

Ťažké je naozaj zmeniť názor

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

Kniha I.

Pravda a indície

Eliezer Yudkowsky

Obsah

Predslov	iii
I Predvídateľné chyby	1
1 Čo myslím slovom "rozumnosť"?	2
2 Rozumné cítenie	5
3 Načo pravda? A...	7
4 ...čo je teda skreslenie?	9
5 Dostupnosť	11
6 Príťažujúce podrobnosti	13
7 Klam plánovania	16
8 Ilúzia priehľadnosti: Prečo vám nikto nerozumie	19
9 Očakávame krátke inferenčné vzdialenosti	21
10 Objektív, ktorý vidí svoje chyby	23
II Falošné názory	25
11 Nech názory platia nájomné (v očakávaných vnemoch)	26
12 Bájka o vede a politike	28
13 Viera vo vieru	31
14 Bayesovské džudo	34
15 Vyznávanie a fandanie	35
16 Signály na potlesk	37
17 Názor ako uniforma	39
18 Predstieranie múdrosti	40
19 Nárok náboženstva na nevyvrátiteľnosť	43
III Všímanie si zmätku	46
20 Čo je indícia?	47
21 Vedecká indícia, právna indícia, rozumná indícia	49
22 Koľko indície treba?	51
23 Occamova britva	53
24 Einsteinova drzosť	56
25 Zameraj svoju neistotu	58
26 Cnosť presnosti	60
27 Tvoja sila racionalistu	62
28 Neprítomnosť indície je indíciou neprítomnosti	64
29 Zákon zachovania očakávanej indície	66
30 Spätný pohľad znehodnocuje vedu	68
IV Tajomné odpovede	70
31 Falošné vysvetlenia	71
32 Hádanie učiteľovho hesla	73
33 Veda ako uniforma	76
34 Falošná kauzalita	78
35 Sémantické stopky	81
36 Tajomné odpovede na tajomné otázky	83
37 Márnosť emergencie	86
38 Nehovorte "zložitost'"	88
39 Pozitívne skreslenie: Pozeraj do tmy	91
40 Moja divoká a bezstarostná mladosť	93
41 Neučíme sa z histórie	95
42 Sprístupniť históriu	96
43 Vysvetli / Uctievaj / Ignoruj?	98
44 "Veda" ako zastavovač zvedavosti	100
45 Čo myslím slovom "rozumnosť"?	102
V Medzihra	105
46 Jednoduchá pravda	106

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 I. Predvídateľné chyby

Obsah

1 Čo myslím slovom “rozumnosť”?	2
2 Rozumné cítenie	5
3 Načo pravda? A...	7
4 ...čo je teda skreslenie?	9
5 Dostupnosť	11
6 Priťažujúce podrobnosti	13
7 Klam plánovania	16
8 Ilúzia priehľadnosti: Prečo vám nikto nerozumie	19
9 Očakávame krátke inferenčné vzdialenosti	21
10 Objektív, ktorý vidí svoje chyby	23

Kapitola 1. Čo myslím slovom “rozumnosť”?

Myslím tým nasledujúce:

1. **Epistemická racionalita (rozumné poznanie):** systematické zvyšovanie správnosti vašich názorov. Umenie veriť a vytvárať si nové názory na základe indícií (“aktualizovanie”) tak, že vaša mapa zodpovedá územiu¹ tak dobre, ako môže. Táto zhoda medzi názorom a skutočnosťou sa všeobecne nazýva “pravda” a ja tento názov ochotne používam.
2. **Inštrumentálna racionalita (rozumné konanie):** systematické dosahovanie svojich hodnôt. Umenie vyberanie si akcií, ktoré kormidlujú budúcnosť smerom k výsledkom hodnoteným vyššie na vašom rebríčku preferencií. Toto občas nazývam “víťazstvo”.

Rozumnosť sa teda týka vytvárania pravdivých názorov a robenia víťazných rozhodnutí. “Víťazstvo” tu nemusí prichádzať na úkor druhých; životný projekt sa môže týkať spolupráce alebo sebaobetovania namiesto súťaže. “Vaše hodnoty” tu znamenajú *čokoľvek, na čom vám záleží*, vrátane druhých ľudí. Nie je to obmedzené na *sebecké* alebo *nie spoločné* hodnoty.

Ak vám to pripadá ako úplne dobrá definícia, môžete zvyšok preskočiť. V opačnom prípade pokračujte.

Keď ľudia povedia: “X je rozumné!”, zvyčajne je to iba dôraznejší spôsob ako povedať: “Myslím si, že X je pravda” alebo: “Myslím si, že X je dobré.” Načo je nám teda ďalšie slovo “rozumný”, keď už máme slová “pravdivý” a “dobrý”? Aby sme mohli hovoriť o systematických metódach na dosahovanie pravdy a víťazstva - a systematických spôsoboch, ako za týmito štandardmi zaostávame.

Niekedy experimentálni psychológovia odhalia ľudské uvažovanie, ktoré vyzerá veľmi čudne. Napríklad² niekto hodnotí pravdepodobnosť výroku “Bill hrá džez” ako *nižšiu* než pravdepodobnosť výroku “Bill je účtovník, ktorý hrá džez”. To vyzerá ako čudný úsudok, keďže ľubovoľný konkrétny účtovník hrajúci džez je zrejme hráčom džezu. Na akú vyššiu moc sa však odvolávame tvrdením, že tento úsudok je *nesprávny*?

Experimentálni psychológovia používajú dva zlaté štandardy: *teóriu pravdepodobnosti* a *teóriu rozhodovania*. Nech “P(niečo)” znamená “pravdepodobnosť, že sa niečo stane”. Keďže je všeobecným zákonom teórie pravdepodobnosti, že $P(A) \geq P(A \& B)$, úsudok $P(\text{“Bill hráva džez”}) < P(\text{“Bill hráva džez”} \& \text{“Bill je účtovník”})$ je označený ako nesprávny.

Z technického hľadiska by ste povedali, že takýto pravdepodobnostný úsudok je *nebayesovský*. Názory, ktoré sú v súlade s koherentnou distribúciou pravdepodobnosti a rozhodnutia, ktoré maximalizujú očakávanú pravdepodobnosť koherentnej funkcie úžitku, sa nazývajú “bayesovské”.

Problém, čo v praxi označovať ako “rozumné”, sme týmto celkom nevyčerpali, najmä z dvoch dôvodov:

Po prvé, bayesovské formalizmy vo svojej plnej podobe sú pre väčšinu problémov zo skutočného života výpočtovo nezvládnuteľné. Nikto nedokáže *naozaj* spočítať

¹ pozri: 46 Jednoduchá pravda

² http://lesswrong.com/lw/ji/conjunction_fallacy/

všetku tú matematiku a riadiť sa ňou, rovnako ako nedokážete predpovedať pohyb cien na burze vypočítaním pohybu kvarkov.

Preto existuje celá internetová stránka s názvom "LessWrong" (po slovensky: menej chybné) namiesto jednej strany, ktorá by jednoducho uviedla formálne axiomy a basta. Existuje ešte celé umenie hľadania pravdy a dosahovania hodnôt *zvnútra ľudskej mysle*: musíme spoznať svoje vlastné chyby, prekonať svoje skreslenia, vyvarovať sa sebaklamu, dostať sa do dobrého emocionálneho stavu, aby sme čelili pravde a urobili, čo treba, atď. atď. a tak ďalej.

Po druhé, niekedy sa spochybňuje význam samotnej matematiky. Presné pravidlá teórie pravdepodobnosti sú spochybňované napríklad antropickými problémami³, v ktorých je nejasný počet pozorovateľov. Presné pravidlá teórie rozhodovania sú spochybňované napríklad problémami ako je Newcombov⁴, v ktorých iní aktéri môžu predpovedať vaše rozhodnutie skôr než ho urobíte.

V takýchto prípadoch je márne pokúšať sa problém uzavrieť vymyslením nejakej novej definície slova "rozumný" a hovoriť: "A preto je moja obľúbená odpoveď z definície 'rozumná'." To jednoducho vyvolá otázku, prečo by niekto mal venovať pozornosť vašej definícii. Ja sa nezaujímam o teóriu pravdepodobnosti kvôli tomu, že je to sväté písmo, ktoré nám priniesol Laplace. Zaujímam sa o aktualizovanie názorov bayesovským spôsobom (s occamovskými pôvodnými pravdepodobnosťami⁵), pretože očakávam, že takéto myslenie nás systematicky približuje, veď viete, k *správnosti*, k mape, ktorá odráža územie. (Viac o márnosti argumentovania "z definície" tu⁶ a tu⁷.)

A potom existujú otázky "Ako myslieť", ktoré sa nezdajú dostatočne zodpovedané ani teóriou pravdepodobnosti, ani teóriou rozhodovania - napríklad otázka, aký mať pocit z pravdy, keď už ju máte⁸. Aj tu, pokúsiť sa definovať "rozumnosť" konkrétnym spôsobom nezdôvodňuje odpoveď, iba ju podsúva.

Nejdem sa tu hádať o význam slova⁹, ani keď je tým slovom "rozumnosť". Zmyslom priradovania postupností písmen ku konkrétnym pojmom je umožniť dvom ľuďom *komunikovať*¹⁰ - pomôcť dopraviť myšlienky z jednej mysle do druhej. Pomocou manipulácie, ktoré významy patria ku ktorým slovám, nemôžete zmeniť skutočnosť ani dokázať myšlienku.

Ak teda rozumieme, na ktorý pojem sa *zhruba odvolávam* týmto slovom "rozumnosť", a na ktoré podvýrazmi "rozumné myslenie" a "rozumné konanie", potom *sme komunikovali*: dosiahli sme všetko, čo sa dalo dosiahnuť rozprávaním, ako definovať "rozumnosť". Čo nám ešte zostáva na diskusiu, nie je aký význam priradíme jednotlivým slabikám "ro-zum-nosť"; zostáva nám diskutovať o tom, *ako dobre myslieť*.

Keď sme si toto povedali, je dôležité dodať, že ja a niektorí ďalší ľudia považujeme za prinajmenšom *sporné* všetky konštrukcie "rozumnosti", ktoré ju robia *nie normatívnou*:

Napríklad, ak poviete: "Rozumný (epistemicky) názor je X, ale pravdivý názor je Y," potom pravdepodobne používate slovo "rozumný" na označenie niečoho iného než na čo myslím ja. (Napríklad "rozumnosť" by mala byť *konzistentná po úvahe* - "rozumne" sa pozrieť na indíciu a "rozumne" zväžiť, ako vaša myseľ

³ <http://www.anthropic-principle.com/primer.html>

⁴ pozri: 291 Newcombov problém a lutovanie rozumnosti

⁵ pozri: 23 Occamova britva

⁶ pozri: 151 Podobenstvo o bolehlave

⁷ pozri: 169 Argumentovanie "z definície"

⁸ pozri: 2 Rozumné čítanie

⁹ pozri: 160 Háďanie sa o definíciách

¹⁰ pozri: 162 Argumentovanie štandardným používaním

spracováva túto indíciu, by nemalo viesť k dvom rôznym záverom.) Podobne, ak zistíte, že hovoríte: "Rozumné (inštrumentálne) je urobiť X, ale správna vec je urobiť Y," potom takmer určite používate niektoré zo slov "rozumné" alebo "správne" spôsobom, s ktorým nesúhlasím.

V takomto prípade - alebo v hociktorom inom prípade, kde hrozí spor - by ste mali dosadiť konkrétnejšiu reč¹¹: "Sebe by som najviac prospel útekem preč, ale dúfam, že by som sa aspoň pokúsil odtiahnuť to dievča z kolajnic" alebo "Kauzálna teória rozhodovania je zvyčajne formulovaná tak, že by ste podľa nej mali vziať dve krabice v Newcombovom probléme¹², ale ja by som radšej mal milión dolárov."

Slovo "rozumný" má svoje možné úskalia, ale je kopec *nie* hraničných prípadov, kde slovo "rozumný" veľmi dobre *komunikuje*, o čo človeku ide; podobne aj "nerozumný". V takýchto prípadoch sa ho nebojím používať.

Človek by si však mal dať pozor, aby to slovo nepoužíval *príliš*. Nedostávame žiadne body za to, že ho iba nahlas vyslovujeme. Ak priveľa hovoríte o Ceste, nedosiahnete ju.

Pôvodný článok s komentármi¹³

¹¹ pozri: 164 Nech sú vaše slová tabu

¹² pozri: 291 Newcombov problém a ľutovanie rozumnosti

¹³ http://lesswrong.com/lw/31/what_do_we_mean_by_rationality/

Kapitola 2. Rozumné cítenie

O "rozumnosti" ľudia radi hovoria, že odporuje všetkým emóciám - že všetky naše žiaľe a všetky naše radosti sú automaticky proti logike, pretože sú to *city*. Pritom je dosť čudné, že v teórii pravdepodobnosti neviem nájsť žiadnu vetu dokazujúcu, že by som sa mal tváriť ľadovo chladne a bez výrazu.

Je teda rozumnosť nezávislá od cítenia? Nie; naše emócie vyplývajú z našich modelov skutočnosti. Ak budem veriť, že môjho mŕtveho brata našli živého, budem šťastný; ak sa zobudím a uvedomím si, že to bol sen, budem smutný. P. C. Hodgell povedal: "Čo môže byť zničené pravdou, nech sa zničí." Šťastie môjho snívajúceho ja odporovalo pravde. Môj smútok po zobudení je rozumný; nie je žiadna pravda, ktorá by ho zničila.

Rozumnosť začína tým, že sa pýtame, aký je svet, ale šíri sa ako vírus do každej ďalšej myšlienky, ktorá závisí na tom, aký si myslíme, že svet je. Keď hovorím o vašich názoroch na to, "aký je svet", myslím tým hocičo, čo veríte, že existuje v skutočnosti, hocičo, čo buď existuje alebo neexistuje, ľubovoľný prvok množiny "vecí, ktoré môžu spôsobiť iné veci." Ak veríte, že vo vašej skrini je škriatok, ktorý vám zväzuje šnúrky na topánkach, potom je to názor na to, aký je svet. Vaše topánky sú skutočné - môžete ich zdvihnúť. Ak existuje niečo, čo môže chytiť a zviazať vaše šnúrky, aj to musí byť skutočné, časť rozsiahlej siete príčin a následkov, ktorú voláme "vesmír".

Hnevať sa na škriatka, ktorý vám zviazal šnúrky, zahrňa stav mysle, ktorý nie je *iba* o tom, aký je svet. Predpokladajme, že ako buddhistu, alebo pacienta po lobotómii, alebo skrátka veľmi flegmatického človeka by vás zistenie, že máte zviazané šnúrky, nenahnevalo. Neovplyvňovalo by to vaše očakávania o svete - stále by ste očakávali, že otvoríte skriňu a nájdete šnúrky zviazané. Váš hnev ani pokoj by tu nemali ovplyvniť váš odhad, lebo čo sa deje vo vašej skrini, nezávisí od emocionálneho stavu mysle; aj keď môže byť náročné myslieť takto jasne.

Ale pocit hnevu je spojený so stavom mysle, ktorý *je* o tom, aký je svet; hneváte sa, *pretože* si myslíte, že vám škriatok zviazal šnúrky. Kritérium rozumnosti sa šíri ako vírus, od počiatkovej otázky, či vám škriatok zviazal alebo nezviazal šnúrky, až do výsledného hnevu.

Stať sa rozumnejším - získavať lepšie odhady o tom, aký je svet - môže pocity zmierniť *alebo* zosilniť. Niekedy utekáme pred silnými citmi tým, že popierame fakty, že cívame pred pohľadom na svet, ktorý viedol k tejto silnej emócií. Ak je to tak, potom ako budete študovať zručnosti rozumnosti a cvičiť sa v nepopieraní faktov, vaše *city* zosilnejú.

V začiatkoch som si nikdy nebol celkom istý, či je *v poriadku* cítiť niečo silne - či je to dovolené, či je to správne. Nemyslím si, že tento zmätok vznikol len z môjho mladistvého nechápania rozumnosti. Všimol som si podobné ťažkosti u ľudí, ktorí sa ani nepokúšajú byť racionalistami; keď sú šťastní, pochybujú, či naozaj smú byť šťastní, a keď sú smutní, nie sú si celkom istí, či pred touto emóciou utekať alebo nie. Prinajmenšom od Sokratových čias, a asi aj dávno pred ním, ste mohli pôsobiť vzdelane a sofistickovane tak, že ste nikdy pred nikým nedali najavo, že vám na niečo veľmi záleží. Cítiť je *hanba* - v slušnej spoločnosti sa to skrátka nerobí. Mali by ste vidieť, ako čudne sa na mňa ľudia pozerajú, keď si uvedomia, ako veľmi mi záleží na rozumnosti. Nemyslím si, že je to kvôli nezvyčajnej téme, ale že nie sú zvyknutí vidieť duševne zdravých dospelých ľudí, ktorým na *niečom* viditeľne záleží.

Teraz však viem, že na silných pocitoch nie je nič zlé. Odkedy som si osvojil pravidlo: "Čo môže byť zničené pravdou, nech sa zničí", uvedomil som si aj:

“Čo pravda živí, nech prekvitá.” Keď sa stane niečo dobré, som šťastný, a nemám v hlave zmätok, či je rozumné, aby som bol šťastný. Keď sa stane niečo hrozné¹, neutekám pred smútkom hľadaním falošnej útechy a falošnej nádeje. Predstavujem si minulosť a budúcnosť ľudstva, desiatky miliárd smrtí počas dejín, biedu a strach, hľadanie odpovedí, trasúce sa ruky siahajúce nahor z tolkej krvi, čím by sme sa jedného dňa mohli stať, až z hviezd urobíme svoje mestá, všetku tú temnotu a všetko to svetlo - viem, že to nikdy nedokážem naozaj pochopiť, a že to neviem povedať slovami. Napriek celej svojej filozofii sa stále hanbím priznávať silné pocity, a vám je pravdepodobne nepríjemné počúvať o nich. Ale teraz už viem, že cítiť je rozumné.

Pôvodný článok s komentármi²

¹ <http://yudkowsky.net/yehuda.html>

² http://lesswrong.com/lw/hp/feeling_rational/

Kapitola 3. Načo pravda? A...

Niektoré komentáre na tomto blogu sa týkali otázky, prečo by sme mali hľadať pravdu. (Našťastie väčšina nespochybňovala, čo to je pravda¹.) Naša motivácia nastaviť svoje myšlienky na rozumnosť, ktorá určuje, či je dané nastavenie “dobré” alebo “zlé”, sa odvíja z dôvodu, prečo vlastne tú pravdu chceme hľadať.

Je napísané: “Prvou cnosťou je zvedavosť.” Zvedavosť je jedným dôvodom hľadať pravdu, a hoci nie je jediným dôvodom, má istú zvláštnu a obdivuhodnú čistotu. Ak je vaším motívom zvedavosť, budete otázky uprednostňovať podľa toho, ako tieto otázky samotné šteklija váš osobný estetický zmysel. Zamotanejšia otázka s väčšou pravdepodobnosťou neúspechu môže byť hodna viac úsilia než jednoduchšia, skrátka preto, lebo je zábavnejšia.

Tuším, že niektorí ľudia namietnu, že zvedavosť je emócia a preto je “nie rozumná”. Ja označujem emóciu za “nie rozumnú” vtedy, keď sa opiera o mylné názory, alebo presnejšie, o nerozumné poznávacie správanie: “Ak sa k tvojej tvári blíži železo, o ktorom veríš, že je horúce, avšak ono je chladné, Cesta odporuje tvojmu strachu. Ak sa k tvojej tvári blíži železo, o ktorom veríš, že je chladné, avšak ono je horúce, Cesta odporuje tvojmu pokoji.” A naopak, emócia vyvolaná správnymi názormi alebo rozumnými poznávacími úvahami je “rozumná emócia”; čo má tú výhodu, že aj pokoj môžeme považovať za emocionálny stav a nie za privilegovaný štandard. Keď si ľudia myslia, že “emócie” a “rozum” sú protiklady, mám podozrenie, že v skutočnosti myslia na Systém 1 a Systém 2 - rýchle úsudky založené na vnímaní verzus pomalé úsudky založené na uvažovaní. Úsudky založené na uvažovaní nie sú vždy pravdivé, ani úsudky založené na vnímaní nie sú vždy nepravdivé; je teda veľmi dôležité rozoznávať toto rozdelenie od “rozumnosti”. Oba systémy môžu slúžiť pravde, alebo proti nej bojovať, podľa toho, ako sa použijú.

Okrem čirej emocionálnej zvedavosti, aké ďalšie môžu byť motívy túžiť po pravde? No, možno chcete dosiahnuť nejaký konkrétny cieľ v skutočnom svete, napríklad postaviť lietadlo, a preto potrebujete vedieť nejakú konkrétnu pravdu o aerodynamike. Alebo prostejšie, chcete čokoládové mlieko a preto chcete vedieť, či v miestnych potravinách majú čokoládové mlieko, aby ste sa rozhodli, či pôjdete tam alebo niekam inam. Ak chcete pravdu kvôli tomuto, potom budete otázky uprednostňovať podľa očakávaného úžitku ich informácie - nakoľko prípadné odpovede ovplyvnia vaše rozhodnutia, nakoľko na vašich rozhodnutiach záleží, nakoľko očakávate, že nájdete odpoveď, ktorá zmení vaše rozhodnutia oproti terajším.

Hľadať pravdu iba pre jej praktické využitie sa môže zdať neslušné - nemali by sme túžiť po pravde iba kvôli nej samotnej? - avšak takéto hľadanie je mimoriadne dôležité, pretože vytvára vonkajšie overovacie kritérium: ak sa vaše lietadlo zrúti z oblohy, alebo ak pôjdete do obchodu a nebudú mať čokoládové mlieko, je to znamenie, že ste niečo spravili zle. Dostávate spätnú väzbu, ktoré spôsoby myslenia fungujú, a ktoré nie. Čistá zvedavosť je nádherná vec, ale nemusí vydržať dosť dlho na to, aby overila svoje odpovede, keď už príťažlivosť tajomstva pominula. Zvedavosť ako ľudská emócia tu bola už dávno pred starovekými Grékmi. Na cestu Vedy však ľudstvo pevne postrčilo uvedomenie, že niektoré spôsoby myslenia odhaľujú názory, pomocou ktorých môžeme *meniť svet*. Čo sa týka čistej zvedavosti, túto túžbu rovnako dobre uspokojovalo aj rozprávanie o bohoch a hrdinoch pri táboráku, a nikto si neuvedomil, že je na tom niečo zle.

¹ pozri: 46 Jednoduchá pravda

Existujú aj motívy hľadania pravdy iné ako zvedavosť a pragmatizmus? Tretím dôvodom, ktorý mi napadá, je morálka: Veríte, že hľadať pravdu je vznešené a dôležité a hodnotné. Hoci aj takýto ideál pripisuje pravde vnútornú hodnotu, tento stav mysle sa veľmi odlišuje od zvedavosti. Byť zvedavý, čo je za oponou, nie je rovnaký pocit ako veriť, že máte morálnu povinnosť tam pozrieť. V tomto druhom stave mysle máte väčší sklon veriť, že aj niekto *iný* by sa mal pozrieť za oponu, prípadne ho karhať, ak úmyselne zavrie oči. Preto označujem ako "morálku" názor, že hľadanie pravdy je pragmaticky dôležité "pre spoločnosť", a preto je to povinnosť nás všetkých. Pri tejto motivácii budete otázky uprednostňovať podľa vašich ideálov o tom, ktoré pravdy sú najdôležitejšie (nie najužitočnejšie ani najzaujímavejšie); alebo podľa vašich morálnych ideálov o tom, kedy a za akých okolností je povinnosť hľadať pravdu najsilnejšia.

Mám sklon podozrievať morálku ako motiváciu k rozumnosti, *nie* preto, že by som odmietal tento morálny ideál, ale preto, že privoláva určité problémy. Je príliš ľahké osvojiť si, ako naučené morálne povinnosti, spôsoby myslenia, ktoré sú hroznými prešlapmi pri tanci. Predstavte si pána Spocka z seriálu *Star Trek*, naivný archetyp rozumnosti. Spockov emocionálny stav je vždy nastavený na "pokoj", aj keď je to celkom neprimerané. Často udáva veľa platných číslíc pri pravdepodobnostiach, ktoré sú hrubo nekalibrované. (Napríklad: "Kapitán, ak zamierite s Enterprise priamo do tejto čiernej diery, naša pravdepodobnosť prežitia je iba 2,234%." Napriek tomu, v deviatich prípadoch z desiatich Enterprise vyviazne. Aký tragický blázon udáva štyri platné číslice pre údaj, ktorý je o dva rády mimo?) Napriek tomu si podľa tohto známeho obrazu mnoho ľudí predstavuje povinnosť byť "rozumný" – nečudo, že sa s ňou vrelo nestotožňujú. Urobiť z rozumnosti morálnu povinnosť znamená dať jej všetky hrozné stupne voľnosti svojvoľného kmeňového zvyku. Ľudia dôjdu k nesprávnej odpovedi, a potom rozhorčene namietajú, že konali správne, namiesto aby sa zo svojej chyby poučili.

A predsa, ak máme *zlepšiť* svoje zručnosti rozumnosti, prekročiť výkonové štandardy lovcov a zberačov, potrebujeme rozvážne názory na to, ako správne myslieť. Keď si pre seba píšeme nové myšlienkové programy, začínajú v Systéme 2, systéme rozvážnosti, a až pomaly – ak vôbec – sa cvičením dostanú do nervových obvodov riadiacich Systém 1. Ak teda existujú určité druhy myslenia, o ktorých zistíme, že sa im chceme *vyhnúť* – napríklad skreslenia – musia byť na úrovni Systému 2 reprezentované ako zákazy myslieť takýmto spôsobom; ako deklarovaná povinnosť vyhýbania sa.

Ak chceme poznať pravdu, najúčinnnejšie to dosiahneme myslením určitými spôsobmi namiesto iných; a toto sú techniky rozumnosti. Niektoré z týchto techník rozumnosti zahŕňajú prekonávanie určitého druhu prekážok, skreslenia...

Pôvodný článok s komentármi²

² http://lesswrong.com/lw/go/why_truth_and/

Kapitola 4. ...čo je teda skreslenie?

Skreslenie je určitý druh prekážky na ceste k nášmu cieľu dosiahnuť pravdu – považujeme ho za “prekážku”, pretože máme pravdu ako cieľ – ale existujú aj mnohé prekážky, ktoré nie sú “skreslenia”. Keby sme rovno začali otázkou: “Čo je to skreslenie?”, išli by sme na to z nesprávnej strany. Ako hovorí príslovie: “Existuje štyridsať druhov bláznovstva, ale iba jeden druh zdravého rozumu.” Pravda je úzky terč, malá oblasť konfiguračného priestoru, ktorú chceme zasiahnuť. “Lúbi ma alebo nelúbi?” môže byť binárna otázka, ale “ $E=MC^2$ ” je malá bodka v priestore všetkých rovníc, tak ako vyhrávajúci žreb lotérie v priestore všetkých žrebov. Chyba nie je výnimočná situácia; práve úspech je *a priori* taký nepravdepodobný, že si vyžaduje vysvetlenie.

Nezačíname morálnou povinnosťou “znižovať skreslenia”, pretože skreslenia sú škaredé a zlé a Skrátko To Nerobíme. S takýmto zmýšľaním môže človek skončiť, ak nadobudol deontologickú povinnosť “rozumnosti” spoločenskou osmózou, ktorá vedie ľudí k snahe vykonávať techniky bez docenenia ich dôvodov. (A práve toto je škaredé a zlé a Skrátko To Nerobíme, podľa knihy *To nemyslíte vážne, pán Feynman*, ktorú som čítal ako dieťa.)

My skôr chceme dosiahnuť pravdu, z hocijakého dôvodu, a na ceste k svojmu cieľu nachádzame rôzne prekážky. Tieto prekážky si nie sú navzájom celkom nepodobné – existujú napríklad prekážky súvisiace s nedostatkom dostupnej výpočtovej sily alebo s vysokou cenou informácií. Zhodou okolností jedna veľká skupina prekážok vyzerá, že má určitú spoločnú povahu – tvorí zhluk v priestore prekážok na ceste k pravde – a tento zhluk sme nazvali “skreslenia”.

Čo je to skreslenie? Vieme sa pozrieť na tento empirický zhluk a nájsť presný test členstva? Možno zistíme, že to naozaj nevieme vysvetliť lepšie, než ukázať na pár reprezentatívnych príkladov a dúfať, že poslucháč porozumie. Ak ste vedec, ktorý práve začína skúmať oheň, môže byť omnoho múdrejšie ukázať na táborák a povedať: “Oheň je tamtá oranžovosvetlá horúca vec,” namiesto povedania: “Definujem oheň ako alchymickú transmutáciu substancií, ktoré uvoľňujú flogiston.” Ako som povedal v Jednoduchéj pravde¹, nemali by ste niečo ignorovať iba preto, že to neviete definovať. Ja nedokážem spamäti zacitovať rovnice všeobecnej teórie relativity, ale napriek tomu, ak skočím z útesu, spadnem. A to isté môžeme povedať o skresleniach – nedopadnú na nás o nič menej tvrdo, ak sa ukáže, že nevieme presne definovať, čo to “skreslenie” vlastne je. Môžeme teda ukázať na omyl konjunkcie, prehnanú dôveru, heuristiku reprezentatívnosti, ignorovanie základnej miery a povedať: “Takéto veci.”

S prihliadnutím na horeuvedené, zdá sa, že označujeme ako “skreslenia” tie prekážky na ceste k pravde, ktoré nespôsobuje cena informácie, ani obmedzená výpočtová sila, ale tvar nášho vlastného myšlienkového mechanizmu. Tento mechanizmus je napríklad evolučne optimalizovaný na ciele, ktoré aktívne bránia presnosti poznania; napríklad mechanizmus na vyhrávanie hádok prispôsobený na politické situácie. Alebo výberový tlak presnosť poznania odkláňa; napríklad veríme tomu, čomu veria druhí, aby sme spoločensky zapadli. Alebo, pri klasických heuristikách a skresleniach, mechanizmus funguje podľa známeho algoritmu, ktorý prináša nejaký úžitok, ale produkuje aj systematické chyby: samotná heuristika dostupnosti nie je skreslenie, ale spôsobuje známe, presne opísateľné skreslenia. Naše mozgy robia niečo nesprávne, a po mnohých pokusoch a náročnom premýšľaní niekto opíše tento problém spôsobom, ktorému

¹ pozri: 46 Jednoduchá pravda

System 2 dokáže porozumieť; potom to nazývame "skreslenie". Aj keby sme nemali žiaden lepší spôsob poznávania, stále je to chyba, ktorú pomenovateľným spôsobom vytvára konkrétny druh poznávacieho mechanizmu - nie pretože toho mechanizmu máme primálo, ale kvôli samotnému tvaru tohto mechanizmu.

Skreslenia odlišujeme od chýb vychádzajúcich z poznávacieho obsahu, ako sú prijaté názory alebo prijaté morálne povinnosti. Tie nazývame skôr "chyby" než "skreslenia", a je omnoho ľahšie ich napraviť, keď si ich na sebe všimneme. (Zdrojom tejto chyby, alebo zdrojom zdroja tejto chyby, však v konečnom dôsledku môže byť nejaké skreslenie.)

Skreslenia odlišujeme od chýb spôsobených poškodením jednotlivého ľudského mozgu, alebo prijatými kultúrnymi normami; skreslenia spôsobuje mechanizmus, ktorý majú všetci ľudia spoločný.

Platón nemal "skreslenie", keď nepoznal teóriu relativity - nemal túto informáciu ako získať, jeho nevedomosť nespôboval tvar jeho myšlienkového mechanizmu. Ale ak Platón veril, že z filozofov by boli lepší králi, pretože on sám bol filozof - a ak tento názor vznikol vďaka všeobecnému politickému inštinktu prispôbenému na sebaapresadzovanie, a nie preto, lebo Platónovi jeho otecko hovoril, že každý má morálnu povinnosť presadzovať vládu svojej profesie, ani preto, lebo Platón v detstve čuchal priveľa lepidla - potom to bolo skreslenie, či už na to Platóna niekedy niekto upozornil alebo nie.

Nemusí byť ľahké napraviť skreslenia. Možno sa ani napraviť nedajú. Ale keď sa pozrieme na svoj vlastný myšlienkový mechanizmus a vidíme kauzálne vysvetlenie pomenovateľnej triedy chýb; a keď to vyzerá, že tento problém spôsobuje vyvinutý tvar tohto mechanizmu, a nedostatok tohto mechanizmu alebo jeho konkrétny zlý obsah; potom to nazývame skreslenie.

Ja osobne vidím našu snahu ako nadobúdanie osobných zručností rozumnosti, ako zlepšovanie techniky hľadania pravdy. Úlohou je dosiahnuť pozitívny cieľ pravdy, nie vyhnúť sa negatívnemu cieľu chyby. Priestor chýb je rozsiahly, nekonečné množstvá nekonečne rozmanitých chýb. Je ťažké opísať taký obrovský priestor: "Čo je pravdou o jednom jablku, nemusí byť pravdou o druhom jablku; preto sa o jednom jablku dá povedať viac než o všetkých jablkách na svete." Priestor úspechu je užší, preto sa o ňom dá viac povedať.

Aj keď (ako vidíte) nie som proti diskusii o definíciách, mali by sme si pamätať, že to nie je náš prvoradý cieľ. Sme tu preto, aby sme pokračovali vo veľkej ľudskej snahe o pravdu: pretože vedomosti zúfalo potrebujeme, a navyše, sme zvedaví. Pre tento cieľ sa usilujeme prekonávať rôzne prekážky na ceste, či už ich voláme "skreslenia" alebo nie.

Pôvodný článok s komentármi²

² http://lesswrong.com/lw/gp/whats_a_bias_again/

Kapitola 5. Dostupnosť

Heuristika dostupnosti je posudzovanie frekvencie alebo pravdepodobnosti udalosti podľa toho, ako ľahko si vieme spomenúť na jej príklady.

Známa štúdia z roku 1978 od Lichtensteina, Slovic, Fischhoffa, Laymana a Combsa, "Odhadovaná frekvencia smrteľných udalostí", študovala chyby pri vyčíslovaní vážnosti rizík alebo odhadovaní, ktoré z dvoch nebezpečenstiev nastáva častejšie. Pokusné osoby si mysleli, že nehody spôsobujú zhruba rovnako veľa úmrtí ako choroby; mysleli si, že vražda je častejšou príčinou smrti než samovražda. V skutočnosti choroby spôsobujú asi 16-krát viac smrtí ako nehody a samovražda je dvakrát častejšia než vražda.

Zrejma hypotéza pre tieto skreslené názory je, že o vraždách sa hovorí viac ako o samovraždách – preto si človek skôr spomenie, že počul o vražde, než o samovražde. Nehody sú dramatickejšie než choroby – možno preto si ich ľudia skôr zapamätajú, alebo si na ne skôr spomenú. V roku 1979 nasledujúca štúdia od Combsa a Slovic ukázala, že skreslené odhady pravdepodobnosti silno korelujú (0,85 a 0,89) so skreslenými frekvenciami správ v dvoch novinách. To nerieši, či si na vraždy ľahšie spomíname, pretože sa o nich viac píše, alebo či o vraždách noviny viac píše, pretože sú farbistejšie (a teda sa lepšie zapamätajú). Ale každopádne tu účinkuje skreslenie dostupnosti. Selektívne spravodajstvo je jedným z hlavných zdrojov skreslení dostupnosti. V pravom prostredí ste väčšinu z toho, o čom ste vedeli, zažili osobne; alebo ste o tom počuli priamo od súkmeňovca, ktorý to videl. Medzi vami a danou udalosťou bola zvyčajne nanajvyš jedna vrstva selektívneho spravodajstva. Na dnešnom internete môžete vidieť správy, ktoré cestou k vám prešli rukami šiestich blogerov – šesť postupných filtrov. V porovnaní s našimi predkami žijeme vo väčšom svete, v ktorom sa omnoho viac stane a omnoho menšia časť z toho dôjde k nám – omnoho silnejší efekt výberu, ktorý môže vytvárať omnoho väčšie skreslenia dostupnosti.

V skutočnom živote je nepravdepodobné, že niekedy stretnete Billa Gatesa. Vďaka selektívnemu spravodajstvu médií však môžete byť v pokušení porovnávať svoje životné úspechy s ním – a utrieť príslušnú hedonistickú pokutu. Objektívna frekvencia Billa Gatesa je 0,000 000 000 15, ale počujete o ňom omnoho častejšie. Na druhej strane, 19% ľudí na tejto planéte žije z menej než jedného dolára denne, ale pochybujem, že každý piaty článok, ktorý čítate na internete, napísali oni.

Používanie dostupnosti asi vytvára skreslenie absurdity¹; nespomínate si na udalosti, ktoré sa nestali, a preto ich pravdepodobnosť považujete za nulovú. Ak v poslednej dobe nenastali záplavy (ale pravdepodobnosti sa stále dajú pomerne ľahko vypočítať), ľudia si odmietajú kupovať poistenie proti záplavám dokonca aj keď je silno dotované a jeho cena je hlboko pod skutočnou poistnou hodnotou. Kunreuther a kolektív (1993) naznačuje, že za podceňovanie hrozby záplavy môže "neschopnosť jednotlivcov predstaviť si záplavy, ktoré sa nikdy nestali... Ľudia na záplavových územiach sú do veľkej miery väzňami svojej skúsenosti... Nedávne záplavy zrejme nastavujú hornú hranicu výšky škody, o ktorej si menežeri myslia, že by ich mala znepokojovať."

Burton a kol. (1978) uvádza, že keď sa postavia priehradami a protizáplavové bariéry, zníži sa frekvencia záplav a tým sa zrejme vytvorí falošný pocit bezpečnosti, ktorý vedie k zníženej opatrnosti. Hoci postavenie priehrad znižuje frekvenciu záplav, škoda *pripadajúca na jednu záplavu* je potom o toľko väčšia, že priemerné ročné škody *narastú*. Múdry človek by zo spomienky na malé riziká

¹ http://lesswrong.com/lw/j4/absurdity_heuristic_absurdity_bias/

extrapoloval možnosť veľkých rizík. Namiesto toho sa zdá, že zážitok malých rizík nastavuje vnímanú hornú hranicu rizika. Spoločnosť dobre chránená proti malým nebezpečenstvám nekoná žiadne akcie proti veľkým nebezpečenstvám, keď sú pravidelné malé záplavy odstránené, začne sa na záplavových rovinách stavať. Spoločnosť vystavená pravidelným malým nebezpečenstvám berie tieto malé nebezpečenstvá ako hornú hranicu nebezpečenstva, chráni sa proti pravidelným malým záplavám, ale nie proti občasným veľkým.

Pamäť nie je vždy dobrým radcom ohľadom pravdepodobností v minulosti, a už vôbec nie v budúcnosti.

Burton, I., Kates, R. a White, G. 1978. *Životné prostredie ako nebezpečenstvo*. New York: Oxford University Press.

Combs, B. a Slovic, P. 1979. Príčiny smrti: Skreslené novinové spravodajstvo a skreslené úsudky. *Journalism Quarterly*, 56: 837-843.

Kunreuther, H., Hogarth, R. a Meszaros, J. 1993. Neistota poistenca a zlyhanie trhu. *Journal of Risk and Uncertainty*, 7: 71-87.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M. a Combs, B. 1978. Vnímaná frekvencia smrteľných udalostí. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), November: 551-78.

Pôvodný článok s komentármi²

² <http://lesswrong.com/lw/j5/availability/>

Kapitola 6. Priťahujúce podrobnosti

Iba potvrdzujúci detail, s cieľom dať umeleckú dôveryhodnosť inak nudnému a nepresvedčivému rozprávaniu...

—Pooh-Bah v opere Mikado od Gilberta a Sullivana

Klam konjunkcie¹ je, keď ľudia hodnotia pravdepodobnosť $P(A\&B)$ ako vyššiu než pravdepodobnosť $P(B)$, napriek vete, že $P(A\&B) \leq P(B)$. Napríklad v jednom experimente z roku 1981, 68% pokusných osôb hodnotilo ako pravdepodobnajúšie, že “Reagan poskytne federálnu podporu slobodným matkám a zoškrta federálnu podporu lokálnym vládam” než že “Reagan poskytne federálnu podporu slobodným matkám”.

Dlhá séria múdro navrhnutých pokusov, ktoré vyvrátili alternatívne hypotézy a definitívne potvrdili štandardnú interpretáciu², potvrdila, že klam konjunkcie nastáva preto, lebo “dosadzujeme hodnotenie reprezentatívnosti za hodnotenie pravdepodobnosti”. Pridaním ďalších podrobností môžete dosiahnuť, že výsledok bude vyzerat *typickejši* pre proces, ktorý ho vytvára. Tvrdenie, že Reagan podporí slobodné matky, môžete urobiť dôveryhodnejším, keď *pridáte* tvrdenie, že Reagan *ešte aj* zoškrta podporu lokálnym vládam. Nedôveryhodnosť jedného tvrdenia sa vyváži dôveryhodnosťou druhého tvrdenia; “spriemerujú sa”.

Čiže: Pridanie podrobnosti môže urobiť scenár NAPOHLAD DÔVERYHODNEJŠÍ, hoci sa tým celková udalosť nevyhnutne stane MENEJ PRAVDEPODOBNOU.

Ak je to tak, potom by *čisto hypoteticky* mali existovať futuristi, ktorí splietajú nehorázne dôveryhodné a podrobné dejiny budúcnosti, alebo ľudia, ktorí zhltnú obrovské balíky nepodložených tvrdení, ak sa zostavia okolo niekoľkých presvedčivo znejúcich tvrdení. Ak vám predložia klam konjunkcie v nahom, priamom porovnaní, možno sa vám pri tomto konkrétnom probléme podarí vedome sa opraviť. Ale to je iba nalepenie náplaste na problém, nie jeho oprava vo všeobecnosti.

V pokuse z roku 1982³, kde profesionálni predpovedači systematicky priradovali vyššie pravdepodobnosti “Rusko napadne Poľsko, s následným prerušením diplomatických vzťahov medzi USA a ZSSR” oproti “prerušenie diplomatických vzťahov medzi USA a ZSSR”, každá pokusná skupina dostala iba jeden z výrokov. Akú stratégiu mohli použiť títo predpovedači, ako skupina, aby odstránili klam konjunkcie, keď žiaden z nich jednotlivu priamo o tomto porovnaní nevedel? Keď žiaden z nich ani nevedel, že tento pokus bude o klame konjunkcie? Ako boli zlepšiť svoje odhady pravdepodobnosti?

Náplasť na jeden špeciálny chyták nerieši problém vo všeobecnosti. Chyták je príznakom, nie chorobou. Bolo by to rovnako hlúpe ako, povedzme, zakazovat vreckové nožíky v lietadlách.

Čo mohli predpovedači urobiť, aby sa vyhli klamu konjunkcie, aj keď porovnanie priamo nevideli, ani nevedeli, že ich niekto ide skúšať z klamu konjunkcie? Zdá sa mi, že by si museli všimnúť slovo “a”. Museli by si naň dávať pozor - nielen dávať pozor, ale uskočiť pred ním. Aj keby nevedeli, že ich potom výskumníci budú skúšať konkrétne z klamu konjunkcie. Museli by si všimnúť konjunkciu *dvoch detailov* a byť *šokovaní* tou drzosťou, že ich niekto žiada o podporu takejto

¹ http://lesswrong.com/lw/ji/conjunction_fallacy/

² http://lesswrong.com/lw/jj/conjunction_controversy_or_how_they_nail_it_down/

³ http://lesswrong.com/lw/ji/conjunction_fallacy/

šialene komplikovanej predpovede. A museli by takú pravdepodobnosť *podstatne* penalizovať – aspoň o štyri rády, podľa detailov experimentu.

Možno by pomohlo, keby sa predpovedači zamysleli nad možnými dôvodmi, prečo by USA a Sovietsky Zväz mohli prerušiť diplomatické vzťahy. Scenár nie je “USA a Sovietsky Zväz náhle bezdôvodne prerušia diplomatické vzťahy” ale “USA a Sovietsky Zväz prerušia diplomatické vzťahy z nejakého dôvodu”.

A čo pokusné osoby, ktoré hodnotili “Reagan poskytne federálnu podporu slobodným matkám a zoškrtná federálnu podporu lokálnym vládam”? Opäť, mali byť šokované slovom “a”. Navyše, mali absurdity *sčítavať* – absurdita nech je logaritmus pravdepodobnosti, takže sa dá sčítať – a nie priemerovať. Mali si pomyslieť: “Reagan môže ale nemusí zoškrtnať podporu lokálnym vládam (1 bit), ale vyzerá veľmi nepravdepodobne, že by podporil slobodné matky (4 bity). Celková absurdita: 5 bitov.” Alebo trebárs: “Reagan by nepodporil slobodné matky. Celé to neprichádza do úvahy. Ten druhý výrok to už len zhoršuje.”

Podobne, predstavte si šesťstennú kocku so štyrmi zelenými a dvoma červenými stenami. Pokusné osoby mali stavať⁴ na jednu z postupností (1) “ČZČČČ”, (2) “ZČZČČČ” alebo (3) “ZČČČČČ”, či sa objaví niekde medzi 20 hodmi kocky. 65% pokusných osôb si vybralo “ZČZČČČ”, ktoré je jednoznačne horšie ako “ČZČČČ”, pretože ľubovoľná postupnosť obsahujúca “ZČZČČČ” obsahuje aj “ČZČČČ”. Ako mohli pokusné osoby konať lepšie? Všimnúť si, že jedna postupnosť obsahuje druhú? Možno; ale to je iba náplast, ktorá nerieši základný problém. Presným výpočtom pravdepodobností? To by základný problém určite vyriešilo, ale nie vždy môžete pravdepodobnosť presne vypočítať.

Pokusné osoby heuristicky prehrali tým, že si mysleli: “Aha! Postupnosť 2 má vyšší pomer zelených k červeným! Mal by som stavať na postupnosť 2!” Aby heuristicky vyhrali, pokusné osoby by si mali pomyslieť: “Aha! Postupnosť 1 je *krátka!* Mal by som stavať na postupnosť 1!”

