriť, že by ste jednoducho pozreli na svoju zložito vypracovanú nepravdivú mapu a skrátka by ste vedeli, že je to nepravdivá mapa bez očakávanej korelácie s územím, a preto by ste, napriek jej zložitej konštrukcii nedokázali do nej investovať svoju dôverčivosť...

Ak si budete stále hovoriť, že reflexívna konzistencia prevládne a spôsobí, že prestanete veriť na základnej úrovni, akonáhle si na meta-úrovni uvedomíte, že mapa neodráža...

Keď naozaj príde na lámanie chleba... možno to naozaj nedokázate.

Keď príde na úmyselný sebaklam, musíte *veriť vo svoju vlastnú neschopnosť!*

Povedzte sami sebe, že toto úsilie je odsúdené na neúspech... *a ono bude!*

Je toto sila pozitívneho myslenia, alebo sila negatívneho myslenia? Tak či onak, vyzerá to ako rozumná opatrnosť.

Pôvodný článok s komentármi¹⁰

⁷ pozri: 108 Moorov paradox

⁸ pozri: 13 Viera vo vieru

⁹ pozri: 107 Viera v sebaklam

¹⁰ http://lesswrong.com/lw/1o/dont_believe_youll_selfdeceive/

Časť XI. Príliš pohodlné výhovorky

Obsah

110 Správne použitie pokory	157
111 Tretia možnosť	160
112 Ale stále je tu šanca, nie?	162
113 Klam sivej	164
114 Absolútna autorita	167
115 Ako ma presvedčiť, že $2 + 2 = 3$	171
116 Nekonečná istota	173
117 0 a 1 nie sú pravdepodobnosti	176

Kapitola 110. Správne použitie pokory

Všeobecne sa uznáva, že dobrá veda si vyžaduje nejaký druh pokory. Aký *druh*, to už je kontroverzné.

Predstavte si kreacionistu, ktorý hovorí: “Ale kto môže naozaj vedieť, či evolúcia je správna? Je to len teória. Mali by ste byť pokornejší a mať otvorenú myseľ.” Je toto pokora? Evolucionista tu praktizuje veľmi priberčivú nedostatočnú dôveru, odmieta zapojiť mohutné množstvá indícií v prospech záveru, ktorý mu je nepohodlný. Povedal by som, že či už toto nazvete slovom “pokora” alebo nie, je to nesprávny krok v tanci.

A čo inžinier, ktorý pokorne navrhuje havarijné poistky do mechanizmu, hoci si je čertovsky istý, že mechanizmus sa nepokazí? Toto mi pripadá ako dobrý druh pokory. Z historického hľadiska sme už počuli o inžinieroch, ktorí si boli čertovsky istí, že sa nový stroj nepokazí, a potom sa aj tak pokazil.

A čo študent, ktorý pokorne robí skúšku správnosti na svojej písomke z matematiky? Opäť, toto by som označil ako dobrú pokoru.

A čo študent, ktorý povie: “Bez ohľadu na to, koľko skúšok správnosti urobím, nikdy si nemôžem byť istý, že moje odpovede v písomke sú správne,” a preto neurobí ani jednu skúšku správnosti. Aj keď táto voľba pochádza z podobnej emócie ako u predchádzajúceho študenta, je menej múdra.

Odporúčite študovať usilovnejšie a študent odpovie: “Nie, to by u mňa nefungovalo; ja nie som jeden z tých chytrých ľudí ako vy; chudáčik ako ja nemôže dúfať v nič lepšie.” Toto je spoločenská skromnosť, nie pokora. Súvisí s regulovaním postavenia v tlupe, nie s vedeckým procesom. Ak niekoho žiadate, aby bol “pokornejší”, štandardne si vaše slová spojí so spoločenskou skromnosťou – čo je intuitívny, každodenný, dedične daný pojem. Vedecká pokora je nedávny a zriedkavo používaný objav, a nie je vrodene spoločenská. Vedecká pokora je niečo, čo by ste mohli robiť aj sám v skafandri, vzdialený celé svetelné roky od Zeme, a nikto by sa nepozeral. Dokonca aj keby ste mali absolútnu záruku, že vás už nikto nikdy nebude kritizovať bez ohľadu na to, čo o sebe poviete alebo si pomyslíte. Aj tak by ste urobili skúšku správnosti vo výpočte, ak ste múdry.

Ten študent povie: “Ale ja som videl ako druhí študenti robia skúšky správnosti a potom sa aj tak ukázalo, že to majú zle. A vôbec, čo ak vďaka problému indukcie tentokrát $2+2=5$? Bez ohľadu na to, čo urobím, nebude mať istotu.” Znie to veľmi hlboko a veľmi pokorne. Ale nie je náhoda, že tento študent chce rýchlo odovzdať písomku, ísť domov a hrať počítačové hry.

Koniec nejakej éry vo fyzike sa zvyčajne neohlasuje hromom a trúbkami; častejšie začína niečím, čo vyzerá ako malá, maličká chyba... Ale keďže fyzici majú takú arogantnú predstavu, že ich modely by mali fungovať *vždy*, nielen *väčšinou*, všímajú si tieto malé chyby. Zvyčajne po podrobnejšom preskúmaní chybička zmizne. Len občas sa chyba rozšíri až do bodu, keď vyhodí do vzduchu celú teóriu. Preto je napísané: “Ak nehľadáš dokonalosť, zastavíš sa ešte pred prvým krokom.”

Pomyslite však na tú spoločenskú trúfalosť, snažiť sa mať *vždy* pravdu! Mám vážne podozrenie, že keby Veda tvrdila, že evolučná teória platí *väčšinou*, ale nie *vždy* – alebo keby Veda pripustila, že v niektoré dni Zem možno je plochá, ved' kto to môže vedieť naisto – potom by vedci mali lepšiu povest' v

spoločnosti. Veda by bola vnímaná ako menej konfrontačná, pretože by sme sa nemuseli hádať s ľuďmi, ktorí tvrdia, že Zem je plochá – bolo by dosť miesta na kompromis. Keď sa veľa hádate, ľudia vás vnímajú ako konfrontačných. Ak opakovane odmietať kompromisy, je to ešte horšie. Berte to ako otázku postavenia v tlupe: vedci si iste zaslúžili nejaké osobitné postavenie výmenou za také spoločensky užitočné nástroje ako lieky a mobilné telefóny. Avšak toto spoločenské postavenie neospravedľňuje ich požiadavku, aby sa na verejných školách učili *iba* vedecké myšlienky o evolúcii. Kňazi majú predsa takisto vysoké spoločenské postavenie. Vedci si žiadajú viac, než si zaslúžia – vyhrali trochu postavenia, a odrazu si myslia, že sú náčelníkmi celej tlupy! Mali by byť viac pokorní, a občas robiť kompromisy.

Zdá sa, že mnohí ľudia majú pomerne nejasné predstavy o “pokore rozumu”. Je nebezpečné mať normu, ktorej rozumiete iba nejasne; váš myšlienkový obraz môže mať toľko stupňov voľnosti, že sa dokáže prispôbiť, aby ospravedlnil takmer hocikaký skutok. Tam, kde ľudia majú nejasné myšlienkové modely, ktoré možno použiť na zdôvodnenie hocičoho, zvyčajne dospejú k tomu, že uveria tým veciam, ktorým už na začiatku veriť chceli. To je také pohodlné, že ľudia sa zvyčajne neradi vzdávajú neurčitosti. Ale cieľom našej etiky je, aby nás riadila, nie aby bola nami riadená.

“Pokora” je často nepochopená cnosť. To neznamena, že by sme mali zahodiť pojem pokory, ale že by sme ho mali používať opatrne. Možno pomôže, ak sa pozrieme na *činy* odporúčané nejakým “pokorným” typom myslenia a opýtame sa: “Robí ma takéto konanie silnejším alebo slabším?” Ak myslíte na problém indukcie aplikovaný na most, ktorý má zostať stáť, dôjsť k záveru, že nič nie je isté, bez ohľadu na to, koľko bezpečnostných opatrení urobíte, môže znieť rozumne. Ak však uvážite rozdiely v skutočnom svete medzi pridaním pár káblov navyše alebo pokrčením plecami, zdá sa jasné, čo z toho robí most silnejším.

Prevažná väčšina odvolávok na “pokoru rozumu”, ktoré som videl, boli zámenkami na pokrčenie plecami. Ten, kto si kúpi žreb lotérie a hovorí: “Ale nemôžeš *vedieť*, že nevyhrám.” Ten, kto neverí v evolúciu a hovorí: “Ale nemôžeš mi *dokázať*, že je to pravda.” Ten, kto odmieta čeliť napohľad zložitému problému a hovorí: “Asi je príliš ťažké na vyriešenie.” Problémom je motivovaný skepticizmus, čiže sklon nepotvrďovať – skúmať oveľa prísnejšie tie tvrdenia, ktorým nechceme veriť. Pokora, ako je najčastejšie dezinterpretovaná, je úplne všeobecnou výhovorkou niečomu neveriť; vzhľadom na to, že aj tak si nemôžeme byť *istí*. Vyvarujte sa úplne všeobecných výhovoriek!

Ďalším problémom pokory je, že sa príliš ľahko *vyznáva*. V knihe “Breaking the Spell” (“Odčarovanie”) Dennett ukazuje, že hoci mnohým náboženským tvrdeniam je veľmi ťažké uveriť, pre človeka je ľahké uveriť, že by v ne *mal* veriť. Dennett toto nazýva “viera vo vieru”. Čo by to znamenalo naozaj predpokladať, naozaj veriť, že tri sa rovná jednej? Omnoho ľahšie je veriť, že by ste *mali* nejakو veriť, že tri sa rovná jednej, a reagovať príslušne v primeraných chvíľach v kostole. Dennett naznačuje, že mnoho z “náboženskej viery” by sa malo študovať ako “náboženské vyznávanie” – to, čo si ľudia myslia, že by tomu mali veriť, a o čom vedia, že by to mali hovoriť.

Je príliš ľahké na každý protiargument odpovedať: “Nuž samozrejme, môžem sa mýliť.” A potom, keď už ste si splnili svoju povinnosť zohnúť kolená smerom ku Skromnosti, keď ste jej vzdali vyžadovanú úctu, môžete pokračovať svojím pôvodným smerom a nezmeniť ani chlpu.

Vždy je tu pokušenie tvrdiť čo najviac vecí s čo najmenšou námahou. Pokušenie spojiť všetky prichádzajúce informácie spôsobom, ktorý nám umožní zmeniť naše názory a najmä naše *konanie* čo najmenej. John Kenneth Galbraith povedal: “Keď má človek na výber medzi zmenou názoru a dôkazom, že názor zmeniť netreba,

prakticky každý sa pustí do dokazovania." A čím *nepohodľnejšie* je zmeniť názor, tým viac úsilia ľudia na tento dôkaz vynaložia.

Ale viete, ak už ste tak či tak odhodlaní *urobiť* to isté, nemá zmysel sa takto neuveriteľne snažiť o zdôvodnenie. Často som zažil, ako ľudia dostali nové informácie, napohľad ich prijali, a potom starostlivo vysvetlili, prečo aj tak urobia presne tú istú vec, ktorú si predtým naplánovali, akurát s odlišným zdôvodnením. Zmyslom rozmýšľania je *utvárať* naše plány; ak sa chcete tak či tak držať pôvodných plánov, načo sa vôbec unúvať toľkým zdôvodňovaním? Keď sa stretnete s novou informáciou, tá ťažká časť je *aktualizovať, reagovať*, a nenechať túto informáciu len tak zmiznúť v čiernej diere. Nesprávne pochopená pokora je úžasnou čiernou dierou – jediné, čo treba urobiť, je pripustiť, že sa možno mýlite. Preto je napísané: "Byť pokorný znamená prijať konkrétne opatrenia v očakávaní vlastných chýb. Priznať svoju omylnosť a potom s ňou nič neurobiť nie je pokorné; je to chvanie sa svojou skromnosťou."

Pôvodný článok s komentármi¹

¹ http://lesswrong.com/lw/gq/the_proper_use_of_humility/

Kapitola 111. Tretia možnosť

Viera v Deda Mráza dáva deťom pocit úžasu a nabáda ich, aby sa pekne správali v nádeji, že dostanú darčeky. Ak bude viera v Deda Mráza zničená pravdou¹, deti stratia svoj pocit úžasu a prestanú sa pekne správať. Preto, hoci je viera v Deda Mráza fakticky nepravdivá, je to vznešená lož, ktorej čistý zisk by sme mali zachovať z utilitariánskych dôvodov.

Toto je klasicky nazývané falošná dilema², klam vylúčenia stredy, alebo klam výberu balíka³. Aj keby sme prijali faktické a morálne predpokladu horeuvedeného argumentu, stále to nestačí. Ani predpoklad, že taktika Deda Mráza (povzbudzovanie detí, aby verili v Deda Mráza) je lepšia než nulová taktika (nerobiť nič), nedokazuje, že dedomrázizmus je *najlepšia zo všetkých možných alternatív*. Aj iné taktiky by mohli poskytnúť deťom pocit úžasu, napríklad vziať ich na sledovanie štartu Space Shuttle alebo zásobiť ich románmi science fiction. Podobne (ak si dobre spomínam), ponúkať deťom úplatky za dobré správanie povzbudzuje deti, aby sa správali dobre *iba* vtedy, keď sa dospelí pozerajú; zatiaľ čo chválenie bez úplatkov vedie k bezpodmienečne dobrému správaniu.

Vznešené lži bývajú vo všeobecnosti klamy výberu balíka; a odpoveďou na klam výberu balíka je, že ak jeho predpokladaný prínos naozaj potrebujeme, vieme si zostrojiť tretiu možnosť, ako ho získať.

Ako získame tretie možnosti? Prvým krokom k získaniu tretej možnosti je rozhodnúť sa, že ju ideme hľadať, a posledným krokom je rozhodnutie prijať ju. Znie to samozrejme, a predsa sa väčšina ľudí zasekne na týchto krokoch, a nie počas procesu hľadania. Odkiaľ pochádzajú falošné dilemy? Niektoré vznikajú poctivo, pretože je kognitívne náročné vidieť lepšie možnosti. Ale jednou továrňou na falošné dilemy je zdôvodňovanie pochybných rozhodnutí poukazovaním na predpokladané výhody oproti ničnerobeniu. V takomto prípade zdôvodňujúci *nechce* žiadnu tretiu možnosť; nájdenie tretej možnosti by zničilo toto zdôvodnenie. Posledná vec, ktorú chce dedomrázista počuť, je, že chvála funguje lepšie než úplatky, alebo že kozmická loď dokáže inšpirovať rovnako ako lietajúci sob.

Najlepšie býva nepriateľom dobrého. Ak je vaším cieľom *naozaj* pomáhať ľuďom, potom je lepšia možnosť dôvodom na oslavu – až nájdeme túto lepšiu stratégiu, budeme môcť ľuďom pomáhať účinnejšie. Ale ak je cieľom zdôvodniť jednu konkrétnu stratégiu *tvrdením*, že *pomáha ľuďom*, tretia možnosť je nepriateľský argument⁴, konkurent.

Moderná kognitívna psychológia vníma rozhodovanie ako hľadanie možností. V skutočnom živote nestačí možnosti porovnávať, v prvom rade ich potrebujete vytvárať. Pri mnohých problémoch je možností tak veľa, že potrebujete kritérium, kedy sa v hľadaní zastaviť. Ak si chcete kúpiť dom, nemôžete porovnať každý dom v meste; v nejakom bude sa musíte prestať obzerať a rozhodnúť sa.

Ale čo keď sa naše vedomé motívy hľadania – kritériá, ktoré si môžeme pripustiť – nezhodujú s nevedomými vplyvmi? Keď vykonávame údajne altruistické hľadanie, hľadanie lepšej altruistickej taktiky, a nájdeme stratégiu, ktorá druhým pomáha, ale nás znevýhodňuje – nuž, neprestaneme hľadať *tam*; pokračujeme v hľadaní. Sami sebe však povieme, že hľadáme stratégiu, ktoré prinesie ešte väčší altruistický úžitok, samozrejme. Predpokladajme však, že nájdeme taktiku,

¹ pozri: 258 Dvanásť cností rozumnosti

² http://en.wikipedia.org/wiki/False_dilemma

³ http://en.wikipedia.org/wiki/Package-deal_fallacy

⁴ pozri: 47 Politika zabíja myslenie

ktorá má nejaký obhájitelný úžitok, ale zhodou okolností *zároveň* vyhovuje aj nám osobne. Potom hľadanie ihneď ukončíme! V skutočnosti budeme pravdepodobne *odporovať* každej rade, aby sme opäť začali hľadať – vyhovoríme sa trebárs na nedostatok času. (Z nejakého dôvodu máme vždy kognitívne zdroje na hľadanie zdôvodnení našich existujúcich rozhodnutí.)

Dajte si pozor, keď si všimnete, že argumentuje v prospech rozhodnutia, ktoré je *obhájitelné* namiesto optimálneho; alebo že má nejakú výhodu oproti ničnerobeniu, namiesto aby bolo najvýhodnejšou zo všetkých možných akcií.

Falošné dilemy sa časti vyskytujú ako zdôvodnenie neetických pravidiel, ktoré sú, obrovskou zhodou okolností, veľmi pohodlné. Klamanie je napríklad často omnoho pohodlnejšie než hovorenie pravdy; názor, ktorý ste mali na začiatku, je pohodlnejší než aktualizácia. Odtiaľ pochádza obľúba argumentov v prospech vznešených lží; slúžia ako obrana pôvodného názoru – nenájdete vznešeného klamára, ktorý by vypočítal novú vznešenú lož; vždy si držia tú lož, s ktorou začali. Radšej toto hľadanie rýchlo ukončiť!

Ak to chcete robiť lepšie⁵, opýtajte sa rovno: *Keby som videl, že existuje lepšia možnosť než moje terajšie rozhodnutie, tešil by som sa z hĺbky srdca, alebo by som cítil malý záblesk váhania, než by som sa vzdal?* Ak sú odpovede “nie” a “áno”, dajte si pozor na to, že ste asi nehládali tretiu možnosť.

Čo vedie k ďalšej dobre otázke, ktorú môžete sami sebe položiť: *Strávil som päť minút so zavretými očami brainstormujúc divoké a tvorivé možnosti, snažiac sa vymyslieť lepšiu možnosť?* Musí to byť päť minút skutočného času, lebo inak by ste iba žmurkli – zavreli oči a opäť ich otvorili – a povedali: “Veru hej, hľadal som iné možnosti, ale žiadne nie sú.” Žmurknutie je dobrou čiernou dierou⁶, do ktorej môžete odhodiť svoje povinnosti. Odporúčam použiť skutočné, fyzické hodiny.

A čo sa týka tých divokých a tvorivých možností – dávali ste si dobrý pozor na to, aby ste nevymysleli niečo dobré? Mali ste v kútiku mysle tajné úsilie zaručiť, aby každá možnosť, nad ktorou sa zamyslíte, bola očividne zlá?

Je úžasné, ako veľa vznešených klamárov a im podobných je ochotných prijať etické priestupky – pri všetkom povinnom plači nad svojou agóniou svedomia – hoci nestrávil ani päť minút na hodinách hľadaním inej možnosti. Existujú myšlienkové hľadania, pri ktorých si tajne želáme, aby sme neuspeli; a keď je vidina úspechu nepohodlná, ľudia prijímajú prvú dostupnú výhovorku, aby sa vzdali.

Pôvodný článok s komentármi⁷

⁵ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

⁶ pozri: 110 Správne použitie pokory

⁷ http://lesswrong.com/lw/hu/the_third_alternative/

Kapitola 112. Ale stále je tu šanca, nie?

Pred rokmi som hovoril s niekým, kto mimochodom poznamenal, že neverí v evolúciu. Ja som povedal: "Toto nie je devätnáste storočie. Keď Darwin prvýkrát navrhol evolúciu, mohlo byť rozumné pochybovať o nej. Ale toto je dvadsiate prvé storočie. Dokážeme čítať gény. Ľudia a šimpanzy majú spoločných 98 % DNA. Vieme, že ľudia a šimpanzy sú príbuzní. Je to jasné."

On: "Možno je tá DNA podobná iba náhodou?"

Ja: "Pravdepodobnosť niečoho takého je asi dve na sedemsto päťdesiat milióntu k jednej."

On: "Ale stále je tu šanca, nie?"

Dobre, je tu pár dôvodov, prečo si moje minulé ja nemôže v tejto konverzácii nárokovať jasné morálne víťazstvo. Jeden z dôvodov je, že si nepamätám, odkiaľ som vzal ten vzorec $2^{(750\,000\,000)}$, ale pravdepodobne je to meta-rádovo sedí. Druhým dôvodom je, že moje minulé ja nepoužilo pojem kalibrovannej istoty. Zo všetkých prípadov v histórii ľudstva, keď človek vypočítal, že niečo má šancu $2^{(750\,000\,000)} : 1$, nepochybne sa mýlil vo viac než jednom z $2^{(750\,000\,000)}$ prípadov. Napríklad ten odhad spoločných génov bol opravený na 95 %, nie 98 % - a aj to sa možno týka iba 30 000 známych génov a nie celého genómu, a v tom prípade to ani meta-rádovo nesedí.

