

# Racionalita a biasy

# Najskôr hádanka

- Mám pravidlo (zatiaľ tajné) na základe ktorého formujem trojice čísel.
- Čísla 2, 4, 6 sú príkladom, ktorý spĺňa toto pravidlo
- Hádanka funguje takto: dávajte mi trojice čísel aj vám poviem či splňajú pravidlo.
- Ked' si budete myslieť, že poznáte moje pravidlo, tak sa ho pokúste uhádnuť. Ak ho uhádnete, vyhrávate. Trojíc vám dám kol'ko chcete.

# Čo je to racionalita? (podľa Yudkowskeho)

1. **Epistemic rationality:** systematically improving the accuracy of your beliefs.
  - Snaha spresňovať náš obraz o svete
2. **Instrumental rationality:** systematically achieving your values.
  - Snaha vyberať akcie tak, aby smerovali budúcnosť k dôsledkom, ktoré majú pre nás vyššiu preferenciu

Predstavte si to ako dve úlohy racionálneho agenta:

1. Spoznávať prostredie
2. Dosiahnuť cieľ'

# Epistemic rationality

- Vy ste mozog
- Váš mozog používa vstupy zo zmyslov na to aby si formuloval model sveta.
- Zmätenie / prekvapenie znamená, že ste dostali vstup, ktorý je nekonzistentný s vašou predstavou o svete. Neznamená to, že je niečo zlé so svetom, len s vašim modelom o svete.
- Kúzelnícke triky sú prekvapivé práve vďaka tomu, že popierajú našu predstavu o fungovaní sveta, nie fungovanie sveta.
- Zmätenie pomáha zlepšovať model nášho sveta (2-4-6 task, HPMOR ch. 8).

# Instrumental rationality

- Snaha vyberať akcie tak, aby smerovali budúcnosť k dôsledkom, ktoré majú pre nás vyššiu preferenciu
- Hľadáme optimálne kroky k tomu aby sme maximalizovali nejakú užitočnosť.
- Neznamená to “účel svätí prostriedky” alebo dosiahnuť cieľ za každú cenu.
- Neznamená to automaticky ani na úkor ostatných
- Všetko záleží od našej „objective function“, ktorá pravdepodobne zahŕňa aj život v okolitom svete (snáď tu nie je žiadny sociopatický tyran, ktorý chce ovládnuť celú galaxiu)

# Čo je to “Bias” ~ Systematická odchýlka

- Predstavte si že máte urnu obsahujúcu neznámy počet bielych a červených guličiek (70 a 30)
- Náhodne bez vracania vyberiete 10 z nich a na základe ich farby spravíte odhad.
- Ak ich bude 7 a 3, tak viete celkom dobre spraviť odhad. Ak 6 a 4, tak trochu horšie, ale cca dobre.
- Ak pokus zopakujete veľa krát, tak v priemere budete celkom presní.
- Čím viac sa učíte (opakujete pokus), tým presnejší máte odhad.

# Ale čo ak sú biele guličky t'ažšie a klesnú na dno?

- Biele guličky klesnú dolu a je teda väčšia pravdepodobnosť, že vyberiete červenú.
- Vybraná vzorka teda bude nereprezentatívna a bude v nej nejaká konzistentná chyba.
- Tento *bias* sa volá štatistický
- Ak je metóda učenia sa o svete vychýlená, tak s postupom učenia sa odhad nespresňuje. Môže sa dokonca zhoršovať. - Toto je celkom strašidelné: Čím viac sa učíme, tým je náš obraz sveta nepresnejší. Racionalita je snaha bojovať s biasmi.
- Príklad o moriakovi.

# Cognitive (Kognitívny) bias

- Odteraz ak poviem bias, tak mám na mysli cognitive bias. Inak poviem.
- Analogicky k štatistickému biasu.
- Štatistický bias skresľuje vzorku tak, že nereprezentuje verne populáciu.
- Kognitívny bias skresľuje spôsob akým uvažujeme.
- Čo ak veríme, že červené guľôčky majú nejakú úžasnú schopnosť. Môžeme trpieť optimizmom (optimism bias, wishful thinking) a môžeme nadhodnocovať pravdepodobnosť výskytu.

# Úloha

- Predstavte si že ste stretli plachú osobu. Čo je pravdepodobnejšie, že je to knihovník alebo predavač?