Mali by cítiť silnejší *emocionálny dopad* Occamovej britvy – cítiť *každú* pridanú podrobnosť ako bremeno, dokonca aj jeden hod kockou navyše.

Kedysi som sa rozprával s niekým, kto bol očarený nejakým neopatrným futuristom. (Takým, ktorý pridával množstvo dobre znejúcich detailov.) Pokúšal som sa vysvetliť, prečo nie som rovnako zhypnotizovaný týmito úžasnými, neuveriteľnými teóriami. Vysvetlil som teda klam konjunkcie, konkrétne ten pokus “prerušenie vzťahov +/- napadnutie Poľska”. On na to: “Dobre, ale ako to súvisí s...” Ja: “Je pravdepodobnejšie, že sa vesmíry replikujú *z nejakého dôvodu*, než že sa replikujú *pomocou čiernych dier*, pretože *vyspelé civilizácie stavajú čierne diery*, pretože *vesmíry sa vyvíjajú tak, aby to robili.*” On: “Ach.”

Až dovtedy necítil tieto podrobnosti navyše ako bremeno navyše. Naopak, to boli potvrdzujúce detaily, dodávajúce dôveryhodnosť rozprávaniu. Nieкто vám predloží balík zvláštnych myšlienok z ktorých *jedna* je, že vesmíry sa replikujú. Potom predloží podporu *pre tvrdenie*, že *vesmíry sa replikujú*. To však nie je podpora pre ten balík⁵, aj keď to celé podáva ako jeden príbeh.

Musíte tieto podrobnosti rozmotáť. Musíte chytiť každú z nich samostatne a opýtať sa: “Ako vieme *túto* podrobnosť?” Nieкто vám vykreslí obrázok úpadku ľudstva do nanotechnologickej vojny, kde Čína odmietne dodržiavať medzinárodné dohody o kontrole, nasledujú preteky vo vývoji zbraní... Počkaj chvíľu – ako vieš, že to bude Čína? Máš tam vo vrecku krištáľovú guľu, alebo sa len tešíš, že si futurista? Odkiaľ pochádzajú všetky tieto podrobnosti? Odkiaľ pochádza *táto konkrétna* podrobnosť?

⁴ http://lesswrong.com/lw/ji/conjunction_fallacy/

⁵ http://en.wikipedia.org/wiki/Package-deal_fallacy

Pretože je napísané:

Ak môžeš uľahčiť svoje bremeno, musíš to urobiť.

Niet takej slamky, ktorá by nemala silu zlomiť ti chrbát.

Pôvodný článok s komentármi⁶

⁶ http://lesswrong.com/lw/jk/burdensome_details/

Kapitola 7. Klam plánovania

Medzinárodné letisko v Denveri¹ otvorili o 16 mesiacov neskôr a jeho cena prekročila pôvodný plán o 2 miliardy dolárov (videl som aj údaj 3,1 miliardy dolárov). Eurofighter Typhoon², spoločný vojenský projekt niekoľkých európskych krajín, dokončili o 54 mesiacov neskôr za cenu 19 miliárd dolárov namiesto 7 miliárd. Dom opery v Sydney³ je asi najznámejšia oneskorená stavba všetkých čias, pôvodne odhadovaný na rok 1963 za 7 miliónov dolárov, naozaj dokončený v roku 1973 za 102 miliónov dolárov.

Sú to izolované pohromy, ktorým venujeme pozornosť vďaka selektívnej dostupnosti⁴? Sú to príznaky nesprávne motivovanej byrokracie alebo vlády? Áno, veľmi pravdepodobne. Ale existuje aj príslušné kognitívne skreslenie opakovane potvrdené pokusmi, v ktorých plánovali jednotlivci.

Buehler a kol. (1995) požiadal svojich študentov o odhad, kedy si (študenti) myslia, že dokončia svoje osobné akademické projekty. Konkrétne sa výskumníci pýtali na odhad času, kedy si študenti myslia, že ich osobný projekt bude hotový s pravdepodobnosťou 50%, 75% a 99%. Chcete hádať, koľko študentov stihlo skončiť do vlastného odhadu úrovne pravdepodobnosti 50%, 75% a 99%?

- 13% pokusných osôb dokončilo svoj projekt v čase, ktorému priradili pravdepodobnosť 50%;
- 19% dokončilo v čase, ktorému priradili pravdepodobnosť 75%;
- a iba 45% (menej ako polovica!) dokončilo v čase odhadovanej pravdepodobnosti 99%.

Ako píše Buehler a kol. (2002), “Výsledky pre pravdepodobnosť 99% sú osobitne šokujúce: Napriek požiadavke, aby urobili vysoko konzervatívny odhad, predpoveď, ktorú prakticky naisto cítia, že ju splnia, dôvera študentov vo svoje časové odhady vysoko prevyšovala ich výsledky.”

Všeobecnejšie sa tento jav nazýva “klam plánovania”. Klam plánovania je, keď si ľudia myslia, že vedia plánovať, ha ha.

Návod k problému algoritmu plánovania objavil Newby-Clark a kol. (2000), ktorý zistil, že:

- Keď požiadame pokusné osoby o predpovede založené na realistickom “najpravdepodobnejšom” scenári; alebo
- Keď požiadame pokusné osoby o vysnívaný “najlepší možný” scenár...

...získané výsledky od seba *nemožno rozoznať*.

Keď ľudí žiadame o “realistický” scenár, predstavia si, že všetko pôjde presne podľa plánu, žiadne *nečakané* meškania ani *nepredvídané* katastrofy – rovnaká predstava ako ich “najlepší možný prípad”.

Ako sa ukazuje, skutočnosť zvyčajne dáva výsledky o čosi horšie než “najhorší možný prípad”.

Na rozdiel od väčšiny kognitívnych skreslení, pri klame plánovania poznáme dobrú heuristiku na odstránenie skreslenia. Nebude fungovať na chaos rozmerov

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

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

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

⁴ pozri: 5 Dostupnosť

medzinárodného letiska v Denveri, ale bude fungovať na mnohé osobné plány a dokonca aj na nejaké organizačné veci malého rozsahu. Jednoducho použije "pohľad zvonka" namiesto "pohľadu zvnútra".

Ľudia majú sklon zostavovať svoje predpovede tak, že pomyslia na čiastkové jedinečné vlastnosti danej úlohy, a vytvoria si scenár, ako túto úlohu zamýšľajú dokončiť – čo je to, čo si zvyčajne predstavujeme ako *plánovanie*. Keď chcete niečo urobiť, musíte si naplánovať kde, kedy, ako; zistiť, koľko času a koľko prostriedkov potrebujete; predstaviť si kroky od začiatku po úspešný záver. Toto všetko je "pohľad zvnútra", neberie do úvahy nečakané zdržania a nepredvídané katastrofy. Ako sme už videli, požiadať ľudí, aby si predstavili "najhorší možný prípad" stále nestačí na prekonanie ich optimizmu – nepredstavujú si dosť Murphyho zákona.

Vonkajší pohľad je, keď sa úmyselne *vyhýbate* mysleniu na osobitné jedinečné vlastnosti daného projektu, a iba sa opýtate, ako dlho trvalo dokončiť *zhruba* podobné projekty v minulosti. Je to proti intuícii, pretože pohľad zvnútra má omnoho viac detailov – je tu pokušenie myslieť si, že starostlivo na mieru ušitá predpoveď, ktorá zohľadní všetky dostupné údaje, dá lepšie výsledky.

Pokusy však ukazujú, že čím podrobnejšia je predstava pokusných osôb, tým optimistickejšie (a menej presné) sa stávajú. Buehler a kol. (2002) požiadal skupinu pokusných osôb, aby popísali veľmi konkrétne plány svojich vianočných nákupov – kde, kedy a ako. Táto skupina v priemere očakávala, že dokončí nakupovanie viac ako týždeň pred Vianocami. Druhej skupiny sa jednoducho opýtal, kedy očakávajú, že dokončia svoje vianočné nákupy, a priemerná odpoveď bola 4 dni. Obe skupiny skončili v priemere 3 dni pred Vianocami.

Podobne, Buehler a kol. (2002) v opise medzikultúrnej štúdie zistili, že japonskí študenti očakávali, že svoje eseje dokončia 10 dní pred termínom. V skutočnosti ich dokončili 1 deň pred termínom. Keď sa ich pýtali, kedy dokončovali podobné úlohy v minulosti, odpovedali: "1 deň pred termínom". Toto je sila pohľadu zvonka oproti pohľadu zvnútra.

Podobne sa zistilo, že skúsení ľudia mimo organizácie, ktorí poznajú menej podrobností, ale môžu sa oprieť o relevantné spomienky, sú často omnoho menej optimistickí a omnoho presnejší než skutoční plánovači a realizátori.

Existuje teda dosť spoľahlivý spôsob, ako opraviť klam plánovania, ak robíte niečo, čo sa *zhruba* podobá na referenčnú triedu predchádzajúcich projektov. Skrátka sa opýtajte, ako dlho trvali podobné projekty v minulosti, a nezohľadňujte pritom *žiadne* osobitné vlastnosti tohto projektu. Ešte lepšie, opýtajte sa skúseného človeka mimo organizácie, ako dlho trvali podobné projekty.

Dostanete odpoveď, ktorá bude znieť príšerne dlho, a bude jasné, že neodráža pochopenie špeciálnych dôvodov, prečo táto konkrétna úloha zaberie menej času. Tá odpoveď je pravdivá. Vyrovajte sa s tým.

Buehler, R., Griffin, D. and Ross, M. 1994. Skúmanie "klamu plánovania": Prečo ľudia podceňujú čas dokončenia svojej úlohy. *Journal of Personality and Social Psychology*, 67: 366-381.

Buehler, R., Griffin, D. and Ross, M. 1995. Je načase: Optimistické predpovede v práci a láske. Pp. 1-32 in *European Review of Social Psychology, Volume 6*, eds. W. Stroebe and M. Hewstone. Chichester: John Wiley & Sons.

Buehler, R., Griffin, D. and Ross, M. 2002. Vnútri klamu plánovania: Príčiny a dôsledky optimistických časových predpovedí. Pp. 250-270 in Gilovich, T.,

Griffin, D. and Kahneman, D. (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, U.K.: Cambridge University Press.

Newby-Clark, I. R., Ross, M., Buehler, R., Koehler, D. J. and Griffin, D. 2000. Ľudia sa sústredia na optimistické a ignorujú pesimistické scenáre pri predpovedaní času dokončenia úlohy. *Journal of Experimental Psychology: Applied*, 6: 171-182.

Pôvodný článok s komentármi⁵

⁵ http://lesswrong.com/lw/jg/planning_fallacy/

Kapitola 8. Ilúzia priehľadnosti: Prečo vám nikto nerozumie

Pri skreslení spätného pohľadu¹ ľudia, ktorí poznajú výsledok nejakej situácie, veria, že tento výsledok sa dal ľahko predpovedať. Keď už výsledok poznáme, prerozprávame si celú situáciu² vo svetle tohto výsledku. Aj keď nás na to upozornia, nedokážeme si toto vysvetlenie odmyslieť natoľko, aby sme vžili do kože niekoho, kto nevie to, čo my.

S tým úzko súvisí *ilúzia priehľadnosti*. Vždy vieme, čo myslíme *svojimi* vlastnými slovami, a preto očakávame, že to budú vedieť aj druhí. Keď čítame, čo sme sami napísali, ľahko sa nám text správne vysvetľuje, keď vieme, čo sme tým naozaj mysleli. Ťažké je vžiť sa do kože niekoho, kto musí vysvetľovať naslepo, riadený iba slovami.

June odporučí Markovi reštauráciu; Mark sa tam navečeria a zistí, že (a) jedlo nie je nič moc a obsluha je podpriemerná (b) jedlo je vynikajúce a obsluha bezchybná. Potom Mark zanechá June na odkazovači nasledujúcu správu: "June, práve som dovečeral v reštaurácii, ktorú si mi odporučila, a musím povedať, že bola úžasná, fakt úžasná." Keysar (1994) predložil skupine pokusných osôb scenár (a) a 59 % považovalo Markovu správu za sarkastickú *a mysleli si, že June by tento sarkazmus vnímala*. Iné pokusné osoby dostali scenár (b) a iba 3 % z nich si mysleli, že June bude vnímať Markovi správu ako sarkastickú. Keysar a Barr (2002) naznačujú, že pokusné osoby si vypočuli skutočnú správu zo záznamu. Keysar (1998) ukázal, že ak pokusným osobám povieme, že reštaurácia bola hrozná, *ale Mark chcel zaprieť svoju reakciu*, pokusné osoby veria, že June nebude vnímať sarkazmus v (tej istej) správe:

To, že si [June] všimne sarkazmus, predpovedali s rovnakou pravdepodobnosťou, keď sa [Mark] pokúšal skryť svoju negatívnu skúsenosť, ako keď mal pozitívnu skúsenosť a bol úprimný. Účastníci teda brali Markov *komunikačný zámer* ako priehľadný. Akoby predpokladali, že June bude vnímať ten zámer, ktorý Mark chcel, aby vnímala.

(Uvedená formulácia je od Keysara a Barra 2002.)

"Hus visí vysoko" je archaické anglické slovné spojenie, ktorý sa v modernom jazyku nepoužíva. Keysar a Bly (1995) povedali jednej skupine pokusných osôb, že "hus visí vysoko" znamená, že budúcnosť vyzerá dobre; druhej skupine pokusných osôb povedali, že "hus visí vysoko" znamená, že budúcnosť vyzerá pochmúrne. Pokusných osôb sa potom pýtali, ktorý z týchto dvoch významov by *neinformovaný* poslucháč skôr pripísal tomuto slovnému spojeniu. Každá skupina si myslela, že poslucháč by chápal toto slovné spojenie "štandardne".

(Ďalšie testované slovné spojenia boli "išiel na niekoho ako ujo", "ísť pozdĺž paluby" a "vyložiť do levandulí". Ach, angličtina, aký pôvabný jazyk.)

Keysar a Henly (2002) testovali kalibráciu rečníkov: Zvyknú rečníci podceňovať, preceňovať, alebo správne odhadovať, ako často im poslucháči rozumejú? Rečníci dostali nejednoznačné vety ("Muž naháňa ženu na bicykli") a rozlišujúce obrázky (muž beží za ženou, ktorá ide na bicykli), potom ich požiadali, aby vyslovili tieto slová pred poslucháčom, a potom ich požiadali o odhad, koľko poslucháčov pochopilo zamýšľaný význam. Rečníci si mysleli, že boli správne pochopení v

¹ http://lesswrong.com/lw/il/hindsight_bias/

² pozri: 34 Falošná kauzalita

72 % prípadov, ale naozaj boli pochopení v 61 % prípadov. Keď poslucháč nepochopil, rečníci si v 46 % prípadov mysleli, že pochopil; keď poslucháč pochopil, rečníci si iba v 12 % prípadov mysleli, že nepochopil.

Ďalšie pokusné osoby, ktoré *počuli* vysvetlenie, nevykazovali toto skreslenie a očakávali pochopenie poslucháčov iba v 56 % prípadov.

Ako poznamenávajú Keysar a Barr (2002), dva dni pred útokom Nemecka na Poľsko, Chamberlain poslal list s úmyslom vysvetliť, že Británia sa v prípade invázie zapojí do boja. Tento list, formulovaný zdvorilým diplomatickým jazykom, Hitler pochopil ako zmierlivý - a tanky vyštartovali.

Nebudte príliš rýchli pri odsudzovaní tých, ktorí nepochopili vaše dokonale jasné vety, vyslovené alebo napísané. Je možné, že vaše slová boli dvojznačnejšie než si myslíte.

Keysar, B. (1994). Iluzórna priehľadnosť zámeru: Používanie jazykovednej perspektívy v texte³. *Cognitive Psychology*, 26, 165-208.

Keysar, B. (1998). Používatelia jazyka ako riešitelia problémov: Aký problém nejednoznačnosti vlastne riešia? In S. R. Fussell and R. J. Kreuz (Eds.), *Social and Cognitive Psychological Approaches to Interpersonal Communication*, 175-200. Hillsdale, NJ: Erlbaum.

Keysar, B., & Barr, D. J. (2002). Sebaukotvenie v konverzácii: Čo používatelia jazyka nerobia a "mali by". (2002). In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. (pp. 150-166). Cambridge University Press.

Keysar, B., & Bly, B. (1995.) Intuície o priehľadnosti slovných spojení: Môže niekto udržať tajomstvo tým, že všetko vyklopí?⁴ *Journal of Memory and Language*, 34, pp. 89-109.

Keysar, B., & Henly, A. S. (2002). Ako rečníci preceňujú svoju účinnosť⁵. *Psychological Science*, 13, 207-212.

Pôvodný článok s komentármi⁶

³ http://psychology.uchicago.edu/people/faculty/keysar/6_cogpsych94.pdf

⁴ http://psychology.uchicago.edu/people/faculty/keysar/7_keysarbly95.pdf

⁵ http://psychology.uchicago.edu/people/faculty/keysar/20_keysarhenly02.pdf

⁶ http://lesswrong.com/lw/ke/illusion_of_transparency_why_no_one_understands/

Kapitola 9. Očakávame krátke inferenčné vzdialenosti

Životné prostredie, na ktoré je *Homo sapiens* evolučne prispôsobený¹ (skrátene: “praveké prostredie”), sa skladalo z bánd² po nanajvýš 200³ lovcov a zberačov, bez písma. Všetky zdedené vedomosti sa odovzdávali slovne, naspamäť.

V takomto svete sú všetky základné poznatky všeobecne známe. Každá informácia, ktorá nie je prísne súkromná, je verejná, bodka.

V pravekom prostredí ste sa ťažko od hocikoho vzdialili o viac ako *jeden inferenčný krok*. Keď ste objavili novú oázu, nemuseli ste svojim súkmeňovcom vysvetľovať, čo je to oáza alebo prečo je dobré piť vodu alebo ako sa chodí. Iba vy ste vedeli, kde sa tá oáza nachádza; to bola súkromná vedomosť. Ale každý mal základy na to, aby pochopil váš popis oázy, pojmy potrebné na rozmyšľanie o vode; to boli všeobecné vedomosti. Keď ste v pravekom prostredí vysvetľovali veci, takmer *nikdy* ste nemuseli vysvetľovať použité pojmy. V krajnom prípade ste museli vysvetliť *jeden* nový pojem, nie dva alebo viac naraz.

V pravekom prostredí neexistovali abstraktné vedné odbory s hromadami starostlivo zozbieraných dôkazových materiálov, zovšeobecnenými do elegantných teórií prenášaných písomne knihami, ktorých závery sú *stovky inferenčných krokov vzdialené* od všeobecne spoločných základných predpokladov.

Ak v pravekom prostredí niekto niečo povedal bez zrejmeho podkladu, bol to klamár alebo hlupák. Nepomysleli ste si: “Hej, tento človek možno má dobre podložené vzdelanie, o ktorom nikto z mojej bandy nikdy nepočul,” pretože v pravekom prostredí sa dalo spoľahnúť na to, že toto sa nestáva.

A naopak, ak ste povedali niečo očividne jasné a druhý človek to tak nevidel, potom *on* bol hlupák, alebo bol úmyselne tvrdohlavý, aby vás vytočil.

A na dôvažok, ak niekto niečo povedal bez zrejmeho podkladu a *očakával*, že tomu uveríte – a správal sa rozhorčene, keď ste neuverili – to musel byť *šialenec*.

Spolu s ilúziou priehľadnosti⁴ a sebaukotvením⁵ to podľa mňa vysvetľuje *veľa* z povestných ťažkostí, ktoré má väčšina vedcov pri komunikácii s laickou verejnosťou – dokonca aj s vedcami z iných odborov. Keď sledujem zlyhania pri vysvetľovaní, zvyčajne vidím, že vysvetľujúci urobil *jeden* krok vzad, keď by potreboval urobiť dva alebo viac krokov vzad. Alebo že poslucháči predpokladajú, že veci by mali byť jasné na jeden krok, keď na ich vysvetlenie treba dva alebo viac krokov. Obe strany konajú, akoby očakávali veľmi krátke inferenčné vzdialenosti od všeobecných vedomostí po každú novú vedomosť.

Keď sa biológ rozpráva s fyzikom, môže zdôvodniť evolúciu povedaním, že je to “najjednoduchšie vysvetlenie”. Ale nie každý na tejto Zemi bol naočkovaný legendami z dejín vedy, od Newtona po Einsteina, ktoré dávajú slovnému spojeniu “najjednoduchšie vysvetlenie” jeho úžasný dôraz: sú to Slová Moci, ktoré sa vyslovujú pri zrode teórií a sú vytesané na ich náhrobných kameňoch. Niekomu inému môže: “Ale to je najjednoduchšie vysvetlenie!” znieť ako zaujímavý ale sotva drvivý argument; nepripadá mu to ako až taký mocný nástroj

¹ http://en.wikipedia.org/wiki/Evolutionary_psychology#Environment_of_evolutionary_adaptedness

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

³ http://en.wikipedia.org/wiki/Dunbar%27s_number

⁴ pozri: 8 Ilúzia priehľadnosti: Prečo vám nikto nerozumie

⁵ <http://lesswrong.com/lw/kf/selfanchoring/>

na porozumenie firemnej politiky alebo na opravu pokazeného auta. Biológ je očividne zalúbený do svojich vlastných nápadov, príliš povýšený na to, aby sa otvoril alternatívnym vysvetleniam, ktoré znejú rovnako dôveryhodne. (Keď to znie dôveryhodne mne, malo by to znieť dôveryhodne každému príčetnému členovi mojej bandy.)

A z pohľadu biológa, on chápe, ako evolúcia po prvýkrát môže znieť trochu divne - ale keď niekto odmieta evolúciu ešte po tom, čo mu biológ vysvetlil, že to je najjednoduchšie vysvetlenie, nuž, je jasné, že nevedci sú skrátka hlupáci a nemá zmysel rozprávať sa s nimi.

Jasný argument musí vystavať inferenčnú *cestu*, začínajúc od toho, čo obecenstvo už *pozná alebo prijíma*. Ak sa nevrátite dostatočne ďaleko naspäť, rozprávate sa iba sami so sebou.

Ak v hociktorom bode niečo tvrdíte bez očividného zdôvodnenia argumentmi, ktoré ste predtým predložili, obecenstvo si skrátka pomyslí, že ste obeťou nejakej sekty.

Toto sa stane, aj keď si dovoľíte *viditeľne* prikladať nejakému argumentu väčšiu váhu než si v *tej chvíli* zaslúži v očiach obecenstva. Napríklad, ak budete rozprávať, akoby ste si mysleli, že "jednoduchšie vysvetlenie" je pádny argument pre evolúciu (čo naozaj je), namiesto iba tak trochu zaujímavej myšlienky (ako to znie niekomu, kto nebol vychovaný k úcte voči Occamovej britve).

Aha, a radšej by ste nemali nijako naznačovať, že *si myslíte*, že pracujete tucet inferenčných krokov ďalej od toho, čo obecenstvo vie, alebo že *si myslíte*, že máte špeciálne základné poznatky, ktoré oni nemajú. Obecenstvo nevie nič o evolučno-psychologickom argumente pre kognitívne skreslenie podceňovať inferenčné vzdialenosti, ktoré vedie k dopravným zápcham v komunikácii. Budú si skrátka myslieť, že sa vyvyšujete.

A ak si myslíte, že dokážete pojem "systematické podceňovanie inferenčných vzdialeností" vysvetliť stručne, iba pár slovami, mám pre vás jednu zlú správu...

Pôvodný článok s komentármi⁶

⁶ http://lesswrong.com/lw/kg/expecting_short_inferential_distances/

Kapitola 10. Objektív, ktorý vidí svoje chyby

Súvisí s kapitolou: Čo je indícia?¹

Svetlo opustí Slnko, narazí do vašich šnúrok na topánkach a odrazí sa; niektoré fotóny vojdú do zreničiek vašich očí a dopadnú na vašu sietnicu; energia fotónov vyvolá nervové impulzy; nervové impulzy sa prenesú do oblastí mozgu spracovávajúcich obraz; tam sa optická informácia spracuje a rekonštruuje na 3D model, ktorý sa rozozná ako rozviazaná šnúrka na topánke; a tak veríte, že vaše šnúrky sú rozviazané.

Toto je tajomstvo *rozumného uvedomovania* – celý tento proces previazania nie je kúzlo², a vy ho dokážete *pochopiť*. Dokážete *pochopiť*, ako vidíte svoje šnúrky na topánkach. Dokážete *rozmyšľať* o tom, aké druhy myšlienkových procesov budú vytvárať názory odzrkadľujúce skutočnosť, a ktoré myšlienkové procesy nebudú.

Myši dokážu vidieť, ale nedokážu pochopiť videnie. Vy dokážete pochopiť videnie a preto dokážete robiť veci, ktoré myši nedokážu. Chvíľku sa tomu čudujte³, lebo to naozaj je zázrak.

Myši vidia, ale nevedia, že majú zrkovú kôru, takže nevedia korigovať optické ilúzie. Myš žije v myšlienkovom svete, ktorý obsahuje mačky, diery, syr a pasce na myši – ale nie myšacie mozgy. Ich fotoaparát nedokáže odfoťiť vlastný objektív. Ale my, ako ľudia, sa dokážeme pozrieť na zdanlivo neskutočný obrázok⁴ a uvedomiť si, že časť toho, čo vidíme, je samotný objektív. Nemusíte vždy veriť vlastným očiam, ale musíte si uvedomiť, že *máte* oči – musíte mať oddelené myšlienkové vedierka pre mapu a pre územie, pre zmysly a pre skutočnosť. Aby ste si nemysleli, že je to všedná schopnosť, spomeňte si, aká zriedkavá je v ríši zvierat.

Celá myšlienka vedy je, jednoducho povedané, reflektívne uvažovanie o spoľahlivejšom procese, aby obsah vašej mysle odzrkadľoval obsah sveta. To je jedna z vecí, ktoré myši nikdy nevynájdu. Keď sa zamyslíme nad záležitosťou “robenia opakovaných pokusov na falzifikovanie teórií”, uvidíme, *prečo* to funguje. Veda nie je oddelené magistérium⁵, vzdialené od skutočného života a porozumenia bežných smrteľníkov. Veda nie je niečo, čo platí iba vnútri laboratória⁶. Veda samotná je pochopiteľný proces v tomto svete, ktorý koreluje mozgy so skutočnosťou.

Veda dáva *zmysel*, keď sa nad ňou zamyslíte. Myši však nevedia rozmyšľať o myslení, a preto nemajú vedu. Človek by nemal prehliadať tento zázrak – alebo potenciálnu moc, ktorú nám dáva ako jednotlivcom, nielen vedeckým spoločnostiam.

Pripúšťam, že pochopiť stroj na myšlienky môže byť *o čosi zložitejšie* než pochopiť parný stroj – ale nie je to *principiálne* odlišná úloha.

Raz som šiel na EFNNet-ovský kanál #philosophy a opýtal som sa: “Veríte, že v najbližších 20 rokoch nastane jadrová vojna? Ak nie, prečo nie?” Jeden človek mi na túto otázku odpovedal, že neočakáva jadrovú vojnu do 100 rokov, pretože

¹ pozri: 20 Čo je indícia?

² pozri: 36 Tajomné odpovede na tajomné otázky

³ pozri: 44 “Veda” ako zastavovač zvedavosti

⁴ <http://www.richrock.com/gifs/optical-illusion-wheels-circles-rotating.png>

⁵ pozri: 19 Nárok náboženstva na nevyvrátenosť

⁶ pozri: 183 Mimo laboratória

“tí, ktorí môžu rozhodovať o jadrovej vojne, o ňu momentálne nemajú záujem.”
“Ale prečo to predlžuješ na 100 rokov?” opýtal som sa. “Číra nádej,” odpovedal.

Keď sa zamyslíme nad celým týmto myšlienkových procesom, vidíme, že myšlienka na jadrovú vojnu robí daného človeka nešťastným, a vidíme, že jeho mozog preto tento názor odmieta. Ale ak si predstavíte miliardu svetov - Everettove vetvy alebo Tegmarkove duplikáty⁷ - tento myšlienkový proces nebude systematicky korelovať⁸ optimistov s vetvami, v ktorých jadrová vojna nenastane. (Niektorí chytrí teraz určite povie: “Aha, ale pretože mám nádej, budem v práci makať o čosi viac, tým pozdvihnem globálnu ekonomiku, a tým zabránim krajinám sklzánuť do zloštného a beznádejného stavu, v ktorom je jadrová vojna možná. Takže tieto dve udalosti predsa navzájom súvisia.” V tomto bode musíme vytiahnuť Bayesovu vetu⁹ a odmerať náboj previazanosti kvantitatívne. Vaša optimistická povaha nemôže mať *taký* veľký dopad na svet; nemôže samotná znížiť pravdepodobnosť jadrovej vojny o 20% alebo o koľko vaša optimistická povaha posunula váš názor. Posunúť svoj názor o veľké množstvo kvôli udalosti, ktorá nesie iba maličký náboj previazanosti, stále pokazí vaše kreslenie mapy.)

Pýtať sa, ktorý názor vás robí šťastnými, to je pozeranie dovnútra, nie von - povie vám to niečo o vás samotných, ale nie je to indícia previazaná s prostredím. Nemám nič proti šťastiu, ale malo by vychádzať z¹⁰ vášho obrazu sveta, nie z manipulácie s myšlienkovými farbičkami.

Keď vidíte toto - keď vidíte, že nádej posúva vaše myšlienky *prvého rádu* neprimerane ďaleko - ak dokážete pochopiť svoj mozog ako stroj na kreslenie máp, ktorý má svoje chyby - potom môžete uplatiť reflexívnu opravu. Mozog je pokazený objektív, ktorým vidíme skutočnosť. To platí o myšiacích aj ľudských mozgoch. Ale ľudský mozog je pokazený objektív, ktorý dokáže pochopiť svoje vlastné kazy - svoje systematické chyby, svoje skreslenia - a uplatniť na ne opravy druhého rádu. Toto v *praxi* robí tento pokazený objektív omnoho mocnejším. Nie dokonalým, ale omnoho mocnejším.

Pôvodný článok s komentármi¹¹

⁷ <http://arxiv.org/abs/astro-ph/0302131>

⁸ pozri: 20 Čo je indícia?

⁹ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

¹⁰ pozri: 2 Rozumné čítanie

¹¹ http://lesswrong.com/lw/jm/the_lens_that_sees_its_flaws/

Časť II. Falošné názory

Obsah

11	Nech názory platia nájomné (v očakávaných vnemoch)	26
12	Bájka o vede a politike	28
13	Viera vo vieru	31
14	Bayesovské džudo	34
15	Vyznávanie a fandanie	35
16	Signály na potlesk	37
17	Názor ako uniforma	39
18	Predstieranie múdrosti	40
19	Nárok náboženstva na nevyvrátiteľnosť	43

Kapitola 11. Nech názory platia nájomné (v očakávaných vnemoch)

Takto začína prastaré podobenstvo:

Ak v lese padne strom a nikto ho nepočuje, vznikne pritom zvuk? Jeden povie: "Áno, vznikne, pretože to spôsobí vibrácie vo vzduchu." Druhý povie: "Nie, nevznikne, pretože to žiaden mozog zvukovo nespracuje."

Predstavme si, že po páde daného stromu títo dvaja pôjdu spolu do lesa. Bude jeden očakávať, že uvidí strom padnutý napravo, a druhý očakávať, že uvidí strom padnutý naľavo? Predpokladajme, že pred pádom stromu títo dvaja nechali vedľa stromu zaznamenávač zvuku. Bude jeden, keď si prehrá záznam, očakávať, že začuje niečo iné ako druhý? Predstavme si, že by všetky mozgy na svete napojili na elektroencefalograf; očakával by jeden, že uvidí iné čiary než druhý? Hoci sa títo dvaja hádajú, jeden hovorí: "Nie" a druhý hovorí: "Áno", neočakávajú žiadne odlišné vnemy. Títo dvaja si myslia, že majú rôzne modely sveta, ale nijako sa nelíšia v očakávaní, čo sa im *stane*.

Je lákavé pokúsiť sa túto triedu omylov odstrániť trvaním na tom, že jediným legitímnym zdrojom názoru je očakávanie zmyslového vnemu. Svet však v skutočnosti obsahuje mnohé, čo priamo nevnímame. Nevidíme atómy, z ktorých sa skladá tehla, ale tie atómy tam naozaj sú. Pod nohami máte podlahu, ale tú podlahu *nevnímate* priamo; vidíte svetlo *odrazené* od podlahy, či skôr vidíte to, čo z tohto svetla spracovala vaša sietnica a zraková kôra. Odvodiť existenciu podlahy z pohľadu na podlahu je krok naspäť do nevidených príčin vnemu. Môže to vyzeráť ako veľmi krátky a priamy krok, ale stále je to krok.

Stojíte navrchu vysokej budovy, vedľa ručičkových hodín s hodinovou, minútovou a tikajúcou sekundovou ručičkou. V ruke máte kolkársku guľu a pustíte ju zo strechy. Pri ktorom tiknutí hodín budete počuť zvuk dopadu kolkárskej gule na zem?

Aby ste odpovedali presne, musíte použiť názory ako *Gravitácia Zeme je 9,8 metra za sekundu za sekundu a Táto budova je asi 120 metrov vysoká*. Tieto názory nie sú neverbálne očakávania zmyslového vnemu; sú to slovné výroky. Asi nie je veľkým prehánaním opísať tieto názory ako vety zložené zo slov. Avšak tieto dva názory majú inferenčný *dôsledok*, ktorý je priamym zmyslovým očakávaním – ak je sekundová ručička hodín na čísle 12, keď pustíte guľu, očakávate, že ju uvidíte na čísle 1, keď o päť sekúnd neskôr začujete úder. Aby sme predpokladali zmyslové vnemy čo najpresnejšie, musíme spracovávať názory, ktoré nie sú očakávaniami zmyslových vnemov.

Je veľkou silou druhu *Homo sapiens*, že sa dokážeme lepšie než ľubovoľný iný živočíšny druh na svete naučiť modelovať nevidené. Je to zároveň jedna z našich veľkých slabostí. Ľudia často veria na veci, ktoré sú nielen nevidené, ale aj neskutočné.

Ten istý mozog, ktorý stavia siete vydedukovaných príčin zmyslových vnemov, dokáže stavať aj siete príčin, ktoré so zmyslovými vnemmi nie sú spojené alebo sú slabo spojené. Alchymisti verili, že flogiston spôsobuje oheň – mohli by sme si ich mysle veľmi zjednodušene znázorniť nakreslením malého uzla s názvom "Flogiston" a šípky od tohto uzla k ich zmyslovému vnemu praskajúceho táboráku – lenže tento názor nedáva žiadne predpovede; linka od flogistonu

k vnemu sa nastaví vždy až po danom vneme, namiesto aby ten vnem vopred ohraničila. Alebo povedzme, že vás profesor postmodernej angličtiny učí, že známy spisovateľ Wulky Wilkinsen je v skutočnosti "postutopický". Čo by ste podľa toho mali očakávať od jeho kníh? Nič. Tento názor, ak ho tak môžeme nazvať, sa nijako nespája so zmyslovými vnemmi. Ale radšej by ste si mali zapamätať tvrdenie, že "Wulky Wilkinsen" má vlastnosť "postutopický", aby ste to vedeli zopakovať pri blížiaci sa písomke. Podobne, ak "postutopický" autori zobrazujú "koloniálne odcudzenie"; ak sa písomka bude pýtať, či Wulky Wilkinsen zobrazuje koloniálne odcudzenie, radšej by ste mali odpovedať áno. Tieto názory sú spojené jeden s druhým, hoci stále nie sú spojené so žiadnym očakávaným vnemom.

Dokážeme vybudovať celé siete názorov spojených iba navzájom - nazvime ich "vznášajúce sa" názory. Je to jedinečná ľudská chyba medzi živočíšnymi druhmi, prekrútenie schopnosti *Homo sapiens* budovať všeobecnejšie a pružnejšie siete názorov.

Racionalistická cnosť *empirizmu* spočíva k ustavičnému pýtaniu sa, aké vnemy naše názory predpovedajú - alebo ešte lepšie, zakazujú. Veríte tomu, že príčinou ohňa je flogiston? Čo potom očakávate, že sa kvôli tomu stane? Veríte, že Wulky Wilkinsen je postutopický? Čo potom očakávate, že kvôli tomu uvidíte? Nie, nie "koloniálne odcudzenie"; aký vnem budete mať? Veríte, že keď v lese padne strom a nikto ho nepočuje, aj tak vznikne zvuk? Aký vnem sa vám kvôli tomu musí prihodiť?

Ešte lepšie je opýtať sa: aký vnem *nemôžete* mať? Veríte, že *elan vital* vysvetľuje tajomnú živosť živých bytostí? Čo potom tento názor *nedovoľuje*, aby sa stalo - čo by definitívne vyvrátilo tento názor? Prázdna odpoveď znamená, že váš názor *neobmedzuje* vnemy; dovoľuje, aby sa vám stalo *hocičo*. Vznáša sa.

Keď sa hádate o napohľad faktickej otázky, vždy myslíte na to, o akom rozdielnom očakávaní sa hádate. Ak nedokážete nájsť rozdiel v očakávaní, pravdepodobne sa hádate o označeniach vo vašej sieti názorov - alebo ešte horšie, o vznášajúcich sa názoroch, parazitoch na vašej sieti. Ak neviete, aké vnemy vyplývajú z toho, že Wulky Wilkinsen je postutopický, môžete sa hádať donekonečna. (Môžete o tom aj donekonečna publikovať články.)

Najdôležitejšie je, nepýtajte sa, čomu veriť - pýtajte sa, čo očakávať. Každá otázka o názore by mala vyplývať z otázky o očakávaní, a táto otázka o očakávaní by mala byť v centre pátrania. Každý pokus o názor by mal začať ako výsledok konkrétneho pokusu o očakávanie, a mal by naďalej platiť nájomné vo forme budúcich očakávaní. Ak sa z názoru vyklúje neplatič, vyhodte ho.

Pôvodný článok s komentármi¹

¹ lesswrong.com/lw/i3/making_beliefs_pay_rent_in_anticipated_experiences/

Kapitola 12. Bájka o vede a politike

Za Rímskej ríše sa občiansky život delil na frakcie Modrých a Zelených. Modrý a Zelení sa navzájom vraždili v súbojoch, prepadoch, skupinových bitkách, povstaniach. Procopius o týchto bojovných frakciách povedal: "Tak v nich vyrastá nenávisť voči svojim bližným, ktorá nemá príčinu a nikdy nevyhasne ani nezmizne, pretože neustúpi ani putám manželstva, ani príbuzenstva, ani priateľstva, a je to rovnaké ešte aj keď tí, ktorí majú odlišné farby, sú bratmi či inými príbuznými." Edward Gibbon napísal: "Podpora niektorej frakcie sa stala nevyhnutnou pre každého kandidáta na občiansky alebo cirkevný post."

Kto boli títo Modrý a Zelení? Boli to športoví fanúšikovia - prívrženci modrého alebo zeleného tímu pretekárskych kočov.

Predstavte si budúcu spoločnosť, ktorá ujde do rozsiahlej siete podzemných jaskýň a zapečatí všetky vchody. Nie je dané, či utekajú pred chorobou, vojnou alebo žiarením; predpokladajme, že sa prvým Podzemšťanom podarí pestovať jedlo, nájsť vodu, recyklovať vzduch, vyrobiť svetlo a prežiť, a že ich potomkom sa darí a časom vytvoria mestá. Zo sveta hore zostali iba povesti napísané na zdrapoch papiera; jeden z takýchto zdrapov papiera opisuje *oblohu*, rozsiahly otvorený vzdušný priestor nad veľkou podlahou bez hraníc. Obloha má nebeskú farbu a obsahuje zvláštne vznášajúce sa predmety podobné ohromným balíkom bielej bavlny. Význam slova "nebeská" je však kontroverzný; niektorí tvrdia, že označuje farbu známu ako "modrá", iní že označuje farbu známu ako "zelená".

V raných dobách podzemnej spoločnosti Modrý a Zelení riešili svoj spor otvoreným násilím; dnes vládne prímerie - mier zrodený z rastúceho pocitu márnosti. Zmenili sa kultúrne zvyky; je tu veľká a úspešná stredná trieda, ktorá vyrástla vďaka účinnému dodržiavaniu zákona a odvykla si od násilia. Školy poskytujú určitý historický nadhľad; ako dlho trvali bitky medzi Modrými a Zelenými, koľkí zomreli, ako málo sa v dôsledku toho zmenilo. Niektorí sa otvárajú zvláštnej novej filozofii, že ľudia sú ľudia, či už Modrý alebo Zelení.

Konflikt nezmizol. Spoločnosť sa naďalej delí na Modrých a Zelených a existuje "Modrý" a "Zelený" postoj k takmer každej aktuálnej politicky alebo kultúrne významnej téme. Modrý presadzujú dane z osobného príjmu, Zelení presadzujú dane z obchodného obratu; Modrý presadzujú prísnejšie manželské zákony, kým Zelení chcú zjednodušiť rozvody; Modrých podporujú najmä mestské centrá, zatiaľ čo vzdialenejší farmári a predajcovia vody zvyknú byť Zelení; Modrý veria, že Zem je veľká guľatá skala v strede vesmíru, Zelení že je to plochá kamenná doska krúžiaca okolo ďalšieho objektu nazývaného Slnko. Nie každý Modrý či každý Zelený má "Modrý" či "Zelený" názor na všetky veci, ale sotva by ste našli mestského kupca, ktorý by veril, že obloha je modrá a zároveň by presadzoval osobné dane a voľnejšie manželské zákony.

Podzemie je stále polarizované; mier je neistý. Pár ľudí si úprimne myslí, že Modrý a Zelení by sa mali kamarátiť, a dnes už je bežné, že Zelení chodia nakupovať do Modrého obchodu, alebo že Modrý chodia do Zelenej krčmy. Z prímeria pôvodne zrodeného z vyčerpania potichu vyrastá duch tolerancie, dokonca aj priateľstva.

Jedného dňa Podzemím otrásie menšie zemetrasenie. Šesťčlennú turistickú výpravu tento otras zastihne pri prezeraní trosiek pradávnych obydlí v horných jaskyniach. Cítia krátky pohyb skaly pod nohami, jedna turistka spadne a oškrie si koleno. Skupina sa rozhodne vrátiť, v obave pred ďalšími zemetraseniami. Cestou späť jeden člen zachytí závan čohosi zvláštneho vo vzduchu, vôňu

prichádzajúcu z dávno nepoužívanej chodby. Ignorujúc dobre mienené varovania svojich súputníkov, táto osoba vezme lampáš a vojde do chodby. Kamenná cesta mieri vyššie... a vyššie... až skončí v jame vykrojenej zo sveta, na mieste, kde všetky skaly končia. Dialka, nekonečná dialka, sa rozprestiera bez konca; námestie schopné obsiahnuť tisíce miest. V nepredstaviteľnej výške pálčivá iskra, taká jasná, že na ňu nemožno priamo pozerieť, osvetľuje všetok viditeľný priestor, ako obnažené vlákno akejsi obrovskej žiarovky. Vo vzduchu visia bez podpory veľké nepochopiteľné chumáče bielej bavlny. A ten rozľahlý svietiaci strop nad tým všetkým... jeho *farba*... je...

Tu sa história rozvetvuje podľa toho, ktorý člen turistickej výpravy sa rozhodol putovať chodbou na povrch.

Aditya Modrá stála pod modrou oblohou celú večnosť a pomaly sa usmievala. Nebol to príjemný úsmev. Bola v ňom nenávisť a zranená pýcha; pripomínala si každú hádku, ktorú kedy mala so Zeleným, každé súperenie, každú prekážku v kariére. "*Mala si pravdu po celý čas,*" šepkala jej obloha zhora, "*a teraz to môžeš dokázať.*" Aditya tam chvíľu stála, nasávala túto správu, jasala nad ňou, a potom sa vrátila do kamennej chodby, aby to oznámila celému svetu. Ako Aditya kráčala, stískala dľaň v zaťatú päť. "*Prímerie,*" povedala, "skončilo."

Barron Zelený dlhé sekundy nechápavo hľadel na chaotické farby. Keď prišlo pochopenie, zasiahlo ho ako úder baranidlom do žalúdka. Z očí mu vytryskli slzy. Barron si spomenul na masaker v Cathay, kde Modrá armáda zmasakrovala všetkých obyvateľov Zeleného mesta vrátane detí; spomenul si na dávneho Modrého generála Annasa Rella, ktorý označil Zelených za "jamu plnú chorôb; mor, ktorý treba vyčistiť"; pomyslel na záblesky nenávisti, ktoré vídaval v očiach Modrých, a niečo sa v ňom zlomilo. "*Ako môžeš byť na ich strane?*" zakričal Barron na oblohu a potom začal plakať; pretože vedel, stojac pod nenávisťou modrou žiarou, že vesmír bol vždy miestom zla.

Charles Modrý zaskočene uvažoval nad modrou oblohou. Ako profesor zmiešanej univerzity Charles starostlivo zdôrazňoval, že názory Modrých aj Zelených sú rovnako platné a zaslúžia si toleranciu: Obloha je metafyzický konštrukt a nebeská je farba, ktorú možno vnímať viac ako jedným spôsobom. Charles sa nakrátko zamyslel, keby na danom mieste stál Zelený, či by nad sebou nevidel zelený strop; alebo či ten strop azda nebude zelený o takomto čase zajtra; nemohol však na to stavať ďalšie prežitie civilizácie. Toto je iba nejaký druh prírodného javu, ktorý nijako nesúvisí s morálnou filozofiou ani spoločnosťou... ľahko by sa však dal dezinterpretovať, obával sa Charles. Charles si vzdychol a vrátil sa do kamennej chodby. Zajtra sa sem vráti sám a cestu zablokuje.