Ale myslím si, že odpoveď toho druhého je stále veľmi smiešna.

Nespomínam si, čo som povedal potom - asi niečo ako "Nie" - ale pamätám si tento prípad, pretože mi dal niekoľko vhľadov do zákonov myslenia, ako im rozumejú neosvietení.

Prvýkrát mi napadlo, že ľudská intuícia robí kvalitatívny rozdiel medzi "Nie je šanca" a "Veľmi maličká šanca, na ktorú sa oplatí myslieť". Vidíte v debate o lotérii¹ na Overcoming Bias, kde niekto povedal: "Je veľký rozdiel medzi nulovou šancou na výhru a epsilonovou šancou na výhru," a ja som odpovedal: "Nie, ten rozdiel je v ráde epsilonov; ak o tom pochybuješ, nech sa epsilon rovná $1 / 10^{(10^{100})}$."

Problém je, že teória pravdepodobnosti nám občas dovolí vypočítať šancu, ktorá je naozaj príliš malá na to, aby bola hodna myšlienkového priestoru, ktorý zaberá - ale v tom čase už ste ju vypočítali. Ľudia si mýlia mapu a územie, takže na inštinktívnej úrovni myslenie na symbolicky opísanú pravdepodobnosť sa zdá ako "pravdepodobnosť, na ktorú sa oplatí myslieť", aj keď tento symbolický popis odkazuje na číslo také malé, že keby to bolo zrnko prachu, nevideli by ste ho. Dokážeme použiť slová na opísanie takýchto malých čísel, ale nie pocity - takýto malý pocit neexistuje, neaktivoval by dostatok neurónov a neuvolnil by dostatok neurotransmiterov, aby sme ho cítili. To je dôvod, prečo ľudia kupujú žreby v lotérii - nikto nedokáže precítiť malosť takejto malej pravdepodobnosti.

Ale čo mi pripadá ešte fascinujúcejšie, je kvalitatívne rozlišovanie medzi "istými" a "neistými" argumentmi, kde pokiaľ argument nie je istý, máte povolené ho ignorovať. Napríklad, ak je pravdepodobnosť nula, musíte sa daného názoru vzdať, ale ak je pravdepodobnosť $10^{(10^{100})}$, môžete si ho ponechať.

¹ pozri: 310 Nová vylepšená lotéria

Dobre, toto je slobodná krajina a nikto by vás nemal posadiť za mreže kvôli nelegálnemu rozmýšľaniu, ale ak ste sa rozhodli ignorovať argument, ktorý hovorí, že pravdepodobnosť je $10^{(10^{100})}$, prečo neignorovať aj argument, ktorý hovorí, že pravdepodobnosť je nula? Chcem tým povedať, keďže ste sa tak či tak rozhodli ignorovať indície, prečo je o toľko horšie ignorovať isté indície než neisté indície?

Často som v živote zistil, že som sa poučil² z do očí bijúcich zlých príkladov druhých ľudí, keď som ich zovšeobecnil na menej jasné prípady. V tomto prípade je doplňujúcim ponaučením to, že ak nemôžete ignorovať pravdepodobnosť $1 / 10^{100}$ len preto, že sa vám zachcelo, nemôžete ignorovať ani pravdepodobnosť 0,9 preto, že sa vám zachcelo. Je to ten istý šmyklavý svah.

Spomeňte si na tento príklad, ak sa niekedy pristihnete, že si myslíte: "Ale nemôžete mi *dokázať*, že sa mýlim." Ak ste rozhodnutí ignorovať pravdepodobnostný argument, prečo neignorovať aj dôkaz?

Pôvodný článok s komentármi³

² pozri: 52 Obrátená hlúposť nie je inteligencia

³ http://lesswrong.com/lw/ml/but_theres_still_a_chance_right/

Kapitola 113. Klam sivej

Súvisí s kapitolou: Tsuyoku Naritai¹

Sofistik: "Svet nie je čiernobiely. Nikto nekoná čisté dobro ani čisté zlo. Všetko je sivé. Preto nikto nie je lepší než nikto iný."

Zetét: "Kedže poznáš iba sivú, došiel si k záveru, že všetky sivé majú rovnaký odtieň. Vysmievaš sa z jednoduchosti dvojfarebného pohľadu, ale sám si ho nahradil jednofarebným..."

—Marc Stiegler, Dávidov prak

Neviem, či táto Sofistikova chyba má oficiálny názov, ale ja ju nazývam klam sivej. Videli sme ju vo včerašom článku – ten, ktorý veril, že šanca dve na sedemsto päťdesiat miliónu k jednej proti nemu znamená "stále je tu šanca". Všetky pravdepodobnosti boli preňho jednoducho "nejasné" a to znamenalo, že má právo ignorovať ich, ako sa mu zachce.

"Mesiak sa skladá zo zeleného syra" a "Slnko sa skladá najmä z vodíka a hélia" sú oba neisté výroky, ale nemajú rovnakú neistotu.

Všetko sú odtiene sivej, ale existujú také svetlé odtiene sivej, že je to takmer biela, a také tmavé odtiene sivej, že je to takmer čierna. A aj keby nie, stále môžeme porovnávať odtiene a hovoriť "tento je tmavší" a "tento je svetlejší".

Pred rokmi bolo jedným z čudných formujúcich momentov mojej kariéry racionalistu prečítanie tohto odstavca z knihy *Hráč hier* od Iaina M. Banksa, najmä zvýraznená veta:

System viny nerozoznáva žiadnych nevinných. Ako pri každom mocenskom aparáte, ktorý si myslí, že každý je buď zaňho alebo proti nemu, my sme proti nemu. Aj ty by si bol, keby si nad tým rozmýšľal. Samotný spôsob, akým rozmýšľaš, ťa radí medzi jeho nepriateľov. To nemusí byť tvoja chyba, pretože **každá spoločnosť vtlača niektoré svoje hodnoty tým, ktorí v nej vyrastajú, ale pointa je, že niektoré spoločnosti sa tento účinok snažia maximalizovať a niektoré minimalizovať**. Ty pochádzaš z tej druhej a žiadajú ťa, aby si sa vysvetlil niekomu z tej prvej. Vykrúcanie sa bude omnoho ťažšie než si dokážeš predstaviť; neutralita je pravdepodobne nemožná. Nemôžeš si vybrať nemať postoje, ktoré máš; nie sú to nejaké oddelené veci, ktoré by si nejako mohol odpojiť od zvyšku tvojej bytosti; sú funkciou tvojej existencie. Ja to viem a oni to vedia; mal by si sa s tým radšej zmieriť.

Teraz mi nepíšte zlostné komentáre, že keď spoločnosti vtlačajú menej svojich hodnôt, potom každá nasledujúca generácia musí viac pracovať, aby sa dostala zo štartu. To nie je to, čo som si z tohto odseku odniesol.

Z tohto odseku som si odniesol niečo, čo v retrospektíve vyzerá natolko zrejmé, že som to mohol nájsť na stovkách miest; ale niečo v tomto odseku spôsobilo, že mi to docvaklo.

Bol to celý pojem Kvantitatívnej Cesty aplikovanej na problémy života ako sú morálne úsudky a hľadanie osobného sebazdokonaľovania. Že aj keď nedokázate niečo len tak zapnúť a vypnúť, aj tak budete mať sklon to zvyšovať alebo znižovať.

¹ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

Je to také samozrejme, že sa to neoplatí spomínať? Hovorím, že to nie je príliš samozrejmé, pretože nejeden bloger povedal o *Overcoming Bias*: "Je to nemožné, nikto nemôže úplne odstrániť skreslenia." Nezaujíma ma, či je dotyčný profesionálny ekonóm, je jasné, že ešte nepochopil, ako sa Kvantitatívna Cesta týka každodenného života a vecí ako osobné sebazdokonaľovanie. To, čo nedokážem *odstrániť*, sa mi ešte stále môže oplatiť *redukovať*.

Alebo si vezmite túto výmenu medzi Robinom Hansonom² a Tylerom Cowenom³. Robin Hanson povedal, že sa snaží dať aspoň 75 % váhy predpisom ekonomickej teórii oproti vlastnej intuícii: "Snažím sa väčšinou iba priamočiaro aplikovať ekonomickú teóriu, a pridávať k nej iba málo osobného alebo kultúrneho hodnotenia." Tyler Cowen odpovedal:

Podľa mňa neexistuje nič také ako "priamočiaro aplikovať ekonomickú teóriu"... teórie sa vždy aplikujú cez naše osobné a kultúrne filtre a nijako inak to nemôže byť.

Áno, ale môžete sa snažiť minimalizovať tento efekt, alebo môžete robiť veci, ktoré ho iste zvýšia. A *ak* sa ho snažíte minimalizovať, potom si myslím, že v mnohých prípadoch nie je nerozumné nazvať výstup "priamočiarym" - dokonca aj v ekonómii.

"Každý je nedokonalý." Mohandas Gándhí bol nedokonalý, aj Josif Stalin bol nedokonalý, ale nebol to rovnaký odtieň nedokonalosti. "Každý je nedokonalý!" je výborný príklad nahradenia dvojfarebného pohľadu jednofarebným. Ak povie: "Nikto nie je dokonalý, ale *niektorí ľudia sú menej nedokonalí ako iní*," možno nezožnete potlesk⁴; ale tým, ktorí sa snažia robiť veci lepšie⁵, ste ukázali nádej. Napokon, nikto nie je *dokonale* nedokonalý.

(Vždy keď mi niekto povie: "Perfekcionizmus ti škodí," odpoviem: "Myslím si, že je okej byť nedokonalý, ale nie taký nedokonalý, že si to druhí ľudia všimnú.")

Podobná je hlúposť tých, ktorí hovoria: "Každá vedecká paradigma prináša nejaké svoje predpoklady, podľa ktorých interpretuje pokusy," a potom sa správajú, ako by tým dokázali, že veda je na rovnakej úrovni ako šamanstvo. Každý svetonázor vtlača niečo zo svojej štruktúry do svojich pozorovaní, ale pointa je, že existujú svetonázory, ktoré sa snažia toto vtlačanie minimalizovať, a sú svetonázory, ktoré sa v tom vyžívajú. Neexistuje biela, ale existujú odtiene sivej, ktoré sú omnoho svetlejšie než iné, a je hlúposť zaobchádzať s nimi, akoby boli všetky na rovnakej úrovni.

Ak mesiac posledných pár miliárd rokov obiehal okolo Zeme, ak ste ho posledných pár rokov videli na oblohe, a ak očakávate, že ho uvidíte na jeho určenom mieste v jeho určenej fáze aj zajtra, toto nie je istota. A ak očakávate, že neviditeľný drak⁶ vylieči vašej dcéry rakovinu, ani toto nie je istota. Ale sú to pomerne odlišné stupne neistoty - očakávanie, že sa veci opäť stanú rovnakým spôsobom, aký ste v minulosti predpovedali na dvanásť desatinných miest, verzus očakávanie, že sa stane niečo, čo *porušuje* doteraz pozorovaný poriadok. Nazvať obidvoje "vierou" sa mi zdá trochu príliš nepresné⁷.

Je to veľmi zaujímavé, psychologicky - celá tá vec, že: "Aj veda je založená na viere, tak vidíš!" Typicky to hovoria ľudia, ktorí tvrdia, že viera je *dobrá* vec. Prečo teda hovoria: "Aj veda je založená na viere!" tým zlostným triumfálnym tónom, a nie ako kompliment? A to pomerne *nebezpečný* kompliment, človek

² <http://www.overcomingbias.com/2007/12/economist-judgm.html>

³ <http://marginalrevolution.com/marginalrevolution/2007/12/how-my-views-di.html>

⁴ pozri: 16 Signály na potlesk

⁵ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

⁶ pozri: 13 Viera vo vieru

⁷ pozri: 26 Cnosť presnosti

by si pomyslel, z ich pohľadu. Ak je veda založená na "viere", potom je veda rovnakého typu ako náboženstvo – možno ich priamo porovnávať. Ak je veda náboženstvom, potom je to náboženstvo, ktoré uzdravuje chorých a odhaľuje tajomstvá hviezd. Dávalo by zmysel povedať: "Kňazi vedy dokážu jasne, verejne, overiteľne chodiť po Mesiaci ako zázrak založený na viere, a kňazi tvojej viery toto nedokážu." Ste si istí, že chcete ísť týmto smerom, veriaci? Keď sa nad tým lepšie zamyslíte, možno by ste celé to "Aj veda je náboženstvo!" radšej odvolali.

Je to zvláštna dynamika: Pokúšate sa vyčistiť svoj odtieň sivej a dostanete ho do bodu, kde je pomerne svetlý, načo niekto vstane a povie hlboko urazeným tónom: "Ale to nie je biela! Je to sivá!" Jedna vec je, keď niekto povie: "Nie je to také svetlé ako si myslíš, pretože sú tu konkrétne problémy X, Y a Z." Ale iná vec je, keď niekto zlostne povie: "To nie je biela! Je to sivá!" bez poukázania na konkrétne tmavé škvrny.

V tomto prípade začnem podozrievať, že sa jedná o psychológiu ešte nedokonalejšiu než je bežné – že niekto možno uzavrel diabolskú zmluvu so svojimi vlastnými chybami a teraz odmieta počuť o hocijakej možnosti zlepšenia. Keď si niekto nájde výhovorku, prečo sa nepokúšať zlepšiť, často odmieta pripustiť, že by sa niekto iný *mohol* pokúsiť zlepšiť, a každý spôsob zlepšenia je odvtedy jeho nepriateľom, a každé tvrdenie, že je možné pohnúť sa dopredu, je preňho urážkou. A tak jedným dychom pyšne povedia: "Som rád, že som sivý," a nasledujúcim dychom zlostne: "A ty si tiež sivý!"

Ak neexistuje čierna a biela, stále existuje svetlejšia a tmavšia, a nie všetky sivé sú rovnaké.

Dodatok: G⁸ pripomína Asimovovu Relativitu nesprávneho⁹: "Keď si ľudia mysleli, že zem je plochá, mýlili sa. Keď si ľudia mysleli, že zem je guľa, mýlili sa. Ale ak si myslíš, že myslieť si, že zem je guľatá je rovnako nesprávne ako myslieť si, že zem je plochá, potom je tvoj pohľad ešte nesprávnejší než oni obaja dohromady."

Pôvodný článok s komentármi¹⁰

⁸ http://lesswrong.com/lw/mm/the_fallacy_of_gray/hho

⁹ <http://chem.tufts.edu/AnswersInScience/RelativityofWrong.htm>

¹⁰ http://lesswrong.com/lw/mm/the_fallacy_of_gray/

Kapitola 114. Absolútna autorita

Príde k vám niekto a povýšenie povie: “Veda vlastne nič naozaj *nevie*. Máte akurát *teórie* – nemôžete vedieť *naisto*, že máte pravdu. Vy vedci ste zmenili názor na to, ako funguje gravitácia – čo ak zajtra rovnako zmeníte názor na evolúciu?”

Hľadte na tú hlbokú kultúrnu priepasť¹. Ak si myslíte, že ju prekleniete niekoľkými vetami, čaká vás trpké sklamanie.

Vo svete neosvietených existujú iba autority a neautority. Čomu možno veriť, tomu možno veriť; čomu nemožno veriť, to môžete v pohode zahodiť. Existujú dobré zdroje informácií a zlé zdroje informácií. Ak vedci čo len raz vo svojej histórii zmenili svoje príbehy, potom veda nemôže byť skutočnou Autoritou a už nikdy jej nemožno dôverovať – ako keď pristihnete svedka pri rozpore vo výpovedi, alebo keď prichytíte zamestnanca ako kradne z pokladne.

Navyše, dotýčny považuje za samozrejmé, že zástanca nejakej myšlienky ju musí obhajovať proti všetkým možným protiargumentom² a nesmie nič pripustiť. Všetky tvrdenia sú hodnotené podľa toho. Ak ešte aj *zástanca* vedy pripúšťa, že veda je menej než dokonalá, potom musí byť úplne bezcenná.

Keď niekto prežil celý svoj život zvyknutý na istotu, nemôžete mu len tak povedať: “Veda je pravdepodobnosť, rovnako ako všetko iné poznanie.” Prvú polovicu vašej vety vezme ako priznanie viny; druhú polovicu odmietne ako chabý pokus obviňovať všetkých ostatných, aby ste sa vyhli odsúdeniu.

Priznali ste, že nie ste dôveryhodní – chod' teda preč, veda, a viac nás neobťažuj!

Jedným zrejším zdrojom tohto myšlienkového vzorca je náboženstvo, ktorého písma údajne pochádzajú od Boha; preto by priznanie ľubovoľnej chyby v nich úplne zničilo ich autoritu; preto je každá stopa pochybnosti hriechom, a vyhlasovať istotu³ je *povinné* bez ohľadu na to, či ste si istí alebo nie.

Ale obávam sa, že tradičný školský systém je tu tiež na vine. Učiteľ vám povie isté veci a vy im musíte veriť a musíte ich na skúške naspäť odrecitovať. Keď však niečo v triede povie študent, nemusíte sa tomu podriaďovať – (zdá sa, že) máte slobodu súhlasiť alebo nesúhlasiť a nikto vás za to nepotrestá.

Obávam sa, že táto skúsenosť mapuje oblasť viery na spoločenské oblasti *autority*, *príkazov*, *zákona*. V spoločenskej oblasti je kvalitatívny rozdiel medzi absolútnymi zákonmi a neabsolútnymi zákonmi, medzi príkazmi a odporúčaniami, medzi autoritami a neautoritami. Zdá sa, že existuje presné poznanie a nepresné poznanie, tak ako presné predpisy a nepresné predpisy. Presným autoritám sa treba podrobiť, kým nepresné odporúčania možno poslúchať alebo ignorovať podľa osobného rozhodnutia. A veda, keďže sama pripúšťa, že má možnosť chyby, musí patriť do tej druhej kategórie.

(Na okraj poznamenávam, že vidím určitú podobnosť s tými, ktorí si myslia, že ak nedostanete autoritatívnu pravdepodobnosť napísanú na kúsku papiera od učiteľa v triede, alebo z nejakého iného nespochybniteľného zdroja, potom vaša neistota⁴ nie je vecou bayesovskej teórie pravdepodobnosti. Niekto by mohol – *ach!* – nesúhlasiť s vašim odhadom pôvodnej pravdepodobnosti. Tak sa zdá tým,

¹ pozri: 9 Očakávame krátke inferenčné vzdialenosti

² pozri: 48 Debaty o pravidlách by nemali vyzeráť jednostranne

³ pozri: 15 Vyznávanie a fandanie

⁴ pozri: 25 Zameraj svoju neistotu

ktorí nie sú celkom osvietení, že bayesovské pôvodné pravdepodobnosti patria do kategórie názorov, ktoré navrhujú študenti, a nie do kategórie názorov, ktoré prikazujú učitelia – nie je to pravé *poznanie*.)

Táto hlboká kultúrna priepasť medzi autoritatívnou cestou a kvantitatívnou cestou je veľmi otravná pre tých, ktorí ponad ňu hľadajú z racionalistickej strany. Je tu niekto, kto verí, že má *poznanie* *spoľahlivejšie* než je púhy pravdepodobnostný odhad vedy – ako sú odhady, že mesiac zajtra vyjde na svojom určenom mieste a v určenej fáze, tak ako to bolo každú pozorovanú noc od vynálezu astronomických záznamov, a ako to predpovedajú fyzikálne teórie, ktorých predchádzajúce predpovede boli úspešne potvrdené na štrnást desiatich miest. A čo je to *poznanie*, ktoré ten neosvietený vyzdvihuje nad to naše, a prečo? Pravdepodobne je to nejaký plesnivý starý zvitok, ktorý bol vyvrátený jedenástimi spôsobmi v nedeľu, v pondelok, a v každý ďalší deň v týždni. Napriek tomu je (podľa ich slov) *spoľahlivejší* než veda, pretože si nikdy nepripustí chybu, nikdy nezmení názor, bez ohľadu na to, koľkokrát ho vyvráti. Pohadzujú slovom “istota” ako tenisovú loptou, používajú ho zľahka ako pierko – zatiaľ čo vedci sú zaťažení povinnosťou pochybovať, a zápasia, aby dosiahli aspoň štipku pravdepodobnosti. “Ja som dokonalý,” hovoria bez ohľadu na svet, “iste som veľmi vysoko nad vami, ktorí sa stále musíte namáhať, aby ste sa zdokonalili.”