# Príklad cognitive biasov

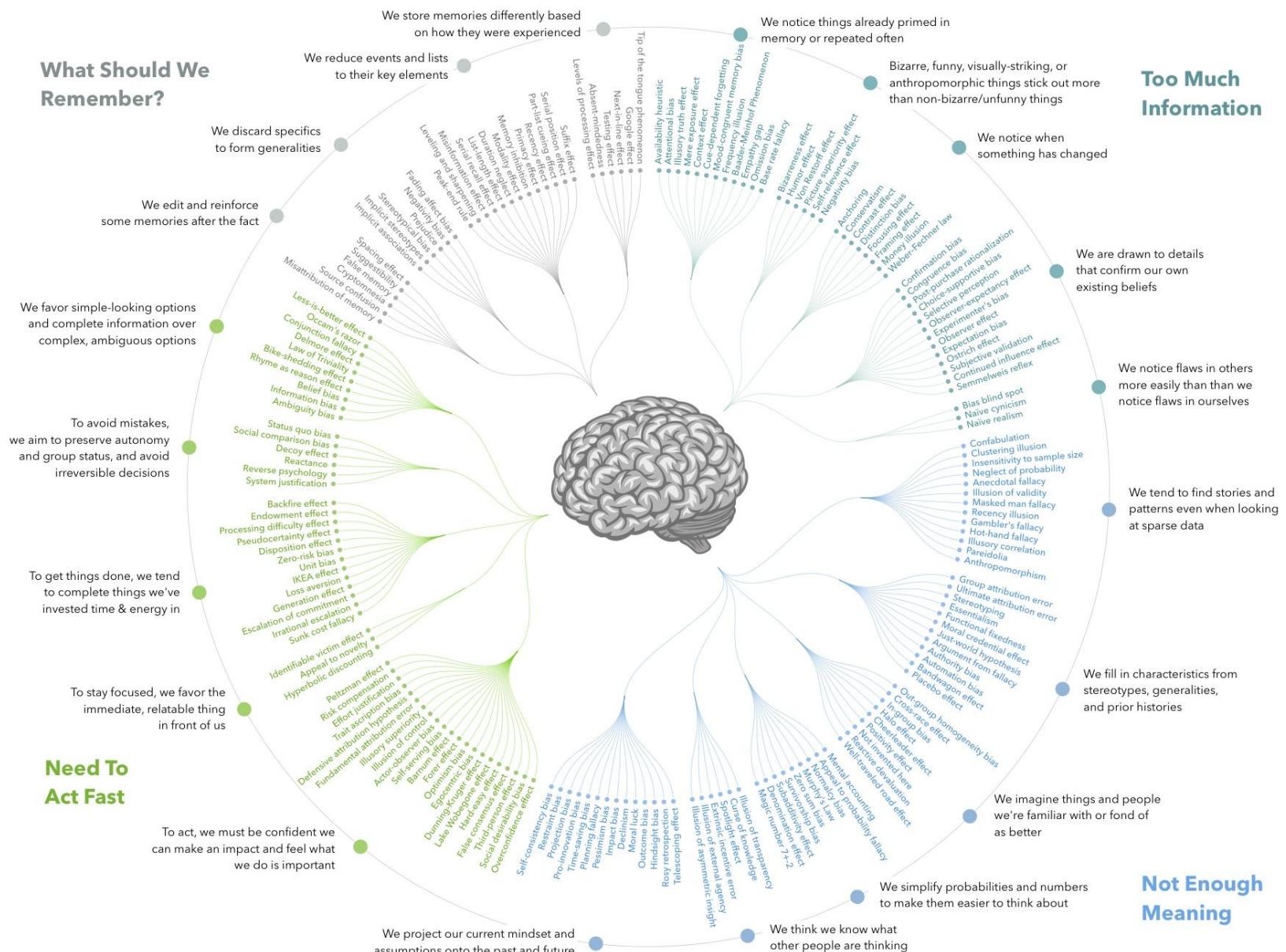
- Úloha: Predstavte si že ste stretli plachú osobu. Čo je pravdepodobnejšie, že je to knihovník alebo predavač?
- Ak ste odpovedali knihovník, tak ste sa pravdepodobne stali obeťou biasu: *base rate neglect*.
- Postavili ste odhad na tom ako charakteristika dobre opisuje objekt a nie na tom ako častá je táto charakteristika v populácií. Predavačov je v populácií tak veľa, že aj tých plachých je viac ako plachých knihovníkov (v USA je vraj 75 x toľko predavačov ako knihovníkov)

# Iný bias: *sunk cost fallacy*

- Tendencia cítiť záväzok k veciam, do ktorých ste venovali nejaké zdroje v minulosti namiesto toho aby ste sa poučili a posunuli sa ďalej.
- Ikea to používa na predávanie nábytku
- Dá sa to použiť na to aby vás mali ľudia radšej :)

Biasov existuje strašne veľa

COGNITIVE BIAS CODEX



Problém je, že človek sa nestáva imúnnym voči biasom ak o nich vie.