Daria, kedysi Zelená, sa snažila dýchať uprostred trosiek svojho sveta. *Necúvnem*, vravela si Daria, *neodvrátim zrak*. Celý svoj život bola Zelená a teraz musí byť Modrá. Jej priatelia a jej rodina sa od nej odvrátia. *Hovor pravdu, aj keď sa ti chveje hlas*, hovorieval jej otec; ale jej otec už bol mŕtvy, a jej mama by to nikdy nepochopila. Daria opätovne oblohu pokojný modrý pohľad, pokúsila sa to prijať a nakoniec sa jej dych upokojil. *Mýlila som sa*, povedala si trúchlivo; *nie je to nakoniec také zložitá*. Nájde si nových priateľov a možno jej raz aj rodina odpustí... alebo, predstavovala si s nádychom nádeje, postúpi tú istú skúšku, postaví sa pod tú istú oblohu? "Obloha je modrá," povedala Daria pokusne a nič hrozné sa nestalo; nedokázala sa však usmievať. Daria Modrá smutne vydýchla a vrátila sa do sveta, rozmýšľajúc, čo povie.

Eddin Zelený pozrel na modrú oblohu a začal sa cynicky rehoť. Dejiny jeho sveta sa konečne vyjasnili; ani on nedokázal uveriť, že mohli byť až takí blázni. "Hlúposť," povedal Eddin, "hlúposť, *hlúposť*, a po celý čas to bolo priamo tu." Nenávisť, vraždy, vojny, a po celý čas išlo len o nejakú vec, o ktorej niekto písal tak, ako o hocičom inom. Ani poézia, ani nič krásne, nič na čom by

sa príčetnému človeku mohlo niekedy záležať, iba jedna blbosť, ktorá bola nafúknutá do nenormálnych rozmerov. Eddin sa unavene oprel o ústie jaskyne a rozmýšľal, ako tejto informácii zabráni zničiť celý svet, hoci uvažoval, či by si to vlastne všetci nezaslúžili.

Ferris mimovoľne vzdychol, zamrazený čírym úžasom a potešením. Ferris hladne pozeral okolo seba, postupne sa na všetko zrakom prisával a len neochotne odtrhol, aby pozrel ďalej; modrá *obloha*, biele *oblaky*, šírý neznámy *priestor* plný miest a vecí (a ľudí?), ktoré žiaden Podzemšťan ešte nevidel. "Aha, tak *takúto* farbu to má," povedal Ferris a pokračoval v skúmaní.

Pôvodný článok s komentármi¹

¹ http://lesswrong.com/lw/gt/a_fable_of_science_and_politics/

Kapitola 13. Viera vo vieru

Carl Sagan raz rozprával podobenstvo¹ o človeku, ktorý k nám príde a tvrdí: “V mojej garáži je drak.” Fascinujúce! Odpovedáme, že si toho draka chceme pozrieť – poďme ihneď do garáže! “Ale počkajte,” povie nám dotyčný, “to je *neviditeľný* drak.”

Ako Sagan podotýka, to ešte nerobí túto hypotézu nefalzifikovateľnou. Možno pôjdeme do garáže dotyčného a hoci neuvidíme žiadneho draka, budeme počuť ťažké dýchanie bez viditeľného zdroja; na podlahe sa budú tajomne zjavovať odtlačky nôh; a prístroje ukážu, že čosi v garáži spotrebováva kyslík a vydychuje kysličník uhličitý.

Predpokladajme však, že dotyčnému povieme: “Dobre, navštívime garáž a uvidíme, či budeme počuť ťažké dýchanie,” a dotyčný rýchlo povie: nie, to je *nepočutelný* drak. Navrhujeme odmerať kysličník uhličitý vo vzduchu a dotyčný povie, že tento drak nedýcha. Navrhujeme rozprášiť do vzduchu vrece múky a sledovať, či obklopí obrysy neviditeľného draka, a dotyčný ihneď povie: “Cez tohto draka múka prechádza.”

Carl Sagan použil toto podobenstvo na vykreslenie klasického ponaučenia, že zlé hypotézy vyžadujú veľa rýchlych dodatkov, aby sa vyhli falzifikácii. Ja však toto podobenstvo hovorím kvôli inej pointe: Dotyčný musí mať *niekde* v mysli presný model tejto situácie, pretože dokáže *presne* predpovedať, *z ktorých experimentálnych výsledkov sa potrebuje vyhovoriť*.

Niektorí filozofi sú takýmito scenármi veľmi zmätení a pýtajú sa: “Verí teda dotyčný *naozaj*, že je tam drak, alebo nie?” Akoby ľudský mozog mal na disku dost miesta iba pre jeden názor zároveň! Skutočné mysle sú omnoho zamotanejšie. Ako sme si povedali vo včerajšom článku, existujú rôzne typy názorov; nie každý názor je priame očakávanie². Dotyčný očividne *neočakáva*, že uvidí niečo nezvyčajné, keď otvorí dvere garáže; inak by sa dopredu nevyhováral. Množina výrokových názorov dotyčného môže zároveň obsahovať *V mojej garáži je drak*. Racionalistovi sa môže zdať, že tieto dva názory by sa mali zraziť a pobiť, aj keď sú rôzneho typu. Je však fyzikálny fakt, že môžete napísať “Obloha je zelená!” vedľa obrázku modrej oblohy, a papier kvôli tomu nevzbĺkne plameňom.

Racionalistická cnosť empirizmu nás má chrániť pred týmto druhom chyby. Máme sa stále pýtať svojich názorov, aké vnemy predpovedajú, aby platili nájomné formou očakávania. Avšak problém dotyčného s drakom je hlbší a nedá sa vyliečiť takouto jednoduchou radou. Nie je vyslovene *ťažké* spojiť vieru v draka s očakávaným vnemom v garáži. Ak veríš, že v tvojej garáži je drak, môžeš očakávať, že otvoríš dvere a draka uvidíš. Ak draka neuvidíš, potom to znamená, že v tvojej garáži drak nie je. To je pomerne priamočiare. Môžete si to vyskúšať vo vlastnej garáži.

Nie, táto vec s neviditeľnosťou je príznakom niečoho omnoho horšieho.

Podľa toho, aké ste mali detstvo, možno si pamätáte obdobie, keď ste prvýkrát začali pochybovať o existencii Deda Mráza, ale stále ste verili, že by ste *mali* veriť v Deda Mráza, tak ste tieto pochybnosti skúšali potlačiť. Ako si všimol Daniel Dennett, kde je ťažké v niečo veriť, je často omnoho ľahšie veriť, že *by ste* tomu *mali* veriť. Čo to znamená veriť, že Najvyššia Vesmírna Obloha³ je zároveň dokonale modrá aj dokonale zelená? Táto veta je zmätená; nie je vôbec jasné, čo *by znamenalo* veriť tomu – čomu presne *by ste* verili, keby ste verili.

¹ <http://www.godlessgeeks.com/LINKS/Dragon.htm>

² pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

³ pozri: 12 Bájka o vede a politike

Omnoho ľahšie je veriť, že je *správne*, že je *dobré* a *cnostné* a *užitočné* veriť tomu, že Najvyššia Vesmírna Obloha je zároveň dokonale modrá aj dokonale zelená. Dennett toto nazýva "viera vo vieru".

A tu sa veci skomplikujú, ako to už ľudské mysle zvyknú robiť – myslím si, že ešte aj Dennett príliš zjednodušuje, ako táto psychológia funguje v praxi. Ak totiž veríte, že veríte, nemôžete sami sebe priznať, že iba veríte, že veríte, pretože cnostné je *veriť*, nie *veriť*, že veríte, takže ak iba veríte, že veríte, namiesto aby ste verili, nie ste cnostní. Nikto sám sebe *neprizná*: "Neverím, že Najvyššia Vesmírna Obloha je modrá a zelená, ale verím, že by som tomu mal veriť" – jedine, že by bol nezvyčajne schopný priznať si svoj nedostatok cnosti. Ľudia neveria, že veria, že veria, oni skrátka veria, že veria.

(Ak ste z tohto zmätení, možno vám pomôže študovať matematickú logiku, kde sa vycvičíte robiť ostré rozdiely medzi výrokom P, dôkazom výroku P, a dôkazom že P je dokázateľné. Podobne ostré rozdiely sú medzi P, chcieť P, veriť v P, chcieť veriť v P, a veriť že veríte v P.)

Existujú rôzne druhy viery vo vieru. Môžete veriť vo vieru explicitne; môžete si vo svojom prúde vedomia recitovať vetu: "Je cnostné veriť, že Najvyššia Vesmírna Obloha je dokonale modrá aj dokonale zelená." (A pritom veriť, že tomu veríte, pokiaľ nie ste nezvyčajne schopní priznať si svoj nedostatok cnosti.) Ale sú aj menej explicitné formy viery vo vieru. Možno sa dotýčny s drakom bojí verejného výsmechu, ktorý si predstavuje ako následok verejného priznania, že sa mýlil (hoci racionalista by mu v skutočnosti zablahoželal a ostatní sa mu skôr budú vysmievať, ak bude naďalej tvrdiť, že má draka v garáži). Možno dotýčny s drakom uhýba pred predstavou priznania sebe samému, že žiaden drak nie je, pretože je to v rozpore s jeho sebaobrazom slávneho objaviteľa draka, ktorý videl vo svojej garáži to, čo sa ostatným nepodarilo vidieť.

Keby všetky naše myšlienky boli vedomé vety, s akými narábajú filozofi, ľudská myseľ by bola pre ľudí omnoho zrozumiteľnejšia. Prchavé myšlienkové obrazy, nevyslovené cúvnutia, riadenie sa nepriznanými túžbami – tie nás utvárajú tak isto, ako slová.

Aj keď s Dennettom nesúhlasím v niektorých detailoch a komplikáciách, stále si myslím, že Dennettov pojem *vieru vo vieru* je kľúčovým vhladom potrebným na pochopenie dotýčného s drakom. Potrebujeme však širší pojem *vieru*, neohraničený na verbálne vety. "Viera" by mala zahŕňať aj nevyslovené ovládače očakávania. "Viera vo vieru" by mala zahŕňať aj nevyslovené kognitívne-behaviorálne smernice. Nie je psychologicky realistické povedať: "Dotýčny neverí, že v jeho garáži je drak; verí, že je užitočné veriť, že v jeho garáži je drak." Ale je realistické povedať, že dotýčny *má očakávania akoby* v jeho garáži nebol žiaden drak, ale *má výhovorky akoby* veril, že verí.

Môžete mať obyčajný myšlienkový obraz svojej garáže, bez drakov, ktorý správne predpovedá vaše vnemy po otvorení dverí, a pritom si nikdy verbálne nepomyslieť vetu *V mojej garáži nie je žiaden drak*. Dokonca by som sa staval, že sa vám to stalo – že keď otvárate dvere na svojej garáži alebo na spálni alebo hocikde, a nečakáte žiadnych drakov, že vám hlavou nebehá žiadna takáto verbálna veta.

A aby ste cúvli pred vzdaním sa svojej viery v draka – alebo cúvli pred vzdaním sa svojho *sebaobrazu* ako človeka, ktorý verí v draka – nie je nutné explicitne si myslieť: *Chcem veriť, že v mojej garáži je drak*. Nutné je iba cúvnuť pred predstavou priznania si, že neveríte.

Aby dotýčny s drakom správne predpovedal, z ktorých experimentálnych výsledkov sa bude potrebovať vyhovoriť, musí (a) mať niekde v mysli správny

model riadiaci očakávaní a (b) kognitívne jednáť tak, aby chránil (b1) svoju voľne sa vznášajúcu výrokovú vieru v draka alebo (b2) svoj sebaobraz, ako verí v draka.

Ak niekto verí, že verí v draka, a zároveň verí v draka, problém je omnoho menej vážny. Taký človek bude ochotný ísť s kožou na trh ohľadom experimentálnych predpovedí a možno aj súhlasiť, že sa vzdá viery, ak sa experimentálny predpoklad ukáže nesprávny – aj keď vieru vo veru môže do tohto stále zasiahnuť, ak samotná vieru nie je úplne pevná. Ak si niekto chystá výhovorky *vopred*, zdá sa, že jeho vieru a vieru vo vieru už nie sú v súlade.

Pôvodný článok s komentármi⁴

⁴ http://lesswrong.com/lw/i4/belief_in_belief/

Kapitola 14. Bayesovské džudo

Môžete sa trochu zabaviť s ľuďmi, ktorých očakávania sa rozladili s tým, čomu veria, že veria¹.

Raz som bol na večernej párty, skúšal som jednému mužovi vysvetliť, čím sa živím, a on povedal: “Neverím, že je možná umelá inteligencia, lebo iba Boh dokáže stvoriť dušu.”

V tej chvíli som musel byť zhora osvietený, lebo som okamžite odpovedal: “Chcete tým povedať, že ak vytvorím umelú inteligenciu, vyvráti to vaše náboženstvo?”

On na to: “Čože?”

Povedal som: “No, ak vaše náboženstvo tvrdí, že je nemožné, aby som vytvoril umelú inteligenciu, potom, ak vytvorím umelú inteligenciu, bude to znamenať, že vaše náboženstvo nemá pravdu. Buď vaše náboženstvo pripúšťa, že môžem vytvoriť UI; alebo, ak zostavím UI, vyvráti to vaše náboženstvo.”

Nasledovala pauza, ako si uvedomil, že práve vyslovil hypotézu podliehajúcu falzifikácii, a potom povedal: “No, nemyslel som, že nemôžete vytvoriť inteligenciu, ale že nemôže byť emocionálna rovnakým spôsobom ako my.”

Povedal som: “Ak teda zostavím umelú inteligenciu, ktorá bez toho, že by mala naschvál naprogramovaný taký scenár, začne rozprávať o emocionálnom živote, ktorý bude znieť ako ten náš, to bude znamenať, že vaše náboženstvo sa mýli.”

On: “No, hm, asi sa budeme musieť zhodnúť, že sa v tomto nezhodneme.”

Ja: “Nie, to sa v skutočnosti nedá. Existuje veta o rozumnosti s názvom Aumannova veta o zhode, ktorá dokazuje, že dvaja racionalisti sa nemôžu zhodnúť na tom, že sa nezhodnú. Ak sa dvaja ľudia v niečom nezhodnú, aspoň jeden z nich musí niečo robiť nesprávne.”

Trochu sme diskutovali o tomto. Nakoniec povedal: “No, asi som tým chcel naozaj povedať, že si nemyslím, že dokážete vytvoriť niečo večné.”

Povedal som: “Nuž, to si nemyslím ani ja! Teší ma, že sme dokázali dosiahnuť zhodu, ako vyžaduje Aumannova veta o zhode.” Podal som mu ruku, potriasol mi ju a potom odišiel.

Žena, ktorá stála obďaleč a počúvala náš rozhovor, mi vážne povedala: “Toto bolo krásne.”

“Ďakujem veľmi pekne,” povedal som.

Pôvodný článok s komentármi²

¹ pozri: 13 Viera vo vieru

² http://lesswrong.com/lw/i5/bayesian_judo/

Kapitola 15. Vyznávanie a fandanie

Raz som sa zúčastnil panelovej diskusie na tému: “Sú veda a náboženstvo zlučiteľné?” Jedna zo žien v diskusii, pohanka, bez zastavenia rečnila o tom, ako verí, že Zem vznikla, keď sa v prvotnej priepasti narodila obrovská prvotná krava, ktorá lízaním priviedla na svet prvotného boha, ktorého potomkovia zabili prvotného obra a jeho telo použili na stvorenie Zeme, atď. Príbeh bol dlhý, plný podrobností a absurdnejší než predstava, že Zem leží na chrbte obrovskej korytnačky. A rozprávajúca očividne poznala vedu natolko, aby to vedela.

Doteraz zápasím so slovami, ako opísať, čo som videl, keď táto žena hovorila. Hovorila s... pýchou? Spokojnosťou so sebou samou? S úmyselným vystatovaním?

Táto žena pokračovala v opisovaní jej mýtu stvorenia zdanlivo celú večnosť, aj keď pravdepodobne to bolo iba päť minút. Tá zvláštna pýcha/spokojnosť/vystatovanie očividne nejako súvisela s jej *vedomím*, že jej presvedčenia sú z vedeckého hľadiska škandalózne. A nebolo to preto, že by nenávidela vedu; v rámci panelovej diskusie zastávala názor, že veda a náboženstvo sú zlučiteľné. Hovorila dokonca o tom, že je celkom pochopiteľné, prečo Vikingovia hovorili o prvotnej priepasti, vzhľadom na krajinu, v akej žili - odargumentovala svoje vlastné náboženstvo! - a napriek tomu trvala na tom, že tomuto “verí”, a vyslovovala to so svojráznym uspokojením.

Nie som si istý, či Danielov Dennettov pojem “viery vo vieru¹” siaha tak ďaleko, aby zahrnul aj túto udalosť. Toto bolo ešte čudnejšie. Neprednášala svoj mýtus o stvorení s fanatickou vierou niekoho, kto potrebuje ubezpečiť sám seba. Nesprávila sa tak, akoby očakávala, že presvedčí nás, obecnosť - ani akoby našu vieru potrebovala na potvrdenie vlastnej.

Dennett, okrem pomenovania viery vo vieru, tvrdil aj, že mnohé z toho, čo nazývame “náboženské presvedčenie”, by sa naozaj malo študovať ako “náboženské vyznávanie”. Predstavte si, že mimozemský antropológ študuje skupinu študentov postmodernej angličtiny, ktorí napohľad všetci *veria*, že Wulky Wilkinsen bol postutopický autor. Správna otázka by nemala znieť: “Prečo všetci študenti veria tomuto čudnému názoru?” ale “Prečo všetci píšú túto čudnú vetu na písomke?” Lebo aj keď nejaká veta v podstate nemá zmysel, stále môžete vedieť, že sa od vás očakáva, že ju budete nahlas skandovať.

Myslím si, že Dennett je možno trochu príliš cynický, keď naznačuje, že náboženské vyznávanie je *iba* vyslovovanie názorov nahlas - väčšina ľudí je natolko čestná, že keď povedia náboženskú vetu nahlas, cítia povinnosť povedať túto verbálnu vetu aj vo svojom vlastnom prúde vedomia.

Ale zdá sa, že ani pojem “náboženského vyznávania” celkom nezahŕňa tvrdenie tejto pohanky o viere v prvotnú kravu. Keby ste mali vyznať náboženskú vieru k spokojnosti kňaza, či k spokojnosti spoluveriacich - sakra, keby ste mali uspokojiť svoj vlastný sebaobraz ako veriaceho človeka - museli by ste *predstierať*, že veríte, *omnoho presvedčivejšie*, než robila táto žena. Ako recitovala svoj príbeh o prvotnej krave, stále s tou čudnou vystatovačnou pýchou, ani sa *nesnažila* byť presvedčivá - ani sa nesnažila presvedčiť nás, že berie svoje vlastné náboženstvo vážne. Myslím si, že aj toto ma na tom tak zarazilo. Poznám ľudí, ktorí vedia, že veria smiešnym veciam, ale keď ich vyznávajú, dávajú omnoho viac úsilia do presvedčania seba samých, že svoju vieru berú vážne.

¹ pozri: 13 Viera vo vieru

Nakoniec mi došlo, že táto žena sa nepokúša presvedčiť ani nás, ani seba. Jej recitovanie mýtu o stvorení vôbec *nebolo o mýte o stvorení*. Namiesto toho, svojím útočným päťminútovým blabotáním *fandila pohanstvu*, ako keby držala vlajku na futbalovom zápase. Vlajka, ktorá hovorí: “MODRÍ, DO TOHO!²”, nie je vyjadrením faktu, ani pokusom presvedčiť; nemusí byť presvedčivá – je to pokrik.

Tá zvláštna vystatovačná pýcha... to bolo, akoby pochodovala nahá v sprievode gay pride. (Mimochodom, nemal by som žiadne námietky, keby *naozaj* pochodovala nahá v sprievode gay pride. Lesbismus nie je niečo, čo môže byť zničené pravdou³.) Nebol to len pokrik, ako keď niekto pochoduje, ale pohoršujúci pokrik, ako keď pochoduje nahý – verí, že ju nemožno uväzniť ani kritizovať, pretože to robí pre svoj pochod hrdosti.

Preto jej záležalo na tom, aby hovorila veci, ktoré boli viac než smiešne. Keby sa pokúšala, aby to vyznelo prijateľnejšie, bolo by to ako obliecť sa.

Pôvodný článok s komentármi⁴

² pozri: 12 Bájka o vede a politike

³ pozri: 2 Rozumné čítanie

⁴ http://lesswrong.com/lw/i6/professing_and_cheering/

Kapitola 16. Signály na potlesk

Súvisí s kapitolou a článkom: Sémantické stopky¹, V skutočnosti nechceme vašu účasť²

Na Summite Singularity 2007 jeden z rečníkov požadoval demokratický nadnárodný vývoj UI. Pristúpil som teda k mikrofónu a povedal:

Predstavte si, že skupina demokratických republík vytvorí konzorcium na vývoj UI, a že v celom tom procese bude veľa politikárčenia - niektoré záujmové skupiny budú mať nezvyčajne veľký vplyv, iné budú odstránené - inými slovami, výsledok bude vyzeráť tak ako produkty moderných demokracií. Ako alternatívu si predstavte, že skupina rebelských kockáčov vyvinie UI doma v pivnici, a prikáže tejto UI aby sa opýtala každého na celom svete - keď dodá mobil každému, kto ho nemá - a aby urobila čokoľvek, čo povie väčšina. Ktorá z týchto možností je podľa vás "demokratickejšia", a cítili by ste sa pri niektorej z nich bezpečne?

Chcel som zistiť, či verí v pragmatickú primeranosť demokratického politického procesu alebo či verí v morálnu správnosť hlasovania. Rečník však odpovedal:

Ten prvý scenár znie ako redakčný komentár v časopise Reason a ten druhý znie ako zápleтка hollywoodskeho filmu.

Zmätene som sa opýtal:

Aký druh demokratického procesu ste teda *mali* na mysli?

Rečník odpovedal:

Niečo ako projekt ľudského genómu - to bol medzinárodne sponzorovaný výskumný projekt.

Opýtal som sa:

Ako by rôzne záujmové skupiny riešili svoje konflikty v štruktúre ako bol projekt ľudského genómu?

A rečník povedal:

Neviem.

Táto slovná výmena mi pripomenula výrok³ nejakého diktátora, ktorého sa pýtali, či má niekedy v úmysle posunúť svoj štátik smerom k demokracii:

My veríme, že už máme demokratický systém. Niektoré prvky v ňom zatiaľ chýbajú, napríklad možnosť ľudu vyjadriť svoju vôľu.

Podstatou demokracie je konkrétny mechanizmus, ktorý rieši politické konflikty. Keby všetky skupiny uprednostňovali rovnaké pravidlá, nepotrebovali by sme demokraciu - spolupracovali by sme automaticky. Procesom uznášanania môže byť priame väčšinové hlasovanie, volení zákonodarcovia, alebo hoci aj UI citlivá na názory voličov, ale musí to byť *niečo*. Čo to vôbec *znamená*, žiadať "demokratické" riešenie, ak tým nemyslíme mechanizmus rozhodovania konfliktov?

¹ pozri: 35 Sémantické stopky

² http://lesswrong.com/lw/ja/we_dont_really_want_your_participation/

³ <http://content.time.com/time/magazine/article/0,9171,954853,00.html>

Myslím, že to znamená, že ste povedali slovo “demokracia” a preto sa čaká, že obecenstvo zatlieska. Nie je to ani tak výrokové tvrdenie, ako ekvivalent signálu “Potlesk”, ktorý oznamuje obecenstvu v štúdiu, kedy má tleskať.

Tento príklad je zaujímavý iba tým, že som si pomýlil signál na potlesk s návrhom pravidiel, čoho výsledkom bol následný všeobecný trapas. Väčšina signálov na potlesk je omnoho jasnejšia a dokážete ich odhaliť jednoduchým testom opaku. Predstavte si napríklad, že niekto povie:

Mali by sme zvážiť riziká a príležitosti UI.

Ak urobíte opak tohto tvrdenia, dostanete:

Nemali by sme zvážiť riziká a príležitosti UI.

Keďže tento opak znie *nenormálne*, pôvodné tvrdenie je pravdepodobne normálne, z čoho vyplýva, že nevyjadruje novú informáciu. Existuje veľa legitímnych dôvodov na povedanie vety, ktorá by mimo kontextu nedávala informáciu. “Mali by sme zvážiť riziká a príležitosti UI” môže byť úvodom k diskusnej téme; môže zdôrazniť dôležitosť konkrétneho návrhu ako ich zvažovať; môže kritizovať nevyvážený návrh. Odkázať na normálne tvrdenie môže ohraničenému racionalistovi sprostredkovať novú informáciu – súvislosť nemusela byť samozrejmá. Ale ak nenasleduje *nič* konkrétne, táto veta je pravdepodobne signálom na potlesk.

Som v pokušení urobiť raz prednášku, ktorá nebude obsahovať *nič iné* ako signály na potlesk a čakať, ako dlho obecenstvu potrvá, než sa začne smiať:

Som tu, aby som vám dnes navrhol, aby sme zvážili riziká a príležitosti pokročilej Umelej Inteligencie. Mali by sme sa vyhnúť rizikám a v rámci možností využiť príležitosti. Nemali by sme sa zbytočne vystavovať nebezpečenstvám, ktorým sa môžeme vyhnúť. Aby sme dosiahli tieto ciele, musíme plánovať múdro a rozumne. Nemali by sme konať v strachu a panike, ani sa poddávať technofóbii; nemali by sme sa však riadiť ani slepým nadšením. Mali by sme rešpektovať záujmy všetkých strán, ktorých sa Singularita týka. Musíme sa pokúsiť zabezpečiť, aby z pokročilej technológie vyplynuli výhody čo najväčšiemu počtu jednotlivcov, nie iba hŕstke vyvolených. Musíme sa v rámci možností pokúsiť vyhnúť násilným konfliktom s využitím týchto technológií; a musíme zabrániť jednotlivcom, ktorí by sa chceli zmocniť rozsiahlej ničivej kapacity. Musíme o týchto veciach rozmýšľať vopred, nie až dodatočne, keď už bude príliš neskoro s tým niečo urobiť...

Pôvodný článok s komentármi⁴

⁴ http://lesswrong.com/lw/jb/applause_lights/

Kapitola 17. Názor ako uniforma

Zatiaľ som opísal rozdiely medzi názorom ako ovplyvňovačom očakávania¹, vierou vo vieru², vyznávaním a fandaním³. Spomedzi týchto by sme názory ovplyvňujúce očakávania mohli nazvať “pravé názory” a zvyšné formy “nepravé názory”. Pravý názor môže byť nesprávny alebo nerozumný, napríklad keď niekto naozaj očakáva, že modlitba vylieči jeho choré dieťa, ale o tých zvyšných formách by sa dalo povedať, že to “vlastne nie je názor”.

Ďalšou formou nepravého názoru je názor ako skupinová identifikácia - ako spôsob zapadnutia. Robin Hanson používa ako výbornú metaforu⁴ nosenie nezvyčajného oblečenia, skupinovej uniformy, ako je kňazovo rúcho alebo židovská čiapočka, takže ja to nazvem “názor ako uniforma”.

V pojmoch realistickej psychológie človeka⁵, moslimovia, ktorí narazili lietadlom do World Trade Center, nepochybne vnímali sami seba ako hrdinov brániacich pravdu, spravodlivosť a islamský spôsob života pred odpornými cudzími netvormi, ako vo filme Deň nezávislosti⁶. Iba veľmi neskúsený čudák, taký typ čudáka, ktorý ani netuší, ako svet vnímajú nečudáci, by toto povedal nahlas v krčme v Alabame. Toto jednoducho Američania nehovoria. Američania hovoria, že teroristi “nenávidia našu slobodu” a že narazenie lietadlom do budovy je “zbabelý čin”. Nemôžete v jednej vete vysloviť slová “hrdinské sebaobetovanie” a “samovražedný atentátnik”, dokonca ani za účelom presného popisu, ako nepriatelia vnímajú svet. Samotný pojem odvahy a altruizmu samovražedného atentátnika je nepriateľská uniforma - je to zrejme, lebo takto to hovoria nepriatelia. Zbabelosť a sociopatia samovražedného atentátnika je americká uniforma. Žiadne úvodzovky nie sú dosť dobré, aby ste v nich mohli hovoriť, ako nepriateľ vníma svet; bolo by to ako ísť na Halloween prezlečený za nacistu.

Názor ako uniforma môže pomôcť vysvetliť, ako sa ľudia dokážu *zapáliť* pre nepravý názor. Samotná viera vo vieru⁷ alebo náboženské vyznávanie⁸ by len ťažko vytvorili skutočné, hlboké, mocné emocionálne účinky. Aspoň si to myslím; priznávam, že na toto nie som odborník. Ale zdá sa mi, že je to takto: Ľudia, ktorí prestali očakávať v súlade so svojím náboženstvom, môžu urobiť veľa preto, aby sami seba *presvedčili*, že sú zapálení, a toto zúfalstvo si možno pomýliť so zápalom. Nie je to však ten istý oheň, ktorý mali ako deti.

Na druhej strane, pre človeka je veľmi ľahké skutočne, vášnivo, celou svojou bytosťou patriť do nejakej skupiny, fandiť svojmu oblúbenému športovému tímu⁹. (Na tomto základe spočíva podvod s “republikánmi alebo demokratmi” a podobné falošné dilemy¹⁰ v iných krajinách, ale to je téma na iný článok.) Stotožniť sa s nejakým kmeňom je veľmi silná emócia. Ľudia za ňu zomierajú. A keď sa raz ľudia stotožnia s nejakým kmeňom, budú vyslovovať názory, ktoré sú uniformou tohto kmeňa, s plnou vášňou zapadnutia do tohto kmeňa.

Pôvodný článok s komentármi¹¹

¹ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

² pozri: 13 Viera vo vieru

³ pozri: 15 Vyznávanie a fandanie

⁴ http://lesswrong.com/lw/i6/professing_and_cheering/egb

⁵ pozri: 51 Sú vaši nepriatelia od narodenia zlí?

⁶ <http://www.imdb.com/title/tt0116629/>

⁷ pozri: 13 Viera vo vieru

⁸ pozri: 15 Vyznávanie a fandanie

⁹ pozri: 12 Bájka o vede a politike

¹⁰ pozri: 111 Tretia možnosť

¹¹ lesswrong.com/lw/i7/belief_as_attire/

Kapitola 18. Predstieranie múdrosti

Súvisí s článkom: Proti dospelosti¹

Najhorúcejšie miesto v pekle je vyhradené pre tých, ktorí v čase krízy zostali neutrálni.

—John F. Kennedy, nesprávne citujúc Danteho

Jeden príklad signalizovania dospelosti, ktorému sa oplatí venovať osobitne, je prejavovanie *neutrality* alebo *vyčkávania so záverom* s cieľom signalizovať dospelosť, múdrosť, nestrannosť alebo nadhľad.

Príkladom môže byť včera opísaný prípad mojich rodičov, ktorí mi na teologické otázky ako: “Keď mal staroveký Egypt také dobré záznamy o mnohých iných veciach, prečo nemá žiaden záznam o tom, že tam niekedy boli Židia?” odpovedali: “Ach, keď som mal tvoj vek, tiež som kládol podobné otázky, ale už som z toho vyrástol.”

Iným príkladom môže byť riaditeľ školy, ktorý prichytí dve deti, ako sa bijú na dvore a prísne povie: “Nezáleží na tom, kto bitku začal, ale kto ju skončí.” Samozrejme že záleží na tom, kto bitku začal. Riaditeľ zrejme nemá k dispozícii dobré *informácie* o tomto dôležitom fakte, ale ak je to tak, mal by to *povedať* a *neodmietat dôležitosť* toho, kto udrel prvý. Keby nejaký rodič skúsil udrieť riaditeľa, videli by sme, ako ďaleko by sa na súde dostal s postojom: “Nezáleží na tom, kto začal.” Pre dospelých sú však detské bitky akurát *nepohodlné* a pre ich *pohodlie* je nepodstatné, ktoré dieťa začalo, podstatné je iba čo najrýchlejšie bitku ukončiť.

Som presvedčený, že podobne to funguje v medzinárodnej diplomacii, keď veľmoci prísne hovoria menším skupinám, aby *okamžite* prestali bojovať. Veľmociam nezáleží na tom, kto začal – kto provokoval alebo kto na provokáciu neprimerane reagoval – pretože pre veľmoci je *nepohodlné* iba pokračovanie konfliktu. Nemôžu si Izrael a Hamas skrátka podať ruky?

Toto nazývam “predstieranie Múdrosti”. Samozrejme existuje veľa spôsobov ako skúšať signalizovať múdrosť. Ale skúšať signalizovať múdrosť tým, že odmietam hádať – odmietam sčítať indície – odmietam vysloviť záver – odmietam zaujať stranu – stojím mimo sporu a pozerám zhora povýšeným a pohrdavým pohľadom – inými slovami, signalizujem múdrosť tým, že nič nehovorím a nerobím – tak toto mi pripadá mimoriadne *pozérske*.

Paolo Freire povedal: “Umyť si ruky nad konfliktom medzi mocným a bezmocným znamená postaviť sa na stranu mocného, nie byť neutrálny.” Školský dvor je skvelým miestom pre šikanujúcich a hrozným miestom pre obeť, ak učiteľov nezaujímá, *kto začal*. Podobne v medzinárodnej politike: Svet, v ktorom sa veľmoci odmietajú prikloniť na niektorú stranu a iba žiadajú okamžité prímerie, je skvelý svet pre agresorov a hrozným miestom pre napadnutých. Ale samozrejme je to veľmi pohodlné miesto pre veľmoc alebo riaditeľa školy.

Časť tohto správania možno teda pripísať číremu sebecktvu zo strany toho Múdrehu.

Ale časť z toho súvisí aj so signalizovaním nadhľadu. Napokon – čo by si *ostatní dospelí* pomysleli o riaditeľovi, ktorý by sa naozaj *pridal* na niektorú stranu pri

¹ http://lesswrong.com/lw/yo/against_maturity/

bitke púhych *detí*? Nuž znížilo by to jeho spoločenské postavenie na úroveň *púheho účastníka hádky*!

Podobné to má uctievaný starešina – či už je to generálny riaditeľ, prestížny akademik alebo zakladateľ diskusnej skupiny – ktorého povest spravodlivého závisí od jeho odmietania vysloviť záver tam, kde sa ostatní pridávajú na niektorú zo strán. Strany sa dovolávajú jeho podpory, ale takmer vždy márne; pretože Múdri sú uctievaní ako sudcovia iba pod podmienkou, že takmer nikdy naozaj nesúdia – *tým* by sa z nich stali iba ďalší účastníci sporu, o nič lepší než ktorýkoľvek iný.

(Zvláštne je, že sudcovia v skutočnom systéme zákona môžu opakovane vynášať skutočné rozsudky a nestrácajú tým *automaticky* svoju povest nestrannosti. Možno vďaka všeobecne chápanej norme, že *musia* súdiť, že to je ich náplň práce. Alebo možno preto, lebo sudcovia nemusia opakovane rozhodovať v témach rozdeľujúcich kmeň, na ktorého úcte závisia.)

Existujú *aj* prípady, kde je rozumné počkať s vyslovením záveru, kde ľudia rýchlo vyslovujú závery iba vďaka svojim skresleniam. Ako povedal² Michael Rooney: “Táto chyba sa podobá tej, ktorú stále vidím u začínajúcich študentov filozofie: keď sa stretnú s dôvodmi na skepticizmus, namiesto toho sa stanú relativistami. To znamená, že keď je rozumným postupom počkať s vyslovením záveru, privedla ľudí namiesto toho vysloviť záver, že každý záver je rovnako dôveryhodný ako hociktorý iný.”

Ale ako sa teda môžeme vyhnúť (súvisiacemu, no odlišnému) pseudo-racionalistickému správaniu, keď signalizujeme svoju neskreslenú nestrannosť pomocou falošného tvrdenia, že súčasný stav indícií je vyvážený? “Veď áno, samozrejme existuje veľa zapálených darwinistov, ale ja si myslím, že nemáme dostatok dôkazov na to, aby sme sa mohli definitívne prikloniť k prirodzenému výberu namiesto inteligentného dizajnu.”

Na túto tému si odporúčam zapamätať, že *aj neutralita je definitívny záver*. Neznamená to stáť *nad* vecou. Znamená to vyjadriť definitívny a konkrétny záver, že rovnováha indícií v konkrétnej veci povoľuje *iba* jediný záver a to je neutrálny záver. Keďže *aj* takýto záver môže byť nesprávny, uprednostňovanie neutrality sa dá napadnúť úplne rovnako ako uprednostňovanie hociktorej strany.

Podobne v otázkach politiky. Keď niekto tvrdí, že *aj* strana pro-life *aj* strana pro-choice má dobré argumenty a že by sa v skutočnosti mali pokúsiť o kompromis a vzájomné rešpektovanie, nezastáva tým stanovisko, ktoré je *nad* oboma stranami debaty o interrupciách. Predkladá tým definitívny záver, ktorý je rovnako konkrétny ako povedanie „pro-life!” alebo „pro-choice!”

Mimochodom... toto nie je pozvánka na diskutovanie v komentároch o interrupciách ani o Izraeli a Hamase. *Overcoming Bias* tu nie je na tento druh diskusií; existujú iné diskusné fóra určené na tieto *úplne legitímne témy*. Možno až sa *Less Wrong* raz dostatočne rozrastie... ale nie teraz.

Ale to *neznamená*, že racionalista je príliš dospelý na to, aby hovoril o politike. *Neznamená* to, že racionalista je povznesený nad tieto hlúpe spory, do ktorých sa zapájajú iba politickí straníci a mladiství nadšenci.

Ako opisuje Robin, schopnosť mať potenciálne rozdeľujúce konverzácie je obmedzený zdroj³. Ak si myslíte, že dokážete potiahnuť lano nabok⁴, máte dôvod vynaložiť svoje obmedzené zdroje na pomerne menej časté témy, kde vaša hraničná diskusia umožňuje pomerne väčší hraničný výnos.

² pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

³ <http://www.overcomingbias.com/2009/02/the-cost-of-talking-values.html>

⁴ http://www.overcomingbias.com/2007/05/policy_tugowar.html

Ale veci, ktorým ste dali menšiu prioritu, sú potom otázkou vašich obmedzených zdrojov. *Nie* otázkou vznášania sa nad vecou, pokojne a Múdro.

Dodatok: Moja reakcia⁵ na komentár PG na HN⁶ vyzerá ako zhrnutie, ktoré sa oplatí zopakovať. Je rozdiel medzi:

- Vyslovením neutrálneho záveru;
- Odmietnutím investovať hraničné zdroje;
- Predstieraním, že niektoré z horeuvedeného je známkou hlbkej múdrosti, dospelosti a nadhľadu; s príslušným dôsledkom, že pôvodné strany majú menej nadhľadu a odtiaľto zhora medzi nimi nie sú podstatné rozdiely.

Pôvodný článok s komentármi⁷

⁵ <https://news.ycombinator.com/item?id=489863>

⁶ <https://news.ycombinator.com/item?id=489702>

⁷ http://lesswrong.com/lw/yp/pretending_to_be_wise/

Kapitola 19. Nárok náboženstva na nevyvrátiteľnosť

Najstarší mne známy záznam vedeckého experimentu je, ironicky, príbeh o Eliášovi a Baalových kňazoch¹.

Lud Izraela kolísal medzi Jehovom a Baalom, preto Eliáš oznámil, že vykoná experiment, ktorý to rozhodne – za oných čias dosť nová myšlienka! Baalovi kňazi položia na oltár svojho býka, Eliáš položí na oltár Jehovovho býka, ale žiaden z nich nesmie zapáliť oheň; ktorý Boh je pravý, ten zošle oheň na svoju obeť. Baalovi kňazi slúžili Eliášovi ako kontrolná skupina – rovnaké drevené palivo, rovnaký býk, rovnakí kňazi, avšak vzývajúci nepravého boha. Potom Eliáš na svoj oltár nalial vodu – čím pokazil symetriu experimentu, ale to bolo ešte za dávnych čias – aby znázornil vedomé prijatie bremena dôkazu, ako keď sa požaduje hladina spoľahlivosti 0,05. Oheň prišiel zhora na Eliášov oltár, čo je experimentálne pozorovanie. Prizerajúci sa izraelský ľud kričal: “Hospodin je Boh!” – peer review.

A potom ľudia odtiahli 450 Baalových kňazov k rieke Kišon a podrezali im hrdlá. Je to tvrdé, ale potrebné. Falzifikovanú hypotézu musíte jasne vyradiť, a to rýchlo, skôr než si začne vytvárať ochranné výhovorky. Keby Baalovým kňazom dovolili prežiť, začali by blabotať o tom, že náboženstvo je oddelené magistérium, ktoré nemožno ani dokázať ani vyvrátiť.

Vtedy za starých čias ľudia naozaj verili² svojim náboženstvám namiesto toho, aby iba verili v ne³. Biblickí archeológovia, ktorí hľadali Noemovu archu, si nemysleli, že márnia svoj čas; očakávali⁴, že sa môžu presláviť. Až keď sa im nepodarilo nájsť potvrdzujúce indície – a namiesto nich našli odporujúce indície – urobili religionisti to, čo William Bartley nazval *útočiskom v záväzku*: “Verím, pretože verím.”

Vtedy za starých čias neexistovala predstava, že náboženstvo je oddelené magistérium. Starý Zákon je záznamom kultúrneho prúdu vedomia: história, zákon, mravné prirovnania, a takisto aj modely, ako funguje vesmír. V žiadnej pasáži Starého Zákona nenájdete nikoho hovoriť o nadprirodzenom zázraku zložitosti vesmíru. Ale nájdete kopec vedeckých tvrdení⁵, napríklad že vesmír bol stvorený za šesť dní (čo je metafora pre Big Bang), alebo že zajace prežívajú svoju potravu. (Čo je metafora pre...)

Vtedy za starých čias, keby ste povedali, že miestne náboženstvo “sa nedá dokázať”, upálili by vás na hranici. Jedným zo základných kameňov viery ortodoxného judaizmu je, že Boh sa objavil na hore Sinaj a povedal hromovým hlasom: “Áno, to všetko je pravda.” Z bayesovského pohľadu⁶ je to sakra jednoznačná indícia pre nadľudsky mocnú bytosť. (Hoci to samotné ešte nedokazuje, že táto bytosť je Boh, alebo že táto bytosť je dobrá – mohli to byť mimozemskí pubertáci.) Prevažná väčšina náboženstiev v ľudskej histórii – okrem tých vynájdenej *veľmi* nedávno – rozpráva príbehy o udalostiach, ktoré by tvorili celkom nepochybné indície, keby sa naozaj stali. Nezávislosť náboženstva a faktických otázok je *nedávny* a výlučne *západný* prístup. Ľudia, ktorí napísali pôvodné sväté písma, tento rozdiel ani nepoznali.

¹ http://www.biblia.sk/index.php?akc=biblia_sk&hl_kniha=1ki&strana=18&hl_druh=0

² pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

³ pozri: 13 Viera vo vieru

⁴ pozri: 13 Viera vo vieru

⁵ <http://www.skepticfiles.org/atheist/genesisd.htm>

⁶ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

Rímska ríša zdedila filozofiu od starých Grékov; zaviedla vo svojich provinciách zákon a poriadok; udržiavala úradné záznamy; a vynucovala náboženskú toleranciu. Nový Zákon, vytvorený v období Rímskej ríše, nesie v dôsledku toho stopy modernosti. Nemohli by ste si vymyslieť príbeh o tom, ako Boh úplne zničil mesto Rím (na štýl Sodomy a Gomory), pretože rímski historici by vám oponovali, a vy by ste ich nemohli jednoducho ukameňovať.

Naopak, ľudia, ktorí vymysleli príbehy Starého Zákona, si mohli vymyslieť prakticky čokoľvek, čo sa im zapáčilo. Raní egyptológovia boli úprimne šokovaní, keď nenašli absolútne žiadnu stopu po tom, že by hebrejské kmene niekedy boli v Egypte - neočakávali síce záznam o desiatich Božích ranách, ale očakávali, že nájdu aspoň *niečo*. Ako sa ukázalo, nenašli nič. Zistili, že v údajnej dobe úteku Egypt ovládal väčšinu Kanaánu. To je jedna *obrovská* historická chyba, ale keď neboli žiadne knižnice, nikto vám nemohol oponovať.

Rímska ríša mala knižnice. Preto Nový Zákon neobsahuje tvrdenia o veľkých, okázalých, rozľahlých geopolitických zázrakoch, ako Starý Zákon robí bežne. Namiesto toho Nový Zákon hovorí o menších zázrakoch, ktoré však stále zapadajú do rovnakej schémy indícií. Chlapec spadne a ide mu pena z úst; príčinou je nečistý duch; od nečistého ducha môžeme logicky očakávať, že ujde pred pravým prorokom, ale neujde pred šarlatánom; Ježiš vyhnal nečistého ducha; preto je Ježiš pravý prorok a nie šarlatán. Toto je úplne normálne bayesovské uvažovanie⁷, ak vyjdeme zo základného predpokladu, že epilepsiu spôsobujú démoni (a že koniec epileptického záchvatu dokazuje, že démon ušiel).

Nielenže si náboženstvo zvyklo robiť nárok na faktické a vedecké záležitosti, náboženstvo si zvyklo robiť nárok na *všetko*. Náboženstvo predpisovalo kódex zákona - v časoch pred legislatívnymi orgánmi; náboženstvo predpisovalo históriu - v časoch pred historikmi a archeológmi; náboženstvo predpisovalo sexuálnu morálku - v časoch pred ženskou rovnoprávnosťou; náboženstvo opisovalo formy vlády - v časoch pred ústavou; a náboženstvo odpovedalo aj na vedecké otázky ohľadom biologickej taxonómie alebo vzniku hviezd. Starý Zákon nerozprával o pocite zázraku pri pohľade na zložitosť vesmíru - sústredil sa na definovanie trestu smrti pre ženy, ktoré si obliekali mužské šaty, lebo to sa v danej dobe považovalo za solídny a uspokojujúci obsah náboženstva. Moderná predstava náboženstva ako púhej *etiky* vyplýva z toho, že každú inú oblasť už prebrali vhodnejšie inštitúcie. Etika je to, čo *zostalo*.