Nie je nič jednoduché⁵, čo by ste im mohli povedať – žiadna *rýchla* zdrvivúca odpoveď. Keď pozorne uvažujete, možno sa vám podarí získať si obecnosť, ak je to verejná diskusia. Nanešťastie však nemôžete len tak vyhrnúť⁶: “Hlúpy smrteľník, kvantitatívna cesta je za hranicami tvojho chápania, a názory, ktoré tak ľahkovážne nazývaš ‘istoty’ sú menej isté než tá najmenšia z našich mocných hypotéz.” Je to rozdiel celkového *životného nastavenia*, ktorý nie je ľahké opísať slovami a už vôbec nie rýchlo.

Čo by ste mohli povedať, ako rečník pred obecnosťou? Ťažko povedať... možno:

- “Sila vedy pochádza zo schopnosti zmeniť názor a priznať si, keď sa mýlime. Ak si vy nikdy nepriznáte, že sa mýlite, to ešte neznamená, že robíte menej chýb.”
- “Každý môže *povedať*, že si je absolútne istý. Ťažšie je však nikdy, naozaj nikdy neurobiť žiadnu chybu. Vedci si tento rozdiel uvedomujú, a preto nehovoria, že sú si absolútne istí. To je celé. Neznamená to, že majú nejaký konkrétny dôvod pochybovať o nejakej teórii – aj keby každý kúsok indície ukazoval tým istým smerom, keby sa všetky hviezdy a planéty zoradili ako dominové kocky, aby podporili jednu hypotézu, vedci aj tak nepovedia, že sú si absolútne istí, pretože skrátka majú vyššie kritériá. To neznamená, že by vedci mali menší *nárok* na istotu než povedzme politici, ktorí sú si vždy takí istí všetkým.”
- “Vedci nepoužívajú slová ‘nie som si absolútne istý’ takým spôsobom, ako ste zvyknutí v bežnej konverzácii. Chceme tým povedať, predstavte si, že by ste boli u lekára na krvnom teste, a lekár by prišiel a povedal: ‘Urobili sme nejaké testy, ale nie je absolútne isté, že sa neskladáte zo syra, a je nenulová šanca, že dvadsať víl zložených z inteligentnej čokolády spieva pesničku od Barneyho “I love you” vo vašom hrubom čreve.’ Utekajte odušu, váš lekár sám potrebuje lekára. Keď však vedec povie to isté, znamená to, že si myslí, že tá pravdepodobnosť je taká malá, že by ste ju neuvideli ani v elektrónovom mikroskope, ale je ochotný pozrieť sa na vaše indície v tom extrémne nepravdepodobnom prípade, že by ste nejaké mali.”
- “Boli by ste ochotní zmeniť svoj názor na niektoré veci, ktoré nazývate ‘isté’, keby ste videli dosť indícií. Myslím tým, predstavte si, že by sa samotný Boh

⁵ pozri: 9 Očakávame krátke inferenčné vzdialenosti

⁶ pozri: 9 Očakávame krátke inferenčné vzdialenosti

zniesol z oblakov a povedal by vám, že celé vaše náboženstvo je pravdivé, s výnimkou narodenia z panny. Ak by toto dokázalo zmeniť váš názor, potom nemôžete povedať, že ste si absolútne istí narodením z panny. Z technických dôvodov teórie pravdepodobnosti, ak je teoreticky možné, že by ste o niečom zmenili svoj názor, potom to nemôže mať pravdepodobnosť presne rovnú jednej. Táto neistota môže byť menšia než zrnko prachu, ale musí tam byť. Ak by ste nedokázali zmeniť názor ani keby vám Boh povedal, že je to inak, potom máte problém pripustiť, že sa mýlite, ktorý zrejme prekračuje všetko, čo vám smrteľník ako ja môže povedať.”

V istom zmysle je však zaujímavejšia otázka, čo povedať niekomu, keď *nie ste* pred obecenstvom. Ako začať ten dlhý proces učenia niekoho, ako žiť vo vesmíre bez istoty?

Myslím si, že prvým krokom by malo byť pochopenie, že bez istoty sa žiť *dá* – že aj *keby* ste si, *hypoteticky povedané*, neboli ničím istí, to vám stále neberie schopnosť robiť morálne alebo faktické rozdiely. Parafrázujúc Lisa Bujolda: “Netlač silnejšie, znižuj odpor.”

Jedna z častých *obhajob* Absolútnej Autority je niečo, čo nazývam: “argumentovanie argumentovaním sivou”, a vyzerá to asi takto:

- Morálny relativisti hovoria:
 - Svet nie je čiernobiely, preto:
 - Všetko je sivé, preto:
 - Nikto nie je lepší než druhí, preto:
 - Môžem robiť čokoľvek sa mi zachce a vy ma nemôžete zastaviť, hahahaha.
- My však potrebujete vedieť zastaviť ľudí, aby nevražili.
- Preto musí existovať nejaký spôsob, ako si byť absolútne istí, inak morálni relativisti vyhrajú.

Obrátená hlúposť⁷ nie je inteligencia. Nemôžete dôjsť k správnej odpovedi tým, že otočíte *každý jeden* riadok argumentu, ktorý viedol k nesprávnemu záveru – to dáva bláznovi príliš podrobnú kontrolu nad vami. Každý jeden riadok⁸ musí byť správny, aby matematický argument platil. A neplatí, že keď morálni relativisti povedia: “Svet nie je čiernobiely,” musí to byť nepravda, rovnako ako neplatí, že ak si Stalin myslel, že $2 + 2 = 4$, musí “ $2 + 2 = 4$ ” byť nepravda. Chyba (a stačí, ak urobíte jednu) je v skoku z dvojfarebného pohľadu na jednofarebný, že všetky sivé farby majú rovnaký odtieň.

Bol by to príliš veľký ústupok (vlastne ústupok vo všetkom) súhlasiť s predpokladom, že musíte mať absolútne poznanie absolútne dobrých možností a absolútne zlých možností, aby ste mohli byť morálni. Môžete mať neisté poznanie relatívne lepších a relatívne horších možností, a aj tak si vyberať. V skutočnosti by to mala byť rutina, nie niečo, z čoho robíme drámu.

Myslím tým, áno, keby ste si museli vybrať medzi dvoma alternatívami A a B, a keby sa vám nejako podarilo získať určite správnu dobre kalibrovanú 100%-nú istotu, že A je absolútne a úplne žiadúce, a že B je suma všetkého zlého a odporného, bola by to *postačujúca* podmienka na výber A namiesto B. Ale nie je to *nevyhnutná* podmienka.

⁷ pozri: 52 Obrátená hlúposť nie je inteligencia

⁸ pozri: 6 Pritlačujúce podrobnosti

Aha, a: Logický klam: Odvolávanie sa na dôsledky názoru⁹.

Pozrime sa, čo ešte potrebujú vedieť? No, že existuje celá racionalistická kultúra, ktorá hovorí, že pochybnosti, otázky a priznanie si chyby nie sú hrozné hanebné veci.

Existuje tu celá tá predstava, že môžeme získavať informácie tým, že sa *na veci pozeráme*, namiesto toho, že nás o nich presviedčajú. Keď sa pozriete na veci pozornejšie, niekedy zistíte, že sú iné ako ste si mysleli na prvý pohľad; ale to neznamená, že vám príroda klame, alebo že by ste sa mali prestať pozeráť.

Ďalej existuje pojem kalibrovanej istoty – že “pravdepodobnosť” nie je to isté ako malý ukazovateľ vo vašej hlave, ktorý meria emocionálnu pripútanosť k myšlienke. Je to skôr miera toho, ako často v praxi, v skutočnom živote, ľudia v istom stave presvedčenia hovoria veci, ktoré sú naozaj pravda. Ak vezmete sto ľudí a požiadate ich, aby napísali sto výrokov, ktorými sú si “absolútne istí”, koľko z nich bude správnych? Nie sto.

Dokonca tvrdenia, ohľadom ktorých sú ľudia naozaj fanatickí, majú *omnoho menšiu* pravdepodobnosť byť správnymi než tvrdenia typu “Slnko je väčšie ako Mesiak”, ktoré vyzerajú príliš všedne na to, aby sa nad nimi niekto vzrušoval. Pre každé tvrdenie, ktorým si je niekto „absolútne istý“ pravdepodobne nájdete niekoho, kto si je “absolútne istý” jeho opakom, pretože takéto fanatické vyznania presvedčenia nevznikajú v neprítomnosti opozície. Takže tento malý ukazovateľ v hlavách ľudí, ktorý meria ich emocionálnu pripútanosť k názoru, sa nedá ľahko preložiť na kalibrovanú istotu – dokonca sa ani nespráva monotónne.

A čo sa týka “absolútnej istoty” - nuž, ak povieť, že niečo má pravdepodobnosť 99,9999 %, znamená to, že si myslíte, že by ste mohli vysloviť *milión* rovnako silných nezávislých tvrdení, *jedno za druhým*, čo by vám zabralo asi celý rok, a mýlili by ste sa v priemere jedenkrát. To je dosť neuveriteľné. (A je úžasné uvedomiť si, že túto úroveň istoty dokážeme dosiahnuť výrokom: “Nevyhráš v lotérii¹⁰.”) Nehovorme už teda o pravdepodobnosti 1,0. Keď si uvedomíte, že na fungovanie v živote *nepotrebuje* pravdepodobnosť 1,0, uvedomíte si, aké absolútne smiešne je myslieť si, že by ste ľudským mozgom niekedy mohli dosiahnuť 1,0. Pravdepodobnosť 1,0 nie je iba istota, je to *nekonečná istota*.

Vlastne sa mi zdá, že aby sme sa vyhli nepochopeniu verejnosti, možno by vedci mali hovoriť: “Nie sme si NEKONEČNE istí” namiesto “Nie sme si istí”. Pretože to druhé v bežnej reči naznačuje, že máte konkrétne dôvody pochybovať.

Pôvodný článok s komentármi¹¹

⁹ <http://www.nizkor.org/features/fallacies/appeal-to-consequences.html>

¹⁰ pozri: 310 Nová vylepšená lotéria

¹¹ http://lesswrong.com/lw/mn/absolute_authority/

Kapitola 115. Ako ma presvedčiť, že $2 + 2 = 3$

V kapitole “Čo je indícia?”¹ som napísal:

To je dôvod, prečo racionalisti kladú taký silný dôraz na paradoxne vyzerajúce tvrdenie, že názor je skutočne *hodnotný* iba vtedy, keby bolo principiálne možné presvedčiť vás, aby ste si mysleli niečo iné. Keby vaša sietnica skončila v rovnakom stave bez ohľadu na to, aké svetlo na ňu dopadá, boli by ste slepí... Preto máme frázu “slepá viera”. Ak to, čomu veríte, nezávisí na tom, čo vidíte, potom vás oslepili prakticky rovnako, ako keby vám vypichli oči.

Cihan Baran odpovedal²:

Neviem si predstaviť situáciu, ktorá by urobila $2 + 2 = 4$ nepravdivým. Asi z toho dôvodu je moje presvedčenie, že $2 + 2 = 4$, nepodmienené.

Priznávam sa, že si neviem predstaviť “situáciu”, ktorá by urobila $2 + 2 = 4$ nepravdivým. (Existujú redefinície, ale to nie sú “situácie”, a nehovorí sa v nich o 2, 4, =, ani +.) Ale to nerobí moje presvedčenie nepodmieneným. Viem si pomerne ľahko predstaviť situáciu, ktorá by ma *presvedčila*, že $2 + 2 = 3$.

Predstavme si, že by som jedného rána vstal, vybral si dva štuple z uší a položil ich vedľa ďalších dvoch na mojom nočnom stolíku, a všimol by si, že teraz sú tam tri štuple, pričom žiaden štupeľ sa nezjavil ani nezmizol – v rozpore s mojou uloženou spomienkou, že $2 + 2$ sa má rovnať 4. Navyše, keby som si predstavil celý ten proces vo svojej mysli, vyzeralo by to, že aby z XX a XX vzniklo XXXX, muselo by tam nejaké X pribudnúť odnikiaľ, a navyše by to nesedelo so zvyšnou matematikou, ktorú by som si predstavoval, lebo odobrať XX z XXX by dávalo XX, ale odobrať XX z XXXX by dávalo XXX. To by bolo v rozpore s mojou uloženou spomienkou, že $3 - 2 = 1$, ale tá spomienka by bola absurdná zoči-voči fyzikálnemu a mentálnemu potvrdeniu, že $XXX - XX = XX$.

Ešte by som si to skontroloval na vreckovej kalkulačke, Googli, a povedzme mojej kópii knihy 1984, kde Winston píše, že “Sloboda znamená slobodu povedať, že dva plus dva sa rovná tri.” Toto všetko by mi prirodzene ukázalo, že zvyšok sveta súhlasí s mojimi terajšími predstavami a nesúhlasí s mojou spomienkou, že $2 + 2 = 3$.

Ako som mohol byť taký pomýlený, že som veril, že $2 + 2 = 4$? Napadajú mi dve vysvetlenia: Po prvé, neurologická porucha (možno spôsobená kýchnutím) spôsobila, že sa všetky moje spomienky na sčítanie posunuli o 1 nahor. Po druhé, niekto sa so mnou zahráva, či už pomocou hypnózy alebo počítačovej simulácie. V tom druhom prípade by som si myslel, že je pravdepodobnejšie, že sa dotýčný pohral s mojimi matematickými *spomienkami*, než že by sa $2 + 2$ naozaj rovnalo 4. Žiadne z týchto prijateľne znejúcich vysvetlení by mi nezabránilo všimnúť si, že som veľmi, veľmi, *veľmi* zmätený³.

Čo by ma presvedčilo, že $2 + 2 = 3$ je inými slovami celkom rovnaký druh indície ako ten, čo ma momentálne presvedča, že $2 + 2 = 4$: Krížová paľba indícií z fyzického pozorovania, mentálnej vizualizácie a spoločenského súhlasu.

¹ pozri: 20 Čo je indícia?

² http://lesswrong.com/lw/jl/what_is_evidence/f7h

³ pozri: 27 Tvoja sila racionalistu

Boli časy, keď som nemal poňatia o tom, že $2 + 2 = 4$. Nedošiel som k tomuto *novému* názoru náhodnými procesmi – potom by neexistoval konkrétny dôvod, prečo by si môj mozog uložil “ $2 + 2 = 4$ ” namiesto “ $2 + 2 = 7$ ”. Fakt, že si môj mozog uložil odpoveď, ktorá sa prekvapivo podobá na to, čo sa stane, keď položíš dva štuple do uší vedľa dvoch štuple, si žiada vysvetlenie, aké previazanie vytvára toto zvláštne zrkadlenie mysle a skutočnosti.

Naozaj sú iba dve možnosti ohľadom názoru o fakte⁴ – buď sa tam ten názor dostal pomocou procesu, ktorý zväzuje myseľ so skutočnosťou⁵, alebo nie. Ak nie, potom ten názor nemôže byť správny; nanajvýš ak zhodou okolností. Ak má ten názor trochu vnútornej zložitosti⁶ (vyžaduje si simulovanie počítačovým programom s dĺžkou viac než 10 bitov), priestor možností je dosť veľký na to, aby zhoda okolností zmizla.

Nepodmienené fakty nie sú to isté ako nepodmienené názory. Ak ma zviazaná indícia presvedčila, že fakt je nepodmienený, to neznamená, že som tomuto faktu vždy veril a nepotreboval som zviazanú indíciu.

Verím, že $2 + 2 = 4$ a pripadá mi pomerne ľahké predstaviť si situáciu, ktorá by ma presvedčila, že $2 + 2 = 3$. Konkrétne, ten istý druh situácie, aká má v súčasnosti presvedčať, že $2 + 2 = 4$. Nemusím sa teda báť, že som obeťou slepej viery.

Ak sú v obecnstve nejakí kresťania, ktorí poznajú *Bayesovu vetu* (žiadni numerofóbovia, prosím), môžem sa vás opýtať, aká situácia by vás presvedčila o pravdivosti islamu? Pravdepodobne by to bol ten istý druh situácie, ktorý je kauzálne zodpovedný za vašu terajšiu vieru v kresťanstvo: Keby vás vytiahli kričiacich z maternice moslimskej ženy, a keby vás vychovali rodičia, ktorí vám stále hovorili, že je dobré bezpodmienečne veriť v islam. Alebo je toho viac? Ak áno, aká situácia by vás presvedčila o islame, alebo prinajmenšom o nie-kresťanstve?

Pôvodný článok s komentármi⁷

⁴ pozri: 2 Rozumné čítanie

⁵ pozri: 20 Čo je indícia?

⁶ pozri: 23 Occamova britva

⁷ http://lesswrong.com/lw/jr/how_to_convince_me_that_2_2_3/

Kapitola 116. Nekonečná istota

V kapitole Absolútna autorita¹ som tvrdil, že *nepotrebuje* nekonečnú istotu. “Keby ste si museli vybrať medzi dvoma alternatívami A a B, a keby sa vám nejako podarilo získať určite správnu dobre kalibrovanú 100%-nú istotu, že A je absolútne a úplne žiadúce, a že B je suma všetkého zlého a odporného, bola by to *postačujúca* podmienka na výber A namiesto B. Nie je to *nevyhnutná* podmienka... Môžete mať neisté poznanie relatívne lepších a relatívne horších možností, a aj tak si vyberať. V skutočnosti by to mala byť rutina.”

Nemôžu však byť *nejaké* výroky, ktorým máme právo prideliť nekonečnú istotu? Čo taký výrok, že $2 + 2 = 4$ ²?

Musíme rozlišovať medzi mapou a územím³. Vzhľadom na napohľad absolútnu stabilitu a všeobecnosť fyzikálnych zákonov⁴ je možné, že nikdy, v celej histórii vesmíru, žiadna častica neprekročila lokálnu hranicu rýchlosti svetla. To znamená, že hranica rýchlosti svetla môže byť pravdivá nielen 99 % času, alebo 99,9999 % času, alebo $(1 - 1/10^{(10^{100})})$ času, ale jednoducho *vždy a absolútne pravdivá*.

Ale či my môžeme niekedy mať *absolútnu istotu* v hranicu rýchlosti svetla, to je úplne iná otázka. Mapa nie je územie.

Môže byť úplná pravda, že nejaký študent opísal svoju úlohu, ale či máte o tomto fakte nejaké vedomosti – tobôž *absolútnu istotu* v tento názor – je samostatná téma. Ak si hodíte mincou a nepozriete na ňu, môže byť úplná pravda, že na minci padla hlava, a vy si môžete byť úplne neistí, či na minci padla hlava alebo znak. Stupeň neistoty nie je to isté ako stupeň pravdy alebo frekvencia výskytu.

To isté platí pre matematické pravdy. Je otázka, či možno výrok “ $2 + 2 = 4$ ” alebo “V Peanovej aritmetike $SS0 + SS0 = SSSS0$ ” označiť za *pravdivý* v čisto abstraktnom zmysle, bez ohľadu na fyzické systémy, ktoré vyzerajú, že sa správajú v súlade s Peanovými axiómami. Ale keď už som toto povedal, pôjdem na vec a tipnem si, že v ľubovoľnom zmysle, v ktorom je “ $2 + 2 = 4$ ” pravda, je to vždy a presná pravda, nie iba približne pravda (“ $2 + 2$ sa naozaj rovná 4,000 000 4”) ani pravda v 999 999 999 999 prípadoch z 1 000 000 000 000.

Nie som si celom istým, čo v tomto prípade “pravda” znamená, ale stojím za svojím odhadom. Dôveryhodnosť výroku “ $2 + 2 = 4$ je vždy pravda” je omnoho väčšia než dôveryhodnosť ľubovoľného filozofického názoru na to, čo presne znamená “pravda”, “vždy” a “je” v uvedenom výroku.

To však neznamená, že mám *absolútnu istotu*, že $2 + 2 = 4$. Pozrite predchádzajúcu diskusiu o tom, ako ma presvedčiť, že $2 + 2 = 3$ ⁵, čo možno dosiahnuť použitím rovnakých indícií ako tie, čo ma pôvodne presvedčili, že $2 + 2 = 4$. Možno všetky moje doterajšie skúsenosti boli halucinácie alebo si ich nesprávne pamätám. Neurológovia poznajú aj čudnejšie poruchy mozgu než toto.

Ak teda pripíšeme nejakú pravdepodobnosť výroku “ $2 + 2 = 4$ ”, aká by to mala byť pravdepodobnosť? V prípadoch ako je tento hľadáme dobrú kalibráciu – aby sa vety, ktorým pripíšete “pravdepodobnosť 99 %” ukázali ako pravdivé v 99 prípadoch zo 100. Toto je v skutočnosti omnoho zložitejšie než si myslíte. Vezmite si tisíc ľudí a požiadajte každého z nich, aby povedal desať výrokov, ktorými si je “istý na 99 %”. Myslíte si, že z 1000 výrokov bude nesprávnych v priemere 10?