Neznamená to ani že si ich všimnete v akcii.  
Preto manipulácia funguje tak dobre.

Treba na to veľa tréningu a pozornosti.

# Ako vzniká (cognitive) bias?

Obmedzenia rationality:

1. Nedokážeme pozorovať celú realitu. Musíme jednať na základe neúplnej a neperfektnej informácie.
2. Máme obmedzené výpočtové prostriedky. Vieme preskúmať len obmedzený priestor možností.
3. Smie biasnutý. Nevieme priradiť presné pravdepodobnosti možným výsledkom.

Ak by sme sa mali správať čisto racionálne, tak by sme nevedeli spraviť ani to najmenšie rozhodnutie - Bounded rationality

- good enough decisions - satisficing
- Herbert Simon (Nobelova cena) James March

# Bounded rationality znamená, že musíme robiť nejaké optimalizácie.

- Mozog používa heuristiky – ktoré vedú k nedokonalostiam v uvažovaní.
- Biasy sú vlastne nedokonalosti týchto heuristík
- Bias nie je spôsobený nedostatkom informácií a ani nedostatkom výpočtovej sily. Je spôsobený skratkou, ktorú používame na to aby sme tieto problémy obišli.

# Planning fallacy

- Denver International Airport otvorené o 16 mesiacov neskôr a o 2 miliardy \$ drahšie ako plánovali
- Eurofighter Typhoon 54 mesiacov neskôr a za 19 namiesto 7 miliárd \$
- Sydney Opera House otvorený 1973 namiesto 1963 za 102 miliónov \$ namiesto 7 miliónov \$
- ....
- Počuli ste už o pravidle 80% hotovo?

- Ak odhadujeme náročnosť, tak očakávaný odhad je veľmi podobný odhadu v ktorý dúfame.
- Najhorší odhad (worst case scenario) je ale často oveľa miernejší ako najhorší, ktorý je možný.
- Úlohou nie je vymysliť sekvenciu udalostí, ktorá viedie k zániku vesmíru ale k odhadu pravdepodobnosti nastania takej sekvencie.
- Túto pravdepodobnosť podceňujeme a pravdepodobnosť očakávaného priebehu preceňujeme. Súvis s optimism bias a wishful thinking, len pri plánovaní.
- Ako bojovať s týmto biasom? - “outside view” namiesto “inside view”.
  - Insight view znamená odhad pravdepodobnosti na základe plánu. Má oveľa viac detailov ale zároveň podporuje optimizmus keďže myslíme na to ako niečo spraviť a nie ako sa to môže pokaziť.
  - Outside view (napr. porovnanie s predchádzajúcimi podobnými úlohami) má menej detailov ale porovnáva reálne úlohy bez sústredenia sa na nejaké vybrané podčasti (overfitting?).

# Môj „oblúbený“ bias – Hindsight (Spätný pohľad)

- Ak poznáte odpoveď na nejakú otázku/úlohu/udalosť, tak prisudzujete oveľa väčšiu pravdepodobnosť nastaniu tejto možnosti.
- Stalo sa vám, že Vám niekto vynadal: „Ako si toto mohol spraviť, nebolo ti jasné ako to dopadne?“
- „Po vojne je každý generál.“
- „Racionalizácia“
- Problém je, že tento bias nám bráni byť dostatočne prekvapený skutočnosťou a teda nám bráni nachádzať problémy v našom mentálnom modeli.

# Chcecklist for rationality habits

<https://www.lesswrong.com/posts/ttGbpJQ8shBi8hDhh/checklist-of-rationality-habits>

# Zdroje

- [LessWrong - What Do We Mean By "Rationality"?](#)
- [LessWrong - Biases: An Introduction](#)
- [LessWrong - Planning Fallacy](#)
- [LessWrong - Hindsight Devalues Science](#)