Presnejšie povedané, ľudia *si myslia*, že etika je to, čo zostalo. Vezmite si kultúrny záznam spred 2 500 rokov. Za ten čas ľudstvo ohromne napreduje a kúsky starovekého kultúrneho záznamu sa stávajú čoraz jasnejšie zastaranými. Etika tiež nie je imúnna voči ľudskému pokroku - napríklad dnes sa pozeráme zhora na Bibliu odporúčané praktiky, ako je vlastníctvo otrokov. Prečo si ľudia *myslia*, že v etike je stále dovolené hocičo?

V podstate to vôbec nie je malý etický problém, keď niekto pozabíja tisíce nevinných prvorodených chlapcov, aby tým presvedčil nevoleného faraóna, aby prepustil otrokov, ktorých sa logicky dalo z krajiny odteleportovať. Malo by to byť *jasnejšie*, než porovnateľne drobná vedecká chyba tvrdenia, že lúčne koníky majú štyri nohy. Napriek tomu, ak poviete, že Zem je plochá, ľudia na vás budú pozeráť ako blázna. Ale ak poviete, že Biblia je vašim zdrojom morálky, ženy vás nezačnú fackovať. Pre väčšinu ľudí je predstava rozumnosti daná tým, čo si myslia, že im prejde; myslia si, že im prejde schvaľovanie biblickej morálky; a preto vyžaduje iba zvládnuteľné množstvo sebaklamy, aby prehliadli morálne problémy Biblie. Všetci sa dohodli na tom, že si slona v obývačke nebudú všimáť a takýto stav sa dá istý čas udržať.

⁷ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

Možno jedného dňa ľudstvo pokročí ďalej a kto bude odporúčať Bibliu ako zdroj morálky, bude vnímaný tak, ako keď Trent Lott schvaľoval prezidentskú kampaň Stroma Thurmonda. A potom sa povie, že "skutočnou podstatou" náboženstva bola vždy genealógia, alebo niečo.

Predstava, že náboženstvo je oddelené magistérium, ktoré *nemožno dokázať ani vyvrátiť*, je veľká lož - lož, ktorá sa opakuje znova a znova, aby ju ľudia vyslovovali bez rozmýšľania; ktorá je však, po kritickom preskúmaní, jednoducho nepravdivá. Je to hrubé skreslenie toho, ako náboženstvo historicky vzniklo, ako všetky sväté písma prezentujú svoju vieru, čo sa rozpráva deťom, aby uverili, a v čo väčšina veriacich ľudí na Zemi stále verí. Musíte obdivovať túto číru drzosť, na rovnakej úrovni, ako že *Oceánia bola vždy vo vojne s Eastáziou*. Žalobca vytasí zakrvavenú sekeru, a obžalovaný, na chvíľku zaskočený, sa rýchlo zamyslí a povie: "Ale moju nevinu nemôžete vyvrátiť púhym dôkazom - to je oddelené magistérium!"

A keď nevyjde toto, schmatnite kúsok papiera a načmárajte si vlastnú priepustku z väzenia.

Pôvodný článok s komentármi⁸

⁸ http://lesswrong.com/lw/i8/religions_claim_to_be_nondisprovable/

Časť III. Všímanie si zmätku

Obsah

20 Čo je indícia?	47
21 Vedecká indícia, právna indícia, rozumná indícia	49
22 Koľko indície treba?	51
23 Occamova britva	53
24 Einsteinova drzosť	56
25 Zameraj svoju neistotu	58
26 Cnosť presnosti	60
27 Tvoja sila racionalistu	62
28 Neprítomnosť indície je indíciou neprítomnosti	64
29 Zákon zachovania očakávanej indície	66
30 Spätňý pohľad znehodnocuje vedu	68

Kapitola 20. Čo je indícia?

Veta “sneh je biely” je *pravdivá* vtedy a iba vtedy, ak sneh je biely.
—Alfred Tarski

Povedať o tom, čo je, že to je, alebo o tom, čo nie je, že to nie je,
to je *pravda*.

—Aristoteles, Metafyzika IV

Ak vám tieto dva citáty nepripadajú ako dostatočná definícia “pravdy”, prečítajte si toto¹. Dnes budem hovoriť o “indícii”. (Tiež chcem hovoriť o názoroch ohľadom faktov, nie emócií či morálky, ako je rozlíšené tu².)

Kráčate po ceste, vaše šnúry sa rozviažu. Krátko na to z nejakého zvláštneho dôvodu začnete *veriť*, že vaše šnúry sú rozviazané. Svetlo opustí Slnko, dopadne na vaše šnúry a odrazí sa; niektoré fotóny vojdú do zreničiek vašich očí a dopadnú na vašu sietnicu; energia fotónov spustí nervové impulzy; nervové impulzy sa prenesú do oblastí mozgu spracovávajúcich obraz; a tam sa optická informácia spracuje a zrekonštruje na 3D model, ktorý sa rozozná ako rozviazaná šnúra. Je to postupnosť udalostí, reťaz príčin a následkov, medzi svetom a vašim mozgom, pomocou ktorej uveríte tomu, čomu veríte. Výsledkom tohto procesu je stav *mysle*, ktorý odráža stav vašich skutočných *šnúrok*.

Čo je *indícia*? Je to udalosť previazaná väzbou príčiny a následku s tým, čo chcete vedieť. Ak sú cieľom vášho prieskumu napríklad vaše šnúry na topánkach, potom svetlo vstupujúce do vašich zreničiek je indícia previazaná s vašimi šnúrkami na topánkach. Toto by sa nemalo mýliť s technickým pojmom “previazanosť” používaným vo fyzike – ja tu hovorím o “previazanosti” iba v tom zmysle, že dve veci skončia v korelovaných stavoch kvôli väzbe príčiny a následku medzi nimi.

Nie každý vplyv vytvára takú “previazanosť”, akú potrebujeme pre indíciu. Nepomôže vám mať prístroj, ktorý zapípa vždy, keď doňho zadáte vyhrávajúce čísla v lotérii, pokiaľ stroj takisto zapípa *aj* keď doňho zadáte *prehrávajúce* čísla v lotérii. Svetlo odrazené od vašich topánok by nebolo užitočnou indíciou o vašich šnúrkach, keby fotóny skončili v rovnakom fyzickom stave bez ohľadu na to, či sú vaše šnúry zaviazané alebo rozviazané.

Povedané abstraktne: Aby bola udalosť *indíciou* o ciele nášho skúmania, musí sa stať *rôzne*, spôsobom prepojeným s *rôznymi* možnými stavmi cieľa. (Povedané technicky: Musí existovať Shannonova vzájomná informácia medzi udalosťou indície a cieľom skúmania, relatívne voči vášmu súčasnému stavu neistoty ohľadom oboch z nich.)

Previazanie dokáže byť nákazlivé, *keď sa správne spracováva*, čo je dôvod, prečo potrebujete oči a mozog. Ak sa fotóny odrazia od vašich šnúrok na topánkach a narazia na kameň, kameň sa príliš nezmení. Kameň nebude odrážať šnúry žiadnym užitočným spôsobom; nebude merateľne odlišný podľa toho, či boli vaše šnúry zaviazané alebo rozviazané. To je dôvod, prečo sa kameň nepovažuje za užitočného svedka na súde. Fotografický film zachová previazanosť prichádzajúcich fotónov so šnúrkami na topánkach, preto samotná fotografia môže byť indíciou. Ak vaše oči a mozog fungujú správne, vy sa previazete so svojimi vlastnými šnúrkami.

To je dôvod, prečo racionalisti kladú taký silný dôraz na paradoxne vyzerajúce tvrdenie, že názor je skutočne hodnotný iba vtedy, keby bolo principiálne možné

¹ pozri: 46 Jednoduchá pravda

² pozri: 2 Rozumné cítenie

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. Niektoré systémy názorov hovoria, ako pomerne priehľadný trik na posilnenie seba samých, že určité názory sú skutočne hodnotné iba vtedy, ak v ne veríte *bezpodmienečne* – bez ohľadu na to, čo vidíte, bez ohľadu na to, čo si myslíte. Váš mozog je povinný bez ohľadu na všetko skončiť v rovnakom stave. Preto máme frázu “slepá viera”. Ak to, čomu veríte, nezávisí na tom, čo vidíte, potom vás oslepili rovnako účinne, ako keby vám vypichli oči.

Ak vaše oči a mozog pracujú správne, vaše názory budú previazané s faktmi. *Rozumné myslenie produkuje názory, ktoré samotné sú indíciami.*

Ak váš jazyk hovorí pravdu, vaše rozumné názory, ktoré samotné sú indíciami, môžu pôsobiť ako indícia pre niekoho iného. Previazanosť sa dá prenášať reťazami príčiny a následku – a ak hovoríte a druhý počúva, aj to je príčina a následok. Keď poviete do mobilu “mám rozviazané šnúrky”, delíte sa o svoju previazanosť s kamarátom.

Preto sú rozumné názory nákazlivé, medzi úprimnými ľuďmi, ktorí si navzájom veria, že sú úprimní. A preto tvrdenie, že vaše názory *nie* sú nákazlivé – že veríte zo súkromných dôvodov, ktoré sa nedajú prenášať – je také podozrivé. Ak sú vaše názory previazané so skutočnosťou, potom by medzi úprimnými ľuďmi *mali* byť nákazlivé.

Ak vás model skutočnosti naznačuje, že výsledky vašich myšlienkových procesov by *nemali* byť pre druhých nákazlivé, potom váš model hovorí, že vaše názory samotné *nie* sú indíciami, čo znamená, že *nie* sú previazané so skutočnosťou. Mali by ste použiť reflektívnu opravu a prestať veriť.

Naozaj, pokiaľ *cítite*, na *intuitívnej* úrovni, čo toto všetko *znamená*, potom *automaticky* prestanete veriť. Pretože “môj názor *nie* je previazaný so skutočnosťou” *znamená* “môj názor je nepresný”. Akonáhle prestanete veriť, že “veta ‘sneh je biely’ je pravdivá”, mali by ste (automaticky!) prestať veriť, že “sneh je biely”, inak je niečo veľmi zle.

Podte teda vysvetliť, prečo ten druh myšlienkových procesov, ktoré systematicky používate, produkuje myšlienky, ktoré odrážajú skutočnosť. Vysvetlite, prečo si myslíte, že ste *rozumný*. Prečo si myslíte, že použitím takých myšlienkových procesov, aké používate, myseľ skončí s názorom “sneh je biely” vtedy a práve vtedy, keď sneh je biely. Ak neveríte, že výsledky vašich myšlienkových procesov sú previazané so skutočnosťou, prečo veríte výsledkom vašich myšlienkových procesov? Je to predsa to isté, alebo by malo byť.

Pôvodný článok s komentármi³

³ lesswrong.com/lw/jl/what_is_evidence/

Kapitola 21. Vedecká indícia, právna indícia, rozumná indícia

Predstavte si, že vám váš dobrý priateľ, policajný komisár, prísne dôverne povie, že hlavou zločinu vo vašom meste je Wulky Wilkinsen. Ako racionalista, mali by ste veriť tomuto tvrdeniu? Poviem to takto: ak si pôjdete niečo začať s Wulkyho nástročnou dcérou, označil by som vás za somára. Keďže je obozretné správať sa akoby¹ Wulky mal podstatne vyššiu než štandardnú pravdepodobnosť byť šéfom zločincov, výrok policajného komisára musel byť silná bayesovská indícia².

Náš právny systém nezavrie Wulkyho do väzenia na základe výroku policajného komisára. Nie je to prijateľné ako *právna indícia*. Keby sme zavreli každého človeka, ktorého policajný komisár obviní, že je šéfom zločincov, možno by sme *spočiatku* chytili veľa zločineckých šéfov, plus pár ľudí, ktorých policajný komisár nemá rád. Moc má sklón korumpovať; postupne by sme chytali menej a menej skutočných zločineckých šéfov (ktorí by si dávali väčší pozor na zaistenie anonymity) a viac a viac nevinných obetí (neobmedzená moc priťahuje korupciu tak ako med priťahuje muchy).

To neznamená, že výrok policajného komisára nie je rozumná indícia. Stále má nahnutý pomer pravdepodobnosti³ a boli by ste somár, keby ste si začali s Wulkyho nástročnou dcérou. Ale na *spoločenskej* úrovni, pri dosahovaní spoločenských cieľov, úmyselne definujeme "právnu indíciu" tak, že zahŕňa iba isté druhy indície, ako napríklad osobné pozorovania policajného komisára v noci na 4. apríla. Každá právna indícia by ideálne mala byť rozumnou indíciou, ale nie naopak. Vyžadujeme mimoriadne silné dodatočné štandardy než rozumnú indíciu vyhlásime za "právnu indíciu".

Ako píšem túto vetu, 18. augusta 2007 o 20:33 pacifického času, mám na sebe biele ponožky. Ako racionalista, mali by ste veriť tomuto tvrdeniu? Áno. Mohol by som o tom svedčiť pred súdom? Áno. Je to *vedecký* výrok? Nie, pretože neexistuje žiaden pokus, ktorý môžete sami urobiť, aby ste si to overili. Veda sa skladá zo *zovšeobecnení*, ktoré sa týkajú mnohých konkrétnych situácií, takže môžete robiť nové pokusy, ktoré testujú toto zovšeobecnenie a tým si sami overiť, že toto zovšeobecnenie je pravdivé, bez spoliehania sa na niekoho autoritu. Veda je *verejne reprodukovateľné* poznanie ľudstva.

Tak ako súdny systém, aj veda ako spoločenský proces sa skladá z omylných ľudí. Chceme mať chránený súbor názorov, ktoré sú *mimoriadne* spoľahlivé. A chceme mať spoločenské pravidlá, ktoré podporujú tvorbu takéhoto poznania. Preto prijímame mimoriadne silné dodatočné štandardy než rozumné poznanie ustanovíme za "vedecké poznanie" a pridáme ho do chráneného súboru názorov. Mal by racionalista v historickú existenciu Alexandra Veľkého? Áno. Máme hrubý obraz o starovekom Grécku, nedôveryhodný, ale lepší než maximálna entropia. Sme však závislí na autoritách ako Plutarch; nemôžeme ignorovať Plutarcha a overiť si všetko sami. Historické poznanie nie je vedecké poznanie.

Mal by racionalista veriť, že Slnko vyjde 18. septembra 2007? Áno - nie s úplnou istotou, ale je to dobrá stávka. (Pedanti: interpretujte to, že rotácia a obežná dráha Zeme zostanú zhruba rovnaké voči Slnku.) Je toto tvrdenie, keď tento článok píšem 18. augusta 2007, *vedecký* názor?

Môže sa zdať zvrátené odopierať prídavné meno "vedecký" tvrdeniam ako "18. septembra 2007 vyjde Slnko". Keby veda nevedela robiť predpovede

¹ pozri: 13 Viera vo vieru

² pozri: 178 Intuitívne vysvetlenie Bayesovej vety

³ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

budúcich udalostí - udalostí, ktoré sa ešte *nestali* - potom by bola zbytočná; nevedela by predpovedať žiaden pokus. Predpoveď, že Slnko vyjde, jednoznačne je *extrapoláciou* vedeckých zovšeobecnení. Zakladá sa na modeloch slnečnej sústavy, ktoré si sami môžete pokusne overiť.

Predstavte si však, že zostavujete pokus na overenie predpovede číslo 27 uznávanej teórie Q, v novom kontexte. Nemusíte mať žiaden konkrétny dôvod spochybňovať pravdivosť tohto názoru; chcete ho iba vyskúšať v novom kontexte. Zdá sa nebezpečné povedať *pred* vykonaním pokusu, že existuje "vedecký názor" na výsledok. Existuje "konvenčná predpoveď" alebo "predpoveď teórie Q". Ale ak už poznáte "vedecký názor" na výsledok, načo sa unúvať robiť pokus?

Dúfam, že začínate vidieť, prečo stotožňujem Vedu so *zovšeobecneniami*, a nie s históriou konkrétnych pokusov. Historická udalosť sa stane raz; zovšeobecnenie sa týka mnohých udalostí. História nemôžeme reprodukovať; vedecké zovšeobecnenia áno.

Je moja definícia "vedeckého poznania" *pravdivá*? To nie je dobre formulovaná otázka. Špeciálne štandardy, ktoré stanovujeme vede, sú pragmatické voľby. Nikde vo hviezdach ani v horách nie je napísané, že $p < 0,05$ musí byť štandardom pre vedecké publikácie. Mnohí dnes tvrdia, že 0,05 je príliš slabé a že by bolo *užitočné* znížiť to na 0,01 alebo 0,001.

Možno budúce generácie, na základe teórie, že veda je *verejné reprodukovateľné poznanie* ľudstva, budú ako "vedecké" označovať iba články uverejnené v časopisoch s otvoreným prístupom. Ak si pýtate peniaze za prístup k poznaniu, je to časťou poznania *ľudstva*? Môžeme dôverovať nejakému záveru, ak ľudia musia platiť za to, aby ho mohli kritizovať? Je to *naozaj* veda?

Otázka: "Je to *naozaj* veda?" je zle formulovaná. Je časopis s uzavretým prístupom za 20 000 dolárov ročne *naozaj* bayesovskou indíciou? Tak ako pri súkromnom ubezpečení policajného komisára, že Wulky je hlavou zločinu, myslím, že musíme odpovedať: "Áno." Ale mal by časopis s uzavretým prístupom byť kanonizovaný ako "veda"? Mali by sme ho prijať do špeciálneho chráneného súboru názorov? Osobne si myslím, že vede by lepšie poslúžilo rozhodnutie, že iba otvorené poznanie sa považuje za *verejný reprodukovateľný súbor vedomostí ľudstva*.

Pôvodný článok s komentármi⁴

⁴ http://lesswrong.com/lw/in/scientific_evidence_legal_evidence_rational/

Kapitola 22. Koľko indície treba?

Nedávno¹ som definoval *indíciu* ako “udalosť previazanú väzbou príčiny a následku s tým, čo chcete vedieť” a *previazanosť* ako “musí sa stať rôzne pre rôzne možné stavy cieľa”. Takže, koľko previazanosti – koľko indície – treba, aby podporila nejaký názor?

Začnime otázkou dosť jednoduchou na to, aby bola matematická: ako silne by ste sa potrebovali previazať s lotériou², aby ste ju vyhrali? Predpokladajme, že sa žrebuje zo sedemdesiatich loptičiek bez opakovania, a na výhru potrebujete uhádnuť šesť čísel. Potom máme 131 115 985 možných výherných kombinácií, čiže náhodne vybraný žreb má šancu vyhrať 1 / 131 115 985 (0,000 000 7 %). Aby ste vyhrali lotériu, potrebovali by ste indíciu dosť *vyberavú* na to, aby viditeľne uprednostnila jednu kombináciu pred jej 131 115 984 alternatívami.

Predpokladajme, že existujú nejaké testy, pomocou ktorých môžete pravdepodobnostne rozlíšiť medzi vyhrávajúcimi a prehrávajúcimi číslami v lotérii. Napríklad zadáte kombináciu do malej čiernej krabičky, ktorá vždy zapípa, ak je to vyhrávajúca kombinácia, ale má iba šancu 1/4 (25 %) zapípať, ak je kombinácia nesprávna. Bayesovskými³ slovami by sme povedali, že *pomer podmienených pravdepodobností* je 4 k 1. To znamená, že krabička má 4-krát väčšiu šancu zapípať, ak zadáme správnu kombináciu, v porovnaní so šancou zapípať, ak zadáme nesprávnu kombináciu.

To je stále celá kopa možných kombinácií. Ak napíšete 20 neprávnych kombinácií, krabička zapípa pri 5 z nich čistou náhodou (v priemere). Ak zadáte všetkých 131 115 985 možných kombinácií, krabičke síce naisto zapípa pri jednej vyhrávajúcej kombinácii, ale zapípa aj pri 32 778 996 prehrávajúcich kombináciách (v priemere).

Takže táto krabička vám neumožní vyhrať lotériu, ale je lepšia než nič. Ak použijete túto krabičku, vaše šance vyhrať stúpnu z 1 zo 131 115 985 na 1 z 32 778 997. Trochu ste pokročili smerom k nájdeniu vášho cieľa, pravdy, v rámci širokého priestoru možností.

Predpokladajme, že použijete inú čiernu krabičku, aby ste otestovali kombinácie *dvakrát, nezávisle*. Obe krabičky naisto zapípajú pre vyhrávajúci žreb. Ale šanca, že krabička zapípa pre prehrávajúcu kombináciu, je 1/4 *nezávisle* pre každú krabičku; šanca, že *obe* krabičky zapípajú pre prehrávajúcu kombináciu je teda 1/16. Môžeme povedať, že *kumulatívna* indícia dvoch nezávislých testov má pomer podmienených pravdepodobností 16:1. Počet prehrávajúcich žrebov lotérie, ktoré prejdú cez oba testy bude (v priemere) 8 194 749.

Keďže je 131 115 985 možných žrebov v lotérii, asi ste uhádli, že potrebujete indíciu, ktorej sila je okolo 131 115 985 k 1 – udalosť, alebo sériu udalostí, ktorá má 131 115 985-krát väčšiu šancu nastať pri vyhrávajúcej kombinácii než pri prehrávajúcej. Vlastne aj toto množstvo indície by stačilo iba na to, aby vám dalo šancu *pol na pol* vyhrať lotériu. Prečo? Lebo ak použijete filter takejto sily na 131 miliónov prehrávajúcich lístkov, jeden prehrávajúci lístok, v priemere, prejde týmto filtrom. Vyhrávajúci lístok tiež prejde filtrom. Zostanú vám teda dva lístky, ktoré prešli filtrom, iba jeden z nich je vyhrávajúci. Ak si môžete kúpiť iba jeden lístok, máte šancu vyhrať 50 %.

¹ pozri: 20 Čo je indícia?

² pozri: 309 Lotérie: Plytvanie nádejou

³ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

Lepší spôsob, ako si problém znázorniť: Na začiatku je 1 vyhrávajúci lístok a 131 115 984 prehrávajúcich lístkov, takže vaša šanca vyhrať je 1 : 131 115 984. Ak použijete jednu krabičku, šanca zapípania je 1 pre vyhrávajúci lístok a 0,25 pre prehrávajúci lístok. Takže 1 : 131 115 984 vynásobíme 1 : 0,25 a dostaneme 1 : 32 778 996. Pridanie ďalšej krabičky indície opäť vynásobí šance 1 : 0,25, takže šance sú teraz 1 vyhrávajúci lístok na 8 194 749 prehrávajúcich lístkov.

Je pohodlné merať indíciu v bitoch - nie ako bity na pevnom disku, ale matematické bity, ktoré sú koncepčne iné. Matematické bity sú logaritmy pravdepodobností pri základe 1/2. Napríklad ak sú štyri možné výsledky A, B, C a D, ktorých pravdepodobnosti sú 50%, 25%, 12,5% a 12,5% a ja vám poviem, že výsledok bol "D", dal som vám tri bity informácie, pretože som vás informoval o výsledku, ktorého pravdepodobnosť bola 1/8.

Zhodou okolností je 131 115 984 o čosi viac než 2 na 27. Takže 14 krabičiek, alebo 28 bitov indície - udalosť, ktorá má 268 435 456 : 1-krát väčšiu šancu stať sa, ak je hypotéza lístka pravdivá než ak je nepravdivá - by posunula šance z 1 : 131 115 984 na 268 435 456 : 131 115 984, čo sa vykráti na 2 : 1. Šanca 2 ku 1 znamená dve šance na výhru pre každú šancu na prehru, čiže *pravdepodobnosť* výhry s 28 bitmi indície je 2/3. Pridanie ďalšej krabičky, ďalších 2 bitov indície, by posunulo šance na 8 : 1. Pridanie ešte ďalších dvoch krabičiek by posunulo šance na výhru na 128 : 1.

Ak teda chcete licenciu na *silné presvedčenie*, že vyhráte lotériu - ktoré tu svojvoľne definujeme ako pravdepodobnosť omylu menšiu než 1% - 34 bitov indície ohľadom vyhrávajúcej kombinácie by malo stačiť.

Vo všeobecnosti sa pravidlo na zvažovanie "kolko indície treba" riadi podobným vzorom: Čím väčší je *priestor možností*, v ktorom leží daná hypotéza, alebo čím menej pravdepodobná táto hypotéza vyzerá *a priori* v porovnaní s jej susedmi, alebo čím viac si chcete byť istí, tým viac indície potrebujete.

Nemôžete vzdorovať týmto pravidlám; nemôžete si vytvoriť presné názory na základe nedostatočnej indície. Povedzme, že máte 10 krabičiek v rade a začnete zadávať kombinácie do krabičiek. Nemôžete sa zastaviť pri prvej kombinácii, ktorá dostane zapípanie od všetkých 10 krabičiek a povedať: "Ale veď šanca, že by sa toto stalo pre prehrávajúcu kombináciu je milión ku jednej! Môžem ignorovať tieto teoretické bayesiánske pravidlá a skončiť tu." V priemere, na jeden vyhrávajúci lístok prejde takýmto testom aj 131 prehrávajúcich. Ak vezmeme do úvahy priestor pravdepodobností a prvotnú nepravdepodobnosť, skočili ste k príliš silnému záveru na základe nedostatočnej indície. To nie je samoúčelná byrokratická regulácia, to je matematika.

Samozrejme, aj tak môžete *veriť* na základe nedostatočnej indície, ak je to váš rozmar; ale nebudete môcť veriť *presne*. Je to ako keby ste sa pokúšali šoférovať auto bez paliva, pretože neveríte na nejaké hala-bala predstavy, že na jazdenie treba palivo. Nebolo by omnoho *zábavnejšie* a lacnejšie, keby sme sa rozhodli zrušiť zákon, že autá potrebujú palivo? Nebolo by to očividne lepšie pre každého? Nuž, môžete skúsiť, ak je to váš rozmar. Môžete dokonca aj zavrieť oči a predstierať, že auto sa hýbe. Ale aby ste *naozaj* docestovali k presným názorom, na to potrebujete palivo indícií, a čím ďalej chcete ísť, tým viac paliva potrebujete.

Pôvodný článok s komentármi⁴

⁴ lesswrong.com/lw/jn/how_much_evidence_does_it_take/

Kapitola 23. Occamova britva

Súvisí s kapitolou: Príťažujúce podrobnosti¹

Čím zložitejšie je vysvetlenie, tým viac indície potrebujete už len aby ste ho našli v priestore názorov. (V Tradičnej Rozumnosti sa to často formuluje zavádzajúco² ako “Čím zložitejšie je tvrdenie, tým viac indície potrebujete, aby ste zaň argumentovali.”) Ako môžeme merať zložitosť vysvetlenia? Ako môžeme určiť, koľko indície potrebujeme?

Occamova britva sa často formuluje ako: “Najjednoduchšie vysvetlenie, ktoré je v súlade s faktmi.” Robert Heinlein odpovedal, že najjednoduchšie vysvetlenie je: “Pani z dolného konca ulice je bosorka; ona to urobila.”

Vidno, že dĺžka anglickej vety nie je dobrým spôsobom, ako merať “zložitosť”. A “súlady” s faktmi, daný iba tým, že ich teória *nezakazuje*, tiež nestačí.

Prečo presne je dĺžka anglickej vety zlým meradlom zložitosti? Pretože keď vyslovíte nejakú vetu, používate *označenia* pre pojmy, ktorým poslucháč rozumie – ten poslucháč má už zložitosť uloženú v sebe. Prestavme si, že celú tú Heinleinovu vetu skrátime na “Pzdkujbotu!”, takže sa celé vysvetlenie dá vyjadriť jediným slovom; alebo ešte lepšie, dajme mu ľubovoľnú krátku nálepku, napríklad: “Fnord!” Zredukovala sa tým zložitosť? Nie, pretože musíte poslucháčovi dopredu vysvetliť, že “Pzdkujbotu!” je skratka z: “Pani z dolného konca ulice je bosorka; ona to urobila.” A samotné slovo “bosorka” je nálepka pre isté mimoriadne tvrdenia – to, že ho už všetci poznáme, ešte neznamená, že samotný pojem je jednoduchý.

Obrovský elektrický blesk príde z oblohy, niečo trať a severskí domorodci povedia: “Možno sa nejaký naozaj mocný činiteľ nahneval a hodil blesk.” Ľudský mozog je najzložitejším artefaktom v známom vesmíre. Ak nám *hnev* pripadá jednoduchý, je to preto, lebo nevidíme všetky tie nervové obvody, ktoré túto emóciu implementujú. (Pokúste sa vysvetliť, prečo je *Saturday Night Live* zábavné, mimozemšťanovi bez zmyslu pre humor. Necíťte sa však nadradení; vy zase nemáte zmysel pre *fnord*.) Ľudia, ktorí vymysleli hypotézu Thora, konateľa hromu, si neuvedomovali zložitosť hnevu, a vôbec zložitosť inteligencie.

Vysvetliť *človeku* Maxwellove rovnice trvá dlhšie než vysvetliť mu Thora. Ľudia nemajú zabudovaný slovník pre integrály tak, ako máme zabudovaný slovník pre *hnev*. Musíte vysvetliť svoj jazyk, a jazyk za týmto jazykom, a samotný pojem matematiky, než začnete hovoriť o elektrine.

A predsa sa zdá, že by mal existovať nejaký zmysel, v ktorom sú Maxwellove rovnice *jednoduchšie* než ľudský mozog alebo Thor, konateľ hromu.

Existuje: Je *omnoho* ľahšie (ako sa ukazuje) napísať počítačový program, ktorý simuluje Maxwellove rovnice, v porovnaní s počítačovým programom, ktorý simuluje inteligentnú emocionálnu myseľ ako je Thor.

Formalizmus Solomonoffovej indukcie meria “zložitosť opisu” dĺžkou najkratšieho počítačového programu, ktorý vypíše tento opis ako svoj výstup. Ak hovoríte o “najkratšom počítačovom programe”, ktorý niečo urobí, musíte upresniť priestor počítačových programov, čo si vyžaduje jazyk a interpret. Solomonoffova indukcia používa Turingove stroje, alebo skôr bitové reťazce, ktoré definujú Turingove stroje. Čo ak sa vám Turingove stroje nepáčia? V tom prípade máte iba konštantnú penaltu zložitosti za vytvorenie svojho

¹ pozri: 6 Príťažujúce podrobnosti

² pozri: 24 Einsteinova drzosť

univerzálneho Turingovho stroja, ktorý interpretuje taký kód, aký mu zadáte, v tom programovacom jazyku, ktorý sa vám páči. Rôzne induktívne formalizmy sú relatívne voči sebe penalizované v najhoršom prípade konštantnou veličinou zodpovedajúcou veľkosti univerzálneho interpretera daného formalizmu.

V lepších (podľa mňa) verziách Solomonoffovej indukcie počítačový program nevytvára deterministickú predpoveď, ale priraduje pravdepodobnosti reťazcom. Napríklad by sme mohli napísať program na vysvetlenie vyváženej mince tak, že by sme napísali program, ktorý priradí rovnakú pravdepodobnosť všetkým 2^N reťazcom dĺžky N . Toto je prístup Solomonoffovej indukcie k *súlade* s pozorovanými údajmi. Čím vyššiu pravdepodobnosť program priradí pozorovaným údajom, tým viac je program v *súlade* s údajmi. A súčet pravdepodobností musí byť 1, takže program, ktorý je v lepšom "súlade" s jednou možnosťou, musí uberať masu pravdepodobnosti z nejakej inej možnosti, s ktorou je potom v horšom "súlade". Neexistuje žiadna supervyvážená minca, ktorá priradí pravdepodobnosť 100 % hlave a pravdepodobnosť 100 % znaku.

Ako vyvážime súlad s dátami voči zložitosti programu? Keby sme ignorovali penaltu zložitosti a mysleli *iba* na súlad, potom by sme vždy dávali prednosť programom, ktoré tvrdia, že deterministicky predpovedajú dané údaje a priradujú im pravdepodobnosť 100 %. Ak na minci padne "HTTHHT", potom program, ktorý tvrdí, že na minci vždy padá "HTTHHT", je s pozorovanými údajmi v 64-krát väčšom súlade než program, ktorý tvrdí, že minca je vyvážená. Naopak, keby sme ignorovali súlad a brali do úvahy *iba* zložitost', potom by hypotéza "vyváženej mince" vyzerala vždy jednoduchšia než ľubovoľná iná hypotéza. Aj keby minca dopadla "HTHHTHHHTHHHHHTHHHHHT..." Veru, hypotéza vyváženej mince je jednoduchšia a je v súlade s týmito dátami rovnako dobre ako s hocíjakým iným reťazcom 20 hodov mince - ani viac, ani menej - ale vidíme, že iná hypotéza, nie omnoho zložitejšia, je s danými údajmi v omnoho lepšom súlade.

Ak dovoľíte programu uložiť o jeden bit informácie viac, môže tým rozdeliť priestor možností napoly a tým pridať dvakrát väčšiu pravdepodobnosť všetkým bodom v zostávajúcom priestore. To naznačuje, že jeden bit zložitosti programu by mal stáť *aspoň* "dvakrát lepší" súlad. Ak skúsíte vytvoriť počítačový program, ktorý explicitne obsahuje výsledok "HTTHHT", tých šesť bitov, ktoré ste stratili na zložitosti, musí zrušiť všetku dôveryhodnosť získanú 64-násobne lepším súladom. V opačnom prípade by ste skôr či neskôr došli k záveru, že všetky vyvážené mince sú fixné.

Pokiaľ váš program nie je chytrý a *nekomprimuje* údaje, nemalo by vám pomôcť presunúť jeden bit z údajov do popisu programu.

Solomonoffova indukcia predpovedá postupnosti tak, že urobíte súčet cez všetky povolené počítačové programy - ak je povolený ľubovoľný program, Solomonoffova indukcia je nevypočítateľná - kde každý program dostane prvotnú pravdepodobnosť $1/2$ umocnenú na dĺžku jeho kódu v bitoch a každý program je ďalej vážený podľa svojho súladu s doteraz pozorovanými údajmi. To vám dáva váženú zmes odborníkov na predpovedanie budúcich bitov.

Formalizmus minimálnej dĺžky správy je takmer ekvivalentný Solomonoffovej indukcii. Pošlete reťazec popisujúci kód a potom pošlete reťazec popisujúci údaje v tomto kóde. Ktorékoľvek vysvetlenie vedie k najkratšej *celkovej dĺžke*, je najlepšie. Ak si predstavíte množinu povolených kódov ako priestor počítačových programov a jazyk na popis kódov ako univerzálny stroj, potom je minimálna dĺžka správy takmer ekvivalentná Solomonoffovej indukcii. (Takmer, pretože si vyberá *najkratší* program namiesto sumy cez všetky programy.)

To nám umožňuje jasne vidieť problém použitia: "Pani z dolného konca ulice je bosorka; ona to urobila" na vysvetlenie vzoru v postupnosti "01010101".

Ak posielate správu kamarátovi a pokúšate sa vysvetliť pozorovanú postupnosť, museli by ste povedať: "Pani z dolného konca ulice je bosorka; ona to urobila, že postupnosť vyšla 0101010101." Vaše obvinenie z bosoráctva by vám neumožnilo *skrátit* zvyšok správy; stále by ste museli opísať, do najmenšieho detailu, údaje, ktoré jej bosoráctvo spôsobilo.

Bosoráctvo môže byť v súlade s našimi pozorovaniami v tom zmysle, že ich kvalitatívne *povoľuje*; ale to len preto, lebo bosoráctvo *povoľuje všetko*, rovnako ako povedanie "Flogiston!"³ Takže aj keď povieť "bosorka", stále musíte popísať všetky pozorované údaje do najmenšieho detailu. *Nezmenšili ste celkovú dĺžku správy opisujúcej vaše pozorovania* tým, že ste preniesli správu o bosoráctve; jednoducho ste len pridali zbytočný úvod, čím ste zvýšili celkovú dĺžku.

Tá skutočná zákernosť bola skrytá v slove "to" v časti "urobila to bosorka". Čo presne urobila bosorka?

Samozrejme, vďaka skresleniu spätného pohľadu⁴ a ukotvovaniu⁵ a falošným vysvetleniam⁶ a falošnej kauzalite⁷ a pozitívnemu skresleniu⁸ a motivovanému poznávaniu⁹ sa môže sa zdať úplne jasné, že keď je žena bosorka, *samozrejme* spôsobí, že na minci padne 0101010101. Ale o tom už som hovoril.

Pôvodný článok s komentármi¹⁰

³ pozri: 34 Falošná kauzalita

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

⁵ pozri: 78 Ukotvenie a prispôsobenie

⁶ pozri: 31 Falošné vysvetlenia

⁷ pozri: 34 Falošná kauzalita

⁸ pozri: 39 Pozitívne skreslenie: Pozeraj do tmy

⁹ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

¹⁰ lesswrong.com/lw/jp/occams_razor/

Kapitola 24. Einsteinova drzosť

V roku 1919 Sir Arthur Eddington viedol expedície do Brazílie a na ostrov Principe s cieľom pozorovať zatmenie slnka a otestovať tým experimentálnu predpoveď Einsteinovej novej teórie Všeobecnej Relativity. Novinár sa opýtal Einsteina, čo by urobil, keby Eddingtonove pozorovania neboli v súlade s jeho teóriou. Einstein sa preslávil odpoveďou: “Bolo by mi ho ľúto. Teória je správna.”

Vyzerá to ako veľmi zadubený výrok, vzdorujúci kliše Tradičnej Rozumnosti, že experiment je suverénom nad všetkým. Zdá sa, že Einstein mal takú veľkú drzosť, že by odmietol skloniť hlavu a podrobiť sa odpovedi Prírody, ako vedci musia. Kto môže *vedieť*, že jeho teória je správna, ešte pred experimentálnym testom?

Samozrejme sa ukázalo, že Einstein mal pravdu. Ja sa snažím nekritizovať ľudí, keď majú pravdu. Ak si naozaj kritiku zaslúžia, nebudem musieť dlho čakať, kým sa pomýlia.

A možno Einstein nebol až taký zadubený, ako znel...

Aby ste priradili pravdepodobnosť vyššiu než 50 % správne kandidátovi zo súboru 100 000 000 možných hypotéz, potrebujete aspoň 27 bitov indície¹ (plus-mínus). Nemôžete očakávať, že nájdete správneho kandidáta bez takýchto silných testov, pretože pri slabších testoch viac než jeden kandidát prejde všetkými testmi. Ak sa pokúsite použiť test, ktorý má šancu na falošný pozitívny výsledok iba milión k jednej (cca 20 bitov), zostane vám sto kandidátov. Už samotné *nájdanie* správnej odpovede v širokom priestore možností si vyžaduje veľké množstvo indície.

Tradičná Rozumnosť kladie dôraz na dokazovanie: “Ak ma chceš presvedčiť o X, musíš mi dať aspoň Y bitov indície.” Často sklízam do takéhoto vyjadrovania, keď poviem veci ako: “Aby sme *zdôvodnili* vieru v tento výrok s pravdepodobnosťou väčšou než 99 %, potrebujeme 34 bitov indície.” Alebo: “aby ste vašej hypotéze priradili pravdepodobnosť vyššiu ako 50 %, potrebujete 27 bitov indície.” Tradičné vyjadrovanie naznačuje, že začínate s nejakým tušením alebo nejakým súkromným spôsobom uvažovania, ktoré vás privedie k predkladanej hypotéze, a potom musíte zhromaždiť “indície”, aby ste to *potvrdili* - aby ste presvedčili vedeckú komunitu alebo zdôvodnili tvrdenie, že *veríte* vo svoje tušenie.

Ale z bayesiánskeho pohľadu potrebujete množstvo indície² približne rovné zložitosti hypotézy³ už len na to, aby ste túto hypotézu objavili v priestore teórií. Nie je to otázka zdôvodňovania niečoho pred niekým. Ak existuje sto miliónov alternatív, potrebujete aspoň 27 bitov indície, aby ste vôbec zamerali svoju pozornosť výhradne na správnu odpoveď.

To je pravda aj keď váš odhad nazývate “tušenie” alebo “intuícia”. Tušenia a intuície sú skutočné procesy v skutočnom mozgu. Ak váš mozog nedostane na strávenie aspoň 10 bitov skutočne previazanej platnej bayesiánskej indície, potom vám váš mozog nedokáže vyhrať správnu 10-bitovú hypotézu do pozornosti - vedome, nevedome, hocijako. Nevedomé procesy nemôžu nájsť jeden cieľ z miliónov použitím púhych 19 bitov previazanosti o nič viac než vedomé procesy. Tušenia môžu byť záhadou pre toho, kto tuší, ale nemôžu porušovať fyzikálne zákony.

Vidíte, kam toto smeruje: *Vo chvíli, keď Einstein prvýkrát sformuloval svoju hypotézu* - keď mu prvýkrát v hlave naskočili rovnice - musel *už mať* dostatočné

¹ pozri: 22 Koľko indície treba?

² pozri: 22 Koľko indície treba?

³ pozri: 23 Occamova britva

indície z pozorovania, aby vybrali zložité rovnice Všeobecnej Relativity do jeho pozornosti. Inak by ich nemohol mať *správne*.

A teraz, aká je šanca, že by Einstein mal z pozorovania *presne* toľko indície, aby priviedla Všeobecnú Relativitu do jeho pozornosti, ale zdôvodňovala by iba pravdepodobnosť 55 %? Povedzme, že Všeobecná Relativita je 29,3-bitová hypotéza. Aká je šanca, že Einstein nadabil na *presne* 29,5 bitu indície počas svojho štúdia fyziky?

Nie veľká! Ak mal Einstein z pozorovania dosť indície, aby vôbec vybral správne rovnice Všeobecnej Relativity, potom mal pravdepodobne dosť indície, aby si bol *sakra istý*, že Všeobecná Relativita je pravdivá.

V skutočnosti, keďže ľudský mozog nevie dokonale efektívne spracovávať informácie, Einstein mal pravdepodobne *omnoho viac indície*, než by v princípe bolo treba, aby dokonalý bayesiánc priradil Všeobecnej Relativite obrovskú dôveryhodnosť.

“Bolo by mi ho ľúto; teória je správna” neznie zďaleka tak pohoršujúco, keď sa na to pozriete z tohto pohľadu. A pamätajte, že Všeobecná Relativita *bola* správna, z celého toho rozsiahleho priestoru možností.

Pôvodný článok s komentármi⁴

⁴ lesswrong.com/lw/jo/einsteins_arrogance/

Kapitola 25. Zameraj svoju neistotu

Budú výnosy dlhopisov rásť, klesať alebo zostanú rovnaké? Ak ste televízny expert a vašou prácou je vysvetliť tento výsledok dodatočne, nie je dôvod obávať sa. Bez ohľadu na to, ktorá z týchto troch možností sa ukáže ako pravdivá, budete vedieť vysvetliť, prečo tento výsledok dokonale sedí s vašou obľúbenou teóriou trhu. Nie je dôvod myslieť na tieto tri možnosti ako navzájom si nejakou odporujúce, ako navzájom sa vylučujúce, pretože dostanete plný počet bodov za expertstvo bez ohľadu na to, ktorý výsledok nastane.

Ale moment! Čo ak ste začínajúci televízny expert a nemáte dosť skúseností na to, aby ste si dôveryhodné vysvetlenia vymysleli na mieste. Potrebujete si vopred prichystať poznámky na zajtrajšie vysielanie a na ich prípravu máte obmedzený čas. V tomto prípade by bolo užitočné vedieť, ktorý výsledok naozaj nastane – či budú výnosy dlhopisov rásť, klesať alebo zostanú rovnaké – pretože potom by ste si potrebovali pripraviť iba jednu množinu výhovoriek.

Žiaľ, nikto nedokáže predpovedať budúcnosť. Čo urobíte? Iste nepoužijete “pravdepodobnosti”. Všetci vieme zo školy¹, že “pravdepodobnosti” sú číselká, ktoré sa objavujú vedľa slovných zadaní a tu nie sú žiadne číselká. Ešte horšie, cítite neistotu. Nespomínate si na pocit neistoty, keď ste manipulovali číselká pri slovných zadaniach. *Univerzitné lekcie matematiky* sú pekné čisté miesta a preto sa *matematika samotná* nemôže týkať životných situácií, ktoré nie sú pekné a čisté. Nechceli by ste nevhodne prenášať mysliace zručnosti z jedného kontextu do druhého². Je jasné, že toto nie je otázka “pravdepodobností”.

Každopádne, máte iba 100 minút na prípravu výhovoriek. Nemôžete stráviť celých 100 minút na “rast” a zároveň celých 100 minút na “pokles” a zároveň celých 100 minút na “rovnaké”. Musíte si nejakou určiť priority.

Keby ste potrebovali zdôvodniť svoje časové výdavky pred revíznou komisiou, museli by ste každej možnosti venovať rovnaký čas. Keďže tam nie sú napísané žiadne číselká, nemali by ste dokumentáciu, ktorou by ste zdôvodnili vynakladanie rôznych množstiev času. Akoby ste počuli revízorov: *A prečo ste, pán Finkledinger, strávili presne 42 minút výhovorkou číslo 3? Prečo nie 41 minút, alebo 43? Priznajte sa – nie ste objektívny! Subjektívne zvýhodňujete, čo sa vám zapáči!*

Avšak, uvedomíte si s malým zábleskom úľavy, že vás žiadna revízna komisia nebude hrešiť. To je dobre, pretože zajtra bude veľký oznam Federálneho rezervného systému a zdá sa nepravdepodobné, že by ceny dlhopisov zostali rovnaké. Nechcete minúť 33 vzácnych minút na výhovorku, o ktorej neočakávate, že ju budete potrebovať.

Vašu myseľ to unáša k výhovorkám, ktoré používate v televízii, prečo každá udalosť dôveryhodne zapadá do vašej teórie trhu. Ale čoskoro je jasné, že dôveryhodnosť vám tu nepomôže – všetky tri udalosti sú dôveryhodne vysvetliteľné. Zosúladenie s vašou obľúbenou teóriou trhu vám nepovie, ako si máte rozdeliť čas. Je neprekročiteľná priepasť medzi 100 minútami vášho času, ktoré sú ohraničené a vašou schopnosťou vysvetľovať, ako nejaký výsledok zapadá do vašej teórie, ktorá je neohraničená.