¹ pozri: 114 Absolútna autorita

² pozri: 115 Ako ma presvedčiť, že $2 + 2 = 3$

³ pozri: 46 Jednoduchá pravda

⁴ pozri: 180 Všeobecný zákon

⁵ pozri: 115 Ako ma presvedčiť, že $2 + 2 = 3$

Nebudem teraz diskutovať skutočné pokusy, ktoré sa robili s kalibráciou – nájdete ich v kapitole inej mojej knihy⁶ – pretože som videl, že keď to len tak vyhrknem na nepripravených ľudí, začnú to používať ako Úplne Všeobecný Protiargument⁷, ktorý im akosi naskočí do hlavy vždy, keď potrebujú znevážiť istotu niekoho, koho názor sa im nepáči, ale zabudnú si na to spomenúť, keď uvažujú o svojich vlastných názoroch. Pokúšam sa teda nehovoriť o pokusoch s kalibráciou, jedine ako súčasť štrukturovanej prednášky o rozumnosti, ktorá zahŕňa varovanie pred motivovaným skepticizmom.

Lenže pozorovaná kalibrácia ľudí, ktorí hovoria, že sú si “na 99 % istí” nie je presnosť 99 %.

Predstavte si, že povieť, že ste si na 99,99 % istí, že $2 + 2 = 4$. Tým tvrdíte, že by ste mohli urobiť 10 000 *nezávislých* výrokov, v ktoré máte rovnakú dôveru, a mýlili by ste sa v priemere jedenkrát. Možno pre $2 + 2 = 4$ je tento mimoriadny stupeň istoty možný: “ $2 + 2 = 4$ ” je extrémne jednoduché, matematicky aj empiricky, a všeobecne spoločensky prijímané (nie s vášnivým potvrdením, ale ticho považované za samozrejmosť). Možno tu naozaj môžete mať istotu 99,99 %.

Nemyslím si, že môžete mať istotu 99,99 % pri výrokoch ako je “53 je prvočíslo”. Áno, vyzerá to pravdepodobne, ale dokiaľ by ste si vytvorili protokoly, ktoré vám umožnia tvrdiť 10 000 *nezávislých* výrokov tohto druhu – čiže nie iba sadu výrokov o prvočíslach, ale zakaždým nový protokol – pomýlili by ste sa viac než iba raz. Peter de Blanc má na túto tému zábavnú anekdotu, ktorú vám môže povedať v komentároch.

Lenže mapa nie je územie: ak poviem, že som si na 99 % istý, že $2 + 2 = 4$, neznamená to, že si myslím, že “ $2 + 2 = 4$ ” je pravda s presnosťou 99 %, alebo že “ $2 + 2 = 4$ ” je pravda v 99 prípadoch zo 100. Výrok, ktorému pripisujem takúto istotu je výrok, že “ $2 + 2 = 4$ vždy a presne”, nie výrok “ $2 + 2 = 4$ približne a väčšinou”.

A čo sa týka predstavy, že by ste mohli mať istotu až 100 % o matematickom výroku – no, teraz vážne! Keby ste povedali istotu 99,999 9 %, tvrdili by ste, že by ste mohli urobiť *milión* rovnako dôveryhodných výrokov, jeden za druhým, a mýlili by ste sa v priemere jedenkrát. To by zabralo asi jeden celý rok rozprávania, keby ste povedali jeden výrok každých 20 sekúnd a rozprávali by ste 16 hodín denne.

Ak tvrdíte istotu 99,999 999 999 9 %, dvíhate to na bilión. Bude teraz rozprávať sto ľudských životov a nepomýlite sa pritom ani raz?

Ak tvrdíte istotu $(1 - 1/10^{(10^{100})})$, vaše ego výrazne prevyšuje ego duševne chorých pacientov, ktorí sa považujú za bohov.

Pritom $1/10^{(10^{100})}$ je omnoho menšie než pomerne malé nepredstaviteľne veľké čísla ako je $3^{3^{3^3}}$.

Ale ani istota $(1 - 1/3^{3^{3^3}})$ nie je až tak bližšia k **PRAVDEPODOBNOSTI 1** ako istota 90 %.

Keby všetko ostatné zlyhalo, hypotetickí Temní páni Matrixu, ktorí *práve teraz* zasahujú do toho, ako váš mozog pripisuje dôveryhodnosť *samotnej tejto vete*, nám zahatajú cestu a ušetria nás pliahy nekonečnej istoty.

Som si tým absolútne istý?

⁶ <http://intelligence.org/files/CognitiveBiases.pdf>

⁷ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

⁸ http://lesswrong.com/lw/kn/torture_vs_dust_specks/

Samozrejme, že nie.

Ako raz povedal Rafal Smigrodski:

Povedal by som, že by sme mali dokázať priradiť matematickým pojmom nevyhnutným na odpovedenie Bayesovho pravidla istotu menšiu než 1, a aj tak ho prakticky používať. Nie som si celkom istý, že si vždy musím byť neistý. Možno si môžem byť o niektorých veciach legitímne istý. Ale keď raz nejakému výroku priradím pravdepodobnosť 1, už to neviem nikdy zrušiť. Bez ohľadu na to, čo uvidím alebo čo sa naučím, budem musieť odmietnuť všetko, čo odporuje tejto axióme. Nepáči sa mi predstava, že by som už nikdy nedokázal zmeniť názor.

Pôvodný článok s komentármi⁹

⁹ http://lesswrong.com/lw/mo/infinite_certainty/

Kapitola 117. 0 a 1 nie sú pravdepodobnosti

1, 2 a 3 sú celé čísla, takisto aj -4. Ak budete ďalej počítat nahor alebo počítat nadol, isto stretnete mnoho ďalších celých čísel. Nestretnete však nič, čo by sa volalo “kladné nekonečno” alebo “záporné nekonečno”, takže tie nie sú celé čísla.

Kladné a záporné nekonečno nie sú celé čísla, ale skôr špeciálne symboly na rozprávanie o správaní celých čísel. Ľudia niekedy hovoria veci ako “5 + nekonečno = nekonečno”, pretože ak začnete pri 5 a budete počítat smerom nahor bez zastavenia, budete dostávať väčšie a väčšie čísla bez ohraničenia. Ale z toho nevyplýva, že “nekonečno - nekonečno = 5”. Nemôžete začať počítat od 0 nahor bez zastavenia, a potom počítat nadol bez zastavenia, a nakoniec zistiť, že ste na čísle 5.

Z tohto vidíme, že nekonečno nielenže nie je celé číslo, ale ani sa *nespráva* ako celé číslo. Ak sa nemúdro pokúsite zmiešať nekonečná s celými číslami, budete potrebovať všakovaké špeciálne nové napohľad nekonzistentné správania, ktoré ste nepotrebovali pre 1, 2, 3 a ďalšie *skutočné* celé čísla.

Aj keď nekonečno nie je celé číslo, nemusíte sa báť toho, že budete mať nedostatok čísel. Ľudia už videli päť oviec, milión zrníek piesku, a septilióny atómov, nikto však nikdy nespočítal nekonečne veľa niečoho. To isté platí pre spojitú množstvá - ľudia odmerali milimetre veľké zrnká prachu, metre veľké zvieratá, kilometre veľké mestá, tisíce svetelných rokov veľké galaxie, ale nikto ešte nenameral niečo nekonečne veľké. V skutočnom svete nekonečná veľmi *nepotrebuje*.

(Poznámka pre sofistikovanejších čitateľov: Nemusíte mi písať zložité vysvetlenia rozdielov medzi povedzme ordinálnymi a kardinálnymi číslami. Áno, poznám rôzne pokročilé definície nekonečna z teórie množín, ale nevidím pre ne dobré použitie v teórii pravdepodobnosti. Vid' nižšie.)

Pri zvyčajnom spôsobe zápisu pravdepodobností sú pravdepodobnosti medzi 0 a 1. Minca môže mať pravdepodobnosť 0,5, že na nej padne znak, alebo hlásateľ priradí pravdepodobnosť 0,9, že bude zajtra pršať.

To však nie je jediný možný zápis pravdepodobností. Môžete napríklad transformovať pravdepodobnosti na šance pomocou transformácie $\tilde{S} = (P / (1 - P))$. Pravdepodobnosť 50 % by teda bola šancou 0,5 / 0,5 alebo 1, čo zvyčajne zapisujeme 1 : 1, zatiaľ čo pravdepodobnosť 0,9 by bola šancou 0,9 / 0,1, čo zvyčajne zapisujeme 9 : 1. Aby ste zo šancí urobili naspäť pravdepodobnosti, použijete $P = (\tilde{S} / (1 + \tilde{S}))$ a je to dokonale reverzibilné, takže táto transformácia je izomorfizmus - obojstranné reverzibilné mapovanie. Pravdepodobnosti a šance sú teda izomorfné, a môžete používať jedno alebo druhé podľa toho, čo je pohodlnejšie.

Je napríklad pohodlnejšie používať šance, keď robíte bayesovské aktualizácie. Povedzme, že hodím šesťstennou kockou: Ak padne iné číslo ako 1, je šanca 10 %, že začujem zvonenie, ale ak padne 1, je šanca 20 %, že začujem zvonenie. Hodil som kocku a počul som zvonenie. Aká je *šanca*, že padlo číslo 1? Nuž, počiatočná šanca je 1 : 5 (čo zodpovedá reálnemu číslu $1 / 5 = 0,20$) a pomer podmienených pravdepodobností je 0,2 : 0,1 (čo zodpovedá reálnemu číslu 2) a tieto môžem jednoducho navzájom vynásobiť a dostanem výslednú šancu 2 : 5 (čo zodpovedá reálnemu číslu $2 / 5$ čiže 0,40). Potom to premením späť na pravdepodobnosti, ak chcem, a dostanem $(0,4 / 1,4) = 2 / 7 = \sim 29\%$.

So šancami sa teda pri bayesovských aktualizáciách lepšie pracuje – ak používate pravdepodobnosti, musíte používať Bayesovu vetu¹ v jej zložitejšom tvare. Ale pravdepodobnosti sa lepšie hodia na odpovedanie na otázky ako: “Ak hodím šesťstennou kocou, aká je šanca, že uvidím číslo od 1 do 4?” Môžete sčítať pravdepodobnosti $1/6$ pre každú stranu a dostať $4/6$, nemôžete však sčítať šance $0,2$ pre každú stranu a dostať šancu $0,8$.

Prečo toto všetko hovorím? Aby som ukázal, že “pomery šancí” sú rovnako legitímny spôsob mapovania neistoty na reálne čísla ako “pravdepodobnosti”. Pri niektorých operáciách sú pohodlnejšie pomery šancí, pri iných pravdepodobnosti. Slávny dôkaz nazývaný Coxova veta (plus jej rôzne rozšírenia a upresnenia) ukazuje, že všetky spôsoby reprezentovania neistoty, ktoré dodržiavajú určité rozumne znejúce obmedzenia, sú v konečnom dôsledku izomorfne.

Prečo záleží na tom, či sú pomery šancí rovnako legitímne ako pravdepodobnosti? Pravdepodobnosti sa štandardne zapisujú medzi 0 a 1, pričom aj 0 aj 1 vyzerajú, že by to mali byť pomerne dosiahnuteľné hodnoty – je ľahké vidieť 1 zebra alebo 0 jednorožcov. Keď si však transformujete pravdepodobnosti na pomery šancí, 0 zostane 0, ale 1 ide do kladného nekonečna. Teraz absolútna pravda nevyzerá až tak ľahko dosiahnuteľná.

Reprezentácia, v ktorej sa bayesovské aktualizácie robia ešte ľahšie, je logaritmus šance – takto E. T. Jaynes odporúchal myslieť na pravdepodobnosti. Povedzme napríklad, že pôvodná pravdepodobnosť nejakého tvrdenia je 0,0001 – to zodpovedá logaritmu šance okolo -40 decibelov. Potom vidíte indíciu, ktoré je 100-krát pravdepodobnejšia, ak je tvrdenie pravdivé, než ak je nepravdivé. To je 20 decibelov indície. Výsledný pravdepodobnosť je teda zhruba $-40 \text{ db} + 20 \text{ db} = -20 \text{ db}$, čo znamená, že výsledná pravdepodobnosť je $\sim 0,01$.

Keď transformujete pravdepodobnosti na logaritmy šancí, z 0 sa stane záporné nekonečno a z 1 sa stane kladné nekonečno. Teraz aj nekonečná istota aj nekonečná nepravdepodobnosť vyzerajú dosť nedosiahnuteľne.

Pri pravdepodobnostiach sa zdá, že medzi 0,999 9 a 0,999 99 je rozdiel iba 0,000 09, takže 0,502 je omnoho ďalej od 0,503 než je 0,999 9 od 0,999 99. Aby sme z pravdepodobnosti 0,999 99 dostali pravdepodobnosť 1, vyzerá to, že treba prekonať iba vzdialenosť 0,000 01.

Ale keď to transformujeme na pomery šancí, z 0,502 a 0,503 sa stane 1,008 a 1,012, a z 0,999 9 a 0,999 99 sa stane 9,999 a 99,999. A keď to transformujeme na logaritmy šancí, z 0,502 a 0,503 sa stane 0,03 decibelov a 0,05 decibelov, ale z 0,999 9 a 0,999 99 sa stane 40 decibelov a 50 decibelov.

Keď pracujete s logaritmami šancí, **vzdialenosť medzi ľubovoľnými dvoma stupňami neistoty sa rovná množstvu indície, ktoré by ste potrebovali, aby ste sa dostali z jedného na druhý.** To znamená, že logaritmy šancí nám dávajú prirodzenú mieru vzdialenosti medzi stupňami istoty.

Použitie logaritmov šancí odhaľuje fakt, že dosiahnutie nekonečnej istoty² si vyžaduje nekonečne silnú indíciu, rovnako ako si nekonečná absurdita vyžaduje nekonečne silnú protiindíciu.

Ďalej, všetky druhy štandardných viet o pravdepodobnosti majú špeciálne prípady, ak sa do nich pokúsite vložiť 1-y alebo 0-y – ako sa napríklad stane, ak skúšate robiť bayesovskú aktualizáciu po pozorovaní, ktorému ste priradili pravdepodobnosť 0.

¹ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

² pozri: 116 Nekonečná istota

Navrhujem teda, že by dávalo zmysel povedať, že 1 a 0 nie sú pravdepodobnosti; rovnako ako záporné a kladné nekonečno, ktoré nespĺňajú axiómy poľa, nie sú medzi reálnymi číslami.

Hlavný dôvod, prečo by toto mohlo znepokojiť teoretikov pravdepodobnosti je, že by sme museli znovu odvodiť vety, ktoré sme predtým získali z predpokladu, že môžeme zjednotenie pravdepodobností vybaviť tým, že sčítame všetky kúsky a povieme, že ich súčet je 1.

Lenže v skutočnom svete, keď hodíte kocku, nemáte doslova nekonečnú istotu³, že padne nejaké číslo od 1 do 6. Kocka môže zastať na hrane; môže na ňu padnúť meteor; alebo zasiahnu Temní páni Matrixu a napíšu "37" na jednu stranu.

Keby ste si vyrobili čarovný symbol, ktorý by znamenal "všetky pravdepodobnosti, ktoré som nezvážil", mohli by ste sčítať všetky udalosti vrátane tohto čarovného symbolu a získali by ste čarovný symbol "T", ktorý by znamenal nekonečnú istotu.

Ale ja by som sa radšej opýtal, či existuje nejaký spôsob, ako odvodiť vety bez použitia čarovných symbolov so špeciálnym správaním. To by bolo elegantnejšie. Rovnako ako existujú matematici, ktorí odmietajú uveriť v dvojité negácie alebo nekonečné množiny, ja by som rád bol teoretikom pravdepodobnosti, ktorý neverí v nekonečnú istotu.

PS: Tu je anekdota o "matematickej istote"⁴ od Petra de Blanca. (Povedal som mu, aby to viac nerobil.)

Pôvodný článok s komentármi⁵

³ pozri: 116 Nekonečná istota

⁴ <http://www.spaceandgames.com/?p=27>

⁵ http://lesswrong.com/lw/mp/0_and_1_are_not_probabilities/

Časť XII. Zanechávanie

Obsah

118 Dôležitosť hovorenia: "joj!"	180
119 Ponuka pomätenosti	182
120 Už to konečne vzdaj	184
121 Správne použitie pochybnosti	185
122 Dokážete čeliť skutočnosti	187
123 Meditácia o zvedavosti	188
124 Nikto vám nemôže udeliť výnimku zo zákonov rozumnosti	191
125 Ponechajte ústupovú líniu	193
126 Kríza viery	196
127 Rituál	202

Kapitola 118. Dôležitosť hovorenia: “joj!”

Práve som dočítal príbeh o páde Enronu, *Najbystrejší chlapi v miestnosti*, ktorý týmto vyhráva moje ocenenie za “Najneprimeranejší názov knihy”.

Neprekvapujúcou črtou pomalého rozkladu a náhleho zrútenia Enronu bolo, že výkonní hráči si nikdy nepriznali, že urobili *veľkú* chybu. Keď katastrofa číslo 247 narástla do takých rozmerov, že si vyžadovala skutočnú zmenu pravidiel, povedali si: “Škoda, že to nevyšlo – bol to taký dobrý nápad – ako teraz skryjeme tento problém v našej súvahe?” Namiesto: “Zo spätného pohľadu sa teraz zdá jasné, že to od samého začiatku bola chyba.” Namiesto: “Bol som hlúpy.” Nikdy nenastala chvíľa prelievať slzy, chvíľa na ponížujúce uvedomenie, na priznanie si *základného* problému. Po bankrote, Jeff Skilling, bývalý prevádzkový riaditeľ a nakrátko výkonný riaditeľ Enronu, odmietol radu svojich právnikov využiť piaty dodatok ústavy; svedčil pred kongresom, že Enron bola *výborná* spoločnosť.

Nie každá zmena je zlepšenie, ale každé zlepšenie je nevyhnutne zmena. Ak si priznávame iba malé lokálne chyby, budeme robiť iba malé lokálne zmeny. Motivácia na *veľkú* zmenu vychádza z uvedomenia si *veľkej* chyby.

Ako dieťa som bol vychovaný rovnakým dielom vedou a vedeckou fantastikou, a od Heinleina a Feynmana som sa naučil postupy tradičnej rozumnosti: Teórie musia byť odvážne a vystavovať sa falzifikácii; buď ochotný podstúpiť hrdinskú obeť vzdania sa svojich názorov, keď narazíš na dôkazy o opaku; argumentuj poctivo; snaž sa neklamať sám seba; a ďalšie nejasné slovné spojenia.

Výchova k tradičnej rozumnosti sa pokúša vytvoriť diskutérov, ktorí *jedného dňa* pripustia dôkaz o opaku – mala by existovať *nejaká* hora dôkazov dosť veľká na to, aby vami pohla. To nie je maličkosť; týmto sa odlišuje veda od náboženstva. Ale nie je tu veľký dôkaz na *rýchlosť*; na vzdanie boja *tak rýchlo, ako sa len dá*; na *efektívne* spájanie indícií, aby *minimum* indícií proti stačilo na zničenie vášho obľúbeného názoru.

Bol som vychovaný v tradičnej rozumnosti a považoval som sa za dosť rozumného. Na bayesiánske remeslo (Laplace / Jaynes / Tversky / Kahneman) som sa prepol až v dôsledku... no, to by bol dlhý príbeh. Približne povedané, prepol som sa, keď som si uvedomil, že nejasné slovné spojenia tradičnej rozumnosti mi nestačili zabrániť v urobení veľkej chyby.

Po tom, čo som si konečne a naplno uvedomil svoju chybu, obzrel som sa na cestu, ktorá ma priviedla k môjmu strašnému uvedomeniu. A videl som, že som robil série malých ústupkov, minimálnych priznaní, neochotných ústupkov o každý milimeter územia, pri každej príležitosti si uvedomujúc najmenšie možné množstvo vlastných chýb, priznávajúc porážku iba v malých znesiteľných sústach. Mohol som sa hýbať omnoho rýchlejšie, uvedomil som si, keby som jednoducho zvrieskol: “JOJ!”

A tak som si pomyslel: *Musím sa túto hru naučiť hrať lepšie.*

V priznaní si *veľkej* chyby sa skrýva jedna *mocná výhoda*. Bolí to. Môže to aj zmeniť celý váš život.

Je *dôležité* mať chvíľu na prelievanie slz, chvíľu na ponížujúce uvedomenie. Priznať si *základný* problém, nerozdeľovať ho na stráviteľné sústa drobných chýb.

Neoddávajte sa dráme a nestaňte sa hrdými na priznávanie chýb¹. Iste je lepšie urobiť to správne na prvýkrát. Ale ak už urobíte chybu, potom je najlepšie uvidieť ju celú naraz. Ešte aj z hľadiska príjemnosti, lepšie je prijať jednu veľkú stratu než veľa malých strát. Alternatíva je natahovať súboj so sebou samým po celé roky. Alternatíva je Enron.

Odvtedy som videl druhých, ako robia svoje vlastné série minimálnych ústupkov, neochotne ustupujúc o každý milimeter územia; nikdy si nepriznávajú globálnu chybu, ak stačí lokálna; vždy sa učiac z každej chyby tak málo, ako sa len dá. Čo by mohli napraviť jedným šmahom, keby sa im chcelo, premenia na maličké lokálne záplaty, do ktorých ich treba dotlačiť. Nikdy nepovedia po priznaní si chyby: *bol som blázon*. Robia, čo je v ich silách, aby minimalizovali svoju hanbu slovami: *v princípe som mal pravdu*, alebo *mohlo to fungovať*, alebo *stále sa chcem držať pravej podstaty toho-k-čomu-som-pripútaný*. V danej chvíli bránia svoju hrdosť, čím si zabezpečujú, že danú chybu opäť zopakujú, a opäť budú musieť brániť svoju hrdosť.

To radšej prehltnúť celú horkú pilulku na jeden strašný glg.

Pôvodný článok s komentármi²

¹ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

² http://lesswrong.com/lw/i9/the_importance_of_saying_oops/

Kapitola 119. Ponuka pomätenosti

Keď som bol veľmi mladý – tuším trinásť- alebo štrnásťročný – myslel som si, že som našiel vyvrátenie Kantorovho diagonálneho argumentu, známej vety, ktorá ukazuje, že reálnych čísel je viac než racionálnych. Ach, tie predstavy slávy a úspechu, ktoré mi tancovali v hlave!

Moja myšlienka bola, že ak môžeme každé celé číslo rozložiť na súčet mocnín dvojky, môžeme namapovať prirodzené čísla na množinu podmnožín prirodzených čísel jednoducho tým, že vypíšeme ich dvojkový zápis. Napríklad 13, v dvojkovej sústave 1101, namapujeme na $\{0, 2, 3\}$. Trvalo celý týždeň než mi napadlo, že by som mohol skúsiť *aplikovať* Kantorov diagonálny argument na moju chytrú konštrukciu, a samozrejme sa našiel protipríklad – dvojkové číslo ...1111, ktoré nezodpovedá žiadnemu konečnému celému číslu.

Takže som našiel tento protipríklad a videl, že môj pokus o vyvrátenie bol falošný, rovnako ako moje sny o sláve a úspechu.

Spočiatku som bol trochu sklamaný.

Mysľou mi prešla myšlienka: “Raz tú vetu dostanem! *Jedného dňa* Kantorov diagonálny argument vyvrátim, aj keď mi to na prvýkrát nevyšlo!” Neznášal som tú vetu, že je tak tvrdohlavo pravdivá, že ma obrala o moju slávu a úspech, a začal som hľadať iné vyvrátenie.

A potom som si čosi uvedomil. Uvedomil som si, že som urobil chybu a že teraz, keď som si svoju chybu všimol, nie je absolútne žiaden dôvod podozrievať silu Kantorovho diagonálneho argumentu, o nič viac než ľubovoľnej inej významnej matematickej vety.

Vtedy som uvidel veľmi jasne, že sa mi ponúkala príležitosť stať sa matematickým pomätencom a stráviť zvyšok svojho života písaním zeleným atramentom zlostných listov pre matematických profesorov. (Raz som čítal knihu o matematických pomätencoch.)

Nechcem som, aby toto bola moja budúcnosť, tak som sa krátko zasmial a nechal to tak. Zamával som Kantorovmu diagonálnemu argumentu na rozlúčku so želaním všetkého dobrého a už som ho viac nespochybňoval.

A teraz si nepamätám, či som si pomyslel už vtedy alebo či som si pomyslel až dodatočne... aký to bol hrozne neférový test pre trinásťročné dieťa. Že som musel už v tom veku byť rozumný alebo zlyhať.

Čím chytrejší ste, tým mladší môžete byť, keď po prvýkrát dostanete niečo, čo vyzerá ako naozaj revolučná myšlienka. Mal som šťastie, že som si tú chybu našiel sám; že mi ju nemusel ukázať nejaký iný matematik, a že mi tým nedal vonkajší predmet obviňovania. Možno by som sa aj tak jedného dňa zotavil. Zotavil som sa aj z horších vecí, ako dospelý. Ale keby som sa bol pomýlil tak skoro, vyvinul by som si vôbec túto schopnosť?

Rozmýšľam, koľkí ľudia píšuci zlostné listy zeleným atramentom mali trinásť rokov, keď urobili svoj prvý osudový prešľap. Rozmýšľam, koľkí z nich boli predtým nádejnými myšliami.

Urobil som chybu. To bolo všetko. Nemal som v *skutočnosti*, v *hlbšom zmysle pravdu*; nedosiahol som morálne víťazstvo; nepreukázal som ambíciu ani

skepticizmus ani žiadnu inú úžasnú cnosť; nebola to rozumná chyba; nemal som ani spoločne pravdu ani najmenší zlomok pravdy. Myslel som si myšlienku, ktorú by som si nebol myslel, keby som bol múdrejší, a to je všetko, čo sa vtedy stalo.

Keby som si toto nebol schopný priznať, keby som bol reinterpretoval svoju chybu ako cnosť, keby som bol trval na tom, že som mal aspoň *čiastočnú* pravdu, aby som si zachoval hrdosť, potom by som sa nebol nevzdal. Naďalej by som hľadal chybu v diagonálnom argumente. A skôr či neskôr by som možno nejakú našiel.

Dokiaľ si nepriznáte, že ste sa mýlili¹, nemôžete pokračovať vo svojom živote; váš sebaobraz bude stále pripútaný k tej starej chybe.

Kedykoľvek cítite pokušenie držať sa myšlienky, ktorú by ste si nikdy neboli pomysleli, keby ste boli múdrejší, ponúka sa vám príležitosť stať sa pomätencom – dokonca aj keď nikdy nenapíšete žiaden zlostný list zeleným atramentom. Ak sa s vami nikto neunúva hádať, alebo ak nikdy nikomu svoju myšlienku nepoviete, stále môžete byť pomätencom. *Lipnutie* je tá definujúca časť.

Nie je to pravda. Nie je to pravda v hlbšom zmysle. Nie je to ani polopravda, ani trochu pravda. Nie je to nič len myšlienka, ktorú ste si nikdy nemali pomyslieť. Nie na konci každého tunela je svetlo. Ľudia robia chyby a nie každá z nich je maskovaný úspech. Ľudia robia chyby; to sa stáva, a to je celé. Povedzte "joj"² a pokračujte vo svojom živote.

Pôvodný článok s komentármi³

¹ pozri: 118 Dôležitosť hovorenia: "joj!"

² pozri: 118 Dôležitosť hovorenia: "joj!"

³ http://lesswrong.com/lw/j8/the_crackpot_offer/

Kapitola 120. Už to konečne vzdaj

Casey Serin, 24-ročný programátor webových stránok s nulovými skúsenosťami v obchode s nehnuteľnosťami dlží bankám 2,2 milióna dolárov po tom, čo klamal na žiadostiach o hypotéku, aby si mohol naraz kúpiť 8 rôznych domov v rôznych štátoch. Časť peňazí, ktoré získal ako hypotéku, vybral v hotovosti (žiadal o väčšie sumy než domy naozaj stáli) a minul ich na svoje životné náklady a semináre o obchodovaní s nehnuteľnosťami. Zdá sa, že očakával rast cien na trhu.

To ešte nie je tá smutná časť. Smutná časť je, že *to ešte stále nevzdal*. Casey Serin porážku nepripúšťa. Odmieťa vyhlásiť bankrot, alebo si nájsť prácu; stále si myslí, že dokáže v obchode s nehnuteľnosťami preraziť. Naďalej míňa peniaze na semináre. Pokúsil sa získať hypotéku na deviaty dom. Ako vidíte, nebol to *neúspech*, bola to iba *príležitosť niečo sa naučiť*. [Poznámka prekladateľa: Medzičasom už definitívne skrachoval a prišiel aj o webovú stránku.]

Toto sa stáva, keď sa odmietnete vzdať.

Zatiaľ čo toho správanie môže vyzeráť ako obyčajná hlúposť, pripomína mi to dvoch ekonómov, nositeľov Nobelovej ceny...

...menovite Mertona a Scholesa z firmy Long-Term Capital Management¹.

Počas prvých troch rokov svojej existencie mohlo LTCM prehadzovať zisky lopatou, v roku 1998 sa však neefektivity, z ktorých LTCM ťažilo, začali strácať – viacerí ľudia poznali daný trik, preto prestal fungovať.

LTCM sa odmietlo vzdať. Navyknutí na ročné výnosy 40 %, požíčali si na väčšie a väčšie páky, aby vyžmýkali drobnejšie a drobnejšie rozdiely. Keď začalo ísť pre LTCM všetko dole kopcom, mali kapitál 4,72 miliardy dolárov, páku 124,5 miliardy dolárov, a derivátové pozície za 1,25 bilióna.

Každé povolanie má iný spôsob, ako byť chytrý – treba sa naučiť iné zručnosti a riadiť sa inými pravidlami. Mohli by ste si preto myslieť, že štúdium “rozumnosti” ako všeobecnej disciplíny nebude príliš prispievať k úspechu v skutočnom živote. Mne sa však aj tak zdá, že *ako nebyť hlúpy*, v tom majú viaceré povolania veľa spoločného. Ak sa pustíte do vyučovania *ako nenechať malé chyby prerásť do veľkých*, je to takmer to isté umenie, či už v hedžových fondoch alebo v láske, a jedna dôležitá vec je: Buď pripravený pripustiť, že si prehral.

Pôvodný článok s komentármi²

¹ http://en.wikipedia.org/wiki/Long-Term_Capital_Management

² http://lesswrong.com/lw/gx/just_lose_hope_already/

Kapitola 121. Správne použitie pochybnosti

Raz, keď som rečnil o Ceste¹, poznamenal som, že väčšina organizovaných systémov viery existuje na *únik pred pochybnosťou*. Jeden poslucháč zareagoval, že na jezuitov sa táto kritika nevzťahuje, pretože oni cvičia organizovanú pochybnosť: ich novicom sa vraj hovorí, aby pochybovali o kresťanstve, pochybovali o existencii Boha, pochybovali o pravosti svojho povolania, pochybovali o svojej vhodnosti na večný sľub čistoty a chudoby. Povedal som: *Aha, ale predpokladá sa, že tieto pochybnosti prekonajú, nie?* On povedal: *Nie, majú pochybovať asi preto, aby ich pochybnosti mohli rásť a silnieť.*

Google mi tieto tvrdenia ani nepotvrdil ani nevyvrátil. (Ak mi niekto z vás vie pomôcť, bol by som mu zaviazaný.) Ale mne takýto scenár pripadá fascinujúci, hodný diskusie, bez ohľadu na to, či to je alebo nie je pravda o jezuitoch. *Keby* jezuiti cvičili úmyselné pochybovanie, ako je opísané hore, boli by *preto* cnostnými racionalistami? [Prekladateľ: V diskusii sa táto otázka celkom nevyriešila, ale zdá sa, že pochybovať o všetkom nie je presná formulácia. Od novicov sa vyžadujú dva roky podrobného sebaskúmania, aby vstúpili iba tí, ktorí to myslia vážne.]

Myslím, že by som musel pripustiť, že jezuitov v horeuvedenom (údajnom) scenári nie je primerané opísať ako “unikajúcich pred pochybnosťou”. Ale aj toto (údajné) správanie sa mi zdá vysoko podozrivé. Pre naozaj cnostného racionalistu by pochybnosť nemala byť strašidelná. Horepopísané správanie mi znie ako program na desenzitizáciu na niečo *veľmi* strašidelné, ako keď arachnofóbia vystavujeme pavúkom za starostlivo kontrolovaných podmienok.

Ale aj tak, povzbudzujú svojich novicov, aby pochybovali – správne? Záleží na tom, či sú ich dôvody pomýlené? Nie je to napriek tomu konanie hodné racionalistu?

Zvedavosť sa snaží zničiť sama seba²; niet takej zvedavosti, ktorá by *nechcela* odpoveď. Ale ak získate odpoveď, ak uspokojíte svoju zvedavosť, potom to slávne tajomstvo už viac nebude tajomným.

Rovnako, každá pochybnosť existuje preto, aby zničila nejaký konkrétny názor. Ak sa pochybnosti nepodarí zničiť svoj cieľ, zomiera nenaplnená – ale aj to je nejaké riešenie, nejaké ukončenie, aj keď smutnejšie. Pochybnosť, ktorá nezničí ani sama seba ani svoj cieľ, akoby ani nikdy neexistovala. *Riešenie* pochybnosti, nie samotný akt pochybovania, je to, čo otáča ozubené koleso rozumnosti vpred.

Každé zlepšenie je zmena, ale nie každá zmena je zlepšenie. Každý racionalista pochybuje, ale nie každá pochybnosť je rozumná. Nosenie pochybnosti³ vás nerobí racionalistom o nič viac než nosenie bieleho laboratórneho pláňa lekárom.

Rozumná pochybnosť vzniká z konkrétneho dôvodu – máte nejaký konkrétny dôvod pochybovať, že podozrivý názor je pravdivý. Tento dôvod potom vyžaduje nejakú líniu vyšetrovania, ktorá buď zničí cieľový názor alebo zničí túto pochybnosť. To platí aj pre vysoko abstraktné pochybnosti ako: “Ktovie, či neexistuje jednoduchšia hypotéza, ktorá by tiež vysvetlila tieto údaje.” V tomto prípade vyšetrujete tak, že sa pokúšate vymyslieť jednoduchšie hypotézy. Ako

¹ pozri: 258 Dvanásť cností rozumnosti

² pozri: 258 Dvanásť cností rozumnosti

³ pozri: 17 Názor ako uniforma

toto hľadanie pokračuje dlhšie a dlhšie bez výsledkov, myslíte si, že je menej a menej pravdepodobné, že nasledujúci krok výpočtu bude tým úspešným. Nakoniec cena hľadania prekročí očakávaný úžitok, a vy prestanete hľadať. V tom bude už viac nemôžete tvrdiť, že *užitočne pochybujete*. Pochybnosť, ktorá nevyšetruje, akoby ani neexistovala. Každá pochybnosť existuje preto, aby zničila sama seba, jedným či druhým spôsobom. Nevyriešená pochybnosť je nulová operácia; neotáča kolesom, ani dopredu, ani dozadu.

Keby ste naozaj verili⁴ náboženstvu (a nie iba verili v⁵ náboženstvo), prečo by ste potom hovorili svojim novicom, aby zvažovali pochybnosti, ktoré musia zomrieť nenaplnené? Bolo by to ako hovoriť študentom fyziky, aby do úmoru pochybovali, že revolúcia v 20. storočí mohla byť chybou, a že newtonovská mechanika bola stále správna. Ak o niečom *naozaj* nepochybujete, prečo by ste *predstierali* že áno?

Pretože všetci chceme vyzeráť ako rozumní – a pochybovanie sa *všeobecne považuje* za cnosť racionalistu. Nie je však všeobecne jasné, že na pochybovanie potrebujete konkrétny dôvod, alebo že nevyriešená pochybnosť je nulová operácia. Namiesto toho si ľudia myslia, že je to o *skromnosti*; submisívne správanie, ktoré udržiava hierarchiu postavenia v tlupe – takmer rovnaký problém ako s pokorou, o čom som už písal⁶. Urobiť veľkú verejnú ukážku pochybnosti, aby ste presvedčili sami seba⁷, že ste rozumný, vám je užitočné asi rovnako ako nosiť laboratórny plášť.

Aby ste sa vyhli vyznávaniu⁸ pochybností, pamätajte, že:

- Rozumná pochybnosť existuje preto, aby zničila svoj cieľový názor, a ak svoj cieľový názor nezničí, zomiera nenaplnená.
- Rozumná pochybnosť vzniká z nejakého konkrétného dôvodu, prečo by daný názor mohol byť nesprávny.
- Nevyriešená pochybnosť je nulová operácia.
- Nevyšetrená pochybnosť akoby ani neexistovala.
- Nemali by ste byť hrdí na púhe pochybovanie, môžete však byť právom hrdí, keď ste akurát *dokončili* trhanie svojho obľúbeného názoru na kúsky.
- Aj keď si môže vyžadovať veľa odvahy čeliť svojim pochybnostiam, nezabúdajte nikdy na to, že *ideálna myseľ* by sa pochybnosti v prvom rade nikdy nebála.

Pôvodný článok s komentármi⁹

⁴ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

⁵ pozri: 13 Viera vo vieru

⁶ pozri: 110 Správne použitie pokory

⁷ pozri: 13 Viera vo vieru

⁸ pozri: 15 Vyznávanie a fandanie

⁹ http://lesswrong.com/lw/ib/the_proper_use_of_doubt/

Kapitola 122. Dokážete čeliť skutočnosti

Čo je pravda, to už je pravda.

Priznať si to to nerobí horším.

Nepriznať si to nespôsobí, že to zmizne.

A pretože je to pravda, je to to, s čím sa tu dá interagovať.

Čokoľvek nepravdivé tu nie je, nezažijeme to.

Ludia dokážu zniesť to, čo je pravda,

pretože už to znášajú.

—Eugene Gendlin

Pôvodný článok s komentármi¹

¹ http://lesswrong.com/lw/id/you_can_face_reality/

Kapitola 123. Meditácia o zvedavosti

Prvou cnosťou je zvedavosť

—Dvanásť cností rozumnosti

Ako racionalisti, sme povinní kritizovať sami seba a spochybňovať svoje názory... alebo nie?

Uvedomte si, čo sa s vami stane na psychologickej úrovni, keď začnete slovami: “Je mojou povinnosťou kritizovať svoje vlastné názory.” Roger Zelazny raz ukázal na rozdiel medzi: “chcieť byť spisovateľom” a “chcieť písať”. Mark Twain povedal: “Klasika je niečo, čo by každý chcel mať prečítané, ale nikto to nechce čítať.” Kritizovať sám seba z pocitu povinnosti znamená *chcieť mať preskúmané*, aby ste potom mohli povedať, že vaša viera nie je slepá. To nie je to isté ako *chcieť skúmať*.

Môže to viesť k motivovanému zastaveniu¹ vášho skúmania. Zvážite námietku, potom protiargument na túto námietku, a *tam sa zastavíte*. Toto zopakujete s niekoľkými námietkami, dokiaľ nemáte pocit, že ste splnili svoju povinnosť skúmať, a *tam sa zastavíte*. Dosiahli ste svoj základný psychologický cieľ: zbaviť sa kognitívnej disonancie, ktorá by vznikla z toho, že sa považujete za racionalistu a predsa viete, že ste sa nepokúsili kritizovať svoje názory. Mohli by sme to nazvať nakupovaním racionalistického uspokojenia² – snaha vytvoriť “hrejivý pocit” splnenej povinnosti.

Potom bude vaša uvádzaná úroveň pravdepodobnosti dosť vysoká na to, aby zdôvodnila, prečo sa držíte svojich pôvodných plánov a názorov, ale nie taká vysoká, aby vyvolala nedôverčivosť vašu alebo iných racionalistov.

Keď ste naozaj zvedaví, budú vás priťahovať otázky, ktoré vyzerajú najslubnejšie na spôsobenie zmeny názoru, alebo otázky, ktoré sa najmenej podobajú tým, ktoré ste skúšali predtým. Na konci vaša pravdepodobnostná distribúcia asi *nebude* vyzeráť tak, ako keď ste začali – nastanú zmeny, či nahor alebo nadol; ľubovoľný smer vám je rovnako dobrý, ak ste úprimne zvedaví.

Porovnajte si toto s podvedomým motívom držať svoje otázky na známom území, aby ste mohli svoje skúmanie rýchlo vybaviť, aby ste už *mali preskúmané*, a obnovili tým známu rovnováhu, na ktorej sú založené vaše staré známe plány a názory.

Ohľadom toho, ako si myslím, že by mala vyzeráť pravá zvedavosť, a akú má silu, odporúčam vám Bájku o vede a politike³. Každý z týchto postáv má vykresľovať inú lekciu. Ferris, posledná postava, stelesňuje silu nevinnej zvedavosti: čo je ľahkosť a dychtivé načahovanie sa po indíciách.

Ursula K. LeGuin napísala: “Nevinnosť nemá silu proti zlu. Ale má silu pre dobro.” Nevinná zvedavosť môže sa môže nevinne zvrtnúť; a preto treba výcviku racionalistu a s ním súvisiacej sofistikovanej⁴ čeliť ako nebezpečenstvu, ak sa chceme stať silnejšími⁵. Napriek tomu sa môžeme pokúšať udržať si túto ľahkosť a dychtivé načahovanie nevinnosti.