A predsa... ešte aj v tomto neistom stave mysle, zdá sa, že tieto tri udalosti očakávate rozdielne; očakávate, že niektoré výhovorky budete potrebovať viac

¹ lesswrong.com/lw/i2/two_more_things_to_unlearn_from_school/

² <http://www.aft.org/newspubs/periodicals/ae/issues.cfm>

než iné. A - toto je tá fascinujúca časť - keď pomyslíte na niečo, vďaka čomu sa zdá, že ceny dlhopisov *skôr* pôjdu hore, vtedy cítite, že budete *menej* potrebovať výhovorku pre pokles či zachovanie cien dlhopisov.

Zdá sa priam, akoby tu bol vzťah medzi tým, ako veľmi očakávate každý z týchto troch výsledkov a kolko času chcete stráviť pripravovaním každej výhovorky. Samozrejme, že sa tento vzťah nedá naozaj vyčíslieť. Na prípravu svojho prejavu máte 100 minút, ale v tomto očakávaní nie je žiadne 100 na rozdelenie. (Dôjdete však na to, že *ak* nastane niektorý konkrétny prípad, vaša funkcia úžitku bude logaritmom času stráveného prípravou výhovorky.)

Napriek tomu... vaša myseľ sa stále vracia k myšlienke, že očakávanie je ohraničené, nie tak ako schopnosť vyhovárať sa, ale tak ako čas na prípravu výhovoriek. Možno by sme očakávanie mali brať ako *obmedzený zdroj*, tak ako peniaze. Vaším prvým impulzom je pokúsiť sa získať viac očakávania, ale čoskoro si uvedomíte, že aj keby ste mali viac očakávania, nebudete mať o nič viac času na prípravu výhovoriek. Nie, vašou jedinou možnosťou je *rozdeliť* váš *obmedzený zdroj* očakávania najlepšie ako viete.

Ste si dosť istí, že vás na hodinách štatistiky neučili nič podobné. Nevysvetlili vám, čo máte robiť, keď sa *cítite* tak strašne neisto. Nevysvetlili vám, čo máte robiť, keď vám nedajú žiadne číselká. Dokonca aj keby ste skúsili používať čísla, mohli by ste skúšať hocijaké čísla - nie je žiaden náznak, aký druh matematiky máte používať, ak vôbec nejaký! Možno by ste používali *dvojice* čísel, pravé a ľavé číslo... alebo ktovie čo ešte? (Na prípravu výhovoriek však máte iba 100 minút.)

Keby tak existovalo umenie, ako *zameriť svoju neistotu* - ako *natlačiť* čo najviac očakávania do toho výsledku, ktorý *naozaj nastane*!

Lenže ako by sa volalo takéto umenie? A aké pravidlá by malo?

Pôvodný článok s komentármi³

³ http://lesswrong.com/lw/ia/focus_your_uncertainty/

Kapitola 26. Cnosť presnosti

Čo je pravda o jednom jablku, nemusí byť pravda o druhom jablku; preto možno o jednom jablku povedať viac než o všetkých jablkách na svete.

—Dvanásť cností rozumnosti

V rámci svojej profesie ľudia rozumejú, aká dôležitá je presnosť; automechanik pozná rozdiel medzi karburátorom a radiátorom a nemyslí na ne obe ako na "súčiastky do auta". Lovec-zberač pozná rozdiel medzi levom a panterom. Upratovač neumýva podlahu čistidlom na okná, aj keď tie fľaše pripadajú podobné tomu, kto toto umenie neovláda.

Mimo svojej profesie ľudia často robia tú chybu, že sa snažia čo najviac rozšíriť zmysel slova, aby pokrylo čo najväčšie územie. Nie je snáď slávnostnejšie, múdrejšie, pôsobivejšie rozprávať o *všetkých* jablkách na svete? O koľko vznešenejšie musí byť *vysvetliť ľudské myslenie vo všeobecnosti* než sa rozptylovať menšími otázkami, napríklad ako ľudia vymýšľajú techniky na riešenie Rubikovej kocky. Veru, zriedkavo sa zdá potrebné všimnúť si *konkrétne* otázky; nie je azda všeobecná teória sama osebe dostatočne hodnotným úspechom?

Zvedaví ľudia majú vo zvyku zdvihnúť na pobreží jeden kamienok z milióna a vidieť na ňom niečo nové, niečo zaujímavé, niečo odlišné. Nazvete tieto kamienky "diamanty" a pýtate sa, čo by na nich mohlo byť špeciálne - aké vnútorné kvality môžu mať spoločné okrem lesku, ktorý ste si všimli ako prvý. A potom príde niekto iný a povie: "Prečo nenazvať diamantom aj *tento* kamienok? A čo tento, a čo tento?" Sú nadšení a myslia to dobre. Pretože nazývať niektoré kamienky "diamanty" a iné nie vyzerá nedemokraticky, exkluzívne, elitársky a neholisticky. Vyzerá to... *obmedzene*... ak sa neurazíte. Nie *otvorene*, nie *všeobjímajúco*, nie *spoločensky*.

Možno si myslíte, že je poetické dať jednému slovu mnoho významov a šíriť tak odtiene konotácií navôkol. Ale aj básnik, ak je dobrým básnikom, sa musí naučiť vidieť svet presne. Nestačí prirovnáť lásku ku kvetu. Horúca žiarlivá nekonzumovaná láska nie je to isté ako láska desaťročia zosobášeného páru. Ak potrebujete kvet, ktorý symbolizuje žiarlivú lásku, musíte ísť do záhrady hľadať a všimnúť si drobné rozdiely - nájsť kvetinu s opojnou vôňou, jasnými farbami a trnmi. Aj keď máte v úmysle dávať významom odtiene a šíriť konotácie, musíte pozorne sledovať, ktoré presne významy odtieňujete a konotujete.

Je nevyhnutnou súčasťou racionalistovho umenia - a dokonca aj básnikovho umenia! - úzko sa sústrediť na nezvyčajné kamienky, ktoré majú nejakú zvláštnu vlastnosť. A pozerať na detaily, ktoré majú tieto kamienky - ale iba tieto kamienky! - navzájom spoločné. Toto nie je hriech.

Je absolútne v poriadku, ak moderní evoluční biológovia vysvetľujú *iba* vzory živých tvorov a nie "evolúciu" hviezd alebo "evolúciu" technológie. Žiaľ, niektorí nešťastníci používajú slovo "evolúcia" zároveň na označenie prirodzene vybraných vzorov replikujúceho sa života *aj* na čisto náhodnú štruktúru hviezd *aj* na inteligentne navrhnutú štruktúru technológie. A ako všetci vieme, keď ľudia používajú to isté slovo, musí to byť tá istá vec. Mali by ste automaticky zovšeobecniť všetko, čo si myslíte, že viete o biologickej evolúcii, aj na technológiu. Každý, kto hovorí opak, musí byť nezmyselne pedantný. Nie je predsa možné, aby vaše priepastné ignorantsvo ohľadom modernej evolučnej teórie bolo také veľké, že nepoznáte rozdiel medzi karburátorom a radiátorom. To je nemysliteľné. Nie, to ten *druhý* - viete, ten čo to naozaj matematicky študoval - je príliš hlúpy na to, aby videl súvislosti.

A čo už môže byť cnostnejšie než vidieť súvislosti? Nepochybne, najmúdrejšími spomedzi všetkých ľudí sú guruovia New Age, ktorí hovoria: "Všetko súvisí so všetkým." Ak to niekedy vyslovíte nahlas, mali by ste urobiť pauzu, aby sa každý stihol zotaviť z čistého šoku z tejto Hlbokej Múdrosti.

Existuje triviálne mapovanie medzi grafom a jeho komplementom. Úplný graf, s hranou medzi každými dvoma vrcholmi, obsahuje rovnaké množstvo informácie ako graf, ktorý nemá žiadne hrany. Dôležité grafy sú tie, v ktorých niektoré veci *nie* sú spojené s niektorými inými vecami.

Keď sa nevedomí snažia vyzeráť hlboko, vykresľujú nekonečné slovné porovnania medzi touto témou a tamtou témou, čo je ako toto, čo je ako tamto; dokiaľ ich graf nie je úplne prepojený a zároveň úplne zbytočný. Liekom sú konkrétne vedomosti a štúdium do hĺbky. Keď rozumiete podrobnostiam vecí, vidíte, že *nie* sú jedna ako druhá a začnete nadšene *odoberať* hrany zo svojho grafu.

Podobne, dôležité pojmy sú tie, ktoré nezahŕňajú všetko možné vo vesmíre. Dobré hypotézy dokážu vysvetliť niektoré možné výsledky, ale nie iné.

Je absolútne v poriadku, že Isaac Newton vysvetlil *iba* gravitáciu, *iba* spôsob, ako veci padajú dole – a ako planéty obiehajú okolo Slnka a ako Mesiac spôsobuje príliv – ale *nie* úlohu peňazí v ľudskej spoločnosti, ani ako srdce pumpuje krv. Uškŕňať sa nad presnosťou je spomienkou na starých Grékov, ktorí si mysleli, že vyjsť von a naozaj sa na veci *pozrieť* je manuálna práca, a manuálna práca je pre otrokov.

Ako vyjadril Platón (v *Republike*, knihe VII):

Keby niekto zaklonil hlavu a učil by sa niečo pozeraním na rôzne vzory na strope, iste by ste si mysleli, že uvažuje svojím rozumom, zatiaľ čo on by iba pozeral svojimi očami... Verím, že žiadne štúdium neupriamuje dušu pozeráť nahor, okrem toho, ktoré sa týka skutočnosti bytia a toho, čo nevidíme. Či už niekto pozerá nahor s otvorenými ústami, alebo pozerá nadol so zatvorenými ústami, ak sú to veci zmyslov, o ktorých sa snaží niečo naučiť, vyhlasujem, že sa nikdy nenaučí, pretože žiadna z týchto vecí nepripúšťa poznanie: Hovorím, že jeho duša pozerá nadol, nie nahor, aj keby ležal na chrbte na zemi či na mori!

Mnohí dnes robia podobnú chybu a myslia si, že úzke pojmy sú nehodné a nevznešené a nefilozofické, asi ako povedzme ísť von a pozeráť sa na veci – konanie hodné iba spodiny. Avšak racionalisti – a aj básnici – potrebujú úzke pojmy na vyjadrenie presných myšlienok; potrebujú pojmy, ktoré zahŕňajú iba niektoré veci a vylučujú iné. Nie je nič zlé na sústredení myšlienok, zúžení pojmov, vylúčení možností a zaostrení svojich tvrdení. Naozaj, nie je! Ak budú vaše slová príliš široké, skončíte s niečím, čo nie je pravda, ba ani dobrá poézia.

A radšej mi ANI NESPOMÍNAJTE ľudí, ktorí si myslia, že Wikipédia je "umelá inteligencia", vynález LSD bola "singularita" alebo že korporácie sú "superinteligentné"!

Pôvodný článok s komentármi²

² http://lesswrong.com/lw/ic/the_virtue_of_narrowness/

Kapitola 27. Tvoja sila racionalistu

(Nasledujúca vec sa mi stala v diskusnej miestnosti IRC kedysi dávno, keď som sa ešte potľkal po diskusných miestnostiach IRC. Čas mi zahmlil pamäť a môj popis môže byť nepresný.)

Bol som v diskusnej miestnosti IRC, keď niekto napísal, že jeho známy potrebuje lekársku radu. Jeho známy hovorí, že mal náhle bolesti v hrudníku, tak zavolať sanitku, sanitka prišla, ale zdravotníci mu povedali, že to nič nie je a odišli, a bolesti v hrudníku sa stále zhoršujú. Čo má jeho známy robiť?

Bol som touto historkou zmätený. Pamätal som si, ako som čítal o bezdomovcoch v New Yorku, ktorí si volali sanitky, len aby ich odviezli niekam do tepla, a že ich zdravotníci vždy museli vziať na pohotovosť, hoci aj 27-krát. Pretože keby ich nevzali, majiteľa sanitiek by bolo možné žalovať o veľa a veľa peňazí. Podobne, pohotovosti majú zo zákona povinnosť poskytnúť zdravotnú starostlivosť každému bez ohľadu na jeho schopnosť platiť. (Náklady potom znáša nemocnica a sú to obrovské sumy, takže nemocnice rušia svoje pohotovosti... Človek sa čuduje, načo vôbec máme ekonómov, keď ich jednoducho ignorujeme.) Takže som celkom nerozumel, ako sa popisované udalosti mohli stať. *Každého*, kto by nahlásil náhle bolesti v hrudníku, by okamžite odviezli na pohotovosť.

A tu som ako racionalista sklamal. Spomenul som si na pár prípadov, keď môj doktor úplne odmietol panikáriť pri popise príznakov, ktoré sa mne zdali veľmi hrozivé. A lekárske establishment mal vždy pravdu. Každý jeden raz. Raz aj mňa bolel hrudník a lekár mi trpezlivo vysvetlil, že mu opisujem bolesť hrudného svalu, nie srdcový infarkt. Povedal som teda na IRC: "Pozri, ak zdravotníci povedali tvojmu známemu, že to nič nie je, *naozaj* to nič nie je – keby bola najmenšia šanca vážneho problému, boli by ho zobrali."

Takto sa mi podarilo príbeh vysvetliť v rámci môjho existujúceho modelu, hoci som cítil, že to bolo trochu nasilu...

Neskôr sa dotýčny vrátil do diskusnej miestnosti IRC a povedal, že jeho známy si to celé vymyslel. Zrejme to nebol jeden z jeho spoľahlivejších známych.

Azda som si mohol uvedomiť, že neznámy známy známeho z IRC môže byť menej dôveryhodný¹ než článok uverejnený v odbornom časopise. Žiaľ, veriť je ľahšie než neveriť; veríme inštinktívne, neveriť však vyžaduje vedomé úsilie².

Namiesto toho som silným tlakom donútil svoj model skutočnosti, aby vysvetlil nezrovnalosť, ktorá sa v *skutočnosti nikdy nestala*. A *vedel som*, aké je to trápne. *Vedel som*, že užitočnosť modelu nie je v tom, čo dokáže vysvetliť, ale v tom, čo nedokáže. Hypotéza, ktorá nič nezakazuje, povoľuje všetko a preto nedokáže obmedziť očakávanie³.

Tvoja sila racionalistu je tvoja schopnosť byť viac zmätený výmyslom než skutočnosťou. Ak dokážeš rovnako dobre vysvetliť hocijaký výsledok, máš nulové vedomosti.

Všetci sme z času na čas slabí; smutné je, že som *mohol* byť silnejší. Mal som všetky informácie potrebné na odvodenie správnej odpovede, dokonca som si

¹ <http://www.overcomingbias.com/2007/08/truth-bias.html>

² <http://www.wjh.harvard.edu/~dtg/Gilbert%20et%20al%20%28EVERYTHING%20YOU%20READ%29.pdf>

³ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

všimol problém, ale potom som ho ignoroval. Môj pocit zmätenia bola Nápoveda a ja som túto Nápovedu zahodil.

Mal som venovať viac pozornosti tomu pocitu, že *to bolo trochu nasilu*. Je to jeden z najdôležitejších pocitov, aké môže mať hľadač pravdy, časť tvojej sily racionalistu. Je to chybou dizajnu ľudského rozumu, že sa tento pocit prejavuje ako tichý hlások v pozadí mysle namiesto kvíliacej poplašnej sirény a žiarivého neónového nápisu: "BUĎ JE TVOJ MODEL NESPRÁVNÝ, ALEBO TENTO PRÍBEH NIE JE PRAVDA."

Pôvodný článok s komentármi⁴

⁴ http://lesswrong.com/lw/if/your_strength_as_a_rationalist/

Kapitola 28. Neprítomnosť indície je indíciou neprítomnosti

Z knihy Robyna Dawesa *Rozumná voľba v neistom svete*:

Post-hoc napasovanie indícií do hypotézy bolo súčasťou najsmutnejšej kapitoly dejín Spojených Štátov: uväznenia Američanov japonského pôvodu v internačných táboroch začiatkom druhej svetovej vojny. Keď kalifornský guvernér Earl Warren vypovedal 21. februára 1942 v San Franciscu pred komisiou kongresu, v rámci kladenia otázok mu pripomenuli, že do tej doby sa nevyskytla žiadna sabotáž ani iný typ špionáže, ktorú by mali na svedomí Američania japonského pôvodu. Warren odpovedal: “Som toho názoru, že táto neprítomnosť [podvratnej aktivity] je tým najhrozivejším znamením celej našej situácie. Presvedča ma to azda viac než hociktorý iný faktor, že sabotáže, ktoré prídu, že aktivity piatej kolóny, ktoré prídu, sú načasované rovnako ako bol načasovaný Pearl Harbor... Verím tomu, že sa nás snažia učičkáť do falošného pocitu bezpečia.”

Veźmime si tento Warrenov argument z bayesovského hľadiska¹. Keď vidíme indíciu, hypotézy, ktoré tejto indícii pripisovali vyššiu podmienenú pravdepodobnosť, získajú pravdepodobnosť na úkor hypotéz, ktoré tejto indícii pripisovali nižšiu podmienenú pravdepodobnosť. Toto je jav *relatívnych* podmienených pravdepodobností a *relatívnych* pravdepodobností. Môžete danej indícii pripísať vysokú podmienenú pravdepodobnosť a predsa stratiť pravdepodobnosť v prospech nejakej inej hypotézy, pokiaľ táto iná hypotéza pripísala ešte vyššou podmienenú pravdepodobnosť.

Warren sa snaží argumentovať, že neprítomnosť sabotáže *potvrďuje*, že piata kolóna existuje. Môžeme argumentovať, že piata kolóna *možno* svoje sabotáže odkladá na neskôr. Ale aj tak je vyššia podmienená pravdepodobnosť, že *neexistencia* piatej kolóny spôsobuje neprítomnosť sabotáže.

Nech E je pozorovanie sabotáže, H1 je hypotéza piatej kolóny Američanov japonského pôvodu a H2 je hypotéza, že piata kolóna neexistuje. Bez ohľadu na to, aká je podmienená pravdepodobnosť $P(E|H1)$, že piata kolóna nerobí sabotáže, nemôže byť taká veľká ako pravdepodobnosť $P(E|H2)$, že neexistujúca piata kolóna nerobí sabotáže. Pozorovanie neprítomnosti sabotáže teda zvyšuje pravdepodobnosť, že piata kolóna neexistuje.

Neprítomnosť sabotáže *nedokazuje*, že piata kolóna neexistuje. Neprítomnosť *dôkazu* nie je *dôkazom* neprítomnosti. V logike $A \rightarrow B$, “ak A, tak B”, nie je to isté ako $\sim A \rightarrow \sim B$, “ak nie A, tak nie B”.

Ale v teórii pravdepodobnosti, neprítomnosť *indície* je vždy *indíciou* neprítomnosti. Ak je E binárna udalosť a $P(H|E) > P(H)$, “vidieť E zvyšuje pravdepodobnosť H”; potom $P(H|\sim E) < P(H)$, “nevidieť E znižuje pravdepodobnosť H”. $P(H)$ je váženým priemerom $P(H|E)$ a $P(H|\sim E)$, takže nevyhnutne leží medzi nimi. Ak vám čokoľvek z tohto znie nejasne, prečítajte si Intuitívne vysvetlenie bayesovského uvažovania².

¹ pozri: 178 Intuitívne vysvetlenie Bayesovej vety

² pozri: 178 Intuitívne vysvetlenie Bayesovej vety

V drvivej väčšine okolností z bežného života, príčina nemusí spoľahlivo vykazovať znaky svojej prítomnosti, ale neprítomnosť príčiny bude tieto znaky vykazovať ešte zriedkavejšie. Neprítomnosť pozorovania môže byť silnou indíciou neprítomnosti alebo slabou indíciou neprítomnosti, podľa toho, s akou pravdepodobnosťou daná príčina spôsobuje pozorovanie. Neprítomnosť pozorovania, ktoré by aj tak bolo veľmi zriedkavé (aj keď ho alternatívna hypotéza nedovoľuje vôbec), je veľmi slabou indíciou neprítomnosti (ale aj tak je to indícia). Toto je klam "medzier v zázname skamenelín" - skameneliny sa tvoria iba zriedkavo; je zbytočné roztrubovať neprítomnosť veľmi zriedkavého pozorovania, keď bolo zaznamenaných tak veľa silných pozitívnych pozorovaní. Ale ak nie sú absolútne *žiadne* pozitívne pozorovania, je čas znepokojovať sa; preto Fermiho paradox.

Tvoja sila racionalistu³ je tvoja schopnosť byť viac zmätený výmyslom než skutočnosťou; ak dokážeš rovnako dobre vysvetliť hocijaký výsledok, máš nulové vedomosti. Sila modelu nie je v tom, čo *môže* vysvetliť, ale v tom, čo *nemôže*, lebo iba zákazy obmedzujú očakávanie⁴. Ak si nevšimneš, kedy tvoj model robí indíciu nepravdepodobnou, je to akoby si nemal žiaden model, a tiež akoby si nemal indície, mozog, ani oči.

Pôvodný článok s komentármi⁵

³ pozri: 27 Tvoja sila racionalistu

⁴ pozri: 13 Viera vo vieru

⁵ lesswrong.com/lw/ih/absence_of_evidence_is_evidence_of_absence/

Kapitola 29. Zákon zachovania očakávanej indície

Friedrich Spee von Langenfeld, kňaz, ktorý počúval spovede odsúdených bosoriek, napísal v roku 1631 *Cautio Criminalis* ("obozretnosť v kriminálnych prípadoch"), kde uštipačne opísal rozhodovací strom na odsudzovanie obvinených z bosoráctva: Ak bosorka viedla zlý a nesprávny život, je vinná; ak viedla dobrý a správny život, aj to je dôkaz, pretože bosorky sa pretvarujú a pokúšajú sa vyzeráť zvlášť cnostne. Keď je žena uväznená: ak sa bojí, dokazuje to jej vinu; ak sa nebojí, dokazuje to jej vinu, pretože bosorky zvyčajne predstierajú nevinu a tvária sa statočne. Alebo keď počuje, že bola udaná z bosoráctva, môže sa pokúšať ujsť alebo zostať; ak utekala, dokazuje to jej vinu; ak zostala, diabol ju zadržal, takže nemohla odísť.

Spee robil spovedníka mnohým bosorkám; mal teda príležitosť vidieť každú vetvu obviňovacieho stromu, kde bez ohľadu na to, čo obvinená z bosoráctva povedala alebo urobila, vždy to bolo použité ako dôkaz proti nej. V každom jednotlivom prípade ste však počuli iba jednu z dvoch možných vetiev. Toto je dôvod, prečo si vedci vopred zapisujú svoje experimentálne predpovede.

Nemôžete mať jedno i druhé - z hľadiska teórie pravdepodobnosti, nielen férovosti. Pravidlo, že "neprítomnosť indície je indíciou neprítomnosti"¹ je špeciálnym prípadom všeobecnejšieho zákona, ktorý by som pomenoval Zákon zachovania očakávanej indície: Očakávaná priemerná výsledná pravdepodobnosť po zohľadnení indície sa musí rovnať pôvodnej pravdepodobnosti.

$$P(H) = P(H)$$

$$P(H) = P(H,E) + P(H,\sim E)$$

$$P(H) = P(H|E) \times P(E) + P(H|\sim E) \times P(\sim E)$$

Preto každému očakávaniu indície zodpovedá očakávanie rovnakej protiindície v opačnom smere.

Ak očakávate silnú pravdepodobnosť, že uvidíte slabú indíciu jedným smerom, musí to byť vyvážené slabým očakávaním videnia silnej indície opačným smerom. Ak ste si veľmi istí svojou teóriou a preto očakávate, že uvidíte výsledok zodpovedajúci vašej hypotéze, môže to posilniť váš názor iba o máličko (už je blízky 1); avšak nečakané zlyhanie vašej predpovede by malo (musí) dať vašej sebadôvere silný úder. Musíte očakávať, že budete mať v *priemere rovnakú* sebadôveru ako na začiatku. Inými slovami, samotné *očakávanie*, že uvidíte indíciu - predtým než ju naozaj uvidíte - by nemalo zmeniť vaše pôvodné názory. (Opäť, ak vám toto nie je intuitívne zrejmé, pozrite si Intuitívne vysvetlenie bayesovského rozmýšľania².)

Ak teda tvrdíte³, že "žiadna sabotáž" je indíciou *pre* existenciu piatej kolóny Američanov japonského pôvodu, musíte tomu zodpovedajúco tvrdiť, že vidieť sabotáž by bolo argumentom *proti* existencii piatej kolóny. Ak tvrdíte, že "dobrý a správny život" je indíciou, že žena je bosorka, potom zlý a nesprávny život musí byť indíciou, že nie je bosorka. Ak tvrdíte⁴, že Boh odmieta prejaviť svoju

¹ pozri: 28 Neprítomnosť indície je indíciou neprítomnosti

² pozri: 178 Intuitívne vysvetlenie Bayesovej vety

³ pozri: 28 Neprítomnosť indície je indíciou neprítomnosti

⁴ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

existenciu, aby testoval vieru ľudí, potom zázraky opísané v Biblii musia byť argumentmi proti existencii Boha.

To neznie celkom správne, však? Venujte pozornosť tomuto pocitu, že *je to trochu nasilu*, tomu tichému hlásku v pozadí vašej mysle⁵. Je to dôležité.

Pre skutočného bayesiána je nemožné hľadať indície, ktoré *potvrdzujú* nejakú teóriu. Neexistuje žiaden plán, ktorý by ste mohli vymyslieť, žiadna chytrá stratégia, žiaden prefíkaný nástroj, pomocou ktorého by ste mohli legitímne očakávať, že vaša dôvera v daný výrok bude (v *priemere*) vyššia než predtým. Môžete hľadať indície iba na to, aby ste teóriu *testovali*, ale nie aby ste ju potvrdili.

Toto uvedomenie dokáže veľmi odbremeniť vašu myseľ. Nemusíte sa znepokojovať ohľadom toho, ako interpretovať všetky možné experimentálne výsledky, aby potvrdzovali vašu teóriu. Nemusíte sa zťažovať plánovaním, ako z *každej* štipky indície vyťažiť potvrdenie pre vašu teóriu, pretože viete, že každému očakávaniu indície zodpovedá očakávanie rovnakej protiindície v opačnom smere. Ak sa pokúsite oslabiť protiindíciu možného "abnormálneho" pozorovania, môžete to urobiť iba oslabením podpory "normálneho" pozorovania, v úplne rovnakej miere v opačnom smere. Je to hra s nulovým súčtom. Bez ohľadu na to, ako špekulujete, ako argumentujete, ako strategizujete, nemôžete očakávať, že výsledný plán hry posunie (v *priemere*) vaše názory konkrétnym smerom.

Môžete sa teda posadiť a uvoľniť, zatiaľ čo čakáte na príchod indície.

...ľudská psychológia je *taká* domotaná.

Pôvodný článok s komentármi⁶

⁵ pozri: 27 Tvoja sila racionalistu

⁶ http://lesswrong.com/lw/ii/conservation_of_expected_evidence/

Kapitola 30. Spätný pohľad znehodnocuje vedu

Tento výňatok¹ z Meyersovej knihy *Skúmanie sociálnej psychológie* sa oplatí prečítať celý. Cullen Murphy, redaktor *The Atlantic*, povedal, že spoločenské vedy neprinášajú “žiadne myšlienky či závery, ktoré by sa nedali nájsť v [hociktorej] encyklopédii citátov... Deň za dňom sociológovia idú do sveta. Deň za dňom zisťujú, že správanie ľudí je viacmenej také, ako by ste čakali.”

Samozrejme, celé toto “čakanie” je spätný pohľad². (Skreslenie spätného pohľadu: Osoby, ktoré poznajú správnu odpoveď na otázku, pripisujú omnoho väčšiu pravdepodobnosť tomu, že “by boli” uhádli túto odpoveď, v porovnaní s osobami, ktoré musia hádať bez poznania správnej odpovede.)

Historik Arthur Schlesinger, Jr. zavrhol vedecké štúdie skúseností vojakov z 2. svetovej vojny ako “nemotornú ukážku” zdravého rozumu. Napríklad:

1. Vojaci s vyšším vzdelaním mali väčšie problémy prispôbiť sa než menej vzdelaní vojaci. (Intelektuáli boli menej pripravení na bojový stres než ľudia z ulice.)
2. Vojaci z juhu USA zvládali horúce podnebie South Sea Island lepšie než vojaci zo severu USA. (Južania sú viac zvyknutí na horúce počasie.)
3. Bieli vojaci sa viac snažili byť povýšení na poddôstojníkov než čierni vojaci. (Roky útlaku majú dopad na motiváciu k úspechu.)
4. Černosi z juhu dávali prednosť bielym dôstojníkom z juhu pred bielymi dôstojníkmi zo severu (pretože dôstojníci z juhu mali väčšiu skúsenosť a zručnosť v interakcii s černochochmi).
5. Kým trvali boje, vojaci viac túžili po návrate domov, než keď vojna skončila. (Počas boja vojaci vedeli, že sú v smrteľnom nebezpečenstve.)

Kolko z týchto zistení si myslíte, že by ste *boli vedeli* predpovedať? 3 z 5? 4 z 5? Sú tam prípady, kde by ste boli predpovedali opak - kde by váš model dostal úder³? Na chvíľku sa zamyslite, než budete pokračovať...

V tejto ukážke (od Paula Lazarsfelda prostredníctvom Meyersa) boli všetky horeuvedené zistenia *opakom* toho, čo sa naozaj zistilo. Kolkokrát ste si mysleli, že váš model dostal úder? Kolkokrát ste priznali, že by ste sa boli mýlili? Tak taký dobrý bol váš model skutočnosti. Miera vašej sily racionalistu⁴ je vaša schopnosť byť viac prekvapený fikciou než skutočnosťou.

Samozrejme, pokiaľ som výsledky neotočil ešte raz. Čo si myslíte?

Máte pocit, že vaše myšlienkové postupy v tejto chvíli, keď *naozaj neviete* správnu odpoveď, sú iné než myšlienkové procesy, ktoré ste použili na racionalizovanie ľubovoľnej strany “známej” odpovede?

Daphna Baratz dala vysokoškolákovi dvojicu domnelých zistení, jedno pravdivé (“Počas prosperity ľudia míňajú väčšie percento svojho príjmu než počas krízy”) a jedno, ktoré bolo opakom pravdy. V oboch prípadoch študenti ohodnotili údajne

¹ <http://csml.som.ohio-state.edu/Music829C/hindsight.bias.html>

² http://lesswrong.com/lw/il/hindsight_bias/

³ pozri: 29 Zákon zachovania očakávanej indicie

⁴ pozri: 27 Tvoja sila racionalistu

zistenie ako to, čo "by boli predpovedali". Dokonalý štandardný omyl spätného pohľadu.

Ktorý vedie ľudí k záveru, že nepotrebujeme vedu, pretože by to všetko "boli predpovedali".

(Presne tak, ako by ste boli predpovedali, však?)

Spätný pohľad nás vedie k systematickému podceňovaniu prekvapivosti vedeckých zistení, najmä objavov, ktorým *rozumieme* - tých, ktoré nám pripadajú skutočné, ktoré si vieme spätne dosadiť do svojich modelov sveta. Ak sa vyznáte v neurológii alebo fyzike a čítate si novinky z tejto oblasti, pravdepodobne podceňujete aj prekvapivosť zistení v týchto oblastiach. Toto neférové znehodnocuje príspevok výskumníkov; a čo je horšie, zabraňuje vám to všimnúť si, keď vidíte indíciu, ktorá nezapadá⁵ do toho, čo by ste *naozaj* boli očakávali.

Musíme vynaložiť vedomé úsilie na to, aby sme boli *dostatočne* šokovaní.

Pôvodný článok s komentármi⁶

⁵ pozri: 29 Zákon zachovania očakávanej indície

⁶ http://lesswrong.com/lw/im/hindsight_devalues_science/

Časť IV. Tajomné odpovede

Obsah

31 Falošné vysvetlenia	71
32 Háďanie učiteľovho hesla	73
33 Veda ako uniforma	76
34 Falošná kauzalita	78
35 Sémantické stopky	81
36 Tajomné odpovede na tajomné otázky	83
37 Márnosť emergencie	86
38 Nehovorte "zložitost"	88
39 Pozitívne skreslenie: Pozeraj do tmy	91
40 Moja divoká a bezstarostná mladost'	93
41 Neučíme sa z histórie	95
42 Sprístupniť históriu	96
43 Vysvetli / Uctieľaj / Ignoruj?	98
44 "Veda" ako zastavovač zvedavosti	100
45 Čo myslím slovom "rozumnosť"?	102

Kapitola 31. Falošné vysvetlenia

Kde bolo, tam bolo, bola raz jedna učiteľka fyziky. Jedného dňa zvolala svojich žiakov do triedy a ukázala im široký štvorcový kovový plát vedľa horúceho radiátora. Žiaci položili ruky na plát a zistili, že strana pri radiátore je studená a odvrátená strana je horúca. A učiteľka povedala: *Prečo si myslíte, že je to tak?* Niektorí žiaci tipovali na prúdenie vzduchu, iní tipovali, že plát obsahuje zvláštne kovy. Vymysleli veľa tvorivých vysvetlení; nikto sa neznižil k tomu, aby povedal: “Neviem” alebo: “Zdá sa mi to nemožné.”¹

Správna odpoveď znela, že pred vstupom žiakov do triedy, učiteľka otočila plát naopak.

Zamyslime sa nad žiakom, ktorý horúčkovo koktá: “No, možno je to kvôli tepelnej vodivosti a takým veciam?” Pýtam sa: jej táto odpoveď pravý názor²? Tieto slová sa ľahko vyznávajú³ – vyslovujú hlasným precíteným hlasom. Ale dokážu tieto slová ovplyvňovať očakávanie⁴?

Uvažujme o tomto nevinnom slovíčku “kvôli”, ktoré sa nachádza pred “tepelnej vodivosti”. Uvažujme o *ďalších* veciach, ktoré by sme zaň mohli napísať. Mohli by sme povedať, napríklad, “kvôli flogistonu” alebo “kvôli mágii”.

“Mágia!” vykriknete. “To nie je *vedecké* vysvetlenie!” Isteže, vety “kvôli tepelnej vodivosti” a “kvôli mágii” ľahko rozoznáme ako patriace do rôznych *literárnych žánrov*. “Tepelná vodivosť” je niečo, čo môže povedať povedzme Spock v *Star Treku*, zatiaľ čo “mágia” je niečo, čo povie Giles v *Buffy, Lovkyni Upírov*.

My, ako bayesiánci, si však nevšímame literárne žánre. Pre nás je podstatou modelu jeho vplyv na očakávanie. Ak povie “tepelná vodivosť”, aké vnemy vás to vedie *očakávať*? Za bežných okolností vás to vedie k očakávaniu, že ak položíte ruku na stranu plátu pri radiátore, budete vám táto strana pripadať teplejšia než vzdialená strana. Ak “kvôli tepelnej vodivosti” dokáže zároveň vysvetliť, prečo vám strana pri radiátore pripadá *chladnejšia*, potom dokáže vysvetliť prakticky *hocičo*.

A⁵ v⁶ tejto⁷ chvíli⁸ už⁹ všetci¹⁰ vieme¹¹ (aspoň¹² dúfam¹³, že¹⁴ vieme¹⁵), že keď viete rovnako dobre vysvetliť ľubovoľný výsledok, máte nulovú vedomosť. “Kvôli tepelnému prúdeniu”, použité týmto spôsobom, je maskovaná hypotéza maximálnej entropie. Z hľadiska očakávania je izomorfná povedaniu “mágia”. Znie to ako vysvetlenie, ale nie je.

Predstavte si, že by sme namiesto hádania merali teplotu kovového plátu v rôznych bodoch v rôznom čase. Keby sme videli kovový plát vedľa radiátora,

¹ pozri: 27 Tvoja sila racionalistu

² pozri: 17 Názor ako uniforma

³ pozri: 15 Vyznávanie a fandanie

⁴ pozri: 13 Viera vo vieru

⁵ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

⁶ pozri: 13 Viera vo vieru

⁷ pozri: 14 Bayesovské džudo

⁸ pozri: 15 Vyznávanie a fandanie

⁹ pozri: 17 Názor ako uniforma

¹⁰ pozri: 19 Nárok náboženstva na nevyvrátiteľnosť

¹¹ pozri: 25 Zameraj svoju neistotu

¹² pozri: 27 Tvoja sila racionalistu

¹³ pozri: 28 Neprítomnosť indície je indíciou neprítomnosti

¹⁴ pozri: 30 Spätný pohľad znehodnocuje vedu

¹⁵ pozri: 29 Zákon zachovania očakávanej indície

za normálnych okolností by sme očakávali, že bodové teploty budú zodpovedať rovnováhe rovnice rozptylu vzhľadom na hraničné podmienky dané prostredím. Možno by ste nevedeli presnú teplotu prvého meraného bodu, ale po odmeraní prvých bodov - nie som dosť fyzikálne zdatný, aby som vedel, kolkých presne - by ste mohli urobiť výborný odhad zvyšku.

Skutočný majster umenia používania čísel na obmedzenie očakávania hmotných javov - "fyzik" - by urobil pár meraní a povedal: "Tento plát bol v rovnováhe s prostredím pred dva a pol minútami, bol otočený, a teraz sa opäť približuje k rovnováhe."

Omyl týchto žiakov nebol len v tom, že nedokázali obmedziť očakávanie. Ich hlbší omyl spočíval v domnení, že robia fyziku. Povedali slovo "kvôli", nasledované slovami, ako hovorí Spock v *Star Treku*, a mysleli si, že tým vstúpili do magistéria vedy.

Veru nie. Iba presunuli svoju mágiu z jedného literárneho žánru do druhého.

Pôvodný článok s komentármi¹⁶

¹⁶ lesswrong.com/lw/ip/fake_explanations/

Kapitola 32. Hádanie učiteľovho hesla

Za mladi som čítal populárne fyzikálne knihy ako od Richarda Feynmana *QED: Zvláštna teória svetla a hmoty*. Vedel som, že svetlo sú vlny, zvuk sú vlny, hmota sú vlny. Vo veku deväť rokov som bol hrdý na svoju vedeckú gramotnosť.

Keď som bol starší a začal som si čítať *Feynmanove lekcie z fyziky*, našiel som poklad s názvom “vlnová rovnica”. Dokázal som pochopiť odvodenie tej rovnice, ale ani dodatočne¹ som nevedel uvidieť jej pravdivosť na prvý pohľad. Tak som nad tou vlnovou rovnicou rozmýšľal s prestávkami tri dni, dokiaľ som neuvidel, aká je trápne samozrejma. A keď som ju konečne pochopil, uvedomil som si, že po celý čas, čo som prijímal poctivé ubezpečenia fyzikov, že svetlo sú vlny, zvuk sú vlny, hmota sú vlny, nemal som najmenšie tušenie, čo slovo “vlna” znamená pre fyzika.

Máme inštinktívny sklon myslieť si, že keď fyzik povie: “svetlo sa skladá z vln” a učiteľ povie: “Z čoho sa skladá svetlo?” a žiak povie: “Z vln!”, že ten žiak povedal pravdivý výrok. To je predsa férové, nie? Keď uznáme odpoveď “vlny” ako správnu od fyzika, nebolo by neférové odmietnuť ju od žiaka? Odpoveď “Vlny!” je určite buď *pravdivá* alebo *nepravdivá*, však?

A to je opäť jeden zvyk zo školy, ktorý sa treba odučiť². Slová nemajú definíciu sami od seba. Keď počujem slabiky “bo-bor” a predstavím si veľkého hlodavca, je to fakt o stave mojej mysle, nie fakt o slabikách “bo-bor”. Postupnosť slabík “skladá sa z vln” (alebo “kvôli tepelnej vodivosti³”) nie je *hypotéza*, je to len vzor vibrácií putujúci vzduchom, alebo machuľa na papieri. Môže sa v mysli niekoho *spojiť* s hypotézou, ale samo od seba to nie je správne ani nesprávne. V škole vám však učiteľ dá hviezdičku za *vyslovenie* “skladá sa z vln”, čo musí byť správna odpoveď, pretože učiteľ počul, ako z fyzik vyludzoval tie isté zvukové vibrácie. Keďže slovné správanie (vyslovené alebo napísané) je to, čo vám prinesie hviezdičku, žiaci si začínajú myslieť, že slovné správanie má pravdivostnú hodnotu. Napokon, svetlo sa buď skladá z vln alebo nie, však?

A to vedie k ešte horšiemu zlozvyku. Predstavte si, že vám učiteľka predloží mätúci problém⁴ týkajúci sa kovového plátu vedľa radiátora; vzdialená strana je na dotyk teplejšia než strana pri radiátore. Učiteľka sa opýta: “Prečo?” Ak poviete: “Neviem,” *nemáte* šancu získať hviezdičku – ani len čiarku za aktivitu na hodine. Ale, počas tohto polroka táto učiteľka použila slovné spojenia “kvôli tepelnému prúdeniu”, “kvôli tepelnej vodivosti” a “kvôli vyžarovaniu tepla”. Jedno z nich je asi to, čo učiteľka chce. Poviete: “No, možno kvôli tepelnej vodivosti?”

Toto nie je *hypotéza* o kovovom pláte. To nie je ani pravý názor⁵. Je to pokus *uhádnuť učiteľkine heslo*.

Aj keď si predstavíte symboly rovnice rozptylu (matematiku riadiacu vedenie tepla), neznamená to, že ste vytvorili *hypotézu* o kovovom pláte. Toto nie je škola; my netestujeme vašu pamäť, či dokázate zapísať rovnicu rozptylu. Toto je bayesovské remeslo; my hodnotíme vaše očakávania vnemov. Ak *použijete* rovnicu rozptylu, odmeriate niekoľko bodov teplomerom a potom

¹ <http://www.math.utah.edu/~pa/math/polya.html>

² lesswrong.com/lw/i2/two_more_things_to_unlearn_from_school/

³ pozri: 31 Falošné vysvetlenia

⁴ pozri: 31 Falošné vysvetlenia

⁵ pozri: 17 Názor ako uniforma

skúsíte predpovedať, čo teplomer povie pri nasledujúcom meraní, potom je to jednoznačne spojené s vnemmi. Dokonca aj keď si žiak iba predstavuje, ako niečo *prúdi* a preto prikladá merač k studenejšej strane plátu v snahe odmerať, kam teplo išlo, potom sa tento myšlienkový obraz prúdenia spája so skúsenosťou; riadi očakávanie.

Ak *nepoužívate* rovnicu rozptylu – vkladáním čísel a získavaním výsledkov, ktoré riadia vaše očakávanie konkrétnych vnemov – potom toto spojenie medzi mapou a územím akoby bolo preťaté nožom. Čo zostalo, nie je názor⁶, ale slovné správanie.

V školskom systéme je to všetko o slovnom správaní, či už napísanom na papieri alebo vyslovenom nahlas. Za slovné správanie dostanete hviezdičku alebo prepadnete. Súčasťou odvykania od tohto zlozvyku je uvedomovanie si rozdielov medzi vysvetlením a heslom.

Vyzerá toto príliš drsne? Keď čelíte mäťúcemu kovovému plátu, nemôže “tepelná vodivosť?” byť prvým krokom k nájdeniu odpovede? Možno, ale iba ak nespadnete do pasce myslenia si, že hľadáte heslo. Čo ak nemáte žiadneho učiteľa, ktorý by vám povedal, že je to zle? Potom si môžete myslieť, že “svetlo sú wakalixy” je dobré vysvetlenie, že “wakalixy” je správne heslo. To sa stalo mne, keď som mal deväť rokov – nie preto, že by som bol hlúpy, ale pretože toto sa stáva *automaticky*. Takto ľudia myslia, pokiaľ nie sú vycvičení, aby do tejto pasce *nepadli*. Ľudstvo zostávalo v takýchto jamách zaseknuté celé tisícročia.

Možno, keby sme žiakov vycvičili, že *slová sa nepočítajú, iba ovládače očakávania*, žiaci by sa *nezasekávali* na: “Tepelná vodivosť? Nie? Žeby tepelné prúdenie? Ani to nie?” Možno *potom* by myšlienka “tepelná vodivosť” viedla k skutočne užitočnej ceste, ako:

- “Tepelná vodivosť?”
- Ale to je len fráza – čo to znamená?
- Rovnica rozptylu?
- Ale to sú len symboly – ako ich použijem?
- K akému očakávaniu ma vedie použitie rovnice rozptylu?
- Určite ma to nevedie k očakávaniu, že strana kovového plátu, ktorá je odvrátená od radiátora, bude na dotyk teplejšia.
- Všimam si⁷, že som zmätený⁸. Možno sa tá blízka strana iba *zdá* studenšou, pretože je vyrobená z izolujúcejšieho materiálu a prenáša menej tepla na moju ruku? Skúsím tú teplotu odmerať...
- Okej, to nebolo ono. Môžem skúsiť overiť, či rovnica rozptylu vôbec platí pre tento kovový plát? Či teplo *prúdi* tak, ako zvyčajne, alebo sa tu deje niečo iné?
- Mohol by som priložiť merač k plátu a skúsiť sledovať, ako sa teplo postupne šíri...

Ak *nie* sme dosť prísni ohľadom toho, že: “No, možno kvôli tepelnému prúdeniu?” môže byť falošným vysvetlením, žiak sa veľmi pravdepodobne zasekne na nejakom hesle ako wakalixy. *Toto sa stáva automaticky, stávalo sa to ľudstvu ako celku po celé tisícročia.*

⁶ pozri: 17 Názor ako uniforma

⁷ pozri: 27 Tvoja sila racionalistu

⁸ pozri: 30 Spätný pohľad znehodnocuje vedu

Pôvodný článok s komentármi⁹

⁹ http://lesswrong.com/lw/iq/guessing_the_teachers_password/

Kapitola 33. Veda ako uniforma

Súvisí s kapitolou: Názor ako uniforma¹

Upútavka na film *X-Men* obsahuje hlas, ktorý hovorí: “V každej ľudskej bytosti... sa nachádza genetický kód... pre mutáciu.” Očividne môžete pomocou mutácie získať všemožné šikovné schopnosti. Mutantka Storm, napríklad, má schopnosť vrhať blesky.