¹ pozri: 111 Tretia možnosť

² pozri: 283 Necitlivosť k rozsahu

³ pozri: 12 Bájka o vede a politike

⁴ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

⁵ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

Ako je napísané v Dvanástich cnostiach:

Ak vo svojom srdci veríš, že už vieš, alebo ak si vo svojom srdci neželáš vedieť, potom tvoje vypytovanie bude bez cieľa a tvoje zručnosti bez smeru. Zvedavosť sa snaží zničiť sama seba; neexistuje zvedavosť, ktorá by netúžila po odpovedi.

Neexistuje skratka žiadna dobrá náhrada za úprimnú zvedavosť. "Pálčivé svrbenie poznať je viac než slávnostná prisaha hľadať pravdu." Nemôžete však vytvoriť zvedavosť silou vôle o nič viac než silou vôle donútiť svoje nohy, aby cítili teplo, keď cítia chlad. Občas jediné, čo máme, sú naše púhe slávnostné prisahy.

Čo teda môžete urobiť s povinnosťou? Na začiatok, môžeme sa skúšať zaujímať o naše povinné skúmania – pozorne striehnuť na iskry skutočného záujmu, alebo aspoň skutočnej nevedomosti a túžby vyriešiť ju. Toto ide spolu so striehnutím na možnosti, ktoré sú bolestivé⁶, pred ktorými cúvate – nie všetko z toho je negatívne myslenie.

Malo by vám aj pomôcť meditovať o Zákone zachovania očakávanej indície⁷. Pre každý nový bod skúmania, pre každý kúsok dosiaľ *nevidenej* indície, na ktorú náhle pozriete, očakávaná výsledná pravdepodobnosť by sa mala rovnať vašej pôvodnej pravdepodobnosti. V mikroprocesoch skúmania by váš názor mal byť rovnako pripravený posunúť sa oboma smermi. Nie každý bod musí stačiť na širokú zmenu – na posunutie názoru z pravdepodobnosti 70% na 30% – ale ak je váš terajší názor 70%, mali by ste byť rovnako pripravení znížiť ho na 69% ako zvýšiť ho na 71%. Nemali by ste si myslieť, že viete, ktorým smerom pôjdete (v priemere), lebo podľa zákonov teórie pravdepodobnosti, ak poznáte svoj cieľ, už tam ste. Ak viete skúmať poctivo, takže každý nový bod má naozaj rovnaký potenciál posunúť váš názor nahor alebo nadol, môže vám to pomôcť zostať zaujatý alebo priam zvedavý na mikroprocesy skúmania.

Ak argument, ktorý zvažujete, *nie je* nový, prečo mu vôbec venujete pozornosť? Je to to, kam by ste sa pozerali, keby ste boli úprimne zvedaví? Kritizujete podvedome svoj názor na jeho najsilnejších miestach⁸, namiesto najslabších? Precvičujete si indície?⁹

Ak si dokážete neprecvičovať už známe obrany, a dokážete znížiť svoj názor o jeden malý kúsok z každej novej indície, možno budete dokonca schopní vzdať sa svojho názoru úplne – uvedomiť si, z ktorého smeru vietor indícií veje proti vám.

Ďalší posilňujúci prostriedok na zvedavosť je to, čo som začal volať Tarskeho litánia, čo je v skutočnosti meta-litánia, ktoré sa špecializuje na každý prípad (tak je to vhodné). Napríklad, ak som v napätí z rozmýšľania, či zamknutá krabica obsahuje diamant, potom si namiesto myslenia na krásne dôsledky, ak krabica obsahuje diamant, môžem opakovať Tarskeho litániu:

Ak táto krabica obsahuje diamant,

Chcem veriť, že táto krabica obsahuje diamant;

Ak táto krabica neobsahuje diamant,

Chcem veriť, že táto krabica neobsahuje diamant;

Kiež nie som pripútaný k názorom, ktoré možno nechcem.

⁶ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

⁷ pozri: 29 Zákon zachovania očakávanej indície

⁸ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

⁹ pozri: 92 Jeden argument proti armáde

Potom môžete meditovať o možnosti, že tam nie je diamant, a o následnej výhode, ktorú budete mať, ak budete veriť, že tam nie je diamant, a o následnej nevýhode, ktorú budete mať, ak budete veriť, že tam diamant je. Pozrite si aj Gendlinovu litániu¹⁰.

Ak v sebe dokážete nájsť najmenší kúsok pravej neistoty, potom si ho strážte ako keď lesník udržiava táborák. Ak dokážete dosiahnuť, aby z neho vzbĺkol plameň zvedavosti, urobí vás ľahkým a dychtivým, a dá zmysel vašim otázkam a smer vašim schopnostiam.

Pôvodný článok s komentármi¹¹

¹⁰ pozri: 122 Dokážete čeliť skutočnosti

¹¹ http://lesswrong.com/lw/jz/the_meditation_on_curiosity/

Kapitola 124. Nikto vám nemôže udeliť výnimku zo zákonov rozumnosti

Tradičná Rozumnosť je formulovaná v pojmach *spoločenských pravidiel*, ktorých porušovanie sa interpretuje ako podvádzanie – ako nedodržiavanie noriem spolupráce. Ak chcete, aby som prijal váš názor, ste povinný poskytnúť mi isté množstvo indície. Ak sa z toho skúšate vykrútiť, všetci vieme, že nedodržiavate svoje povinnosti. Teória je povinná robiť vlastné odvážne predpovede, nielen kradnúť predpovede, ktoré si odmakali iné teórie. Teória je povinná vystaviť sa falzifikácii – ak sa pokúša vykrútiť, je to ako pokúšať sa vykrútiť z obávaného iniciačného rituálu; musíte zaplatiť svoje dlhy.

Tradičná Rozumnosť je formulovaná podobne ako zvyky, ktorými sa riadia ľudské spoločnosti, čo zjednodušuje jej ústne odovzdávanie. Ľudia si všímajú spoločenské podvody omnoho spoľahlivejšie než izomorfné porušenia abstraktných logických pravidiel. Vnímanie rozumnosti ako spoločenskej povinnosti však prináša niektoré zvláštne myšlienky.

Napríklad človek vidí, ako veriaci ľudia obraňujú svoju vieru slovami: “Ty zase nevieš odôvodniť, prečo veríš vo vedu!” Inými slovami: “Ako sa opovažuješ kritizovať ma za neodôvodnené názory, pokrytec! Ty to tiež robíš!”

Pre bayesiánca je mozog nástrojom presnosti: spracováva a sústreďuje previazané indície¹ do mapy, ktorá odráža územie². Princípy rozumnosti sú zákonmi³ v rovnakom zmysle ako druhý zákon termodynamiky: získať spoľahlivý názor si vyžaduje vypočítateľné množstvo previazanej indície⁴, rovnako ako spoľahlivo ochladiť obsah chladničky si vyžaduje vypočítateľné minimum voľnej energie.

Zákony fyziky sú v princípe časovo reverzibilné, takže existuje nekonečne malá pravdepodobnosť – neodlíšiteľná od nuly pre všetkých okrem matematikov – že chladnička sa ochladí spontánne a ešte pritom vygeneruje elektrinu. Existuje o čosi väčšia nekonečne malá pravdepodobnosť, že by ste dokázali správne nakresliť podrobnú⁵ mapu ulíc New Yorku, hoci by ste ho nikdy nenavštívili, iba by ste sedeli vo svojej obývačke so zatiahnutými závesmi a bez internetového pripojenia. Ale ja by som sa na to nespoliehal.

Skôr než sa pokúsíte mapovať nevidené územie, nalejte si trocha vody do šálky pri izbovej teplote a najprv čakajte, či spontánne nezamrzne. Iba vtedy si môžete byť istí, že váš trik – ignorovanie nekonečne malých pravdepodobností úspechu – funguje vo všeobecnosti správne. Možno si hneď neuvedomíte, že vaša mapa je nesprávna, najmä ak ste v New Yorku nikdy neboli; ale môžete vidieť, že voda sama od seba nemrzne.

Ak sú pravidlá rozumnosti spoločenskými zvykmi, potom sa zdá možným ospravedlniť správanie X ak poukážete na to, že druhí robia to isté. Nebolo by *férové* žiadať indície od vás, ak ich nedokážeme poskytnúť sami. Uvedomíme si, že nikto z nás nie je lepší než ostatní⁶, ustúpime a láskavo vás ospravedlníme z

¹ pozri: 20 Čo je indícia?

² pozri: 46 Jednoduchá pravda

³ pozri: 180 Všeobecný zákon

⁴ pozri: 22 Koľko indície treba?

⁵ pozri: 6 Pritažujúce podrobnosti

⁶ pozri: 308 Tsuyoku verzus rovnostársky inštinkt

Nikto vám nemôže
udelieť výnimku zo
zákonov rozumnosti

vašej spoločenskej povinnosti poskytnúť indície pre vaše presvedčenie. A potom budeme všetci naveky žiť šťastne v slobode, bratstve a rovnosti.

Ak sú pravidlá rozumnosti matematickými zákonmi, potom bude snaha ospravedlniť presvedčenie bez indícií poukazaním na to, že niekto iný robí to isté, asi rovnako efektívna ako vymenovanie 30 dôvodov, prečo by ste nemali spadnúť z útesu. Aj keby sme všetci odhlasovali, že je neférové, že vaša chladnička potrebuje elektrinu, aj tak nebude fungovať (s pravdepodobnosťou 1). Aj keby sme všetci odhlasovali, že nemusíte navštíviť New York, mapa bude aj tak nesprávna. Pani Príroda je známa svojou ľahostajnosťou voči podobným prosbám, a takisto aj pani Matematika.

Takže – aby sme sa vrátili k spoločenskému jazyku Tradičnej Rozumnosti – nemyslite si, že *vám prejde* tvrdenie, že je okej mať svojvoľný názor na XYZ, pretože iní ľudia tiež majú svojvoľné názory. Ak sa dve strany zmluvy obe správajú rovnako nesprávne, ľudský sudca sa môže rozhodnúť neudelieť pokutu žiadnej z nich. Ale ak dvaja inžinieri navrhnu svoj stroje rovnako nesprávne, ani jeden stroj nebude fungovať. Jedna chyba v návrhu nemôže ospravedlniť druhú. Aj keď *ja* robím XYZ nesprávne, nepomôže vám to, ani vám to nedá výnimku z pravidiel; znamená to akurát, že sme obaja v kýbli.

Čo sa týka ľudského zákona v liberálnych demokraciách, každý má nárok na svoj názor. Čo sa týka zákona Prírody, nemáte nárok na presnosť. Nezatýkame ľudí za to, že si myslia divné veci, prinajmenšom v tých múdrejších krajinách nie. Nikto však nemôže zrušiť zákon⁷, že potrebujete indície⁸ na vytvorenie *presných* názorov.⁹ Ani hlasovanie celej ľudskej rasy nedokáže získať zhovievavosť na súde Prírody.

Fyzici nerozhodujú o fyzikálnych zákonoch, oni len hádajú, aké sú. Racionalisti nerozhodujú o zákonoch rozumnosti, iba hádame, aké sú. Nemôžete “racionalizovať”¹⁰ niečo, čo od začiatku nie je rozumné. Keby sa vám darom mimoriadnej presvedčivosti podarilo presvedčiť všetkých fyzikov na svete, že máte výnimku zo zákonov gravitácie, a skočíte z útesu, spadnete. Dokonca samotné slová “*my nerozhodujeme*” sú príliš antropomorfné. Neexistuje žiadna vyššia moc, ktorá by vám mohla dať výnimku. Existuje iba príčina a následok.

Pamätajte na to, keď budete žiadať aspoň o jednorazovú výnimku. My vám ju *nemôžeme* dať. Od nás to nezávisí.

Pôvodný článok s komentármi¹¹

⁷ pozri: 180 Všeobecný zákon

⁸ pozri: 20 Čo je indícia?

⁹ pozri: 46 Jednoduchá pravda

¹⁰ pozri: 95 Racionalizovanie

¹¹ http://lesswrong.com/lw/k1/no_one_can_exempt_you_from_rationality_laws/

Kapitola 125. Ponechajte ústupovú líniu

Keď obklúčiš nepriateľa,

Vždy mu ponechaj únikovú cestu.

Musí vidieť, že existuje

Alternatíva k smrti.

—Sun C', Umenie vojny

Nezvyšuj tlak, znižuj odpor.

—Lois McMaster Bujold, Komarr

Včera večer som sa rozprával s neracionalistkou, ktorá akosi zablúdila na zhromaždenie miestnych racionalistov. Vyhlásila, že (a) verí, že existujú duše, a (b) že neverí v kryoniku, lebo verí, že duša by nezostala v zmrazenom tele. Opýtal som sa: “Ale ako to vieš?” Zo zmätku, ktorý jej prebleskol tvárou bolo celkom jasné, že jej táto otázka nikdy nenapadla. Nehovorím to v zlom – vyzerala ako milá osoba s absolútne žiadnym výcvikom v rozumnosti, tak ako väčšina zvyšku ľudského druhu. Naozaj musím tú knihu napísať.

Väčšina výslednej konverzácie bola na témy už pokryté na Overcoming Bias – ak vás niečo *naozaj* zaujíma, pravdepodobne *dokážete* vymyslieť dobrý spôsob, ako to otestovať; pokúste sa najprv dosiahnuť presné názory a až potom z nich nechajte vyplynúť svoje emócie – takéto veci. Ale tento rozhovor mi pripomenul jednu vec, ktorú som tu ešte nepreberal.

“Ubezpeč sa,” navrhol som jej, “že si dokážeš predstaviť, ako by vyzeral svet, keby neexistovali žiadne duše, a čo by si vtedy robila. Nemysli na všetky dôvody, prečo to tak nemôže byť, jednoducho to prijmi ako predpoklad a potom si predstav dôsledky. Takže si pomyslíš: ‘No, keby duše neexistovalo, mohla by som sa prihlásiť na kryoniku’ alebo ‘Keby neexistoval Boh, aj tak by som mohla byť naďalej morálna’ namiesto toho, aby to bola príliš strašná predstava. Ako vec debaúcty by si sa mala pokúsiť veriť tomu, čo je pravda, bez ohľadu na to, aké je to nepohodlné, ako som už povedal; ale ako vec ľudskej povahy, pomáha, keď si ten názor urobíme menej nepohodlným, *skôr než* začneme vyhodnocovať indície v jeho prospech.”

Princíp za touto technikou je jednoduchý: Ako radí Sun C' ohľadom vašich nepriateľov, tak musíte aj sami so sebou – ponechajte si ústupovú líniu, aby ste mali menej problémov s ustupovaním. Predstava, že prídeje o svoju prácu, napríklad, môže vyzeráť omnoho strašidelnejšie, keď na to ani nedokážete pomyslieť, než keď ste si prepočítali, ako dlho vám vydržia úspory, pozreli ste si trh práce vo vašej oblasti, a inak ste presne naplánovali, čo urobíte ďalej. Iba potom budete pripravení *poctivo* zhodnotiť pravdepodobnosť, že si udržíte svoju prácu po plánovanom preúšťaní budúci mesiac. Buďte skutočným zbabelcom, a naplánujte si váš ústup do podrobností – predstavte si každý krok – najlepšie ešte skôr než prvýkrát vstúpíte na bojisko.

Nádej je, že si vyžaduje menej odvahy predstaviť si nepohodlný stav vecí *ako myšlienkový experiment*, než uvažovať, *s akou pravdepodobnosťou* je to pravda. Ale potom, čo urobíte to prvé, je ľahšie urobiť to druhé.

Pamätajte, že bayesiánstvo je presné – dokonca aj keď strašidelná predstava vyzerá nepravdepodobne, je stále dôležité spočítať všetky indície za a proti,

presne férovo, aby ste došli k rozumnej kvantitatívnej pravdepodobnosti. Vizualizovať si strašidelnú predstavu *neznamená* priznať, že si hlboko vnútri myslíte, že je pravdivá. Môžete si predstavovať strašidelné názory v rámci dobrého myšlienkového upratovania. "Myšlienka, ktorú si nedokážete pomyslieť vás ovláda viac než myšlienky, ktoré vyslovujete nahlas" - to sa stane, ešte aj keď je tá nemysliteľná myšlienka nepravdivá!

Technika ponechania ústupovej línie si vyžaduje isté minimum sebaapoctivosti, ak sa má používať správne.

Na začiatok: Musíte byť prinajmenšom schopní pripustiť si, *ktoré* myšlienky vás desia, a ku ktorým myšlienkam ste pripútaní. Ale toto je podstatne menej ťažké než férovo spočítať indície za myšlienku, ktorá vás desí. Pomôže vám, keď poviem, že som občas túto techniku sám použil? Racionalista predsa neodmieta všetky emócie. Existujú myšlienky, ktoré ma desia, ale aj tak verím, že sú nepravdivé. Existujú myšlienky, ku ktorým som pripútaný, ale predsa verím, že sú pravdivé. Ale aj tak stále plánujem svoje ústupy, nie preto, že by som plánoval *ustúpiť*, ale pretože plánovanie ústupu vopred mi pomáha rozmyšľať o danom probléme bez pripútanosti.

Ale väčší test sebaúprimnosti je *naozaj* prijať nepohodlný návrh ako predpoklad, a zistiť, ako by ste si s ním *naozaj* poradili. Keď čelíme nepohodlnej myšlienke, náš prvý impulz je myslieť na všetky dôvody, prečo to *vôbec* nemôže byť tak. A tak v sebe nájdete isté množstvo psychologického odporu, keď sa pokúsite predstaviť si, aký presne by bol svet, a čo by ste s tým robili, keby Môj Najdrahší Názor bol nepravdivý, alebo Moja Najväčšia Obava bola pravdivá.

Myslíte na všetkých tých ľuďoch, ktorí hovoria, že bez Boha je morálna nemožná. (A áno, táto téma sa v rozhovoroch vyskytuje; neponúkam tu slameného panáka.) Keby si veriaci vedeli predstaviť svoju *skutočnú* reakciu na to, keby uverili ako fakt, že Boh neexistuje, uvedomili by si, že by vlastne nešli zabíjať bábätká. Mohli by si uvedomiť, že ateisti reagujú na neexistenciu Boha viacmenej rovnako ako by reagovali oni sami, keby tomu uverili. Hovorím to, aby som ukázal, že *je* značne ťažké predstaviť si, ako by *ste* *naozaj* reagovali, keby ste verili opaku niečoho, čomu teraz pevne veríte.

Plus je vždy protiintuitívne uvedomiť si, že áno, ľudia sa cez veci nejako prenesú. Čerství vodiči nie sú o šesť mesiacov takí smutní, ako to dnes očakávajú, atď. Môže byť rovnako protiintuitívne uvedomiť si, že keby sa strašidelná predstava ukázala ako pravdivá, nejako *by* ste sa s ňou vyrovnali. Vodiči sa vyrovnávajú, aj vy by ste sa vyrovnali.

Pozrite si aj Gendlinovu litániu¹ a Tarskeho litániu². Čo je pravda, už je pravda; priznať si to to neurobí horšie. Nemali by ste sa báť iba si *predstaviť* svet, ktorého sa bojíte. Ak je ten svet už skutočný, predstava to nijako nezhorší; a ak *nie je* skutočný, predstava neurobí žiadnu škodu. A pamätajte si, keď si predstavujete, že ak tie strašné veci, ktoré si predstavujete, sú *naozaj* pravdivé - čo môžu byť! - potom by ste tomu veru chceli veriť, a aj to by ste si mali predstaviť; neveriť by vám nepomohlo.

Kolko veriacich ľudí by si udržalo svoju vieru v Boha, keby si vedeli *presne* predstaviť hypotetický svet, v ktorom by Boh nebol a oni sami by sa stali ateistami?

Ponechať ústupovú líniu je mocná technika, ale nie je ľahká. *Úprimná* predstava si nevyžaduje toľko námahy ako *priamo* priznať, že Boh neexistuje, ale aj tak si vyžaduje námahu.

¹ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

² pozri: 123 Meditácia o zvedavosti

Pôvodný článok s komentármi³

³ http://lesswrong.com/lw/o4/leave_a_line_of_retreat/

Kapitola 126. Kríza viery

Súvisí s kapitolami: Vynaložte mimoriadne úsilie¹, Meditácia o zvedavosti², Vyhýbanie sa naozaj slabým miestam vašich názorov³

Nie je tu skutočná kríza viery, pokiaľ by veci nemohli rovnako ľahko ísť ľubovoľným smerom.

—Thor Shenkel

Mnohí na tomto svete si udržiavajú názory, ktorých chyby by odhalilo desaťročné dieťa, *keby* toto desaťročné dieťa počulo daný názor po prvýkrát. Teraz nehovorím o nejakých rafinovaných chybách. Pre nepripútanú⁴ myseľ by bolo detskou hračkou zriechnuť sa ich, *keby* sa skepticizmus desaťročného použil bez vyhýbania⁵. Ako napísal Premise Checker: “Keby sme predstavu boha neboli zdedili až do veku vedy, iba mimoriadne divný človek by vymyslel takú predstavu a tvrdil, že to niečo vysvetľuje⁶.”

A predsa zruční vedeckí špecialisti, dokonca veľkí inovátori vo svojej oblasti, ešte aj v dnešnej dobe neuplatňujú svoj skepticizmus úspešne. Nositeľ Nobelovej ceny Robert Aumann, autor Aumannovej vety o súhlase, je ortodoxný žid: považujem za veľmi pravdepodobné, že Aumann musel v nejakom bode svojho života zapochybovať o svojej viere. A predsa sa mu nepodarilo pochybovať úspešne⁷. Svoje názory meníme zriedkavejšie než si myslíme.⁸

Toto by vás malo vydesiť do špiku kostí. Znamená to, že môžete byť svetovým vedcom *a* mať bayesovskú matematiku v malíčku, *a* predsa nedokázate odmietnuť názor, ktorého absurditu by videli čerstvé desaťročné oči. Ukazuje to tú nezraniteľnú obrannú polohu, ktorú si názor dokáže pre seba vytvoriť, ak dosť dlho hnisal vo vašej mysli.

Čo treba, aby sme porazili chybu, ktorá si okolo seba vybudovala pevnosť?

Nuž, v čase, keď *viete*, že je to chyba, už bola porazená. Dilema nezní: “Ako sa dokážem zbaviť môjho dávneho nesprávneho názoru X?” ale: “Ako môže vedieť, že môj dávny názor X je nesprávny?” Úprimnosť k sebe samému je najjemnejšia, keď si nie sme *istí*, ktorá cesta je tá správna. A tak vzniká otázka:

Ako dokážeme v sebe vytvoriť skutočnú krízu viery, ktorá by rovnako ľahko mohla dopadnúť ľubovoľným smerom?

Náboženstvo je pokus, ktorý si vieme všetci predstaviť. (Čitatelia, ktorí mali rodičov ateistov, zmeškali základnú životnú skúšku, a musia sa uspokojiť s úbohým náhradou myslenia na svojich veriacich priateľov.) Ale ak ste prešli všetky sympatie a vnímate veriacich ako zlých mutantov⁹, potom si nedokázate predstaviť skutočné vnútorné skúšky, ktorým čelia. Nebudete si vedieť položiť otázku:

“Akú všeobecnú stratégiu by veriaci človek mohol nasledovať, aby unikol zo svojho náboženstva?”

¹ pozri: 315 Vynaložte mimoriadne úsilie

² pozri: 123 Meditácia o zvedavosti

³ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

⁴ pozri: 63 Afektívne špirály smrti

⁵ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

⁶ pozri: 23 Occamova britva

⁷ pozri: 121 Správne použitie pochybnosti

⁸ pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁹ pozri: 51 Sú vaši nepriatelia od narodenia zlí?

Som si istý, že pri pohľade na túto výzvu už niektorí zo seba sypú zoznam štandardných ateistických argumentov - "Museli by si priznať, že neexistuje žiadna bayesovská indícia pre existenciu Boha", "Museli by si uvedomiť morálne vyhýbanie, ktorého sa dopúšťajú, aby ospravedlnili správanie Boha v Biblii", "Museli by sa naučiť používať Occamovu britvu..."

ZLE! ZLE, ZLE, ZLE! Tento druh opakovania¹⁰, kde iba vychrlíte body, *ktoré ste si už dávno premysleli*, je *presne* ten druh myslenia, ktorý udržiava ľudí pri v ich súčasných náboženstvách. Ak zostávate v rámci svojich uložených myšlienok¹¹, ak váš mozog dopĺňa samozrejmu odpoveď tak rýchlo, že ju nedokážete vidieť originálne¹², určite nedokážete navodiť krízu viery.

Možno je to otázka toho, že dosť ľudí nečítalo "Gödel, Escher, Bach" v dostatočne mladom veku, ale všimol som si, že veľká časť populácie - dokonca aj technicky zameraných ľudí - má problém nasledovať¹³ argument, ktorý ide takto meta. V moje pesimistickejšie dni pochybujem, či ťava má dva hrby.

Ešte aj keď sú na to vyslovene upozornení, niektorí ľudia asi *nedokážu nasledovať skok* z objektovej úrovne: "Použi Occamovu britvu! Musíš vidieť, že tvoj Boh je nadbytočný názor!" na meta-úroveň: "Skús zabrániť svojej mysli dopĺňať vzor zvyčajným spôsobom!" Pretože rovnako ako vaši racionalistickí priatelia hovoria o tom, že Occamova britva je dobrá vec, a rovnako ako vám Occamova britva okamžite príde na myseľ, rovnako kolektívne schválená náboženská odpoveď je: "Božie cesty sú tajomné a je domýšľavé predstavovať si, že ich dokážeme pochopiť." Takže ak si vy myslíte, že *všeobecná* stratégia hodná nasledovania je: "Použi Occamovu britvu", je to ako keď veriaci hovorí, že *všeobecná* stratégia je mať vieru.

"Ale... ale Occamova britva je naozaj lepšia než viera! To nie je ako uprednostňovať inú chuť zmrzliny! Každý sa môže pozrieť na históriu, že occamovské uvažovanie bolo omnoho produktívnejšie než viera..."

Čo je všetko pravda. Ale pointa je inde. Pointa je, že keď toho hovoríte, chrľíte zo seba štandardné zdôvodnenie toho, čo už je vo vašej hlave. Skutočná výzva krízy viery je zvládnuť prípad, kde je možné, že naše štandardné závery sú *nesprávne*, aj naše štandardné zdôvodnenia sú *nesprávne*. Takže ak štandardným zdôvodnením pre X je "Occamova britva!", a chcete mať krízu viery ohľadom X, mali by ste zapochybovať, čo Occamova britva naozaj odporúča X, či vaše chápanie Occamovej britvy je správne, a - ak chcete mať dostatočne hlboké pochybnosti - či jednoduchosť je tým druhom kritéria, ktoré v minulosti v tomto prípade dobre fungovalo, alebo by sa dalo rozumne *očakávať*, že bude fungovať, atď. Keby ste poradili religionistovi zapochybovať o tom, či "vieru" je dobré zdôvodnenie pre X, potom by ste mali sebe odporučiť vynaložiť rovnako silné úsilie spochybniť svoju vieru, že "Occamova britva" je dobrým zdôvodnením pre X.

(Pomyslite na všetkých tých ľudí, ktorí nerozumejú minimálnej dĺžke popisu alebo formulácii Occamovej britvy pomocou Solomonoffovej indukcie, ktorí si myslia, že Occamova britva vylučuje hypotézu mnohých svetov¹⁴, alebo hypotézu simulácie¹⁵. Potrebovali by zapochybovať o ich formulácii Occamovej britvy a svojej predstave, prečo je jednoduchosť dobrá vec. Ktorékoľvek X ste si práve zdôvodnili slovami "Occamova britva!", stavím sa, že to nie je na rovnakej úrovni occamovského zásahu do terča ako gravitácia.)

¹⁰ pozri: 92 Jeden argument proti armáde

¹¹ pozri: 81 Uložené myšlienky

¹² pozri: 83 Originálny pohľad

¹³ http://lesswrong.com/lw/h6/chronophone_motivations/

¹⁴ pozri: 234 Dekoherencia je jednoduchá

¹⁵ <http://www.simulation-argument.com/simulation.html>

Ak je "Occamova britva!" vaša zvyčajná odpoveď, vaša štandardná odpoveď, odpoveď ktorú dávajú všetci vaši priatelia – potom by ste mali zablokovvať svoj mozog od okamžitého dopĺňania tohto vzoru, ak sa pokúšate navodiť skutočnú krízu viery.

Lepšie je myslieť na pravidlá ako: "Predstav si, čo by povedal skeptik – a potom si predstav, čo by odpovedal na tvoju reakciu – a potom si predstav, čo ďalšie by mohol povedať, na čo by sa ťažšie odpovedalo."

Alebo: "Skús si myslieť tú myšlienku, ktorá najviac bolí."

A nadovšetko, toto pravidlo:

"Vynalož rovnakú úroveň zúfaleho úsilia¹⁶, aké by bolo treba, aby veriaci odmietol svoje náboženstvo."

Pretože ak sa *nepokúšate* takto silno, potom – napriek všetkému, čo si *myslíte* – vaša hlava môže byť stále naplnená vecami rovnako nezmyselnými a smiešnymi ako náboženstvo.

Bez krčovitého krútiaceho úsilia byť rozumný, bez toho druhu úsilia, ktoré by bolo treba na odvrhnutie náboženstva – ako by ste sa odvážili veriť v niečo, keď Robert Aumann verí v Boha?

Niektor (zabudol som, kto) raz konštatoval, že ľudia majú šancu odmietnuť svoju náboženskú vieru iba do istého veku. Potom už majú odpovede na všetky námietky, a je príliš neskoro. Tento druh existencie musíte prekonať. Toto je test vašej sily ako racionalistu, a je to veľmi ťažké; ale ak ním nedokážete prejsť, budete slabší ako desaťročník.

Ale opäť, keď už viete, že nejaký názor je nesprávny, už ste ho porazili. Nehovoríme tu teda o podniknutí zúfaleho krčovitého úsilia o zvrátenie účinkov¹⁷ náboženskej výchovy, *potom* čo ste došli k záveru, že náboženstvo je nepravdivé. Hovoríme tu o zúfalom úsilí *zistiť*, či by ste mali odhodiť svoje reťaze, alebo si ich ponechať. Úprimnosť k sebe samému je najjemnejšia, keď *nevieme*, po ktorej ceste máme ísť – vtedy racionalizácie nie sú *jasnými* hriechmi.

Nie každá pochybnosť si vyžaduje vykonanie kompletnej Krízy Viery. Ale mali by ste o nej uvažovať, keď:

- Nejaký názor bol dlho vo vašej mysli;
- Je obklopený oblakom známych argumentov a vyvrátení;
- Utopili¹⁸ ste v ňom náklady (čas, peniaze, verejné vyhlásenia);
- Tento názor má emocionálne dôsledky¹⁹ (to samotné neznamená, že je nesprávny);
- Je všeobecne previazaný s vašou osobnosťou.

Žiadne z týchto varovných znamení nie je automatické vyvrátenie. Tieto vlastnosti vytvárajú okolo ohrozeného²⁰ názoru rôzne nebezpečenstvá, a robia jeho vyvrátenie veľmi ťažkým, ak je nesprávny. Ale rovnako by platili pre vieru

¹⁶ pozri: 315 Vynaložte mimoriadne úsilie

¹⁷ pozri: 89 Klam pôvodu

¹⁸ http://cs.wikipedia.org/wiki/Utopen%C3%A9_n%C3%A1klady

¹⁹ pozri: 2 Rozumné čítanie

²⁰ pozri: 77 Sektárske antisektárstvo

Richarda Dawkinsa v evolučnú biológiu, ako vo vieru pápeža v katolicizmus. To neznamená, že sa tu rozprávame o rôznych príchutiach zmrzliny. Iba neosvietení si myslia, že všetky silné presvedčenia sú na rovnakej úrovni bez ohľadu na indície, ktoré ich podporujú, len preto, že sú silné. Naším cieľom nie je mať plytké názory, ale mať mapu, ktorá odráža územie.

Zdôrazňujem to, samozrejme, aby ste si mohli priznať: "Môj názor má tieto varovné znamenia," bez toho, že by ste si museli povedať: "Môj názor je nesprávny."

Ale čo tieto varovné znamenia *označujú*, je názor, o ktorom *efektívne pochybovanie si bude vyžadovať viac než len bežné úsilie*. Aby, pokiaľ je naozaj nesprávny, ste ho naozaj odmietli. A kde nedokážete účinne pochybovať, ste slepí, pretože váš mozog bude tento názor držať bezpodmienečne²¹. Keď sietnica vysiela rovnaké signály bez ohľadu na to, aké fotóny do nej vstupujú, nazývame takéto oko slepým²².

Kedy by ste mali podstúpiť Krízu Viery?

Opäť, pomyslite na radu, ktorú by ste dali veriacemu: Ak zistíš, že sa vo vnútri cítiš trochu nestabilne, ale pokúšaš sa racionalizovať dôvody, prečo je názor stále solídny, potom by si pravdepodobne mal vykonať rituál Krízy Viery. Ak je ten názor tak solídne podložený ako gravitácia, nemusíte sa unúvať – ale pomyslite na všetkých tých veriacich, ktorí by zúfalo chceli dôjsť k záveru, že Boh je rovnako solídny ako gravitácia. Skúste si teda predstaviť, čo by nejaký skeptik mohol povedať na váš argument "solídny ako gravitácia". Určite je jedným z dôvodov, prečo môžete v kríze viery zlyhať, to, že si v prvom rade nikdy nesadnete a nezačnete pochybovať – že nikdy nepoviete: "Tu je niečo, na čo musím vynaložiť úsilie, aby som správne pochyboval."

Ak vaše myšlienky začnú byť takéto komplikované, mali by ste ísť na to a vyvolať Krízu Viery. Neskúšajte to robiť chaoticky, neskúšajte to, ak máte iba chvíľku voľného času. Neponáhľajte sa, aby ste to čo najrýchlejšie mali za sebou, aby ste mohli povedať: "Pochyboval som, tak ako bolo mojou povinnosťou." To by nefungovalo pre veriaceho a nefungovalo by to ani pre vás. Predchádzajúci deň si oddýchnite, nech ste v dobrom mentálnom stave. Vyhradte si pár hodín bez prerušenia. Nájdite si tiché miesto, kde si môžete sadnúť. Zbavte svoju myseľ všetkých štandardných argumentov, skúste vnímať od začiatku. A vynaložte zúfale úsilie, aby ste vytvorili skutočnú pochybnosť, ktorá by dokázala zničiť nepravdivý, ale *iba* nepravdivý, hlboko držaný názor.

Prvky techniky Kríza Viery, rozhádzané po mnohých článkoch:

- Vyhýbanie sa naozaj slabým miestam vašej viery²³ – Jedným z prvých pokúšení krízy viery je pochybovať o najsilnejších miestach svojej viery, aby ste si mohli precvičiť²⁴ svoje dobré odpovede. Potrebujete vyhľadať tie najboľavejšie miesta, nie argumenty, ktoré je najpríjemnejšie zvažovať.
- Meditácia o zvedavosti²⁵ – Roger Zelazny raz rozlíšil medzi "chcieť byť autorom" a "chcieť písať", a takisto existuje rozdiel medzi chcieť mať niečo preskúmané a chcieť skúmať. Nestačí povedať: "Je mojou povinnosťou kritizovať vlastné názory"; musíte byť zvedaví a iba neistota dokáže vytvoriť zvedavosť. Pamätajte na Zákon zachovania očakávanej indície²⁶; môže vám to

²¹ pozri: 93 Spodný riadok

²² pozri: 20 Čo je indícia?

²³ pozri: 97 Vyhýbanie sa naozaj slabým miestam vašich názorov

²⁴ pozri: 92 Jeden argument proti armáde

²⁵ pozri: 123 Meditácia o zvedavosti

²⁶ pozri: 29 Zákon zachovania očakávanej indície

pomôcť Aktualizovať postupne²⁷: Pri každom *jednom* bode, ktorý zvažujete, a pri každom prvku nového argumentu a novej indicie, by ste nemali očakávať, že sa vaše názory posunú (v priemere) viac jedným smerom než druhým – takto môžete byť naozaj zakaždým zvedaví, ktorým smerom to pôjde.

- Uložené myšlienky²⁸ a Pirsigov Originálny pohľad²⁹; zabráňte štandardným myšlienkam, aby vtrhli a doplnili vzor.
- Gendlinova litánia³⁰ a Tarskeho litánia³¹: Ľudia dokážu zniesť to, čo je pravda, lebo už to znášajú. Ak je názor pravdivý, bude pre teba lepšie, keď mu budeš veriť, a ak je nepravdivý, bude pre teba lepšie ho odmietnuť. Veriacemu človeku by si odporučil, aby si skúsil naplno a hlboko predstaviť svet, v ktorom neexistuje žiaden Boh, a aby bez výhovoriek naplno pochopil, že *ak* Boh neexistuje, *tak* preňho bude lepšie veriť, že Boh neexistuje. Ak človek nedokáže toto prijať na hlbokej emocionálnej úrovni, nedokáže mať krízu viery. Mali by ste teda vynaložiť úprimné úsilie predstaviť si *alternatívu* svojho názoru, tak ako by najlepší a najväčší skeptici chceli, aby ste si ju predstavili. Pomyslite na úsilie, ktoré musí vynaložiť religionista, aby si predstavil, bez prekrútenia pre svoje vlastné pohodlie, ateistov svetonázor.
- Vynaložte mimoriadne úsilie³²: pojem *isshokenmei*, zúfalé krčovité úsilie byť rozumný, ktoré by bolo treba na prekročenie úrovne Roberta Aumanna a všetkých veľkých vedcov v histórii, ktorí nikdy nezanechali svoje náboženstvá.
- Heuristika pôvodu³³: Mali by ste byť mimoriadne podozrievaví, ak máte mnoho myšlienok zo zdroja, o ktorom teraz viete, že je nedôveryhodný, ale zhodou okolností sa zdá, že všetky tieto myšlienky predsa len boli pravdivé. (Napríklad človek pripustí, že Bibliu napísali ľudia, ale stále sa drží predstavy, že obsahuje neodmysliteľnú etickú múdrosť³⁴.)
- Dôležitosť hovorenia “Joj”³⁵ - naozaj to menej bolí prehltnúť celú horkú tabletku na jeden hrozný glg.
- Singlethink³⁶, opak doublethinku. Pozrite sa na myšlienky, od ktorých cúvate, ktoré sa objavujú v kútiku vaše mysle iba na chvíľku³⁷, než na ne odmietnete myslieť. Ak si uvedomíte, na čo nemyslíte, môžete na to myslieť.
- Afektívne špirály smrti³⁸ a Odolajte šťastnej špirále smrti³⁹. Afektívne špirály smrti sú hlavné generátory nepravdivých názorov, ktoré vyžadujú Krízu Viery, aby ste sa ich striasli. Ale keďže afektívne špirály smrti môžu vzniknúť aj okolo skutočných vecí, ktoré sú naozaj pekné, nemusíte pripustiť, že váš názor je lož, skúste na každom kroku odolať efektu svätožiari - odmietnite falošnú chválu aj naozaj dobrých vecí. Debaty o pravidlách by nemali vyzeráť jednostranne.⁴⁰
- Počkajte s navrhovaním riešení⁴¹, dokiaľ problém nebude prediskutovaný tak dôkladne, ako sa len dá bez ich navrhovania; zdržte svoju myseľ od poznania,

²⁷ pozri: 91 Aktualizujte postupne

²⁸ pozri: 81 Uložené myšlienky

²⁹ pozri: 83 Originálny pohľad

³⁰ pozri: 122 Dokážete čeliť skutočnosti

³¹ pozri: 123 Meditácia o zvedavosti

³² pozri: 315 Vynaložte mimoriadne úsilie

³³ pozri: 89 Klam pôvodu

³⁴ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

³⁵ pozri: 118 Dôležitosť hovorenia: “joj!”

³⁶ pozri: 104 Singlethink

³⁷ pozri: 104 Singlethink

³⁸ pozri: 63 Afektívne špirály smrti

³⁹ pozri: 64 Odolajte šťastnej špirále smrti

⁴⁰ pozri: 48 Debaty o pravidlách by nemali vyzeráť jednostranne

⁴¹ pozri: 88 Odložte navrhovanie riešení

čo bude jej odpoveď⁴²; a pokúšajte sa o to päť minút než sa vzdáte⁴³, aj vo všeobecnosti, ale najmä keď hľadáte diabla uhol pohľadu.

A tieto štandardné techniky sú mimoriadne relevantné:

- Časť o Spodnom riadku⁴⁴ a Racionalizácii⁴⁵, ktorá vysvetľuje, prečo je vždy nesprávne selektívne argumentovať za jednu stranu debaty.
- Pozitívne skreslenie⁴⁶ a motivovaný skepticizmus⁴⁷ a motivované zastavenie⁴⁸, aby ste selektívne neľadali podporu, selektívne neľadali protiargumenty na protiargumenty, a selektívne nezastavili argument skôr než sa stane nebezpečným. Nevšímanie si alternatív⁴⁹ je špeciálnym prípadom zastavenia sa. Špeciálnym prípadom motivovaného skepticizmu je falošná skromnosť⁵⁰, kde hanblivo priznáte, že nikto nemôže vedieť⁵¹ to, čo vy nechcete vedieť. Nepožadujte selektívne priveľa autority⁵² od protiargumentov.
- Pozor na Sémantické stopky⁵³, Signály na potlesk⁵⁴, a výber Vysvetli/Uctievaj/Ignoruj⁵⁵.
- Cíťte váhu Priťažujúcich podrobností⁵⁶; každá podrobnosť je samostatné bremeno, bod krízy.