Prosím ťa, drahý čitateľ, aby si zvážil biologické ústrojenstvo potrebné na generovanie elektriny; biologické adaptácie potrebné na zabránenie zraneniu vlastnou elektrinou; a kognitívne obvody potrebné na presne naladenú kontrolu bleskov. Keby sme v skutočnosti pozorovali nejaký organizmus, ktorý získal všetky tieto schopnosti v *jednej generácii* ako výsledok *mutácie*, naprosto by to vyvrátilo neodarwinovský model prirodzeného výberu. Bolo by to horšie než nájsť skamenelého kráľika v predkambriu. Keby evolučná teória *naozaj* dokázala vysvetliť Storm, bola by schopná vysvetliť čokoľvek² a všetci vieme, čo by z toho vyplývalo.

Komix *X-Men* používa pojmy ako “evolúcia”, “mutácia” a “genetický kód” iba aby sa umiestnil do toho, čo považuje za vedecký *literárny žáner*. Čo ma na tom najviac desí, je predstava, koľko ľudí, najmä v médiách, chápe vedu *len* ako literárny žáner.

Stretávam sa s ľuďmi, ktorí veľmi jednoznačne veria v³ evolúciu a vysmieávajú sa z hlúposti kreacionistov. A napriek tomu nemajú predstavu o tom, čo teória evolučnej biológie dovoľuje a zakazuje. Budú hovoriť o “ďalšom kroku v evolúcii ľudstva”, akoby sa sem prirodzený výber dostal podľa nejakého plánu. Alebo ešte horšie, budú hovoriť o niečom úplne mimo oblasti evolučnej biológie, napríklad o vylepšenom dizajne počítačových čipov, alebo o delení korporácií, alebo o ľuďoch, ktorí sa nahrajú do počítača, a budú *toto* volať “evolúcia”. Keby evolučná biológia mohla zahŕňať toto, mohla by zahŕňať všetko.

Pravdepodobne väčšina ľudí, ktorí *veria* v evolúciu, používa slovné spojenie “kvôli evolúcii⁴”, pretože chcú byť súčasťou vedeckého davu - viera ako vedecká uniforma⁵, ako keď nosíte laboratórny plášť. Keby vedecký dav namiesto toho používal frázu “kvôli inteligentnému dizajnu”, rovnako veselo by používali aj to - pre ich ovládače očakávania by v tom nebol žiaden rozdiel. Povedať “kvôli evolúcii” namiesto “kvôli inteligentnému dizajnu” *pre nich* nezakazuje Storm. Jediným účelom pre nich je identifikovať sa so skupinou.

Stretávam sa s ľuďmi, ktorí sú celkom ochotní baviť sa o myšlienke umelej inteligencie hlúpejšej než človek alebo dokonca aj umelej inteligencie mierne múdrejšej než človek. Zoznámte ich s pojmom silne nadľudskej umelej inteligencie a oni sa náhle rozhodnú, že je to “pseudoveda⁶”. Nie je to preto, že by si mysleli, že majú teóriu inteligencie, ktorá im dovoľuje vypočítať teoretickú hornú hranicu sily optimalizačného procesu. Namiesto toho si spájajú silnú nadľudskú UI s *literárnym žánrom* apokalyptickej literatúry; zatiaľ čo UI bežiacu v malej korporácii si spájajú s literárnym žánrom časopisu *Wired*. Nehovoria na základe modelu myslenia. Neuvedomujú si, že nejaký model *potrebujú*. Neuvedomujú si, že veda *je o* modeloch. Ich zdrvivúca kritika spočíva výlučne v *porovnaní s apokalyptickou literatúrou*, namiesto povedzme známych zákonov,

¹ pozri: 17 Názor ako uniforma

² pozri: 27 Tvoja sila racionalistu

³ pozri: 15 Vyznávanie a fandanie

⁴ pozri: 31 Falošné vysvetlenia

⁵ pozri: 17 Názor ako uniforma

⁶ http://lesswrong.com/lw/io/is_molecular_nanotechnology_scientific/

ktoré zakazujú takýto výsledok. Chápu vedu *iba* ako literárny žáner alebo ako skupinu, do ktorej patria. Táto uniforma⁷ im nepripadá ako laboratórny plášť; nie je to ten futbalový tím, ktorému fandia⁸.

Existuje nejaká časť vedy, na ktorú ste *hrdí*, že v ňu veríte a napriek tomu tento názor nevyužívate pracovne? Mali by ste sa sami seba opýtať, ktoré budúce vnemy táto viera *zakazuje*, aby sa vám stali. To je suma toho, čo ste vstrebali a stalo sa to naozaj vašou súčasťou. Všetko ostatné sú pravdepodobne heslá⁹ a uniformy¹⁰.

Pôvodný článok s komentármi¹¹

⁷ pozri: 17 Názor ako uniforma

⁸ pozri: 15 Vyznávanie a fandanie

⁹ pozri: 32 Hádanie učiteľovho hesla

¹⁰ pozri: 17 Názor ako uniforma

¹¹ http://lesswrong.com/lw/ir/science_as_attire/

Kapitola 34. Falošná kauzalita

Flogiston bol odpoveďou 18. storočia na elementárny oheň gréckych alchymistov. Zapáľte drevo a nechajte ho horieť. Čo je to ten oranžovo svetlý “oheň”? Prečo sa drevo premieňa na popol? Chemici 18. storočia na obe otázky odpovedali: “flogiston”.

...a to je všetko. To bola ich odpoveď: “Flogiston.”

Flogiston uniká z horiacej hmoty ako viditeľný oheň. Ako flogiston uniká, horiaca hmota stráca flogiston a stáva sa z nej popol, “skutočná hmota”. Oheň v uzavretej nádobe zhasne, pretože vzduch sa nasýti flogistonom a už ho viac neprijme. Uhlie zanecháva po spálení málo zvyškov, pretože je to takmer čistý flogiston.

Teória flogistonu sa samozrejme nedala použiť na *predpovedanie* výsledku chemickej reakcie. Najprv ste sa pozreli na výsledok, potom ste použili teóriu flogistonu, aby ste ho *vysvetlili*. Nebolo to tak, že by teoretickí flogistonisti predpovedali, že oheň v uzavretej nádobe zhasne; namiesto toho zapálili oheň v nádobe, videli ako zhasol a potom povedali: “Vzduch musel byť nasýtený flogistonom.” Nemohli ste použiť teóriu flogistonu na povedanie, čo by ste *nemali* uvidieť¹; dokázala vysvetliť všetko.

To bolo v raných časoch vedy. Dlhو si nikto neuvedomil, že je to problém. Falošné vysvetlenia² *neznejú* falošne. To je to, čo ich robí nebezpečnými.

Moderný výskum naznačuje, že ľudia rozmýšľajú o príčinách a následkoch pomocou niečoho ako orientované acyklické grafy (DAGy) bayesovských sietí. Pretože pršalo, chodník je mokrý; pretože je chodník mokrý, šmýka sa.

[Dážď] -> [Chodník je mokrý] -> [Chodník sa šmýka]

Z tohto dokážeme odvodiť - alebo v bayesovskej sieti presne spočítať pravdepodobnosť - že ak sa chodník šmýka, pravdepodobne pršalo; ale ak už vieme, že chodník je mokrý, nová informácia, že chodník sa šmýka, nám nepovie nič nové o tom, či pršalo.

Prečo je oheň horúci a jasný keď horí?

[“Flogiston”] -> [Oheň je horúci a jasný]

Znie to ako vysvetlenie. Je to *zapísané* v rovnakom formáte kognitívnych údajov. Ľudská myseľ však nedokáže automaticky rozoznať, kedy má príčina neobmedzujúcu šípku na následok. Čo je horšie, vďaka skresleniu spätného pohľadu³ sa nám zdá, že príčina obmedzovala⁴ následok, aj keď bola iba upravená⁵, aby s následkom ladila.

Čo je zaujímavé, naše moderné chápanie pravdepodobnostného uvažovania o kauzalite⁶ dokáže presne opísať, čo robili teoretickí flogistonisti nesprávne. Jednou z hlavných inšpirácií pre bayesovské siete bolo uvedomenie si problému dvojitého započítania indícií ak odvodzovanie rezonuje medzi následkom a príčinou. Povedzme napríklad, že som dostal kúsok nespoľahlivej informácie, že chodník je mokrý. To by malo spôsobiť, že si pomyslím, že je o čosi pravdepodobnejšie, že pršalo. Ale ak je pravdepodobnejšie, že pršalo, nie je

¹ pozri: 27 Tvoja sila racionalistu

² pozri: 31 Falošné vysvetlenia

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

⁴ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

⁵ pozri: 29 Zákon zachovania očakávanej indície

⁶ http://books.google.sk/books/about/Probabilistic_Reasoning_in_Intelligent_S.html?id=AvNID7LyMusC

potom pravdepodobnejšie, že chodník je mokrý? A neznamenaloby to, že je pravdepodobnejšie, že chodník sa šmýka? Ale ak sa chodník šmýka, asi je mokrý; a potom by som mal opäť zvýšiť pravdepodobnosť, že pršalo...

Judea Pearl použil ako metaforu algoritmus, ako sa počítajú vojaci v rade. Predstavte si, že stojíte v rade a vidíte dvoch susedných vojakov, jedného pred vami a jedného za vami. To ste dokopy traja vojaci. Opýtate sa vojaka vedľa vás: “Kolkých vojakov vidíš ty?” Obzrie sa a povie: “Troch.” Tak to máme dokopy šesť vojakov. Očividne toto *nie* je správny spôsob, ako to robiť.

Rozumnejšie je opýtať sa vojaka pred vami: “Koľko vojakov je pred tebou?” a vojaka za vami: “Koľko vojakov je za tebou?” Otázka: “Koľko vojakov je pred tebou?” sa dá ďalej odovzdať bezo zmätku. Ak som na začiatku radu, poviem výsledok: “1 vojak vpredu”. Osoba stojaca hneď za mnou dostane správu: “1 vojak vpredu” a odovzdá výsledok: “2 vojaci vpredu” vojakovi za ním. Podobne, každý vojak, ktorý dostane správu: “N vojakov vzadu” od vojaka tesne za ním, ju odovzdá ako: “N+1 vojakov vzadu” vojakovi pred ním. Koľko je vojakov celkovo? Spočítajte dve čísla, ktoré dostanete, pridajte jednotku za seba: toľko je v rade celkovo vojakov.

Pointa je, že každý vojak musí *oddelene* sledovať dve správy, správu o vojakoch vpredu a správu o vojakoch vzadu, a sčítat ich až na záver. Nikdy nepridávate vojakov z prijatej správy o vojakoch vzadu do správy o vojakoch vpredu, ktorú odovzdáte dozadu. A vôbec, celkové množstvo vojakov sa nikdy neodovzdáva ako správa – nikto ho nikdy nepovie nahlas.

Analogický princíp funguje pri dôkladnom pravdepodobnostnom uvažovaní o kauzalite. Ak sa niečo dozviete o tom, či pršalo, z nejakého *iného* zdroja než pozorovania, že chodník je mokrý, toto pošle správu dopredu z [dážď] do [chodník je mokrý] a zvýši naše očakávanie, že chodník je mokrý. Ak pozorujete, že chodník je mokrý, toto pošle správu dozadu nášmu názoru, že pršalo, a táto správa pokračuje z [dážď] do všetkých susedných uzlov *okrem* uzla [chodník je mokrý]. Každý kus indicie započítame presne raz; žiadna správa sa “neodráža” dopredu a dozadu. Presne tento algoritmus nájdete v klasickej knihe Judeu Pearl: “Pravdepodobnostné uvažovanie v inteligentných systémoch: Siete s dôveryhodným odvodzovaním.”⁷

Čo bolo teda zlé na teórii flogistonu? Keď pozorujeme, že oheň je horúci, uzol [oheň] môže poslať správu dozadu do uzla [„flogiston“], čím aktualizujeme svoj názor o flogistone. Ale ak sa stalo toto, nemôžeme to považovať za úspešnú predpoveď teórie flogistonu smerom dopredu. Správa by mala ísť iba jedným smerom a neodrážať sa naspäť.

Žiaľ, ľudia nepoužívajú presné algoritmy na aktualizáciu sietí názorov. Učíme sa o rodičovských uzloch z pozorovania potomkovských uzlov a predpovedáme potomkovské uzly podľa názoroch o rodičovských. Pritom však prísne neoddeľujeme záznamy o správach dozadu a správach dopredu. Pamätáme si len, že flogiston je horúci, čo *spôsobuje*, že oheň je horúci. Takže sa zdá, že teória flogistonu predpovedá horúcosť ohňa. Ešte horšie, vyzerá to, že *flogiston spôsobuje, že oheň je horúci*.

Dokiaľ si nevšimnete, že sa tu nerobia žiadne predpovede *dopredu*, nebudete tento neobmedzujúci kauzálny uzol považovať za “falošný”. Je zakreslený úplne rovnakým spôsobom ako hocikáky iný uzol vo vašej sieti názorov. Vyzerá ako fakt, ako všetky ostatné fakty, ktoré poznáte: *Flogiston spôsobuje, že oheň je horúci*.

Správne navrhnutá UI by si tento problém všimla hneď. Nevyžadovalo by to ani špeciálny kód na tento účel, iba správne záznamy o sieti názorov. (Žiaľ, my ľudia

⁷ http://books.google.sk/books/about/Probabilistic_Reasoning_in_Intelligent_S.html?id=AvNID7LyMusC

nedokážeme prepísať svoj vlastný kód tak, ako by to dokázala správne navrhnutá UI.)

Hovoriť o “skreslení spätného pohľadu⁸” je iba netechnický spôsob, ako povedať, že ľudia presne nerozdeľujú správy dopredu a správy dozadu, čím umožnia, aby správy dopredu boli kontaminované správami dozadu.

Tí, čo kedysi dávno šli cestou flogistonu, sa nesnažili byť hlúpi. Žiaden vedec sa úmyselne nepokúša zaseknúť v slepej uličke. Sú aj vo *vašej* hlave nejaké falošné vysvetlenia? Ak sú, ručím vám za to, že nie sú označené ako “falošné vysvetlenia”, takže hľadanie kľúčového slova “falošný” vo vašich myšlienkach ich neodhalí.

Vďaka “skresleniu spätného pohľadu⁹” takisto nestačí skontrolovať, ako dobre vaša teória “predpovedá” fakty, ktoré už poznáte. Musíte predpovedať zajtrajšok, nie včerajšok. Je to jediný spôsob, ako zamotaná ľudská myseľ môže mať istotu, že posielala čistú správu dopredu.

Pôvodný článok s komentármi¹⁰

⁸ pozri: 30 Spätný pohľad znehodnocuje vedu

⁹ pozri: 30 Spätný pohľad znehodnocuje vedu

¹⁰ http://lesswrong.com/lw/is/fake_causality/

Kapitola 35. Sémantické stopky

A dieťa sa opýtalo:

Odkiaľ pochádza tento kameň?

Odstiepil som ho z veľkého balvanu v strede osady.

Odkiaľ pochádza ten balvan?

Asi sa skotúlal z vysokej hory, ktorá sa týči nad našou dedinou.

Odkiaľ pochádza tá hora?

Rovnako ako všetky skaly: sú to kosti Ymira, prvotného obra.

Odkiaľ pochádza prvotný obor Ymir?

Z veľkej priepasti Ginnungagap.

Odkiaľ pochádza veľká priepasť Ginnungagap?

Túto otázku sa nikdy nepýtaj.

Vezmime si zdanlivý paradox Prvej Príčiny. Veda vytrasovala udalosti späť k Veľkému Tresku, ale prečo sa stal samotný Veľký Tresk? Dá sa pekne odpovedať, že samotný čas začal v okamihu Veľkého Tresku – že pred Veľkým Treskom nie je žiaden tok minút a hodín. Ale aj táto odpoveď predpokladá naše fyzikálne zákony, ktoré samotné majú zložitú štruktúru; to si žiada vysvetlenie. Odkiaľ pochádzajú fyzikálne zákony? Mohli by ste povedať, že sme všetci v počítačovej simulácii, ale potom táto počítačová simulácia musí bežať na fyzikálnych zákonoch nejakého iného sveta – a odkiaľ pochádzajú *tie* fyzikálne zákony?

Vtedy niektorí ľudia povedia: “Boh!”

Ako by si mohol niekto, hoci aj veľmi nábožný človek, pomyslieť, že toto *pomohlo* odpovedať na paradox Prvej Príčiny? Prečo sa automaticky neopýtať: “Odkiaľ pochádza Boh?” Povedať: “Boh je bez príčiny” alebo “Boh stvoril sám seba” nás necháva v celkom rovnakej pozícii ako: “Čas začal pri Veľkom Tresku.” Akurát sa opýtame, prečo vôbec existuje celý ten metasystém alebo prečo niektoré udalosti môžu byť bez príčiny a iné nie.

Mojím cieľom tu nie je diskutovať o zdanlivom paradoxe Prvej Príčiny, ale pýtať sa, ako si niekto môže myslieť, že “Boh!” *rieši* tento paradox. Povedať “Boh!” je spôsob, ako patriť do svojho kmeňa¹, čo dáva ľuďom motív hovoriť to tak často, ako sa len dá – niektorí ľudia to dokonca odpovedia aj na otázku: “Prečo tento hurikán zasiahol New Orleans?” Napriek tomu by som dúfal, že si ľudia všimnú, že pri *konkrétnej* hádanke o Prvej Príčine povedať “Boh!” nepomáha. Tento paradox vďaka tomu nevyzerá o nič menej paradoxne, *aj keby to bola pravda*. Ako je možné, že si to niekto *nevšímne*?

Jonathan Wallace naznačuje, že “Boh!” slúži ako *sémantická stopka* – že to nie je výrokové tvrdenie, ale skôr kognitívna dopravná značka: nerozmýšľajte za tento bod. Povedať “Boh!” *nerieši* tento paradox, ale vysiela kognitívny dopravný signál na zastavenie zrejmejšieho pokračovania reťaze otázok a odpovedí.

Samozrejme, vy ako dobrý a poriadny ateista by ste to nikdy neurobili, však? Lenže “Boh!” nie je *jediná* sémantická stopka, je to len zrejmy prvý príklad.

¹ pozri: 15 Vyznávanie a fandanie

Transhumánne technológie - molekulárna nanotechnológia, pokročilá biotechnológia, umelá inteligencia, a tak ďalej - nastolujú ťažké politické otázky. Akú rolu, ak vôbec nejakú, by mala hrať vláda pri dozore nad rodičovským výberom génov pre svoje dieťa? Majú mať rodičia možnosť úmyselne vybrať gén pre schizofréniu? Ak je zvýšenie inteligencie dieťaťa drahé, mala by vláda pomôcť zabezpečiť jeho dostupnosť, aby zabránila vzniku kognitívnych elít? Môžete navrhovať rôzne inštitúcie, ktoré budú odpovedať na tieto politické otázky - napríklad, že súkromné nadácie by mali poskytovať finančnú pomoc na zvyšovanie inteligencie - ale samozrejme ďalšia otázka je: "Bude táto inštitúcia efektívna?" Ak sa budeme spoliehať, že právna zodpovednosť za výrobok zabráni firmám vytvoriť škodlivú nanotechnológiu, bude to naozaj *fungovať*?

Poznám človeka, ktorého odpoveď na každú z týchto otázok je: "Liberálna demokracia!" A to je celé. To je jeho odpoveď. Ak mu položíte samozrejmu otázku ako: "Ako veľmi sa liberálnym demokraciám historicky darilo pri rovnako zložitých problémoch?" alebo "Čo ak liberálna demokracia urobí niečo hlúpe?", potom ste autokrat, libertopían, alebo iný druh veľmi veľmi zlého človeka. Nikto nesmie spochybňovať demokraciu.

Raz som tento spôsob uvažovania nazval: "božské právo demokracie". Ale presnejšie by bolo povedať, že "Demokracia!" preňho funguje ako sémantická stopka. Keby mu niekto povedal: "Prenechajme to firme Coca Cola!", sám by položil samozrejme ďalšie otázky: "Prečo? Čo s tým urobí firma Coca Cola? Prečo by sme jej mali dôverovať? Darilo sa jej v minulosti pri rovnako zložitých problémoch?"

Alebo predpokladajme, že by niekto povedal: "Američania mexického pôvodu intrigujú ako odstrániť všetok kyslík zo zemskej atmosféry." Pravdepodobne by ste sa opýtali: "Prečo by *toto* robili? Nemusia vari aj Američania mexického pôvodu dýchať? Fungujú vôbec Američania mexického pôvodu ako jednotná konšpirácia?" Ak si nepoložíte tieto samozrejme nasledujúce otázky keď niekto povie: "Korporácie intrigujú ako odstrániť kyslík", potom je "Korporácie!" vaša sémantická stopka.

Dajte si tu pozor, aby ste nevytvorili nový všeobecný protiargument na veci, ktoré sa vám nepáčia: "Ach, to je iba stopka!" Žiadne slovo nie je samo osebe stopkou; otázka je, či má dané slovo takýto účinok na konkrétnu osobu. Mať voči niečomu silné emócie² z toho ešte nerobí stopku. Ja nemám príliš rád teroristov, ani sa nebojím súkromného vlastníctva, to ale neznamená, že "Teroristi!" alebo "Kapitalizmus!" sú pre mňa kognitívne dopravné značky. (Slovo "inteligencia" na mňa kedysi malo takýto účinok, ale už nie.) Poznávacím znamením sémantickej stopky je *neschopnosť zamyslieť sa nad samozrejmom ďalšou otázkou*.

Pôvodný článok s komentármi³

² pozri: 2 Rozumné čítanie

³ http://lesswrong.com/lw/it/semantic_stopsigns/

Kapitola 36. Tajomné odpovede na tajomné otázky

Predstavte si, že sa pozeráte na svoju ruku a neviete nič o bunkách, nič o biochémií, nič o DNA. Naučili ste sa čosi z anatómie pri pitve, takže viete, že vaša ruka obsahuje svaly; ale neviete, prečo sa svaly hýbu namiesto toho, aby iba ležali ako hlina. Vaša ruka je skrátka... vec... a z nejakého dôvodu sa hýbe podľa vašich rozhodnutí. Nie je to kúzlo?

Živočíšne telo nefunguje ako termodynamický stroj ... vedomie učí každého jednotlivca, že je do istej miery podriadený rozhodnutiam svojej vôle. Zdá sa teda, že pohyblivé tvory majú schopnosť okamžite vplývať na isté pohyblivé častice hmoty vo svojich telách silami, ktoré usmerňujú tieto častice, aby vykonali výsledný mechanický účinok... Pôsobenie živočíšneho alebo rastlinného života na hmotu je nekonečne ďaleko za hranicou akéhokoľvek doposiaľ podniknutého vedeckého bádania. Jeho schopnosť usmerňovania pohybov pohyblivých častíc, v každodenne pozorovanom zázraku našej ľudskej slobodnej vôle, i v raste generácie za generáciou rastlín z jediného semena, sa nekonečne odlišuje od ľubovoľného možného výsledku náhodného súbehu atómov... Moderní biológovia opäť začali čosi akceptovať a je to princíp života.

—Lord Kelvin

Toto bola teória *vitalizmu*; že tajomný rozdiel medzi živou a neživou hmotou vysvetľuje *elan vital* alebo *vis vitalis*. Elan vital naplňal živú hmotu a spôsobil, že sa hýbe tak, ako určuje vedomie. Elan vital sa zúčastňoval na chemických reakciách, ktoré púhe neživé častice nemohli podstúpiť - neskoršia Wöhlerova syntéza urey, zložky moču, bola veľkým úderom vitalistickej teórii, pretože ukázala, že púha *chémia* dokáže zopakovať produkt biológie.

Nazývať "elan vital" vysvetlením, hoci aj falošným vysvetlením¹ ako flogiston², pravdepodobne znamená brať ho príliš vážne. Fungoval v prvom rade ako zastavovač zvedavosti³. Opýtali ste sa: "Prečo?" a odpoveď znela: "Lebo elan vital!"

Keď ste povedali "Elan vital!", *znelo* to, akoby ste vedeli, prečo sa vaša ruka hýbe. Mali ste v hlave malý kauzálny diagram⁴, ktorý hovoril ["Elan vital!"] -> [ruka sa hýbe]. V skutočnosti ste však nevedeli o nič viac než predtým. Nevedeli ste napríklad povedať, či vaša ruka bude vytvárať teplo alebo pohlcovať teplo, pokiaľ ste si tento fakt už predtým nepozreli; ak nie, nevedeli ste ho predpovedať. Vaša zvedavosť vyzerala nasýtená, nebola však nakrmená. Keďže môžete povedať: "Prečo? Lebo elan vital!" na ľubovoľné možné pozorovanie, bolo to rovnaké dobré na vysvetlenie všetkých výsledkov, skrytá hypotéza maximálnej entropie, a tak ďalej.

Väčšie ponaučenie sa však skrýva za úctou vitalistov voči elan vital, ich ochota vyhlasovať ho za tajomstvo mimo všetku vedu. Pri stretnutí s drakom Neznámom, vitalisti nevytasili meče do útoku, ale podriadené sklonili hlavy. Boli hrdí⁵ na

¹ pozri: 31 Falošné vysvetlenia

² pozri: 34 Falošná kauzalita

³ pozri: 35 Sémantické stopky

⁴ pozri: 34 Falošná kauzalita

⁵ pozri: 307 Tsuyoku naritai! (Chcem sa stať silnejším)

svoju nevedomosť, urobili z biológie *posvätné* tajomstvo a preto boli neochotní vzdať sa svojej nevedomosti⁶, keď indície začali klopať na dvere.

Tajomstvo Života bolo *nekonečne ďaleko za hranicou vedy!* Nie iba *trochu* ďaleko, podotýkam, ale *nekonečne* ďaleko! Lord Kelvin bol iste mimoriadne emocionálne vzrušený z toho, že *niečo nevedel*.

Avšak nevedomosť existuje iba na mape, nie na území. Ak nerozumiem nejakému javu, je to fakt o mojom stave mysle, nie fakt o samotnom jave. Nejaký jav môže *prípadat'* tajomný nejakej konkrétnej osobe. Neexistujú javy, ktoré sú tajomné sami osebe. Uctievať jav, pretože vyzerá tak zázračne tajomný, znamená uctievať svoju vlastnú nevedomosť.

Vitalizmus mal s flogistonom spoločnú chybu *zabalenia tajomstva ako podstaty*. Oheň bol tajomný a teória flogistonu zabalila toto tajomstvo do tajomnej podstaty nazvanej "flogiston". Život bol posvätným tajomstvom, a vitalizmus zabalil toto tajomstvo do tajomnej podstaty nazvanej "elan vital". Ani jedno z toho nepomohlo modelu sústrediť hustotu pravdepodobnosti⁷ – urobiť niektoré výsledky ľahšie vysvetliteľnými než iné. Toto "vysvetlenie" iba zabalilo otázku do malej tvrdej nepriehľadnej čiernej guľičky.

V Molierovej komédii lekár vysvetľuje schopnosť uspávadla tým, že obsahuje "dormitívnu potenciú". Ten istý princíp. Je chybou ľudskej psychológie, že keď sa stretáme s tajomným javom, skôr predpokladáme tajomné základné podstaty než zložitý proces za tým.

Ale ešte hlbšou chybou je predpokladať, že *odpoveď* môže byť tajomná. Ak sa nejaký jav zdá tajomný, je to fakt o našom stave poznania, nie fakt o samotnom jave. Vitalisti videli tajomnú medzeru vo svojich vedomostiach, tak predpokladali tajomnú vec, ktorá túto medzeru zapĺňa. Keď tak robili, pomýlili si mapu s územím. Všetok zmätok existuje v mysli, nie v zabalených podstatách.

Toto je konečné a celkom všeobecné vysvetlenie, prečo sú v histórii ľudstva ľudia znovu a znovu šokovaní objavom, že neuveriteľne tajomná otázka má netajomnú odpoveď. Tajomnosť je vlastnosť otázok, nie odpovedí.

Preto nazývam teórie ako vitalizmus *tajomnými odpoveďami na tajomné otázky*.

Toto sú príznaky tajomných odpovedí na tajomné otázky:

- Po prvé, vysvetlenie je zastavovačom zvedavosti⁸, nie ovládačom očakávania⁹.
- Po druhé, hypotéza sa neskladá z pohyblivých častí – model nie je konkrétny zložitý mechanizmus, ale prázdna nehybná podstata či sila. Tajomná podstata alebo tajomná sila môže údajne byť tu alebo tam, spôsobiť¹⁰ toto alebo tamto; ale dôvod, prečo sa tajomná sila takto správa, je zahalený v nepreniknuteľnej prázdnote.
- Po tretie, tí, ktorí ponúkajú toto vysvetlenie, opatrujú svoju nevedomosť¹¹; hovoria hrdo o tom, ako tento jav vzdoruje bežnej vede alebo nie je ako púhe všedné javy.
- Po štvrté, *aj potom, čo dostanete odpoveď, daný jav je stále tajomstvom* a má takú istú zázračnú nevysvetliteľnosť akú mal na začiatku.

⁶ pozri: 258 Dvanásť cností rozumnosti

⁷ pozri: 25 Zameraj svoju neistotu

⁸ pozri: 35 Sémantické stopky

⁹ pozri: 11 Nech názory platia najomné (v očakávaných vnemoch)

¹⁰ pozri: 34 Falošná kauzalita

¹¹ pozri: 258 Dvanásť cností rozumnosti

Pôvodný článok s komentármi¹²

¹² http://lesswrong.com/lw/iu/mysterious_answers_to_mysterious_questions/

Kapitola 37. Márnosť emergencie

Súvisí s kapitolou: Viera vo vieru¹

Zlyhania flogistonu² a vitalizmu³ sú historickým⁴ spätným pohľadom⁵. Odvážim sa ísť s kožou na trh a pomenovať nejakú *súčasnú* teóriu, ktorú považujem za analogicky pomýlenú?

Menujem *emergenciu* alebo *emergentné javy* - zvyčajne definované ako štúdium systémov, ktorých správanie na vyššej úrovni vzniká alebo "sa vynára" z interakcie mnohých prvkov na nižšej úrovni. (Wikipédia⁶: "spontánny vznik makroskopických vlastností a štruktúr zložitých systémov, ktoré nie je ľahké odvodiť z vlastností ich zložiek".) Doslovne vzaté, tento popis sedí na každý jav v našom vesmíre nad úrovňou jednotlivých kvarkov, a to je časť problému. Predstavte si, že ukážete na krach burzy a poviete: "Toto nie je kvark!" Znie vám to ako vysvetlenie? Nie? Potom by nemalo ani: "Je to emergentný jav!"

Je to podstatné meno "emergencia", voči čomu protestujem, nie voči slovesu "vynárať sa" [anglicky: "emerge"]. Nie je nič zlé na povedaní "X sa vynára z Y", kde Y je nejaký konkrétny podrobný model s vnútornými pohyblivými časťami. "Vyplýva z" je ďalšie legitímne slovné spojenie, ktoré znamená presne to isté: Gravitácia vyplýva zo zakrivenia časopriestoru podľa konkrétneho matematického modelu všeobecnej relativity. Chémia vyplýva z interakcie medzi atómami podľa konkrétneho modelu kvantovej elektrodynamiky.

Teraz si predstavte, že by som povedal, že gravitácia je vysvetlená "vyplývavosťou" alebo že chémia je "vyplývavostný jav" a tvrdil, že toto je moje vysvetlenie.

Slovné spojenie "vynára sa z" je prijateľné rovnako ako "vyplýva z" alebo "je spôsobené" sú prijateľné, ak za týmto slovným spojením nasleduje nejaký konkrétny model, ktorý hodnotíme podľa jeho vlastných kvalít.

To však *nie* je spôsob, ako sa "emergencia" bežne používa. "Emergencia" sa bežne používa ako plnohodnotné vysvetlenie samo osebe.

Už si ani nepamätám, kolkokrát som som počul ľudí povedať napríklad: "Inteligencia je emergentný jav!" akoby toto vysvetľovalo inteligenciu. Toto použitie spĺňa všetky body zoznamu pre tajomnú odpoveď na tajomnú otázku⁷. Čo ste sa dozvedeli, keď si poviete, že inteligencia je "emergentná"? Nemôžete robiť žiadne nové predpovede. Neviete nič o správaní mysli v skutočnom svete, čo ste nevedeli predtým. Znie to, akoby ste verili v nový fakt, ale neočakávate žiadne odlišné výsledky. Vaša zvedavosť vyzerá nasýtená, nie však nakrmená. Táto hypotéza nemá žiadne pohyblivé časti - nemá podrobný vnútorný model, ktorým sa dá pohybovať. Tí, ktorí ponúkajú hypotézu "emergencie", vyznávajú svoju nevedomosť ohľadom vnútorných častí a sú na ňu hrdí; stavajú vedu "emergencie" do protikladu k ostatným vedám všedných vecí.

A aj potom, čo dostanete odpoveď: "Prečo? Lebo emergencia!", *daný jav stále zostáva tajomný* a má tú istú posvätnú nepreniknuteľnosť, ako mal na začiatku.

¹ pozri: 13 Viera vo vieru

² pozri: 34 Falošná kauzalita

³ pozri: 36 Tajomné odpovede na tajomné otázky

⁴ pozri: 30 Spätný pohľad znehodnocuje vedu

⁵ http://lesswrong.com/lw/il/hindsight_bias/

⁶ <http://cs.wikipedia.org/wiki/Emergence>

⁷ pozri: 36 Tajomné odpovede na tajomné otázky

Zábavné cvičenie je vynechať prídavné meno “emergentný” z každej vety, v ktorej sa nachádza a pozrieť, či tá veta hovorí niečo iné.

- *Predtým:* Ľudská inteligencia je emergentným výsledkom signalizácie neurónov.
- *Potom:* Ľudská inteligencia je výsledkom signalizácie neurónov.
- *Predtým:* Správanie kolónie mravcov je emergentným výsledkom interakcií mnohých jednotlivých mravcov.
- *Potom:* Správanie kolónie mravcov je výsledkom interakcií mnohých jednotlivých mravcov.
- *Ešte lepšie:* Kolónia sa skladá z mravcov. Dokážeme úspešne predpovedať niektoré vlastnosti správania kolónie použitím modelov, ktoré zahŕňajú iba jednotlivé mravce, bez globálnych premenných pre kolóniu, čo ukazuje, že rozumieme, ako toto správanie kolónie vzniká zo správania mravcov.

Ďalšie zábavné cvičenie je nahradiť slovo “emergentný” slovom⁸, ktoré ľudia pôvodne používali ako vysvetlenie⁹ pred vynálezom emergencie:

- *Predtým:* Život je emergentný jav.
- *Potom:* Život je zázračný jav.
- *Predtým:* Ľudská inteligencia je emergentným výsledkom signalizácie neurónov.
- *Potom:* Ľudská inteligencia je zázračným výsledkom signalizácie neurónov.

Nesprostredkuje vari každé tvrdenie celkom rovnaké množstvo vedomostí o správaní daného javu? Nezodpovedá vari každá hypotéza celkom rovnakej množine výsledkov¹⁰?

Slovo “emergencia” sa stalo veľmi obľúbené, rovnako ako slovo “zázrak” bolo kedysi obľúbené. Slovo “emergencia” má pre ľudskú psychológiu hlbokú príťažlivosť, z toho istého dôvodu. Slovo “emergencia” je úžasne ľahkým vysvetlením a znie dobre, keď sa povie; dáva vám posvätné tajomstvo¹¹, ktoré môžete uctievať. Emergencia je obľúbená, pretože je to odpadové jedlo pre zvedavosť. Pomocou emergencie dokážete vysvetliť čokoľvek a presne to aj ľudia robia; lebo je to úžasný pocit vedieť vysvetliť veci. Ľudia sú stále ľuďmi, aj keď absolvujú pár hodín vedy na vysokej. Keď raz nájdú spôsob, ako sa striasť okov¹² ustálenej vedy, pustia sa do rovnakých machinácií ako ich predkovia, oblečú ich do literárneho žánra “vedy”¹³, ale je to stále psychológia toho istého živočíšneho druhu.

Pôvodný článok s komentármi¹⁴

⁸ pozri: 32 Hádanie učiteľovho hesla

⁹ pozri: 31 Falošné vysvetlenia

¹⁰ pozri: 27 Tvoja sila racionalistu

¹¹ pozri: 36 Tajomné odpovede na tajomné otázky

¹² pozri: 258 Dvanásť cností rozumnosti

¹³ pozri: 17 Názor ako uniforma

¹⁴ http://lesswrong.com/lw/iv/the_futility_of_emergence/

Kapitola 38. Nehovorte “zložitosť”

Kde bolo, tam bolo...

Toto je príbeh môjho prvého stretnutia s Marcellom, s ktorým som neskôr rok pracoval na teórii UI; v tej dobe som ho však ešte neprijal ako učňa. Vedel som, že súťažil na celoštátnej úrovni matematických a počítačových olympiád, čo stačilo na upútanie mojej pozornosti; nevedel som však, či sa dokáže naučiť rozmýšľať o UI.

Požiadal som Marcella, aby povedal, ako si myslí, že by UI mohla objaviť spôsob riešenia Rubikovej kocky. Nie vopred naprogramovaným spôsobom, čo je triviálne, ale skôr ako by UI mohla sama objaviť zákony Rubikovho sveta a zistiť, ako ich využívať. Ako by UI mohla *sama vynájsť* pojmy ako “operátor” alebo “makro”, ktoré sú kľúčom na vyriešenie Rubikovej kocky?

V nejakom bode diskusie Marcello povedal: “No, myslím si, že UI potrebuje mať zložitosť, aby vedela urobiť X a zložitosť, aby vedela urobiť Y...”

A ja som povedal: “Nehovor ‘zložitosť’.”

Marcello: “Prečo nie?”

Povedal som: “Zložitosť by nikdy nemala byť cieľom sama osebe. Môžeš potrebovať použiť nejaký konkrétny algoritmus, ktorý pridáva určité množstvo zložitosti, ale samotná zložitosť pre zložitosť veci iba komplikuje.” (Myslel som pritom na všetkých ľuďoch, ktorých som počul obhajovať názor, že internet sa “prebudí” a stane sa UI, keď sa stane “dostatočne zložitým”.)

A Marcello povedal: “Ale musí existovať *nejaké* množstvo zložitosti, ktoré to dokáže.”

Na chvíľu som zavrel oči a skúšal som rozmýšľať, ako to celé vysvetliť slovami. Mne osobne povedať “zložitosť” jednoducho *prípadalo* ako nesprávny krok v tanci UI. Nikto nedokáže rozmýšľať dosť rýchlo na to, aby slovné zväzil každú vetu vo svojom prúde vedomia; to by si vyžadovalo nekonečnú rekurziu. Uvažujeme v slovách, ale náš prúd vedomia je usmerňovaný pod úrovňou slov, nacvičenými zvyškami minulých vhládov a drsných skúseností...

Povedal som: “Čítal si Technické vysvetlenie technického vysvetlenia¹?”

“Áno,” povedal Marcello.

“Dobre,” povedal som. “Povedať ‘zložitosť’ nekoncentruje tvoju masu pravdepodobnosti.”

“Aha,” povedal Marcello, “ako ‘emergencia²’. Ha. Takže... by som teraz mal rozmýšľať o tom, ako by sa X mohlo naozaj stať...”

A vtedy som si pomyslel: “*Tento je možno vzdelávateľný.*”

Zložitosť nie je zbytočný pojem. Sú k nemu matematické definície, ako Kolmogorovská zložitosť alebo Vapnikovská-Červonenkisovská zložitosť. Dokonca aj na intuitívnej úrovni sa často oplatí myslieť na zložitosť – potrebujete

¹ pozri: 292 Technické vysvetlenie technického vysvetlenia

² pozri: 37 Márnosť emergencie

zvážiť zložitost hypotézy a rozhodnúť sa, či je "príliš zložitá" vzhľadom na podporujúce indície alebo pozrieť na dizajn a skúsiť ho zjednodušiť.

Ale pojmy nie sú užitočné alebo neužitočné sami osebe. Iba ich *použitie* je správne alebo nesprávne. V tanečnom kroku, ktorý sa Marcello pokúšal urobiť, sa snažil dostať vysvetlenie zadarmo, dostať niečo za nič. Je to mimoriadne častý chybný krok, prinajmenšom v mojej oblasti. Môžete sa zapojiť do diskusie o umelej všeobecnej inteligencii a sledovať ľudí, ktorí robia to isté, naľavo i napravo, znova a znova - nepretržite preskakujú veci, ktorým nerozumejú, neuvedomujú si, čo práve robia.

Stane sa to v okamihu: umiestnite neovládajúci kauzálny uzol³ za čosi tajomné; kauzálny uzol, ktorý vyzerá ako vysvetlenie⁴, ale nie je. Táto chyba sa vyskytuje pod úrovňou slov. Nevyžaduje si žiadnu mimoriadnu charakterovú vadu; je spôsob, ako ľudia štandardne⁵ myslia už od dávnych čias.

To, čomu sa musíte vyhnúť, je *preskočenie tej tajomnej časti*; musíte zastať pri tajomstve a čeliť mu priamo. Existuje veľa slov, ktorými sa tajomstvá dajú preskakovať a niektoré z nich by boli legitímne v inom kontexte - napríklad "zložitost". Ale podstatou chyby je toto *preskočenie*, bez ohľadu na to, aký kauzálny uzol je za ním. Toto preskočenie nie je myšlienka, iba mikromyšlienka. Musíte venovať dôkladnú pozornosť, aby ste sa pri nej prichytili. A keď sa nacvičíte vyhýbať preskakovaniu, stane sa to vecou inštinktu, nie slovného uvažovania. Musíte *cítiť*, ktoré časti vašej mapy sú stále prázdne, a čo je dôležitejšie, venovať tomuto pocitu pozornosť.

Predpokladám, že v akadémii je obrovský tlak na zametanie problémov pod koberec, aby ste mohli prezentovať článok s dojmom úplnosti. Viac slávy zožnete za napohľad úplný model, ktorý zahŕňa pár "emergentných javov"⁶, než za explicitne neúplnú mapu s nápismi "fakt neviem, ako táto časť funguje" alebo "a potom sa stane zázrak". V časopise by vám možno ten druhý článok ani neprijali, pretože ktovie, či tie neznáme kroky nie sú vlastne tie, kde sa všetko to zaujímavé deje. A áno, občas sa stane, že všetky tie nezázračné časti vašej mapy sa ukážu ako nedôležité. To je cena, ktorú občas platíte za vstup na neznáme územie a snahu riešiť problémy *postupne*. Ale v tom prípade je *ešte dôležitejšie vedieť*, kedy ešte nie ste na konci. Väčšinou sa ľudia na neznáme územie neodvažujú vstúpiť, pretože sa príliš boja, že premárnia svoj čas.

A ak zakladáte revolučnú novú firmu zameranú na UI,⁷ je tam ešte väčší tlak zamiesť problémy pod koberec; inak si musíte priznať⁸, že zatiaľ neviete, ako zostrojiť UI a vaše súčasné životné plány sa rozsypú. Ale možno to teraz zbytočne zložito vysvetlujem⁸, veď preskakovanie je u ľudí normálne⁹; ak hľadáte príklady, sledujte ľudí, ktorí diskutujú o náboženstve alebo filozofii alebo duchovne alebo nejakej vede, v ktorej neboli profesionálne vycvičení.

Marcello a ja sme si pri našej práci na UI vytvorili zvyk: keď sme narazili na niečo, čomu sme nerozumeli, a to bolo dosť často, povedali sme "kúzlo" - napríklad: "X kúzelné urobí Y" - aby sme si sami pripomenuli, že *je tu nevyriešený problém, medzera v našom chápaní*. Je omnoho lepšie povedať "kúzlo" než "zložitost" alebo "emergencia"; tie druhé slová¹⁰ totiž vytvárajú ilúziu porozumenia. Múdrejšie je povedať "kúzlo" a nechať tak sám sebe značku, pripomienku práce, ktorú ešte bude treba dorobiť.

³ pozri: 34 Falošná kauzalita

⁴ pozri: 31 Falošné vysvetlenia

⁵ pozri: 32 Hádanie učiteľovho hesla

⁶ pozri: 37 Márnosť emergencie

⁷ pozri: 122 Dokážeš čeliť skutočnosti

⁸ pozri: 50 Chyba prisudzovania

⁹ pozri: 32 Hádanie učiteľovho hesla

¹⁰ pozri: 32 Hádanie učiteľovho hesla

Pôvodný článok s komentármi¹¹

¹¹ http://lesswrong.com/lw/ix/say_not_complexity/

Kapitola 39. Pozitívne skreslenie: Pozeraj do tmy

Učím v triede a píšem na tabuľu tri čísla: 2 - 4 - 6. "Myslím si pravidlo," hovorím, "ktoré sa týka trojíc čísel. Postupnosť 2 - 4 - 6, zhodou okolností, toto pravidlo spĺňa. Každý z vás má na lavici kôpku papierikov. Napíšte na papierik postupnosť troch čísel, a ja ju označím 'Áno', ak spĺňa toto pravidlo, alebo 'Nie', ak nespĺňa toto pravidlo. Potom môžete napísať ďalšiu trojicu čísel a opýtať sa, či tieto spĺňajú toto pravidlo a tak ďalej. Keď ste si istí, že poznáte to pravidlo, napíšte pravidlo na papierik. Môžete skúšať toľko trojíc, koľko len chcete."

Tu je záznam otázok jedného študenta:

4, 6, 2 - Nie

4, 6, 8 - Áno

10, 12, 14 - Áno

V tomto bode dotyčný študent napísal svoj odhad pravidla. Čo si vy myslíte, že je toto pravidlo? Chceli by ste vyskúšať ešte ďalšiu trojicu, a ak áno, ktorá by to bola? Pred ďalším čítaním sa na chvíľu zamyslite.