Tu na *Overcoming Bias* je naozaj kopec relevantného materiálu. Kríza Viery je iba kritický bod a náhly náraz dlhodobého *isshoukenmei* – celoživotného nekompromisného úsilia byť taký neuveriteľne rozumný, že stúpnete nad úroveň hlúpych prekliatych chýb. Je to, keď dostanete šancu použiť svoje zručnosti, ktoré ste tak dlho precvičovali, naplno proti sebe samému.

Želám vám veľa šťastia proti vášmu súperovi. Majte nádhernú krízu!

Pôvodný článok s komentármi⁵⁷

⁴² pozri: 87 Svoje názory meníme zriedkavejšie, než si myslíme

⁴³ pozri: 313 Použi silnejšiu snahu, Luke

⁴⁴ pozri: 93 Spodný riadok

⁴⁵ pozri: 95 Racionalizovanie

⁴⁶ pozri: 39 Pozitívne skreslenie: Pozeraj do tmy

⁴⁷ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

⁴⁸ pozri: 98 Motivované zastavenie a motivované pokračovanie

⁴⁹ pozri: 111 Tretia možnosť

⁵⁰ pozri: 110 Správne použitie pokory

⁵¹ http://lesswrong.com/lw/kj/no_one_knows_what_science_doesnt_know/

⁵² pozri: 114 Absolútna autorita

⁵³ pozri: 35 Sémantické stopky

⁵⁴ pozri: 16 Signály na potlesk

⁵⁵ pozri: 43 Vysvetli / Uctievaj / Ignoruj?

⁵⁶ pozri: 6 Priťažujúce podrobnosti

⁵⁷ http://lesswrong.com/lw/ur/crisis_of_faith/

Kapitola 127. Rituál

Súvisí s kapitolou: Zlyhania vedy predkov¹

Miestnosť, v ktorej Jeffreyssai prijímal svojich návštevníkov, ktorí neboli *beisutsukai*, bola ticho formálna, bezchybne určená iba tým najkonzervatívnejším vkusom. Slnčné svetlo a vonkajší vzduch prúdili cez mrežu z vylešteného striebra, ktorej pár ostrých hrotov dávalo najavo, že táto stena sa nemá otvárať. Podlaha a steny boli zo skla, dosť hrubého na to, aby skresľovalo do tej miery, že nezáležalo na tom, čo by mohlo byť pod ním. Na povrchu skla boli jemne vyryté vzory bez konkrétneho významu, akoby načmárané rukou dieťaťa s umeleckými sklonmi (čo mimochodom bolo naozaj tak).

Inde v Jeffreyssaiovom dome boli miestnosti v inom štýle; ale ako zistil, toto bolo to, čo väčšina laikov očakávala od Bayesiánskeho Majstra, a on sa rozhodol neosvecovať ich inak. Táto tichá zábava bola napokon jedným z drobných životných potešení.

Návšteva sedela oproti nemu, kolená na vankúši, podpätky za sebou. Bola tu výhradne kvôli záležitosti jej Konšpirácie, a jej odev to ukazoval: Priliehavá kombinéza z ružovej kože zakrývala aj jej ruky – až po kapucňu, ktoré jej prikrývala hlavu a vlasy, hoci tvár bola priama a nezakrytá.

A tak sa Jeffreyssai rozhodol, že ju prijme v tejto izbe.

Jeffreyssai zhlboka vydýchol. “Si si *istá*?”

“Ach,” povedala, “a to si musím byť *absolútne istá* skôr než moja rada dokáže pohnúť tvojimi názormi? Nestačí ti, že som v danej oblasti odborníčka a ty nie?”

Jeffreyssaiove ústa sa v kútiku zdvihli v polovičnom úsmeve. “Odkiaľ ty vôbec vieš tak veľa o pravidlách? Nikdy si nemala ani len Planckovu dĺžku formálneho výcviku.”

“Je vôbec treba sa na to pýtať?” povedala sucho. “Ak je niečo, o čom vy *beisutsukai* milujete rozprávať donekonečna, sú to dôvody, prečo robíte veci.”

Jeffreyssai sa vnútorne striasol pri predstave, že sa niekto pokúša pochytiť rozumnosť tým, že sleduje druhých, ako o nej hovoria...

“A nestriasaj sa vnútorne, keď ma počúvaš,” povedala. “Ja sa nepokúšam byť racionalistka, iba sa s racionalistom snažím vyhrať debatu. V tom je rozdiel, ako určite rozprávaš svojim žiakom.”

Dokáže ma naozaj tak dobre čítať? Jeffreyssai sa pozrel von cez striebornú mrežu, a slnčné svetlo sa odrazilo od plochy hôr. Vždy, vždy zlaté slnčné svetlo zapadalo každý deň na tomto mieste vysoko nad oblakmi. Nemenná vec, toto svetlo. Vzdialené Slnko, ktoré toho svetlo reprezentovalo, o päť miliárd rokov dohorí; ale teraz, v *tejto* chvíli, Slnko stále svieti. A to sa nikdy nezmení. Načo si želať, aby veci zostali navždy rovnaké, keď toto želanie už bolo splnené tak absolútne, ako len želanie môže byť? Paradox stálosti a nestálosti: iba z pohľadu toho druhého existovali veci ako pokrok, alebo strata.

“Vždy si mi radila dobre,” povedal Jeffreyssai. “Toto sa nemenilo. Celý čas, čo sa poznáme.”

Sklonila hlavu, prikývla. Toto bola pravda a nebolo treba vyslovovať jej dôsledky.

¹ pozri: 243 Zlyhania vedy predkov

“Takže,” povedal Jeffreyssai. “Nie kvôli hádke. Iba preto, že chcem poznať odpoveď. *Si si istá?*” Ani nevedel, ako by to mohla *odhadnúť*.

“Dost istá,” povedala, “dlhú dobu sme zhromažďovali štatistiky, a v deväťsto osemdesiatich piatich prípadoch z tisíce ako je ten tvoj...”

Potom sa zasmiala na výraze jeho tváre. “Nie, žartujem. Samozrejme, že si nie som istá. Túto vec vieš rozhodnúť iba ty. Ale *som si istá*, že by si si mal dať pauzu a urobiť to, čo robievate – som si dost istá, že na to máte nejaký rituál, aj keď o ňom nehovoríte s nezasvätenými – keď *veľmi vážne zvažujete* zanechanie dlhodobého predpokladu vašej existencie.”

Bolo naozaj ťažké hádať sa s tým, premýšľal Jeffreyssai, najmä keď vám odborník na danú oblasť povedal, že sa pravdepodobne mýlite.

“Uznávam,” povedal Jeffreyssai. Z jeho úst mala táto veta konečnosť príkazu. *Nie je nutné sa do mnou ďalej hádať: Vyhrala si.*

“Ale prestaň,” povedala. Vstala z vankúša jedným plynulým posunom bez najlepšieho zbytočného pohybu. Nevystatovala sa svojím vekom, ale ani ho neskrývala. Prijala jeho vystretú ruku a zdvihla ju k perám na formálny bozk. “Zbohom, sensei.”

“Zbohom?” zopakoval Jeffreyssai. To naznačovalo vyšší stupeň rozchodu než dovidenia. “Mám v úmysle ťa opäť navštíviť, milady; a ty si tu vždy vítaná.”

Prešla ku dverám bez odpovede. Vo dverách sa zastavila, bez otočenia. “Už to nebude to isté,” povedala. A potom bez toho, že by jej pohyby vyzerali čo len trochu náhlivo, odišla tak rýchlo, že to bolo takmer akoby zmizla.

Jeffreyssai si vzdychol. Ale prinajmenšom odteraz po samotnú skúšku boli jeho činy predpísané, známe množstvá.

Opustil túto formálnu prijímaciu miestnosť, prešiel cez svoju arénu a poslal za svojimi žiakmi poslov s odkazom, že zajtrajšie lekcie musia improvizovať v jeho neprítomnosti, a že neskôr bude skúška.

A potom nerobil nič konkrétne. Prečítal si ďalších sto strán učebnice, ktorú mal požičanú; nebola veľmi dobrá, ale napokon kniha, ktorú požičal na výmenu tiež nebola veľmi dobrá. Túlal sa z jednej izby vo svojom dome do druhej, potulujúc sa kontroloval rôzne sklady, aby videl, či sa niečo ukradlo (chýbal balíček kariet, ale to bolo všetko). Z času na čas sa jeho myšlienky vrátili k zajtrajšej výzve, a on ich nechal plynúť. Vôbec svoje myšlienky neusmerňoval, iba zablokoval každú myšlienku, ktorá mu napadla *predtým*; a nedovolil žiaden druh záveru, ani žiadnu myšlienku ohľadom toho, kam jeho myšlienky možno smerujú.

Slnko zapadlo, a on ho chvíľu pozoroval, myseľ starostlivo nečinná. Bol to čin fantastickej rovnováhy dostať svoju myseľ do nečinnosti bez toho, že tým človek bol posadnutý, alebo že by vynakladal energiu, aby ju takto udržal; pred pár rokmi by sa z toho bol zapotil, ale cvičenie ho už dávno zdokonalilo.

Ďalšie ráno sa zobudil s chaosom z nočného sna čerstvo v mysli, a robiac, čo vedel, aby si tento pocit chaosu udržal aj v pamäti, zišiel schodiskom, potom ďalším schodiskom, potom ešte ďalším schodiskom, až nakoniec prišiel do najmenej módnjej izby v celom svojom dome.

Bola biela. A to bol koniec celej jej farebnej schémy.

Podlž jednej celej steny boli plakety, ktoré si podľa starej a odporúčanej metódy mladý Jeffreyssai veľmi pozorne sám narysoval, vypaľujúc si *pojmy* do svojej

mysle každým dotykom štetca, ktorý písal tieto slová. *To, čo je možné zničiť pravdou, by malo byť zničené. Ľudia dokážu zniesť to, čo je pravda, pretože to už znášajú. Zvedavosť sa snaží zničiť sama seba.* Jedna malá plaketa dokonca neukazovala nič, iba červený vodorovný rez. Symboly mohli znamenať čokoľvek; pružnosť vizuálnej moci, pred ktorej priamym priznaním by zaváhala ešte aj Konšpirácia Bardov.

Pod plaketami boli do steny vyryté dve sady značiek. V stĺpci plus, dve značky. V stĺpci mínus, päť značiek. Sedemkrát vstúpil do tejto miestnosti; päťkrát sa rozhodol nezmeniť svoj názor; dvakrát odišiel ako iný človek. Neexistoval žiaden predpísaný pomer, ani rozsah – to by bolo výsmechom. Ale ak po čase neboli v plusovom stĺpci žiadne značky, mohli ste rovnako dobre priznať, že nemá zmysel mať túto miestnosť, pretože nemala schopnosť, ktorú predstavovala. Buď to, alebo ste sa narodili poznajúc pravdu a správnosť všetkého.

Jeffreyssai si sadol, nie čelom k plaketám, ale čelom od nich, smerom k prázdnej bielej stene. Bolo lepšie nemať vizuálne rozptýlenie.

V mysli si zopakoval meta-mnemoniku a potom rôzne odkazované sub-mnemoniky, siedmych hlavných princípov s šesťdesiatich dvoch konkrétnych techník, ktoré sa s najväčšou pravdepodobnosťou mohli ukázať potrebné pri Rituále Zmeny Náзору. K týmto si Jeffreyssai pridal ďalšiu mnemoniku, ktorá mu pripomínala jeho vlastných štrnásť najzahanbujúcejších prehliadnutí.

Nenadýchol sa zhlboka. Pravidelné dýchanie bolo najlepšie.

A potom si položil tú otázku.

Pôvodný článok s komentármi²

² http://lesswrong.com/lw/us/the_ritual/

Časť XIII. Medzihra

Obsah

128 Sila inteligencie	206
-----------------------------	-----

Kapitola 128. Sila inteligencie

Vo svojich lebkách nosíme asi kilo slizkého, vlhkého, sivého tkaniva, zvlneného ako pokrčený toaletný papier. Keby ste pozreli na túto nechutnú hrču, nepomysleli by ste si, že je to jedna z najmocnejších vecí v známom vesmíre. Keby ste nikdy nevideli učebnicu anatómie, a videli by ste mozog ležať na ulici, povedali by ste “Fuj!” a pokúsili by ste sa nezamazať si od neho topánky. Aristoteles si myslel, že mozog je orgán na chladenie krvi. Vyzerá neškodne.

Pred piatimi miliónmi rokov predkovia levov vládli cez deň a predkovia vlkov sa túlili nocou. Vládnuce šelmy boli vyzbrojené zubami a pazúrami – ostrými, tvrdými čepelami, ovládanými mocnými svalmi. Ich korist' si v sebaobrane vyvíjala pancierové schránky, ostré rohy, jedy, maskovanie. Táto vojna trvala tisíce vekov a nespočetne veľa pretekov v zbrojení. Nejednen porazený vypadol z hry, ale nikde nebolo stopy po víťazovi. Kde jeden druh mal pancier, iný sa vyvinul, aby ho prelomil; kde sa jeden druh stal jedovatým, iný si vyvinul odolnosť voči jedu. Každý druh mal svoje vlastné miesto – veď kto by dokázal žiť v mori, na oblohe, a na zemi zároveň? Neexistovala dokonalá zbraň ani dokonalá obrana, ani dôvod veriť, že čosi také je možné.

Potom prišiel čas Mäkkých Vecí.

Nemali brnenie. Nemali pazúre. Nemali jed.

Keby ste videli film zobrazujúci jadrový výbuch, a keby vám povedali, že to urobil jeden zo živočíchov na Zemi, ani v najdivokejšom sne by ste si nepredstavovali, že za to môžu Mäkké Veci. Napokon, Mäkké Veci nie sú rádioaktívne.

Na začiatku Mäkké Veci nemali ani stíhačky, ani guľomety, ani pušky, ani meče. Ani bronz, ani železo. Ani kladivá, ani nákovy, ani kliešte, ani kovárske dielne, ani bane. Mäkké Veci mali akurát mäkké prsty – príliš slabé na to, aby zlomili strom, tobôž horu. Očividne neškodné. Na sekanie kameňa by bolo treba oceľ, ale Mäkké Veci nevedeli vylučovať oceľ. V prírode neboli žiadne oceľové čepele, ktoré by ich mäkké prsty mohli zdvihnúť. Ich telá nedokázali vytvoriť ani zďaleka dostatočnú teplotu na roztavenie kovu. Celý scenár bol jasne absurdný.

A čo sa týka Mäkkých Vecí manipulujúcich DNA – to už by ani nebolo smiešne. Mäkké prsty nie sú také malé. Na ich úrovni nie je DNA dostupná; to by bolo ako pokúšať sa zdvihnúť atóm vodíka. Iste, *technicky vzaté* všetko tvorí jeden vesmír, *technicky vzaté* sú Mäkké Veci a DNA časti toho istého sveta, s jednotnými zákonmi fyziky, rovnakou veľkou sieťou kauzality. Ale budme realistickí: odtiaľto tam sa dostať nedá.

Aj keby si Mäkké Veci *jedného dňa* dokázali vyvinúť niektorú z týchto schopností, trvalo by im to milióny rokov. Celé veky sme sledovali príliv a odliv Života, a môžeme vám povedať, jeden rok, to nie je ani len tiknutie na hodinách evolučného času. Isteže, *technicky vzaté*, jeden rok je šesťsto biliónov biliónov biliónov biliónov Planckových intervalov. Ale za menej ako šesťsto miliónov biliónov biliónov biliónov Planckových intervalov sa aj tak nikdy nič nestane, takže táto námietka je iba akademická. Mäkké Veci, ktoré práve bežia krížom savanou, nebudú lietať krížom cez kontinent prinajmenšom ďalších desať miliónov rokov; *nikto* nemôže mať tak veľa sexu.

A teraz mi opäť vysvetlite, prečo umelá inteligencia nemôže urobiť nič zaujímavé cez internet, dokiaľ jej človek programátor nepostaví robotické telo.

Všimol som si, že automatická reakcia človeka na slovo “inteligencia” - to, čo mu vyvstane v mysli v prvej polovici sekundy po počutí slova “inteligencia” -

často určuje jeho automatickú reakciu na singularitu. Často sa im pod slovom "inteligencia" vynára pojem *nepraktického intelektuála* – predstava šachového veľmajstra, ktorý si nevie nájsť priateľku, alebo univerzitného profesora, ktorý by mimo akademickej pôdy nedokázal prežiť.

"Aby ste uspeli v práci, potrebujete viac než len inteligenciu," hovoria ľudia, ako keby sila osobnosti sídlila v obličkách, a nie v mozgu. "Inteligencia je bezmocná voči puške," hovoria, ako keby pušky rástli na stromoch. "Kde by umelá inteligencia získala peniaze?" pýtajú sa, ako keby prvému *Homo sapiens* spadli dolárové bankovky z neba a on ich mohol použiť v supermarkete už existujúcom v pralese. Ľudský druh sa nenarodil do trhovej ekonomiky. Včely vám nepredajú med, ak im zaň ponúknete platbu kreditkou. Ľudský druh peniaze vymyslel, a peniaze existujú – pre nás, nie pre myši alebo osy – pretože v ne naďalej veríme.

Stále sa pokúšam vysvetliť ľuďom, že vzorom inteligencie nie je Dustin Hoffman vo filme *Rain Man*, ale že je to človek ako taký, bodka. Je to mäkká vec, ktorá vybuchne vo vákuu, avšak zanechala odtlačky nôh na svojom mesiaci. Vnútri tej sivej hrče je schopnosť prehľadávať cesty cez veľkú sieť kauzality, a nájsť trasu vedúcu k napohľad nemožnému – schopnosť, ktorú niekedy nazývame tvorivosť.

Ľudia – špeciálne investori – sa občas pýtajú, ako, ak Inštitút výskumu strojovej inteligencie úspešne zostaví skutočnú umelú inteligenciu, bude možné výsledky *speňažiť*. To je to, čo nazývame problém uhla pohľadu.

Alebo to je možno niečo hlbšie než obyčajná zrážka predpokladov. S trochou tvorivého myslenia si ľudia dokážu predstaviť, ako by sa dalo cestovať na Mesiace, alebo vyliečiť kiahne, alebo vyrábať počítače. Predstaviť si však trik, ktorý by dokázal dosiahnuť *všetky tieto veci zároveň*, sa im zdá celkom nemožné – napriek tomu, že takáto sila sídli iba pár centimetrov za ich vlastnými očami. Tá sivá vlhká hmota sa stále zdá byť záhadnou samotnej tejto sivej vlhkej hmote.

A tak, keďže ľudia nedokážu celkom vidieť, ako by to celé mohlo fungovať, sa sila inteligencie zdá menej skutočná. Predstaviť si ju je ťažšie než predstaviť si vežu z ohňa, ktorá pošle loď na Mars. Plán návštevy Marsu upútava našu predstavivosť. Ale keby niekto sľuboval aj návštevu Marsu, aj veľkú zjednotenú teóriu fyziky, aj dôkaz Riemannovej hypotézy, aj liek proti obezite, aj liek proti rakovine, aj liek proti starnutiu, aj liek proti hlúposti... to skrátka znie nesprávne.

A veru by aj malo. Je vážnym zlyhaním predstavivosti myslieť si, že inteligencia dokáže iba tak málo. Kto by si dokázal predstaviť, vtedy dávno, čo budú jedného dňa mysle robiť. Možno ešte ani *nevieme*, čo sú naše skutočné problémy.

Ale medzičasom, keďže je ťažké vidieť, ako môže mať jeden proces takéto rozmanité schopnosti, je ťažké si predstaviť, že by sa zároveň vyriešili čo len také bežné problémy ako obezita a rakovina a starnutie.

Lenže, jeden trik už raz vyliečil kiahne a postavil lietadlá a kultivoval pšenicu a skrotil oheň. Naša súčasná veda možno nemá jasno v tom, ako presne ten trik funguje, ale on napriek tomu funguje. Ak dočasne nerozumieme nejakému javu, je to fakt o danom stave našej mysle, nie fakt o danom jave. Prázdna mapa neznamená, že aj územie je prázdne. Aj keď človek celkom nerozumie tej sile, ktorá zanechala odtlačky nôh na Mesiace, tie odtlačky tam napriek tomu stále sú – skutočné odtlačky, na skutočnom Mesiace, umiestnené tam skutočnou silou. Keby človek tejto sile dostatočne hlboko porozumel, mohol by ju vytvárať a usmerňovať. Inteligencia je rovnako skutočná ako elektrina. Akurát je omnoho mocnejšia, omnoho nebezpečnejšia, má omnoho hlbšie dopady na pokračovanie rozvíjajúceho sa príbehu života vo vesmíre... a je o trochu ťažšie dôjsť na to, ako postaviť jej generátor.

Pôvodný článok¹

¹ <http://yudkowsky.net/singularity/power/>