Uvedená úloha sa zakladá na klasickom experimente Petra Wasona, úloha 2 - 4 - 6. Hoci pokusné osoby v tejto úlohe zvyčajne vyjadrovali vysokú dôveru vo svoj odhad, iba 21 % študentov úspešne uhádlo experimentátorove skutočné pravidlo a neskoršie replikácie naďalej ukazujú mieru úspešnosti okolo 20 %.

Táto štúdia sa volá "O neúspešnom eliminovaní hypotéz v pojmovej úlohe" (*Quarterly Journal of Experimental Psychology*, 12: 129-140, 1960). Pokusné osoby, ktoré skúšali vyriešiť úlohu 2 - 4 - 6, zvyčajne skúšajú vygenerovať *pozitívne* príklady namiesto *negatívnych* príkladov - používajú hypotetické pravidlo na vytvorenie reprezentatívneho príkladu a sledujú, či bude označený "Áno".

Preto, keď si niekto vytvorí hypotézu "číslo sa zvyšuje o dve", otestuje trojicu 8 - 10 - 12, dozvie sa, že vyhovuje a sebaisto oznámi toto pravidlo. Keď si niekto vytvorí hypotézu $X - 2X - 3X$, otestuje trojicu 3 - 6 - 9, dozvie sa, že vyhovuje a potom ohlásí toto pravidlo.

V každom prípade je skutočné pravidlo to isté: tri čísla musia byť v rastúcom poradí.

Ale aby ste objavili toto, museli by ste vytvárať trojice, ktoré by *nemali* vyhovovať, napríklad 20 - 23 - 26 a sledovať, či budú tieto označené "Nie". Čo ľudia v tomto experimente nezvyknú robiť. V niektorých prípadoch si pokusné osoby vymyslia, "otestujú" a oznámia pravidlá omnoho komplikovanejšie, než je skutočná odpoveď.

Tento kognitívny jav je zvyčajne zahrnutý pod "sklon potvrdzovať". Mne sa však zdá, že jav testovania *pozitívnych* príkladov namiesto *negatívnych* by sme mali odlišiť od javu zachovávanía pôvodných názorov. Niekedy sa ako synonymum pre "sklon potvrdzovať" používa "pozitívne skreslenie", čo sa k tejto konkrétnej chybe hodí omnoho lepšie.

Kedysi sa zdalo, že teória flogistonu¹ dokáže vysvetliť, prečo plameň v uzavrenej nádobe zhasne (vzduch je presýtený flogistonom a už nedokáže prijať ďalší), ale

¹ pozri: 34 Falošná kauzalita

teória flogistonu by rovnako dobre vedela vysvetliť, prečo plameň *nezhasne*. Aby ste si toto všimli, musíte hľadať negatívne príklady namiesto pozitívnych, pozeraj na nuly namiesto jednotiek; čo ide proti tomu, čo experiment odhalil ako ľudský inštinkt.

Pretože podľa inštinktov ľudia žijú iba v polovici sveta.

Človeka možno učiť o pozitívnom skreslení celé dni a predsa ho v danej chvíli dokáže prehliadnuť. Pozitívne skreslenie nie je niečo, čo robíme kvôli logike alebo hoci len kvôli emocionálnej pripútanosti. Úloha 2 - 4 - 6 je "chladná", logická, nie emocionálne "horúca". A predsa je táto chyba sub-verbálna, na úrovni predstavivosti, inštinktívnej reakcie. Keďže problém nespočíva vo vytvorení vedomého pravidla, ktoré hovorí: "Mysli iba na pozitívne príklady", nemožno ho vyriešiť slovným poznaním: "Mali by sme myslieť aj na pozitívne aj na negatívne príklady." Ktoré príklady vám automaticky naskakujú do hlavy? Musíte sa naučiť, bez slov, nie cik, ale cak. Musíte sa naučiť cuknúť smerom k nule, namiesto od nej.

Už hodnú dobu píšem o tom, že sila hypotézy je v tom, čo *nedokáže* vyriešiť, nie čo *dokáže*² - že ak viete rovnako dobre vysvetliť hocijaký výsledok, máte nulové poznanie. Takže ak chcete zbadať vysvetlenie, ktoré nepomáha, nestačí pomyslieť na to, čo by dokázalo dobre vysvetliť - musíte hľadať aj výsledky, ktoré by *nedokázalo* vysvetliť, a toto je skutočná sila danej teórie.

Takže toto všetko som už povedal a potom som včera spochybnil užitočnosť "emergencie" ako pojmu³. Jeden diskutér uviedol supravodivosť a feromagnetizmus ako príklad emergencie. Odpovedal som, že nesupravodivosť a neferomagnetizmus sú tiež príkladmi emergencie a že práve v tom je háčik. Ale nebudem tohto diskutéra kritizovať! Napriek rozsiahlemu čítaniu o "sklone potvrdzovať", sám som si nevšimol "chyták" v úlohe 2 - 4 - 6, keď som o nej prvýkrát čítal. Je to subverbálna okamžitá reakcia, ktorú treba zmeniť výcvikom. Ja na sebe ešte stále pracujem.

Tak veľa z racionalistových schopností je pod úrovňou slov. Je to náročná práca pokúšať sa sprostredkovať toto Umenie pomocou článkov na blogu. Ľudia s vami budú súhlasiť, a potom, v nasledujúcej vete, nevedomky urobia niečo úplne opačné. Nestážujem sa! Jedným z hlavných dôvodov, prečo sem píšem, je pozorovanie, čo moje slová *nedokázali* sprostredkovať.

Hľadáte práve teraz pozitívne príklady pozitívneho skreslenia, alebo si šetríte časť svojho hľadania na to, čo by ste podľa pozitívneho skreslenia vidieť *nemali*? Pozeráte sa smerom do svetla alebo do tmy?

Pôvodný článok s komentármi⁴

² pozri: 27 Tvoja sila racionalistu

³ pozri: 37 Márnosť emergencie

⁴ http://lesswrong.com/lw/iw/positive_bias_look_into_the_dark/

Kapitola 40. Moja divoká a bezstarostná mladosť

Hovorí sa, že rodičia robia všetky tie veci, ktoré hovoria svojim deťom, aby ich nerobili, pretože práve preto vedia, že je lepšie ich nerobiť.

Kedysi dávno, v nepredstaviteľne vzdialenej minulosti, som bol oddaný Tradičný Racionalista, považoval som sa za zručného podľa ich merítok, nepoznal som však Bayesovu Cestu. Keď mladý Eliezer natrafil na tajomne vyzerajúcu otázku, prikázania Tradičnej Racionality mu nezabránili vymyslieť Tajomnú Odpoveď¹. Je to tá zďaleka najzahanbujúcejšia chyba v mojom živote a stále mnou mykne, keď na ňu pomyslím.

Čo bola moja tajomná odpoveď na tajomnú otázku? To nebudem popisovať, lebo by to bol dlhý a zložitý príbeh. Bol som mladý, púhy Tradičný Racionalista nepoznajúci učenie Tverskeho a Kahnemana. Poznal som Occamovu britvu, nie však klam konjunkcie². Myslel som, že mi prejde, keď budem rozmýšľať zložito, v literárnom štýle zložitých myšlienok, ktoré som čítal vo vedeckých knihách, nevedomujúc si, že správna zložitosť je možná len vtedy, keď je každý jej krok poriadne pripevnený. Dnes, jedna z hlavných rád, ktoré dávam aspirujúcim mladým racionalistom, je: “Nepokúšaj sa o dlhé reťaze úvah alebo zložité plány.”

Netreba povedať nič viac než toto: Aj potom, čo som vymyslel svoju “odpoveď”, mi bol daný jav stále tajomný³ a mal tú istú vlastnosť úžasnej nepreniknuteľnosti, ako mal na začiatku.

Nemýľte sa⁴, mladý Eliezer nebol hlúpy. Všetky chyby, ktorými sa mladý Eliezer previnil, dodnes robia vážení vedci vo vážných časopisoch. Aby sa pred nimi ochránil, potreboval by zručnosť presnejšiu, než akej sa kedy učil ako Tradičný Racionalista.

Vskutku, mladý Eliezer usilovne a svedomito nasledoval príkazy Tradičnej Racionality počas svojho zblúdenia z cesty.

Ako Tradičný Racionalista, mladý Eliezer dbal na to, aby jeho Tajomná Odpoveď odvážne predpovedala budúcu skúsenosť. Konkrétne, očakával som, že v budúcnosti neurológovia objavia, že neuróny využívajú kvantovú gravitáciu, ako hovoril Sir Roger Penrose. To si vyžadovalo, aby si neuróny udržiavali určitú úroveň kvantovej koherencie, a to bolo niečo, čo ste mohli hľadať a buď to nájsť alebo nenájsť. Buď niečo pozorujete alebo nie, však?

Moja hypotéza však nerobila žiadne *spätné* predpovede. Podľa Tradičnej Vedy, spätné predpovede sa nepočítajú – načo sa s nimi teda zaťažovať? Na druhej strane, pre Bayesiánca, ak *dnes* hypotéza nemá lepšiu podmienenú pravdepodobnosť než “neviem”, vyvoláva to otázku, prečo *dnes* veríte niečomu komplikovanejšiemu než “neviem”. Nepoznal som však Bayesovu Cestu, takže som nemyslel na podmienené pravdepodobnosti ani zameriavanie hustoty pravdepodobnosti. Urobil som Falzifikovateľnú Predpoveď; vari nie toto je Zákon?

Ako Tradičný Racionalista, mladý Eliezer si dával pozor, aby neveril v mágiu, mysticismus, uhlíkový šovinizmus, ani nič toho druhu. Hrdo som vyznával⁵ svoju

¹ pozri: 36 Tajomné odpovede na tajomné otázky

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

³ pozri: 36 Tajomné odpovede na tajomné otázky

⁴ pozri: 50 Chyba prisudzovania

⁵ pozri: 15 Vyznávanie a fandanie

Tajomnú Odpoveď: "Je to iba fyzika, rovnako ako zvyšok fyziky!" Akoby ste z mágie mohli urobiť niečo iné než kognitívny izomorf mágie, ak ju pomenujete⁶ kvantová gravitácia. Nepoznal som však Bayesovu Cestu a nevidel som úroveň⁷, na ktorej moja myšlienka bola izomorfná mágii. Prisahal som *vernosť* fyzike, ale to ma nezachránilo; čo už vie teória pravdepodobnosti o vernosti? Vyhýbal som sa všetkému, o čom mi Tradičná Racionalita povedala, že je zakázané, avšak čo zostalo, bola stále mágia.

Niet pochybnosti, že vernosť Tradičnej Racionalite mi pomohla dostať sa z jamy, ktorú som si sám vykopal. Keby som nebol Tradičným Racionalistom, mohol som byť *úplne* v kýbli. Ale Tradičná Racionalita stále nestačila na to, robiť veci *správne*. Iba ma viedla k iným chybám než boli tie, ktoré vyslovene zakazovala.

Keď rozmýšľam na tým, ako moje mladšie ja veľmi starostlivo nasledovalo pravidlá Tradičnej Racionality v procese získavania *nesprávnej* odpovede, vrhá to svetlo na otázku, prečo ľudia, ktorí si hovoria "racionalisti", nevládnu svetu⁸. Potrebujete *sakra veľké množstvo rozumnosti* než vás privedie k niečomu inému ako novým a zaujímavým chybám.

Tradičná Racionalita sa vyučuje skôr ako umenie než ako veda. Čítate si životopisy slávnych fyzikov opisujúcich lekcie, ktoré ich život naučil, a pokúšate sa robiť to, čo vám povedali, aby ste robili. Lenže vy ste nežili ich životy a polovica z toho, čo sa pokúšajú opísať, je inštinkt, ku ktorému boli vycvičení.

Tradičná Racionalita je zostrojená tak, že by bolo prijateľné, aby som strávil 30 rokov nad svojou hlúpou myšlienkou, pokiaľ ju jedného dňa úspešne falzifikujem, budem dosť poctivý k sebe samému ohľadom toho, čo moja teória predpovedala, prijmem jej vyvrátenie, keď príde, a tak ďalej. To stačí na to, aby západka Vedy cvakla vpred, ale je to trochu drsné voči ľuďom, ktorí premárnia 30 rokov svojho života. Tradičná Racionalita je chôdza, nie tanec. Je vytvorená, aby vás *jedného dňa* dostala k pravde a máte more času na privoniavanie ku kvetom pozdĺž cesty.

Tradiční Racionalisti sa môžu zhodnúť na tom, že sa nezhodnú. Tradiční Racionalisti nemajú ten *ideál*, že myslenie je presné umenie, v ktorom za daných indícií existuje iba jeden správny odhad pravdepodobnosti. V Tradičnej Racionalite máte povolené hádať a potom testovať vaše odhady. Skúsenosť ma však naučila, že ak niečo *neviete* a hádate, ukáže sa, že ste sa mýlili.

Bayesova Cesta je tiež nepresným umením, prinajmenšom v tej podobe, ako o nej rozprávam. Tieto články na blogu sú stále iba tápavými pokusmi vložiť do slov lekcie, ktoré by sa lepšie naučili skúsenosťou. Ale prinajmenšom je *založená* na matematike, plus experimentálnych indíciách z kognitívnej psychológie o tom, ako ľudia naozaj myslia. Možno to bude stačiť na dosiahnutie stratosfericky vysokej latky potrebnej na výcvik, ktorý vám umožní robiť to naozaj správne, namiesto púheho obmedzenia sa na nové zaujímavé chyby.

Pôvodný článok s komentármi⁹

⁶ pozri: 33 Veda ako uniforma

⁷ pozri: 31 Falošné vysvetlenia

⁸ pozri: 90 Vedieť o skresleniach môže ľuďom ublížiť

⁹ http://lesswrong.com/lw/iy/my_wild_and_reckless_youth/

Kapitola 41. Neučíme sa z histórie

Kde bolo, tam bolo, za mojej divokej a bezstarostnej mladosti¹, keď som nepoznal Bayesovu Cestu, dal som Tajomnú Odpoveď² na tajomne vyzerajúcu otázku. Bola to postupnosť mnohých chýb, ale jedna chyba mi vyčnieva ako najkritickejšia: Moje mladšie ja si neuvedomilo, že *vyriešené tajomstvo by malo pôsobiť menej mäťuco*. Pokúšal som sa vysvetliť Tajomný Jav – čo pre mňa znamenalo poskytnúť mu príčinu, zapasovať ho do prepojeného modelu skutočnosti. Prečo by toto malo urobiť daný jav menej Tajomným, ak je to jeho podstata? Pokúšal som sa *vysvetliť* Tajomný Jav, nie premeniť ho (nejakou nemožnou alchýmiou) na všedný jav, jav, ktorý by v prvom rade ani nevyzeral, že si vyžaduje nejaké nezvyčajné vysvetlenie.

Ako Tradičný Racionalista som poznal historické príbehy astrológov a astronómie, alchymistov a chémie, vitalistov a biológie. Ale tento Tajomný Jav bol iný. Bolo to niečo *nové*, niečo *zvláštnejšie*, niečo *zložitejšie*, niečo, čo bežná veda nedokázala vysvetliť celé stáročia...

...ako keby hviezdy a hmota a život neboli tajomstvami počas stoviek a tisícov rokov, od úsvitu ľudského myslenia až kým ich veda konečne nevyriešila...

Učíme sa o astronómii a chémii a biológii v školách a pripadá nám, že tieto veci *vždy* boli skutočným územím vedy, že *nikdy* neboli tajomné. Keď sa veda pokúsi zaútočiť na novú Velkú Skladačku, deti danej generácie sú skeptické, pretože nikdy nevideli, že by veda vysvetlila niečo, čo im *pripadalo* tajomné. Veda je dobrá iba na vysvetľovanie *vedeckých* tém, ako sú hviezdy a hmota a život.

Myslel som si, že lekciami histórie je, že astrológovia a alchymisti a vitalisti mali vrodenu charakterovú vadu³, sklon k mysticismu, ktorý ich viedol k tajomným vysvetleniam netajomných vecí. Avšak, keby nejaký jav *naozaj* bol veľmi čudný, mohlo by čudné vysvetlenie byť namieste?

Až dodatočne, keď som začal vidieť všednú štruktúru vnútri tajomstva, som si uvedomil, v koho topánkach som stál. Až vtedy som si uvedomil, ako rozumne znel *svojho času* vitalizmus, aká *prekvapujúca* a *zahanbujúca* bola odpoveď vesmíru: "Život je všedný a nevyžaduje si čudné vysvetlenie."

Čítame si históriu, ale *nežijeme* v nej, *neprežívame* ju. Kiež by som bol ja *osobne* predpokladal astrologické tajomstvá a potom objavil newtonovskú mechaniku, predpokladal alchymistické tajomstvá a potom objavil chémiu, predpokladal vitalistické tajomstvá a potom objavil biológiu. Napadla by mi moja Tajomná Odpoveď a povedal by som si: *Na toto sa už opäť nachytať nedám*.

Pôvodný článok s komentármi⁴

¹ pozri: 40 Moja divoká a bezstarostná mladost

² pozri: 36 Tajomné odpovede na tajomné otázky

³ pozri: 50 Chyba prisudzovania

⁴ http://lesswrong.com/lw/iz/failing_to_learn_from_history/

Kapitola 42. Sprístupniť históriu

Existuje myšlienkový zvyk, ktorý by som nazval *logickým omylom zovšeobecňovania fiktívnej indície*, a ten si jedného dňa zaslúži samostatný článok na blogu. Novinári, ktorí napríklad hovoria o filme *Terminátor* v spravodajstve o UI, zvyčajne neberú *Terminátora* ako proroctvo alebo nemennú pravdu. Ale tento film sa spomína – je dostupný¹ – akoby to bol ilustrujúci historický prípad. Akoby daný novinár videl, ako sa to odohralo na nejakej inej planéte, takže by sa to ľahko mohlo stať aj tu. Viac o tom je v kapitole 6 tohto článku².

Existuje aj opačná chyba k zovšeobecňovaniu fiktívnej indície: nedostatočné ovplyvnenie *historickou* indíciou. Problém zovšeobecňovania fiktívnej indície je, že je to fikcia – nikdy sa to naozaj nestalo. Nepochádza z rovnakej distribúcie ako náš skutočný svet; fikcia sa systematickým spôsobom odlišuje od skutočnosti³. Avšak história sa *stala a mala* by byť dostupná.

V našom pravekom prostredí neexistovali kiná; čo ste videli na vlastné oči, to bola pravda. Je teda niečo čudné na tom, že fikcie, ktoré vidíme v živo sa pohybujúcich obrázkoch, majú na nás priveľký dopad? Naopak, veci, ktoré sa *naozaj stali*, stretávame iba ako atrament na papieri; stali sa, ale my sme ich nikdy *nevideli* stať sa. Nepamätáme sa, že by sa stali nám.

Táto opačná chyba je považovať históriu za púhy príbeh, spracovávať ju rovnakou časťou mozgu ako román, ktorý čítate. Môžete ústami povedať, že je to “pravda” a nie “fikcia”, ale to neznamená, že vás to ovplyvní natoľko, ako by malo. Mnohé skreslenia obsahujú nedostatočné ovplyvnenie suchou, abstraktnou informáciou⁴.

Kde bolo, tam bolo, dal som Tajomnú Odpoveď⁵ na tajomnú otázku, neuviedomujúc si, že robím presne tú istú chybu ako astrológovia hľadajúci tajomné vysvetlenia pre hviezdy, alebo alchymisti hľadajúci magické vlastnosti hmoty, alebo vitalisti predpokladajúci, že nejasný “elan vital” vysvetlí celú biológiu.

Keď som si nakoniec uvedomil, v koho topánkach som stál⁶, bol to náhly šok z nečakaného spojenia s minulosťou. Uvedomil som si, že objav a skaza vitalizmu – o ktorých som čítal iba v knihách – sa *naozaj stali skutočným ľudom*, ktorí ich prežívali viacmenej tak, ako som ja prežíval objav a skazu svojej vlastnej tajomnej odpovede. A tiež som si uvedomil, že keby som bol naozaj *zažil* minulosť – keby som žil počas minulých vedeckých revolúcií osobne, nie iba čítal o nich v historických knihách – pravdepodobne by som *nebol* zopakoval tie isté chyby. Neprišiel by som s *ďalšou* tajomnou otázkou; prvých tisíc lekcii by mi to poučenie vtieklo do hlavy.

Preto (pomyslel som si), aby som dostatočne precítil silu histórie, mal by som skúsiť napodobniť myšlienky, ktoré by mal Eliezer *žijúci* počas histórie – mal by som skúsiť myslieť, akoby sa všetko, o čom som čítal v historických knihách, v skutočnosti stalo mne. (S primeraným zohľadnením skreslenia dostupnosti historických kníh – mal by som si pamätať, že som bol tisíckrát poddaným za

¹ pozri: 5 Dostupnosť

² <http://intelligence.org/files/CognitiveBiases.pdf>

³ <http://www.overcomingbias.com/2007/07/tell-your-anti-.html>

⁴ pozri: 283 Necitlivosť k rozsahu

⁵ pozri: 36 Tajomné odpovede na tajomné otázky

⁶ pozri: 41 Neučíme sa z histórie

každý raz, čo som bol kráľom.) Mal by som sa ponoriť do histórie, predstaviť si život v obdobiach, ktoré som videl iba ako atrament na papieri.

Prečo by som si mal pamätať prvý let bratov Wrightovcov? Nebol som tam. Ale ako racionalista, môžem sa opovážiť *nepamätať* si udalosť, ktorá sa naozaj stala? Je až taký veľký rozdiel medzi videním udalosti na vlastné oči - čo je v skutočnosti kauzálna reťaz zahŕňajúca odrazené fotóny, nie priame spojenie - a videním udalosti cez historickú knihu? Fotóny a historické knihy obe pochádzajú z kauzálnych reťazí od samotnej udalosti.

Musel som prekonať falošnú amnéziu narodenia sa v konkrétnej dobe. Musel som si pripomenúť - sprístupniť⁷ - všetky spomienky, nielen spomienky, ktoré zhodou okolností patrili mne a mojej vlastnej dobe.

Zem zrazu zostarla.

V mojej pôvodnej pamäti, Spojené Štáty vždy existovali - nikdy nebol čas, kedy by neboli žiadne Spojené Štáty. Nepamätal som sa, dovtedy, ako rástla Rímska ríša, ako priniesla mier a poriadok, ako vydržala tolko storočí, až som zabudol, že veci boli kedysi inak; a ako tá ríša predsa padla, barbari dobyli moje mesto, a poznanie, ktoré som zhromaždil, sa stratilo. Moderný svet sa pred mojimi očami stal krehkým; nebol to prvý moderný svet.

Tak veľa chýb, robených znovu a znovu a *znovu*, pretože som si nepamätal, že som ich urobil, v každej dobe, keď som nežil...

Pomyslite na to, že ľudia sa občas pýtajú, či je prekonávanie skreslení dôležité.

Nepamätáte sa, kolkokrát vás vaše skreslenia zabili? Nie? Všimol som si, že náhla strata pamäti často nasleduje po smrteľnej chybe. Ale verte mi, stalo sa to. Pamätám sa; nebol som tam.

Až nabudúce začnete pochybovať o čudnosti budúcnosti, spomeňte si, ako ste sa narodili v tlupe lovcov a zberačov pred desaťtisíc rokmi, keď nikto vôbec nepoznal Vedu. Spomeňte si, ako ste boli šokovaní, do jadra svojej bytosti, keď Veda vysvetlila veľké a strašné posvätné tajomstvá, ktoré ste kedysi tak veľmi uctievali. Spomeňte si, ako ste kedysi verili, že dokážete lietať, keď zjete tie správne huby, potom ste sklamane prijali, že lietať nikdy nebudete, a potom ste leteli. Spomeňte si, ako ste si vždy mysleli, že otroctvo je správne a potom ste zmenili názor. Nepredstavujte si, ako by ste túto zmenu *dokázali* predpokladať⁸, pretože to je strata pamäte. *Spomeňte si*, že v skutočnosti ste to neuhádli. Spomeňte si, ako sa storočie za storočím svet menil spôsobom, ktorý ste nečakali.

Možno vás potom bude menej šokovať, čo sa stane ďalej.

Pôvodný článok s komentármi⁹

⁷ http://en.wikipedia.org/wiki/Availability_heuristic

⁸ http://lesswrong.com/lw/il/hindsight_bias/

⁹ http://lesswrong.com/lw/j0/making_history_available/

Kapitola 43. Vysvetli / Uctievaj / Ignoruj?

Súvisí s kapitolami: Sémantické stopky¹, Tajomné odpovede na tajomné otázky²

Ako sa náš kmeň túla lúkami, hľadajúc ovocné stromy a korisť, z času na čas sa stane, že z oblohy padá voda.

“Prečo občas padá voda z oblohy?” pýtam sa bradatého mudrca nášho kmeňa.

Na chvíľu sa zamyslí, lebo mu táto otázka nikdy predtým nenapadla, a potom povie: “Z času na čas duchovia oblohy bojujú a vtedy z oblohy kvapká ich krv.”

“Odkiaľ sú títo duchovia oblohy?” pýtam sa.

Jeho hlas sa mení na šepot. “Sú starší než čas. Sú dávni, dávni.”

Keď prší a neviete prečo, máte niekoľko možností. Po prvé, môžete sa jednoducho nepýtať prečo – nesledovať túto otázku, alebo si ju v prvom rade ani nikdy nepomyslieť. To je možnosť Ignoruj, ktorú si bradatý mudrc vybral pôvodne. Po druhé, môžete sa pokúsiť vymyslieť nejaké vysvetlenie, to je možnosť Vysvetli, ako urobil bradatý muž v odpovedi na moju prvú otázku. Po tretie, môžete si vychutnávať pocit tajomnosti – to je možnosť Uctievaj.

Avšak, ako ste si iste všimli z tohto príbehu, vždy keď si vyberiete Vysvetli, v najlepšom prípade dostanete vysvetlenie ako sú “duchovia oblohy”. Lenže aj samotné toto vysvetlenie predmetom rovnakej voľby – Vysvetli, Uctievaj alebo Ignoruj? Vždy keď kliknete Vysvetli, veda chvíľu hrkoce, vráti vysvetlenie, a potom vyskočí ďalšie dialógové okno. Ako dobrí racionalisti cítime povinnosť stále klikať Vysvetli, ale vyzerá to ako cesta bez konca.

Kliknete Vysvetli na život a dostanete chémiu; kliknete Vysvetli na chémiu a dostanete atómy; kliknete Vysvetli na atómy a dostanete elektróny a jadrá; kliknete Vysvetli na jadrá a dostanete kvantovú chromodynamiku a kvarky; kliknete Vysvetli na vznik kvarkov a dostanete Veľký Tresk...

Môžeme kliknúť Vysvetli na Veľký Tresk a čakať, kým sa veda prehrkoce cez svoj proces a možno jedného dňa vráti úplne dobré vysvetlenie. Ale to opäť prinesie iba ďalšie dialógové okno. Ak teda budeme klikať dosť dlho, musíme sa dostať k *špeciálnemu* dialógovému oknu, k *novej* možnosti, k Vysvetlieniu Ktoré Nepotrebuje Vysvetlenie, k miestu kde táto reťaz končí... a toto je možno jediné vysvetlenie, ktoré je hodno poznať.

Tak... a práve som klikol Uctievaj.

Nikdy nezabudnite, že existuje mnoho iných spôsobov uctievania okrem zapalovania sviečok okolo oltára.

Keby som povedal: “Uf, to znie ako paradox. Ktovie, aké je riešenie tohto zdanlivého paradoxu?” vtedy by som bol klikol Vysvetli, ktorému občas chvíľu trvá než vyprodukuje odpoveď.

A ak vám celá táto téma pripadá nedôležitá, nesúvisiaca, alebo by ste radšej rozmýšľanie o nej odložili na zajtra, potom ste klikli Ignoruj.

¹ pozri: 35 Sémantické stopky

² pozri: 36 Tajomné odpovede na tajomné otázky

Vyberajte si svoju možnosť múdro.

Pôvodný článok s komentármi³

³ <http://lesswrong.com/lw/j2/explainworshipignore/>

Kapitola 44. “Veda” ako zastavovač zvedavosti

Súvisí s kapitolami: Sémantické stopky¹, Tajomné odpovede na tajomné otázky², Nehovorte “zložitost”³

Predstavte si, že by som v plnom zábere televíznych kamier zdvihol ruky a zaspieval *abracadabra* a spôsobil, že sa zrodí žiarivé svetlo v prázdnom vzduchu nad mojimi vystretými rukami. Predstavte si, že by som vykonal tento čin jasného nepomýliteľného čarodejníctva pod plným dozorom Jamesa Randiho a celej armády skeptikov. Myslím si, že väčšina ľudí by bola *pomerne zvedavá*, čo sa deje.

Teraz si však predstavte, že nejdem do televízie. Nechcem sa deliť o svoju moc, ani o pravdu za ňou. Chcem, aby moje kúzlo bolo tajné. Zároveň však chcem čarovať svoje kúzla kedykoľvek a kdekoľvek sa mi zachce. Chcem začarovať svoje žiarivé svetlo, aby som si mohol vo vlaku čítať knihu – a aby sa tomu nikto nečudoval. Existuje kúzlo na zastavenie zvedavosti?

Veru áno! Keď sa niekto opýta: “Ako si to urobil?”, jednoducho poviem: “Veda!”

Nie je to skutočné vysvetlenie⁴, ale iba zastavovač zvedavosti⁵. Nepovie vám, či to svetlo bude časom jasnejšie alebo bledšie, či zmení odtieň alebo sýtosť farby, a rozhodne vám to nepovie, ako si vyrobiť svoje vlastné podobné svetlo. V skutočnosti *neviete* o nič viac než ste vedeli predtým ako som povedal toto čarovné slovo⁶. Avšak vy sa odvrátite, spokojný, že sa tu nedeje nič nezvyčajné.

Ešte lepšie, ten istý trik funguje aj s obyčajným vypínačom.

Prepnite vypínač a rozsvieti sa žiarovka. Prečo?

V škole vás naučia, že heslo⁷ k žiarovke je “Elektrina!” Dúfam, že teraz sa už vyvarujete označenia žiarovky za “pochopenú” na základe tohto. Umožní vám slovo “Elektrina!” urobiť výpočet na očakávanie vnemov? Prinajmenšom sa treba naučiť veľa ďalších vecí. (Fyzici by mali ignorovať tento odstavec a nahradiť si ho problémom z evolučnej teórie⁸, kde je podstata teórie opäť vo výpočtoch, ktoré vie urobiť iba pár ľudí.)

Keby ste si mysleli, že žiarovka je *vedecky nevysvetliteľná*, upútalo by to *celú* vašu pozornosť. Odložili by ste všetko ostatné, čo ste práve robili, a sústredili by ste sa na žiarovku.

Ale čo znamená fráza “vedecky vysvetliteľná”? Znamená to, že niekto *iný* vie, ako žiarovka funguje. Keď vám povedia, že žiarovka je “vedecky vysvetliteľná”, neviete o nič viac, než ste vedeli predtým; neviete, či bude žiarovka časom jasnejšia alebo bledšia. Ale pretože niekto *iný* to vie, znižuje to vo vašich očiach hodnotu tohto poznania. Stanete sa menej zvedavý.

Keďže toto je ekonomický blog [poznámka prekladateľa: pôvodný blog *Overcoming Bias*], niekto určite povie: “Keby bola tá žiarovka vedeckou záhadou,

¹ pozri: 35 Sémantické stopky

² pozri: 36 Tajomné odpovede na tajomné otázky

³ pozri: 38 Nehovorte “zložitost”

⁴ pozri: 31 Falošné vysvetlenia

⁵ pozri: 35 Sémantické stopky

⁶ pozri: 32 Hádanie učiteľovho hesla

⁷ pozri: 32 Hádanie učiteľovho hesla

⁸ pozri: 129 Mimoszemský boh

mohol by si jej skúmaním získať slávu a peniaze.” Ale ja teraz nehovorím o chamtivosti. Nehovorím o kariérnej ambícii. Hovorím o čistej emócii zvedavosti – o pocite, že ma niečo zaujíma. Prečo by sa *vaša* zvedavosť mala zmenšiť vďaka tomu, že niekto *iný*, nie vy, vie, ako funguje žiarovka? Nie je to protinost? Že nestačí, aby ste to vedeli vy; ostatní ľudia musia byť v nevedomosti, inak z toho nebudete mať radosť?

Poznanie môže slúžiť aj iným veciam okrem zvedavosti, napríklad spoločensky užitočnej technológii. Na tieto inštrumentálne ciele je podstatné, či niekto iný v našom okolí už vie. Ale pre moju vlastnú zvedavosť, prečo by na tom malo záležať?

Navyše, predstavte si dôsledky toho, keby ste dovolili, aby “Niekto iný už vie odpoveď” fungovalo ako zastavovač zvedavosti. Jedného dňa vojdete do svojej obývačky a uvidíte veľkého zeleného slona, ktorý sa vznáša vo vzduchu obklopený aurou strieborného svetla.

“Čo do čerta...?” povie.

A sponad slona zaznie hlas, ktorý povie: “NIEKTO INÝ UŽ VIE, PREČO JE TENTO SLON TU.”⁹

“Aha,” povie, “tak potom je mi to jedno,” a idete do kuchyne.

Ja nepoznám všeobecnú zjednotenú teóriu vesmírnych zákonov fyziky. Nevie ani veľa o ľudskej anatómii, s výnimkou mozgu. Nevedel by na vlastnom tele ukázať, kde mám obličky, ani sa mi narýchlo nevybavuje, čo robí moja pečeň. (Nie som na to hrdý. Žiaľ, pri všetkej tej matematike, ktorú potrebujem naštudovať, pravdepodobne sa anatómii tak skoro nenaučím.)

Mala by ma, pokiaľ ide o čistú *zvedavosť*, viac zaujímať moja nevedomosť ohľadom definitívnych zákonov fyziky než skutočnosť, že neviem veľa o tom, čo sa deje vnútri môjho vlastného tela?

Keby som zdvihol ruky a začaroval svetelné kúzlo, zaujímalo by vás to. Mala by vás o niečo *menej* zaujímať samotná skutočnosť, že som zdvihol ruky? Keď vy zdvihnete svoju ruku a zamávate, tento čin koordinuje (okrem iného) váš mozoček. Stavím sa, že neviete, ako mozoček funguje. Ja viem trochu – hoci len hrubé detaily, ktoré nestačia na výpočet – ale čo z toho? Čo na tom záleží, ak vy neviete? Prečo by mal existovať dvojaký meter zvedavosti pri mágii a pohyboch ruky?

Pozrite sa na seba v zrkadle. Viete, na čo sa pozeráte? Viete, čo na vás pozerá spoza vašich očí? Viete, čo ste? Niektoré z týchto odpovedí Veda pozná a niektoré z nich Veda nepozná. Ale prečo by vašej zvedavosti malo záležať na tomto rozdiely, ak vy neviete?

Viete, ako fungujú vaše kolená? Viete, ako boli vyrobené vaše topánky? Viete, prečo váš monitor svieti? Viete, prečo je voda mokrá?

Svet okolo vás je plný otázok. Robte si priority, ak musíte. Ale nestážujte sa, že krutá Veda pripravila svet o tajomstvá. S takýmto uvažovaním by som vedel spôsobiť, aby ste prehliadli slona vo vlastnej obývačke.

Pôvodný článok s komentármi¹⁰

⁹ <http://godescalc.wordpress.com/2012/06/24/overlooked-elephant/>

¹⁰ http://lesswrong.com/lw/j3/science_as_curiositystopper/

Kapitola 45. Čo myslím slovom “rozumnosť”?

Súvisí s kapitolami: Hádanie učiteľovho hesla¹, Umelé pridanie².

Klasický článok Drewa McDermotta: “Umelá inteligencia stretáva prirodzenú hlúposť” kritizuje programy UI, ktoré sa pokúšajú reprezentovať pojmy ako *šťastie* je stav mysle pomocou sémantickej siete:

[ŠŤASTIE] -(JE)-> [STAV MYSLE]

A samozrejme vo vnútri uzla “ŠŤASTIE” nie je nič; je to iba holý symbol LISPu so sugestívnym anglickým názvom.

Preto, hovorí McDermott: “Dobrym testom disciplinovaného programátora je pokúsiť sa namiesto takýchto kľúčov použiť vygenerované symboly a pozrieť, či ešte stále obdivuje svoj systém. Keby sme napríklad STAV MYSLE premenovali na G1073...”, mali by sme JE(ŠŤASTIE, G1073), “čo vyzerá o dosť pochybnejšie.”

Alebo, aby som trochu preformuloval túto myšlienku: Keby ste náhodnými symbolmi nahradili *všetky* sugestívne anglické názvy, nemali by ste najmenšiu šancu vydedukovať, čo vlastne to G1071(G1072, G1073) malo znamenať. Mal tým program UI reprezentovať hamburgery? Jablká? Šťastie? Ktovie. *Ak vymažete tie sugestívne anglické názvy, nedorastú naspäť.*

Predstavte si, že vám fyzik povie, že “Svetlo sú vlny³” a vy mu *veríte*. Teraz máte v hlave malú sieť, ktorá hovorí JE(SVETLO, VLNY). Ak sa vás niekto opýta: “Čo je to svetlo?”, viete mu odpovedať: “Vlny!”

Ako hovorí McDermott: “Celý problém je, aby si počujúci všimol, čo mu povedali. Nie ‘pochopil’, ale ‘všimol si’.” Predstavte si, že vám namiesto toho fyzik povie: “Svetlo sú malé krivolaké veci.” (Mimochodom, to nie je pravda.) *Všimli* by ste si nejaký rozdiel v očakávaných vnemoch⁴?

Ako si uvedomiť, že by ste nemali dôverovať svojmu domnelému poznaniu, že “svetlo sú vlny”? Jedna skúška, ktorú môžete použiť, je opýtať sa: “Dokázal by som túto vedomosť *obnoviť*, keby mi ju nejako vymazali z hlavy?”

Toto je podobná myšlienka, ako keď znečitateľníte mená sugestívne nazvaných symbolov LISPu vo vašom programe UI a skúsite, či niekto iný dokáže zistiť, na čo údajne “odkazovali”. Je to podobná myšlienka ako keď si všimnete, že Umelý Matematik⁵ by síce dokázal prehrávať záznam Plus(Sedem, Šesť) = Trinásť, ale nedokázal by ho obnoviť, keby ste mu ho vymazali z pamäte, dokiaľ by ho nejaký človek zase nezadal do databázy. Takisto, ak zabudnete, že “svetlo sú vlny”, nedokážete túto vedomosť získať naspäť žiadnym iným spôsobom než tým pôvodným - opýtať sa fyzika. Nedokážete túto vedomosť vytvoriť sami spôsobom, ktorým ju pôvodne vytvorili fyzici.

Rovnaká skúsenosť, ktorá nás vedie k formulovaniu názoru, spája tento názor s ostatným poznaním a zmyslovými vstupmi a pohybovými výstupmi. Ak vidíte bobra, ktorý obhrýza kmeň, potom viete, ako táto vec-ktorá-obhrýza-kmene vyzerá a dokážete ju rozoznať v budúcich situáciách, či už ju nazvete “bobor”

¹ pozri: 32 Hádanie učiteľovho hesla

² pozri: 144 Umelé pridanie

³ pozri: 32 Hádanie učiteľovho hesla

⁴ pozri: 11 Nech názory platia nájomné (v očakávaných vnemoch)

⁵ pozri: 144 Umelé pridanie

alebo nie. Ale ak získate svoje názory o bobroch od niekoho iného, kto vám povie fakty o "bobroch", možno nebude vedieť rozoznať bobra, keď ho uvidíte.

To je to hrozné nebezpečenstvo skúšania *povedať* Umelej Inteligencii fakty, ktoré by sa nedokázala sama naučiť. To je aj to hrozné nebezpečenstvo skúšania *povedať* niekomu o fyzike, čo si nedokáže sám overiť. Lebo to, čo fyzici myslia slovom "vlna" nie je "malá krivolaká vec" ale čisto matematický pojem.

Ako hovorí Davidson, keby ste verili, že "bobry" žijú v púšti, majú čisto bielu farbu a v dospelosti vážia 150 kíl, nemali by ste o *bobroch* žiadne názory, pravdivé ani nepravdivé. Váš názor o "bobroch" nie je dosť dobrý na to, aby vôbec mohol byť nesprávny. Ak nemáte dosť vnemov na obnovenie názorov, keď sa vymažú, máte dosť vnemov na to, aby ste tento názor vôbec s niečím spojili? Wittgenstein: "Koleso, ktorým možno točiť aj keď sa nič iné nehýbe s ním, nie je súčasťou mechanizmu."

Takmer hneď ako som začal čítať o UI - ešte skôr než som čítal McDermotta - som si uvedomil, že by bol *naozaj dobrý nápad* vždy si položiť otázku: "Ako by som obnovil túto vedomosť, keby mi ju nejako vymazali z hlavy?"

Čím hlbšie je vymazanie, tým prísnejší je test. Keby mi z hlavy vymazali všetky dôkazy Pytagorovej vety, vedel by som ju znovu dokázať? Myslím, že áno. Keby mi z hlavy vymazali všetky informácie o Pytagorovej vete, všimol by som si, že Pytagorovu vetu treba opäť dokázať? To už je ťažšie tvrdiť bez podkladov; ale keby ste mi dali pravouhlý trojuholník s odvesnami 3 a 4 a povedali mi, že dĺžka prepony sa dá vypočítať, myslím si, že by som to dokázal vypočítať, pokiaľ by som stále vedel zvyšok matematiky.

A čo samotný pojem *matematického dôkazu*? Keby mi o tom nikto nikdy nebol povedal, dokázal by som *toto* znovu objaviť na základe ostatných mojich názorov? Boli časy, keď ľudstvo nemalo takýto pojem. Niekoľko ho musel vymyslieť. Čo bolo to, čo si všimol? Všimol by som si, keby som videl niečo rovnako nové a rovnako dôležité? Dokázal by som myslieť tak ďaleko mimo vychodených kolají⁶?

Kolko zo svojich vedomostí by ste dokázali obnoviť? Z akého hlbokého vymazania? To nie je iba test na vyhodenie nedostatočne prepojených názorov. Je to spôsob, ako vstrebať *fontánu poznania*, nie iba jeden fakt.

Bača postaví počítačový systém⁷, ktorý funguje tak, že vhodí kameň do vedra vždy keď ovca vyjde z ohrady a vyberie kameň vždy keď sa ovca vráti. Ak vy, ako učeň, nerozumiete tomuto systému - ak je to kúzlo, ktoré funguje bez zjavnej príčiny - potom nebudete vedieť, čo máte urobiť, ak vám náhodou do vedra padne jeden kameň navyše. To, čo nedokážete sami urobiť, nedokážete ani *prerobiť*, keď to situácia vyžaduje. Nemôžete sa vrátiť k zdroju, doladiť jeden z parametrov nastavenia a znovu vygenerovať výstup, ak nemáte zdroj. Ak je pre vás "Dva plus dva rovná sa štyri" iba holý fakt a potom sa jeden z prvkov zmení na "päť", ako zistíte, že "dva plus päť rovná sa sedem", ak vám iba *povedali*, že "dva plus štyri rovná sa šesť"?

Ak nájdete rastlinu, z ktorej vypadne semienko vždy, keď nad ňou preletí vták, nenapadne vám, že by ste mohli túto rastlinu použiť na čiastočnú automatizáciu počítačového ovieca. Napriek tomu, že ste sa naučili niečo, čo pôvodný autor mohol použiť na vylepšenie svojho vynálezu, vy sa neviete vrátiť k zdroju a znovu ho vytvoriť.

Keď obsiahnete zdroj myšlienky, samotná táto myšlienka sa môže meniť spolu s vami, ako budete nadobúdať nové vedomosti a nové zručnosti. Keď obsahujete zdroj myšlienky, stane sa naozaj vašou časťou a bude rásť spolu s vami.

⁶ pozri: 82 Kolaje "mimo vychodených kolají"

⁷ pozri: 46 Jednoduchá pravda

Snažte sa stať zdrojom každej myšlienky, ktorú je hodno myslieť. Ak táto myšlienka pôvodne prišla zvonku, postarajte sa, aby rovnako prichádzala aj zvnútra. Stále sa pýtajte: "Ako by som obnovil túto myšlienku, keby sa vymazala?" Keď máte odpoveď, predstavte si, že by sa vymazala aj *táto* vedomosť. A keď nájdete fontánu, pozrite sa, čo iné z nej ešte vyteká.

Pôvodný článok s komentármi⁸

⁸ http://lesswrong.com/lw/la/truly_part_of_you/

Časť V. Medzihra

Obsah

46 Jednoduchá pravda	106
----------------------------	-----

Kapitola 46. Jednoduchá pravda

Pamätám si úlohu, ktorý som napísala o existencializme. Učiteľka mi ju vrátila s päčkou. Podčiarkla všetky slová pravda a pravdivý v eseji, bolo ich asi dvadsať, a vedľa každého napísala otáznik. Chcela vedieť, čo myslím slovom pravda.

—Danielle Egan (novinárka)

Autorov predslov:

Cieľom tejto eseje je obnoviť naivný pohľad na pravdu.

Niektó vám povie: “Moja zázračná masť ťa dokáže zbaviť rakoviny pľúc za púhé tri týždne.” Odpoviete: “Neukázala azda klinická štúdia, že toto tvrdenie nie je pravda?” Dotyčný odvetí: “Takéto chápanie ‘pravdy’ je veľmi naivné. Čo rozumieš slovom ‘pravda’?”

Mnohí ľudia pri takejto otázke nevedia, ako dostatočne presne odpovedať. Napriek tomu by nebolo múdre zahodiť pojem “pravda”. Boli časy, keď nikto dostatočne presne nepoznal rovnice gravitácie, a predsa, kto skočil z útesu, ten spadol.

Často som videl – najmä v internetových diskusiách – ako uprostred rozhovoru niekto povie “X je pravda” a potom sa začne debatovať o správnom používaní slova “pravda”. Táto esej *nie* je myslená ako encyklopedický odkaz pre takú debatu. Namiesto toho dúfam, že si debatujúci prečítajú túto esej a potom sa vrátia k tomu, o čom diskutovali predtým, než niekto spochybnil podstatu pravdy.

V tejto eseji kladiem otázky. Ak vidíte niečo, čo vyzerá ako celkom samozrejmá odpoveď, pravdepodobne je to odpoveď, ktorú som mal na mysli. Samozrejmá možnosť nemusí byť *vždy* tá najlepšia možnosť, avšak niekedy, šokujúco, *je*. Ak narazím na samozrejmú odpoveď, neprestanem hneď hľadať ďalej, ale ak hľadám ďalej a tá napohľad samozrejmá odpoveď *stále* vyzerá samozrejme, necítim sa previniť, že sa jej držím. No iste, každý si *myslí*, že dva plus dva je štyri, každý *hovorí*, že dva plus dva je štyri, a aj v bežnej drine každodenného života sa každý *správa* ako keby dva plus dva bolo štyri, ale čomu sa dva plus dva *naozaj, v konečnom dôsledku* rovná? Nakolko to dokážem posúdiť, je to štyri. Je to stále štyri, aj keď túto otázku vyrieknem vážnym, zlovestným tónom. Príliš jednoduché, hovoríte? Možnože v tomto prípade život *nemusí* byť zložitý. Nebolo by to osviežujúce?

Ak patríte k tým šťastným ľuďom, ktorým sa otázka zdá od začiatku jednoduchá, dúfam, že sa vám bude zdať jednoduchá aj na konci. Ak zistíte, že vás zarazili hlboké a významné otázky, pamätajte, že ak viete, ako presne systém funguje, a vedeli by ste si ho sami poskladať z vedier a kameňov, nemala by to pre vás byť záhada.

Ak vám hrozí zmätenie, keď interpretujete metaforu ako metaforu, skúste brať všetko *úplne doslova*.

* * *

Predstavme si, že som bača, v dobe pred zaznamenanou históriou formálnej matematiky a mám problém zrátať si ovečky. Moje ovečky spávajú v uzavretej ohrade, a tá ohrada je dosť vysoká na to, aby ovečky chránila pred vlkami blúdiacimi nocou. Každé ráno musím ovečky vypustiť z ohrady na pašu; každý večer musím ovečky nájsť a zavrátiť ich do ohrady. Ak nejaká ovečka zostane vonku, ráno nájdem jej telo zabitú a ohlodanú vlkami. Nechce sa mi celé hodiny

prehľadávať lúky, hľadajúc poslednú stratenú ovečku, keď si myslím, že už sú asi všetky v ohrade. Niekedy to vzdám skôr a zvyčajne mi to prejde, ale asi každý desiaty raz nájde ráno ďalšiu mŕtvu ovečku.

Kiež by existoval nejaký spôsob, ako zistiť, či sa ovečky stále pasú, pohodlnejší než hľadanie! Skúšal som niekoľko metód: vrhal som vešteckými paličkami môjho kmeňa; cvičil som svoje duchovné schopnosti, aby som ovečky našiel jasnovidectvom; starostlivo som vymenovával dôvody, prečo veriť, že všetky ovečky sú v ohrade. Nič sa však nezmenilo. Asi každý desiaty raz, keď predčasne ukončím hľadanie, ráno nájdem mŕtvu ovečku. Môžem si uvedomovať, že moje metódy nefungujú, môžem starostlivo zdôvodňovať každú svoju chybu, ale dilema zostáva rovnaká. Buď strávim hodinu prehľadávaním každého možného zákutia, kde väčšinu času žiadna ovečka nezostala; alebo idem spať skôr a prídem tak v priemere o jednu desatinu ovce.

Jedného večera som sa cítil zvlášť unavený. Vrhol som vešteckými paličkami a veštecké paličky povedali, že všetky ovečky sa vrátili. Vizualizoval som si každé možné zákutie a nevybavila sa mi žiadna ovečka. Stále som si nebol celkom istý, tak som sa pozrel do ohrady a zdalo sa, že je v nej veľa ovečiek, tak som si zrekapituloval svoje úsilie a rozhodol som sa, že som bol mimoriadne usilovný. To rozptýlilo moju úzkosť a šiel som spať. Ďalšie ráno som našiel *dve* mŕtve ovečky. Čosi sa vo mne zlomilo a začal som uvažovať tvorivo.

V ten deň sa od brány košiara ozývali hlasné údery kladiva.

Nasledujúce ráno som bránu na ohrade iba odchýlil, a ako každá ovečka vyšla z ohrady, vhodil som kamienok do vedierka priklincovaného vedľa brány. Popoludní, ako každá ovečka vošla, jeden kamienok som z vedierka vybral. Keď vo vedierku nezostal žiaden kameň, môžem prestať hľadať, a uložiť sa k spánku. Je to vynikajúci nápad. Bude to *revolúcia* bačovstva.

Tolko teória. V praxi bolo treba veľa vylepšení, než táto metóda začala fungovať spoľahlivo. Niekoľkokrát som celé hodiny hľadal a nenašiel žiadnu ovečku, a predsa nasledujúce ráno neboli žiadne straty. Každý z týchto prípadov si vyžadoval dôkladné premýšľanie, aby som zistil, kde môj vedierkový systém zlyhal. Po návrate z jedného neplodného hľadania som sa zamyslel a uvedomil som si, že som ráno začal s vedierkom, ktoré už nejaké kamene obsahovalo; zdá sa, že to bol zlý nápad. Inokedy som počas dňa náhodne hádzal kamienky do vedierka, aby som zahnal nudu; to bol tiež zlý nápad, ako som si uvedomil po pár hodinách hľadania. Ale cvičil som sa v remesle kamienkopočítania a stal som sa veľmi zručným kamienkopočítateľom.

Jedného popoludnia popri vyšliapanej ceste k mojim pastvinám kráčal bohato vyobliekaný muž v bielom rúchu, sandáloch a obleku.

“Môžem vám pomôcť” pýtam sa.

Muž si vyberie z vrečka odznak a odklopí ho, čím nado všetku pochybnosť dokáže, že je Markos Sophisticus Maximus, poslanec Rumského senátu. (Človek by sa mohol zamyslieť, či sa taký odznak dá ukradnúť, ale moc týchto odznakov je taká veľká, že keby ich použil niekto iný, v okamihu by sa *premenil* na Markosa.)

“Volajte ma Mark,” povie. “Prišiel som skonfiškovať čarovné kamienky v mene Senátu. Takéto mocné čarovné predmety sa nesmú dostať do rúk nevzdelancom.”

“Ten prekliaty učeň,” zašomrem popod fúzy, “zase niečo natáral dedinčanom.” Potom pozriem na Markovu prísnu tvár a vzdychnem si. “To nie sú čarovné kamienky,” poviem nahlas. “Sú to obyčajné kamienky, ktoré som zdvihol zo zeme.”

Markovo tvárou prebehne záblesk zmätenia, ale potom sa opäť rozžiaril. "Prišiel som pre čarovné vedierko!" vyhlási.

"To nie je čarovné vedierko," poviem unavene. "Predtým som doňho odkladal špinavé ponožky."

Mark sa tvári zmätene. "Kde je teda to kúzlo?" opýta sa.

Zaujímavá otázka. "To sa ťažko vysvetľuje," odpoviem.

Môj učeň Autrey, zaujatý udalosťou prichádza a ponúka svoje vysvetlenie: "Je to v hladine kamienkov vo vedierku. Existuje čarovná hladina kamienkov, ktorú musíte mať celkom správne, inak to nefunguje. Ak vhodíte viac kamienkov, alebo nejaké vyberiete, vedierko už nebude mať čarovnú hladinu. Práve teraz je čarovná hladina," Autrey nazrie do vedierka, "zhruba v jednej tretine."

"Chápem!" povie Mark vzrušene. Vytiahne z vrečka vlastné vedierko a hromadu kamienkov. Vezme za hrst kamienkov a vloží ich do vedierka. Potom pozrie do vedierka, koľko je tam kamienkov. "Aha," povie, "čarovná hladina tohto vedierka je v polovici. Správne?"

"Nie!" odsekne Autrey. "Hladina v polovici nie je čarovná. Čarovná hladina je zhruba v jednej tretine. Polovica je určite nečarovná. Navyše používate nesprávne vedierko."

Mark zmätene pozrie na mňa. "Povedali ste predsa, že vedierko nie je čarovné?"

"Nie je," poviem. Okolo mňa vyjde z brány ovečka, tak prihodím do vedierka ďalší kameň. "Navyše, ja sledujem ovce. Rozprávajte sa s Autreyom."

Mark s podozrením hľadá na kamienok, ktorý som práve prihodil, ale rozhodne sa odložiť túto otázku na neskôr. Otočí sa k Autreyovi a hrdo sa vystrie. "Toto je slobodná krajina," povie, "samozrejme, pod benevolentnou diktatúrou Senátu. Môžem ukladať hocijaké kamienky do hocijakého vedierka."

Autrey sa nad tým zamyslí. "Nie, nemôžete," povie nakoniec, "lebo v tom nebude kúzlo."

"Pozrite," povie Mark trpezlivo. "Pozorne som vás sledoval. Pozreli ste sa do svojho vedierka, skontrolovali ste hladinu kamienkov, a vyhlásili ste túto hladinu za čarovnú. Urobil som presne to isté."

"Tak to nefunguje," povie Autrey.

"Aha," povie Mark. "Hladina kamienkov v *mojom* vedierku nie je čarovná, ale vo *vašom* vedierku áno. To tvrdíte? V čom je vaše vedierko lepšie od môjho, há?"

Nuž, povie Autrey, keby sme vyprázdnili vaše vedierko, a potom presypali všetky kamienky z môjho vedierka do vášho, potom by vaše vedierko malo čarovnú hladinu. Existuje aj systém, ako skontrolovať, či má vaše vedierko čarovnú hladinu, pokiaľ vieme, že ju má moje vedierko; nazývame to operácia porovnania vedierok.

Prejde ďalšia ovečka a ja prihodím ďalší kameň.

"Práve tam vhodil ďalší kameň!" povie Mark. "Budete azda tvrdiť, že aj táto nová hladina je magická? Mohol by som do vášho vedierka prihadzovať kamene, dokiaľ nebude jeho hladina rovnaká ako u mňa, a potom by sa naše vedierka zhodovali. Porovnávate moje vedierko s vaším, aby ste určili, či je podľa vás hladina 'čarovná' alebo nie. Nuž, z môjho pohľadu vaše vedierko nie je čarovné, pretože nemá rovnakú hladinu kamienkov ako moje. Ha!"

“Moment,” povie Autrey, “nechápete, že...”

“Slovom ‘čarovná hladina’ jednoducho myslíte hladinu kamienkov vo svojom vlastnom vedierku. A keď ja poviem ‘čarovná hladina’, tak tým myslím počet kamienkov v mojom vedierku. Takže keď vy pozriete na moje vedierko a poviete, že nie je čarovné, slovo ‘čarovné’ tu znamená rôzne veci pre rôznych ľudí. Musíte upresniť, *čia* čarovnosť to je. Môžete povedať, že moje vedierko nemá ‘Autreyovu čarovnú hladinu’ a ja poviem, že vaše vedierko nemá ‘Markovu čarovnú hladinu’. Týmto sa vyrieši domnelý paradox.”

“Ale...” povie Autrey bezmocne.

“Rôzni ľudia môžu mať rôzne vedierka s rôznymi hladinami kamienkov, čo dokazuje, že celé vaše poňatie ‘čarovnosti’ je naprosto subjektívne a svojvoľné.”

“Mark,” poviem, “povedal vám niekto, na čo tie kamienky *slúžia*?”

“*Slúžia*?” povie Mark. “Myslel som, že sú proste čarovné.”

“Keby tie kamienky nič nerobili,” povie Autrey, “náš auditor efektivity procesov ISO 9000 by túto procedúru z našej dennej práce odstránil.”

“Kto je váš auditor?”

“Darwin,” povie Autrey.

“Hm,” povie Mark. “Charles má povest' prísneho auditora. Takže tieto kamienky požehnávajú stáda a zvyšujú tak počet oviec?”

“Nie,” poviem. “Čaro kamienkov je v tomto: keď pozrieme do vedierka a vidíme, že vedierko je bez kamienkov, vieme, že aj pastviny sú bez oviec. Keď nepozrieme do vedierka, musíme hľadať až do zotmenia, aby náhodou nejaká ovca nezostala vonku. Alebo ak skončíme predčasne, niekedy nájdeme na druhý deň mŕtvu ovcu, lebo vlci roztrhajú všetky zabudnuté ovce. Ak sa pozrieme do vedierka, vieme kedy sú všetky ovce doma, a môžeme ísť spať bez strachu.”

Mark nad tým uvažuje. “Znie to dosť nedôveryhodne,” povie nakoniec. “Rozmýšľali ste nad použitím vešteckých paličiek? Veštecké paličky sú neomylné; prinajmenšom každý, kto povie, že sú omylné, je upálený na hranici. A to je mimoriadne bolestivá smrť. Z toho vyplýva, že veštecké paličky sú neomylné.”

“Vy môžete používať veštecké paličky, ak chcete,” poviem.

“Ach, dobré nebesá, ani náhodou,” povie Mark. “Paličky fungujú neomylné, absolútne dokonale v každom prípade, ako sa patrí na také požehnané nástroje. Ale čo keby na druhý deň bola nejaká ovca mŕtva? Ja používam veštecké paličky iba vtedy, keď neexistuje možnosť, že by sa ukázal ich omyl. Inak by ma mohli upáliť zaživa. Ako teda funguje vaše čarovné vedierko?”

Ako funguje vedierko...? Najlepšie bude začať tým najjednoduchším prípadom. “Asi takto,” poviem, “predstavte si, že pastviny sú prázdne, ale vedierko nie. Potom by sme strácali hodiny hľadaním ovce, ktorá tam nie je. A keby boli ovce na pastvinách, ale vedierko by bolo prázdne, potom by sme Autrey a ja išli skôr spať, a na druhý deň by sme našli mŕtvu ovcu. Preto je prázdne vedierko čarovné vždy vtedy a iba vtedy, keď sú pastviny prázdne...”

“Moment,” povie Autrey. “To mi znie ako zbytočná tautológia. Nie sú snáď prázdne vedierko a prázdne pastviny samozrejme tá istá vec?”

“Nie je to zbytočné,” poviem. “Tu je analógia: Logik Alfred Tarski raz povedal, že veta ‘Sneh je biely’ je pravdivá vždy vtedy a iba vtedy, keď sneh je biely. Ak

rozumieš tomuto, potom by si mal chápať, prečo je prázdne vedierko čarovné vždy vtedy a iba vtedy, keď sú pastviny bez oviec.”

“Moment,” povie Mark. “Toto sú *vedrá*. S *ovcami* nemajú nič spoločné. Vedierka a ovce sú samozrejme úplne odlišné. Neexistuje spôsob, ako by ovce mohli interagovať s vedierkami.”

“Odkiaľ teda podľa vás pochádza to kúzlo?” pýta sa Autrey.

Mark uvažuje. “Povedali ste, že môžete porovnať dve vedierka, aby ste zistili, či majú rovnakú hladinu... Chápem, ako môžu vedierka interagovať s vedierkami. Možno keď máte veľké množstvo vedierok a všetky majú rovnakú hladinu, *to* vytvára celé kúzlo. Budem to nazývať koherentistická teória čarovných vedierok.”

“Zaujímavé,” povie Autrey. “Viem, že môj majster pracuje na systéme s viacerými vedierkami – tvrdí, že môžu fungovať lepšie vďaka ‘redundancii’ a ‘korekcii chýb’. To mi znie ako koherentizmus.”

“To nie je to isté...” začnem namietat.

“Vyskúšajme si koherentistickú teória kúzla,” povie Autrey. “Vidím, že máte vo vrecku ešte päť vedierok. Podám vám vedierko, ktoré používame, a vy potom naplníte vaše vedierka na rovnakú hladinu...”

Mark s hrôzou cúvne. “Stop! Tieto vedierka sa v mojej rodine odovzdávali celú generáciu a vždy mali rovnakú hladinu! Ak prijmem vaše vedierko, moja zbierka vedierok sa stane menej koherentná a kúzlo sa stratí!”

“Ale vaše *terajšie* vedierka nemajú s ovcami nič spoločné!” namieta Autrey.

Mark vyzerá podráždene. “Pozrite, ako som už vysvetlil, samozrejme neexistuje spôsob, ako by ovce mohli interagovať s vedierkami. Vedierka môžu interagovať iba s inými vedierkami.”

“Ja vhodím kameň vždy, keď vyjde ovca,” podotknem.

“Keď vyjde ovca, vhodíte kameň?” povie Mark. “A ako to súvisí s našou témou?”

“Je to interakcia medzi ovcou a kameňmi,” odpoviem.

“Nie, to je interakcia medzi kameňmi a *vami*,” povie Mark. “Kúzlo nepochádza z oviec, pochádza z *vás*. Ovce samozrejme nie sú čarovné. Kúzlo musí *odniekadiaľ* pochádzať, aby sa dostalo do vedra.”

Ukážem na drevený mechanizmus postavený na bráne. “Vidíte ten kus plachty, ktorá visí z tej drevenej hračky? Ešte stále sa s tým hráme – zatiaľ to nefunguje spoľahlivo – ale keď tadial prejde ovca, potiahne plachtu. Keď sa plachta odhrnie, zo zásobníka vypadne kamienok a spadne do vedierka. Takto potom Autrey a ja nebudeme musieť vkladať kamienky osobne.”

Mark zmraští obočie. “Nerozumiem tomu... je tá plachta čarovná?”

Pokrčím plecami. “Objednal som si ju cez internet od spoločnosti Prirodzený výber. Tá látka sa volá Zmyslová modalita.” Zastavím sa, vidiac neveriace výrazy Marka a Autreya. “Pripúšťam, že tie názvy znejú tak trochu ako new age. Pointa je, že prechod ovce spustí reťaz príčiny a účinku, ktorá končí kamienkom vo vedre. *Potom* už môžete porovnávať vedierka s vedierkami, atď.”

“Tomu stále nerozumiem,” povie Mark. “Nemôžete strčiť ovcu do vedierka. Iba kamene idú do vedier, a je samozrejmé, že iba kamene môžu interagovať s inými kameňmi.”

“Ovce interagujú s vecami, ktoré interagujú s kameňmi...” hľadám prirovnanie. “Predstavte si, že sa pozriete na svoje šnúrky na topánkach. Fotón vyletí zo Slnka, potom prejde zemskou atmosférou, potom sa odrazí od vašej šnúrky, potom prejde zreničkou vášho oka, potom dopadne na sietnicu, potom ho pohltí čapík alebo tyčinka. Energia fotónu spôsobí, že pripojený neurón vyšle signál, čo spôsobí, že ďalšie neuróny vyšlú signál. Vzor aktivácie neurónov vo vašej zrakovej kôre môže interagovať s vašimi myšlienkami o vašich šnúrkach, keďže vaše myšlienky o šnúrkach sú tiež uložené v neurónoch. Ak rozumiete tomuto, mali by ste chápať, ako prechádzajúca ovca spôsobuje, že kamienok padne do vedierka.”

“A v ktorom presne okamihu počas tohto procesu sa kamienok stane čarovným?” povie Mark.

“To... hm...” Teraz začínam byť popletený *ja*. Potrasiem hlavou, aby som sa zbavil pavučín. Keď som sa dnes ráno zobudil, toto všetko mi pripadalo celkom jednoduché, a systém kamienkov a vedierka sa medzičasom nestal o nič zložitejším. “Je to o dosť ľahšie na pochopenie, keď si pamätáte, že zmyslom celého systému je sledovať ovce.”

Mark smutne vzdychne. “No nič... je zrejmé, že neviete. Možno sú všetky kamienky čarovné už od začiatku, ešte predtým než sa dostanú do vedierka. Tento názor budeme nazývať pankamienkizmus.”

“Ha!” povie Autrey s hlasom plným pohrdania. “To sú len zbožné želania! Nie všetky kamienky boli stvorené ako rovné. Kamienky vo *vašom* vedierku *nie* sú čarovné. Sú to len kusy kameňa!”

Mark sa zatvári prísne. “Teraz,” vykrikuje, “teraz vidíte, po akej nebezpečnej ceste kráčate! Keď raz poviete, že niektorí ľudia majú čarovné kamienky a iní nie, vaša pýcha vás pohltí! Budete sa považovať za lepších od druhých, a potom padnete na nos! Mnohí ľudia v dejinách mučili a vraždili, pretože si mysleli, že ich vlastné kamienky sú najlepšie!” V Markovom hlase zaznie nádych blahosklonnosti. “Uctievať hladinu kamienkov ako ‘čarovnú’ znamená, že existuje absolútna hladina v Najvyššom Vedierku. Dnes už však nikto v Najvyššie Vedierko neverí.”

“Po prvé,” poviem, “ovce nie sú absolútne kamene. Po druhé, ja si nemyslím, že moje vedierko naozaj obsahuje ovce. Po tretie, neuctievam hladinu svojho vedierka ako čarovnú - občas ju zmením - a robím to *preto*, lebo mi záleží na ovciach.”

“Navyše,” povie Autrey, “ak si niekto myslí, že mať absolútne kamienky by mu *dalo* právo mučiť a vraždiť, je to omyl, ktorý s vedierkami nijako nesúvisí. Riešite tu nesprávny problém.”

Mark sa upokojí. “Predpokladám, že od púhych bačov nemožno očakávať nič viac. Vy zrejme veríte aj tomu, že sneh je biely, však?”

“Hmmm... áno?” povie Autrey.

“A vôbec vám neprekáža, že aj *Stalin* veril, že sneh je biely?”

“Hmmm... nie?” povie Autrey.

Mark sa neveriacky zahľadí na Autreya, a nakoniec pokrčí plecami. “Skúsme predpokladať, iba pre účely tejto diskusie, že vaše kamienky sú čarovné a moje nie sú. Viete mi povedať, v čom je ten rozdiel?”

“Moje kamienky *reprezentujú* ovce!” povie Autrey víťazoslávne. “Vaše kamienky nemajú vlastnosť reprezentatívnosti, preto nefungujú. Chýba im zmysel. Len

sa na ne pozrite. Nemajú auru sémantického obsahu, sú to púhe kamene. Potrebujete vedierko so špeciálnou schopnosťou kauzality.”

“Ach!” povie Mark. “Špeciálna schopnosť kauzality, namiesto kúzla.”

“Presne,” povie Autrey. “Nie som poverčivý. Predpokladať kúzla, v dnešnej dobe, by bolo v medzinárodnej bačovskej komunite neprijateľné. Zistili sme, že predpokladať kúzla jednoducho nefunguje ako vysvetlenie pre bačovské fenomény. Takže keď vidím niečo, čomu nerozumiem, a chcem to vysvetliť pomocou modelu bez vnútorných detailov, ktorý neumožňuje žiadne predpovede, ani so spätnou platnosťou, použijeme predpoklad špeciálnej schopnosti. Ak nefunguje ani to, rozhodneme sa to nazývať emergentný fenomén.”

“Aké špeciálne schopnosti má toto vedierko?” pýta sa Mark.

“Hm,” povie Autrey. “Možno je toto vedierko nasiaknuté *vzťahovosťou* voči pastvinám. To by vysvetľovalo, prečo funguje – keď je vedierko prázdne, *znamená* to, že pastviny sú prázdne.”

“Kde ste našli toto vedierko?” pýta sa Mark. “A ako ste prišli na to, že má vzťahovosť k pastvinám?”

“Je to *obyčajné vedro*,” poviem. “Predtým som s ním liezol po stromoch... Nemyslím si, že to *treba* tak komplikovať.”

“Rozprávam sa s Autreyom,” povie Mark.

“Vedierko treba spojiť s pastvinami, a kamienky s ovcami, pomocou čarovného rituálu... prepáčte, pomocou emergentného procesu so špeciálnymi schopnosťami kauzality... ktoré môj majster objavil,” vysvetlí Autrey.

Autrey sa potom pokúša opísať rituál, a Mark prikyvuje s múdрым porozumením.

“Musíte vhodit kamienok *zakaždým*, keď cez bránu vyjde ovca?,” pýta sa Mark. “Vybrať kamienok *zakaždým*, keď sa ovca vráti?”

Autrey prikyvuje. “Hej.”

“To musí byť veľmi ťažké,” povie Mark súcitne.

Autrey sa rozžiari, pijúc Markov súcit ako dážd. “Presne tak!” povie. “Je to *mimoriadne* emocionálne náročné. Keď malo vedierko istý čas rovnakú hladinu, tak... si na tú hladinu zvyknete.”

Ovca vyjde z brány a prejde okolo nich. Autrey ju vidí, zastaví sa, dvihne kamienok a podrží ho vo vzduchu. “Ajhľa!” vyhlási, “prešla ovca! Teraz musím vhodit kameň do tohto vedra, moje drahé vedierko, a zničiť tým hladinu, ktorú si si tak dlho udržiavalo...” Prejde ďalšia ovca. Autrey ju nezbadá, taký je zaujatý svojím vystúpením, takže jeden kameň prihodím ja. Autrey pokračuje: “...pretože to je najťažšou skúškou pravého baču, vhodit kameň, akokoľvek to bolí, akokoľvek miloval pôvodnú hladinu. Iba tí najlepší spomedzi bačov splnia úlohu tak krutú...”

“Autrey,” poviem, “ak raz chceš byť veľkým bačom, nauč sa držať hubu a hádzať kamienky. Žiaden rozruch. Žiadne divadlo. Proste to urob.”

“A tento rituál,” povie Mark, “spája kamienky s ovcami podľa čarovných zákonov podobnosti a súvisu, ako bábika voodoo.”

Autrey sa mykne a obzrie. “Prosím vás! Nevolajte to podobnosť a súvis. My bačovia sme proti poverčivosti. Použite slovo ‘intencionalita’ alebo niečo podobné.”

“Môžem sa pozrieť na ten kameň?” opýta sa Mark.

“Iste,” poviem. Vezmem jeden kamienok z vedierka a hodím ho Markovi. Potom sa zohnem, zdvihnem zo zeme ďalší kameň a vložím ho do vedra.

Autrey na mňa začudovane pozerá. “Nepokazil si to práve?”

Pokrčím plecami. “Myslím, že nie. Keby som to pokazil, zistíme to, ak zajtra nájdeme mŕtvu ovcu, alebo ak budeme celé hodiny hľadať a žiadnu ovcu nenájdeme.”

“Ale...” povie Autrey.

“Naučil som ťa všetko, čo ty vieš, ale nie všetko, čo viem ja,” poviem.

Mark skúma kamienok, sústredene naň hľadí. Drží nad ním svoju ruku a mrmle akési slová, potom potrasie hlavou. “Necítim žiadnu čarovnú moc,” povie. “Prepáčte. Necítim žiadnu intencionalitu.”

“Kameň má intencionalitu iba vnútri čarovného, teda emergentného vedra,” povie Autrey. “Inak je to púhy kameň.”

“To nie je problém,” poviem. Vyberiem kamienok z vedierka a odhodím ho. Potom prídem k Markovi, potľapkám ho po ruke držiacej kamienok a poviem: “Vyhlasujem túto dlaň za súčasť čarovného vedra!” Potom sa vrátim k bráne.

Autrey sa smeje: “Tak toto je od teba totálne zlomyseľné.”

Prikývnem, lebo je to naozaj tak.

“Ale bude to naozaj fungovať?” pýta sa Autrey.

Opäť prikývnem, dúfajúc, že mám pravdu. Kedysi som to skúšal s dvoma vedierkami, a v princípe by nemal byť žiaden rozdiel medzi Markovou rukou a vedierkom. Dokonca aj keď Markova ruka obsahuje *elan vital*, odlišujúci živú hmotu od neživej, trik by mal fungovať rovnako dobre, ako keby bol Mark mramorová socha.

Mark pozerá na svoju ruku, mierne nervózne. “Takže... teraz má kameň opäť intencionalitu?”

“Hej,” poviem. “Neberte si do ruky viac kameňov, ani neodhadzujte tento jeden, inak porušíte rituál.”

Mark vážne prikývne. Potom pokračuje v skúmaní kamienka. “Teraz chápem, ako sa vaše stáda tak rozrastajú,” povie. “Vďaka moci tohto vedra vám stačí vhadzovať kamene a ovce sa budú vracat' z polí. Môžete začať iba s pár ovcami, vypustíte ich, a potom naplníte vedro až po okraj než sa vrátia. A keby vás starostlivosť o toľko ovce začala zafažovať, môžete ich všetky vypustiť, potom vysypať takmer všetky kamene z vedra, takže sa vráti iba pár... opäť zvýšite ich počet, keď príde čas strihania... dobré nebesá, človeče! Uvedomujete si úžasnú silu tohto rituálu, ktorý ste objavili? Predstavte si jeho dôsledky; ľudstvo sa posunie vpred o desaťročie... nie, o storočie!”

“Takto to nefunguje,” poviem. “Ak pridáte kameň, keď ovca neodišla, alebo odoberiete kameň, keď ovca neprišla, porušíte tým rituál. Moc nezostane v kameňoch, ale vyprchá, ako keď praskne mydlová bublina.”

Mark sa tvári hrozne sklamane. “Určite?”

Prikývnem. “Skúšal som to a nefungovalo to.”

Mark ťažko vzdychne. "Ach, táto... *matematika*... vyzerala taká mocná a užitočná, až kým... Nuž dobre. Zbohom, pokrok ľudstva."

"Mark, ten nápad je *skvelý*," povie Autrey povzbudzujúco. "Mňa to vôbec nenapadlo a pritom je to také samozrejmé... ušetrilo by to *ohromné* množstvo námahy... *musí* existovať spôsob, ako váš plán zachrániť! Mohli by sme skúšať rôzne vedrá, hľadať také, ktoré uchová kúzlo... teda intencionalitu v kameňoch aj bez rituálu. Alebo vyskúšať iné kamene. Možno naše kamene jednoducho nemajú správne vlastnosti, a preto nemajú *vnútornú* intencionalitu. Čo keby sme skúšali použiť kamene vytesané do tvaru malých ovečiek? Alebo napísať na kamene slovo 'ovca', to by mohlo stačiť."

"To nebude fungovať," predpovedám sucho.

Autrey pokračuje. "Možno potrebujeme organické kamene namiesto kremíkových... alebo treba použiť drahé drahokamy. Ceny drahokamov sa zdvojnásobuje každých osemnásť mesiacov, takže keby ste teraz kúpili za hrst lacných drahokamov a počkali, o dvadsať rokov by bolo naozaj drahé."

"Vy ste skúšali pridávať kamene, aby ste vytvorili viac oviec, a nefungovalo to?" pýta sa ma Mark. "Čo presne ste robili?"

"Zobral som za hrst dolárových bankoviek. Potom som tie bankovky schoval pod perinu, jednu za druhou. Vždy keď som schoval ďalšiu bankovku, vzal som z krabičky jednu spinku na papier a uložil ju na kôpku. Dával som si pozor, aby som v hlave nedržel ich počet, takže som iba vedel, že mám 'veľa' bankoviek a 'veľa' spiniek. Potom, keď boli všetky bankovky ukryté pod perinou, pridal som na kôpku ešte jednu spinku na papier, čo je ekvivalentné prihodeniu jedného kamienka navyše do vedra. Potom som začal vyberať bankovky spod periny a ukladať spinky späť do krabičky. Keď som skončil, jedna spinka bola navyše."

"Čo taký výsledok znamená?" pýta sa Autrey.

"Znamená to, že trik nefungoval. Keď som porušil rituál tým jediným chybným krokom, moc nezostala, ale hneď vyprchala; kôpka spiniek a zbierka bankoviek sa už nemiňali v rovnakom čase."

"Toto ste *naozaj* skúsili?" pýta sa Mark.

"Áno," hovorím. "Naozaj som vykonal pokus, aby som overil, či výsledok zodpovedá mojim teoretickým predpovediam. Mám sentimentálny vzťah k vedeckej metóde, aj keď to vyzerá absurdne. Navyše, čo keby som sa mylil?"

"Keby to *bolo* fungovalo," povie Mark, "boli by ste vinný z falšovania peňazí! Predstavte si, že by to robil každý; ekonomika by sa zrútila! Každý by mal miliardy dolárov, ale za tie peniaze by sa nedalo nič kúpiť!"

"Vôbec nie," odpoviem. "Podľa rovnakej logiky, kde pridanie ďalšej spinky na kôpku vytvorí ďalšiu dolárovú bankovku, vytvorenie ďalšej dolárovej bankovky by vytvorilo ďalší tovar a služby v hodnote jedného dolára."

Mark krúti hlavou. "Falšovanie peňazí je aj tak zločin... Nemali ste to skúšať."

"Bol som si *dost* istý, že sa to nepodarí."

"Aha!" povie Mark. "Očakávali ste neúspech! *Neverili* ste, že to naozaj dokážete!"

"Veru tak", pripúšťam, "odhadli ste moje očakávania so zarážajúcou presnosťou."

“Nuž a v tom je problém”, odvetí Mark rezko. “Kúzla sú poháňané vierou a silou vôle. Ak neveríte, že to dokážete urobiť, potom to nedokážete. Musíte zmeniť svoju vieru vo výsledok pokusu; tým sa zmení výsledok samotný.”

“To je smiešne,” poviem nostalgicky, “že presne toto mi Autrey povedal, keď som mu hovoril o metóde kameňa a vedra. Že preňho je úplne absurdné tomu uveriť, a preto to jemu fungovať nebude.”

“Ako ste ho presvedčili?” vyzvedá Mark.

“Povedal som mu, aby držal hubu a riadil sa pokynmi,” poviem, “a keď metóda fungovala, Autrey v ňu začal veriť.”

Mark sa zmätene mračí. “To nedáva zmysel. Nerieši to základnú dilemu, či bolo skôr vajce alebo sliepka.”

“Ale rieši. Metóda vedra funguje bez ohľadu na to, či v ňu veríte alebo nie.”

“To je *absurdné!*” vyprskne Mark. “Neverím v existenciu kúziel, ktoré fungujú bez ohľadu na to, či v ne veríte alebo nie!”

“Aj ja som to hovoril,” pridá sa Autrey. “Očividne som sa mýlil.”

Mark sústredene mraští tvár. “Ale... ak ste neverili v kúzla, ktoré fungujú bez ohľadu na to, či v ne veríte alebo nie, prečo potom metóda vedra fungovala, keď ste v ňu neverili? Verili ste v kúzla, ktoré fungujú bez ohľadu na to, či v ne veríte alebo nie, bez ohľadu na to, či veríte v kúzla, ktoré fungujú bez ohľadu na to, či v ne veríte alebo nie?”

“Ja... *asi...* nie...” povedal Autrey neisto.

“Ak ste teda neverili v kúzla, ktoré fungujú bez ohľadu na to, či... vydržte chvíľku, musím si to nakresliť ceruzou na papier...” Mark fanaticky čmára, skepticky pozerá na výsledok, otáča papier hore nohami, potom sa vzdá. “Nechajte tak,” povie. “Kúzla sú pre mňa dosť ťažké na pochopenie; metakúzla sú už na mňa priveľa.”

“Mark, nemyslím si, že rozumiete umeniu kamienkopočítania,” hovorím. “To nie je o používaní kamienkov na ovládanie oviec. Ide o to, aby ovce ovládali kamienky. V tomto umení netreba začať tým, že veríte v jeho fungovanie. Naopak, toto umenie najprv funguje, a až potom začnete veriť, že funguje.”

“Lebo tomuto veríte,” povie Mark.

“Tomuto verím,” odpoviem, “*pretože* je to naozaj tak. Súlad medzi skutočnosťou a mojimi názormi vzniká tak, že skutočnosť ovláda moje názory, nie naopak.”

Prejde ďalšia ovca, čo spôsobí, že prihodím ďalší kamienok.

“Aha! Teraz sme sa dostali na koreň problému,” povie Mark. “Čo je to tá takzvaná ‘skutočnosť’? Rozumiem, čo sa myslí tým, že nejaká hypotéza je elegantná, alebo falzifikovateľná, alebo v súlade s indíciami. Pripadá mi, že nazývať nejaký názor ‘pravdivý’ alebo ‘skutočný’ alebo ‘ozajstný’, je jednoducho len rozdiel medzi povedaním, že si niečo myslíte, a že si niečo naozaj naozaj myslíte.”

Zastavím sa. “No...” poviem pomaly. “Úprimne, tiež si nie som celkom istý, odkiaľ pochádza táto ‘skutočnosť’. Nedokážem si vytvoriť svoju vlastnú skutočnosť v laboratóriu, takže tomu zatiaľ dostatočne nerozumiem. Ale niekedy som silne presvedčený, že sa niečo stane, a potom sa stane niečo celkom iné. Potrebujem názov pre toto čosi, čo určuje výsledky mojich experimentov, takže to nazývam ‘skutočnosť’. Táto ‘skutočnosť’ je čosi odlišné než moje hypotézy.”

Lebo aj keď mám jednoduchú hypotézu, v ktorej prospech svedčí mnoho mne známych indícií, aj tak ma niekedy výsledok prekvapí. Potrebujem teda rôzne názvy pre veci, ktoré určujú moje predpovede, a pre vec, ktorá určuje výsledky mojich experimentov. To prvé volám 'názory' a to druhé volám 'skutočnosť'."

Mark si odfrkne. "Ani neviem, prečo sa unúvam počúvať tieto jasné nezmysly. Čokoľvek povie o tejto takzvanej 'skutočnosti', je jednoducho ďalší názor. Dokonca aj váš názor, že najprv existuje realita a až potom naše názory, je názor. Z toho logicky nevyhnutne vyplýva, že skutočnosť neexistuje; existujú iba názory."

"Moment," hovorí Autrey, "mohli by ste mi zopakovať ten záver. Stratil som sa pri tej ostrej odbočke v strede."

"Bez ohľadu na to, čo hovoríte o skutočnosti, je to iba ďalší názor," vysvetľuje Mark. "Z toho totálne nevyhnutne vyplýva, že neexistuje žiadna skutočnosť, iba názory."

"Aha," poviem. "Takisto platí, že bez ohľadu na to, čo jeme, musíme to jesť ústami. Z toho vyplýva, že neexistuje žiadne jedlo, iba ústa."

"Presne tak," povie Mark. "Všetko, čo jete, musí byť vo vašich ústach. Ako by mohlo existovať nejaké jedlo mimo úst? Samotná táto myšlienka je nezmyselná, čo dokazuje, že 'jedlo' je nekonzistentný pojem. Preto sme všetci na smrť hladní; žiadne jedlo neexistuje."

Autrey pozerá na svoje brucho. "Ale ja *nie* som na smrť hladný."

"Aha!" kričí Mark víťazoslávne. "A čím ste vyslovili samotnú túto námietku? Svojimi *ústami*, priateľu! *Ústami*! Aký lepší dôkaz neexistencie jedla by ste si mohli žiadať?"

"Čo to tu počujem o hladovaní?" opýta sa drsný chrapľavý hlas priamo spoza nás. Autrey a ja zostaneme v pohode, už sme si týmto prešli. Mark vyplašene vyskočí pol metra do vzduchu.

Inšpektor Darwin sa usmeje zovretými perami, potešený z úspešného prekvapenia, a spraví si malú čiarku vo svojom poznámkovom bloku.

"To je len metafora!" povie Mark rýchlo. "Nemusíte mi kvôli tomu brať ústa, ani nič také..."

"Načo sú vám *ústa*, ak neexistuje žiadne *jedlo*?" pýta sa Darwin zlostne. "No nič. Nemám čas na takéto *hlúposti*. Prišiel som na inšpekciu *oviec*."

"Stádam sa darí, pane," hovorím. "Od januára nezomrela žiadna ovca."

"Výborne. Pridelujem vám 0,12 bodu *úspechu*. A teraz, čo tu robí táto *osoba*? Je nevyhnutnou súčasťou tejto *operácie*?"

"Zatiaľ sa mi zdá, že by bol ľudstvu užitočnejší, keby ho zavesili z balóna ako závažie," hovorím.

"Ajaj," povie Autrey potichu.

"Úžitok *ľudstva* ma vôbec *nezaujíma*. Nech sa vyjadří *sám*."

Mark sa rýchlo narovná. "Tento púhy *bača*," povie, ukazujúc na mňa, "tvrdí, že existuje niečo také ako skutočnosť. To ma uráža, pretože ja viem s hlbokým a trvalým presvedčením, že žiadna pravda neexistuje. Koncept 'pravdy' je iba strategický manéver ľudí, ktorí chcú vnucovať svoje názory druhým. Každý

kultúra má inú 'pravdu' a 'pravda' žiadnej kultúry nie je nadradená iným. Čo som práve povedal, platí v každom čase a na každom mieste, a trvám na tom, aby ste súhlasili."

"Počkajte chvíľu," hovorí Autrey. "Ak nič nie je pravda, prečo by som mal veriť tomu, keď hovoríte, že nič nie je pravda?"

"Nepovedal som, že nič nie je pravda..." hovorí Mark.

"Ale áno," preruší ho Autrey, "počul som vás."

"...povedal som, že 'pravda' je výhovorka používaná niektorými kultúrami na vnučovanie svojich názorov druhým. Keď poviete, že niečo je 'pravda', myslíte tým len to, že by pre vašu sociálnu skupinu bolo výhodné, keby sa tomu verilo."

"A toto, čo ste práve povedali," hovorím, "to je pravda?"

"Absolútna a jednoznačná pravda!" povie Mark precítene. "Ľudia si tvoria svoje vlastné skutočnosti."

"Počkajte," povie Autrey, opäť znejúci zmätene, "povedať, že ľudia si vytvárajú svoje vlastné skutočnosti, je logicky úplne iná vec ako povedať, že neexistuje žiadna pravda, čo je stav, ktorý si ani neviem koherentne predstaviť, azda pretože ste stále nevysvetlili, ako presne to funguje..."

"A zase to tu máme," povie Mark podráždene, "pokúšate sa aplikovať svoj západný koncept logiky, rozumnosti, rozumu, koherencie a vnútornej konzistencie."

"No super," šomre Autrey, "tak sme zároveň načali ešte *tretiu*, úplne nesúvisiacu a odlišnú tému..."

"Nie je nesúvisiaca," hovorí Mark. "Pozrite, beriete to zo zlého uhla, ak chápete moje výroky ako hypotézy a snažíte sa odvodiť ich dôsledky. Mali by ste ich brať ako úplne všeobecné výhovorky, ktoré používam, keď niekto povie niečo, čo sa mi nepáči. Nemá to byť model toho, ako funguje vesmír, ale skôr ako karta 'môžete slobodne vyjsť z väzenia'. *Kľúčom* je uplatňovať tieto výhovorky *selektívne*. Keď poviem, že neexistuje nič také ako pravda, uplatňujem to iba na *vaše* tvrdenie, že čarovné vedierko funguje bez ohľadu na to, či v to verím alebo nie. *Neuplatňujem* to na *svoje* tvrdenie, že neexistuje nič také ako pravda."

"Hmmm... prečo nie?" vyzvedá Autrey.

Mark si unavene vzdychne. "Autrey, myslíte si, že ste prvý človek, ktorý kladie túto otázku? Ktorý sa pýta, ako naše vlastné názory môžu mať zmysel, keď žiadne názory nemajú zmysel? Tú istú vec sa pýtajú mnohí študenti, keď sa stretnú s touto filozofiou, o ktorej vás musím informovať, že má veľa priaznivcov a rozsiahlu literatúru."

"Aká je teda odpoveď?" pýta sa Autrey.

"Nazývame to 'problém reflektivity'," vysvetlí Mark.

"Ale aká je *odpoveď*?" trvá na svojom Autrey.

Mark sa povznesene usmieva. "Verte mi, Autrey, nie ste prvý človek, ktorého napadla takáto jednoduchá otázka. Nemá zmysel predkladať nám ju ako nejaké triumfálne vyvrátenie."

"Ale aká je tá *skutočná odpoveď*,?"

"Teraz by som rád presunul k téme, ako logika zabíja malé tulenie mláďatká..."

“Zabíjate čas,” vyštekne inšpektor Darwin.

“Navyše, nesledujete ovce,” poviem a prihodím ďalší kamienok.

Inšpektor Darwin pozerá na dvoch diskutérov, očividne neochotných ustúpiť zo svojich pozícií. “Počujte,” povie Darwin, tentokrát prívetivejšie. “Mám jednoduchý návrh, ako vyriešiť vašu dišputu. Vy tvrdíte,” povie, ukazujúc na Marka, “že názory ľudí menia ich osobnú skutočnosť. A vy pevne veríte,” jeho prst sa otočí smerom na Autreya, “že Markove názory *nedokážu* zmeniť skutočnosť. Takže nechajme Marka naozaj pevne veriť, že vie lietať, a nech skočí z tohto útesu. Mark uvidí sám seba odletieť ako vták a Autrey ho uvidí padnúť dole a rozpleštiť sa, obaja budete spokojní.”

Zastavíme sa a uvažujeme o tom.

“Znie to rozumne...” povie Mark nakoniec.

“Útes je hneď tu,” konštatuje inšpektor Darwin.

Autrey sa tvári mimoriadne sústredene. Nakoniec vykrikuje: “Počkajte! Keby to bola pravda, my všetci by sme sa už dávno odseparovali každý do svojho vlastného vesmíru, a v tom prípade by všetci ostatní ľudia boli iba výtvormi našej predstavivosti... a potom nemá zmysel nám niečo dokazovať...”

Z blízkeho útesu zaznie dlhý, postupne tichnuci výkrik, nasledovaný tupým šplechnutím a tichom. Inšpektor Darwin vo svojom zápisníku nalistuje stranu zobrazujúcu súčasný genofond a zapíše si o čosi nižšiu frekvenciu pre Markove alely.

Autrey vyzerá mierne zdesene. “Bolo to naozaj nutné?”

“N nutné?” povie inšpektor Darwin nechápavo. “Proste sa to *stalo*... Nerozumiem celkom vašej otázke.”

Autrey a ja sa vrátíme k nášmu vedierku. Je čas zaháňať ovečky. Na túto časť nesmiete zabudnúť. Pretože inak, aký by to celé malo zmysel?

Pôvodný článok s komentármi¹

¹ <http://yudkowsky.net/rational/the-simple-truth>