Josh Brekel

The Base-Rate-Fallacy Fallacy

When attempting to avoid the base-rate fallacy leads to error

The University of Utah

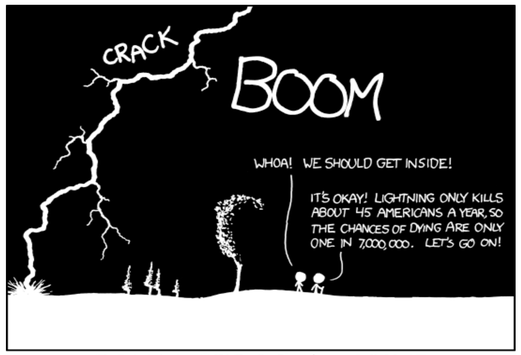
Medical Diagnosis Case*

One in 1,000 people have some disease, D. You have a test for the presence of D with the accuracy described here:

Disease Test Accuracy		Test Result: Positive	Test Result: Negative
Disease D	Present	0.9	0.1
	Absent	0.1	0.9

You randomly select a person to test. The test yields a positive result. How confident should you be that this person has D? Answer: a little less than 1%. According to Tversky and Kahneman, people tend to neglect base rates and focus on the accuracy statistics about the test. This is the **Base-Rate Fallacy.**

*Adapted from Titelbaum (2022) and Royall (1997).



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Munroe, Randall. *xkcd*, "Conditional Risk." https://xkcd.com/795

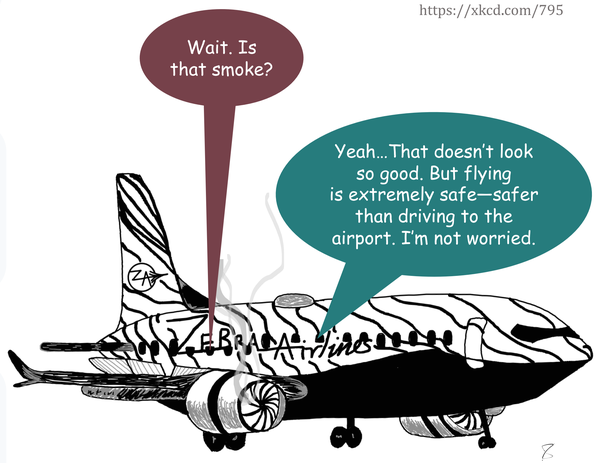
Bayes's Theorem (Odds Form)

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\text{Base Rate}}{1 - \text{Base Rate}} \times \frac{\Pr(E|H)}{\Pr(E|\neg H)}$$

H: hypothesis of interest
E: some evidence (stated in the form of a proposition)
Base Rate: prevalence of H in the population of interest

Epistemic Base-Rate-Fallacy Fallacy

Occurs when agents neglect evidence from likelihoods due to their focus on a low base rate (i.e., prior probability).



Base-Rate Neglect (& Fixation?)

Neglect

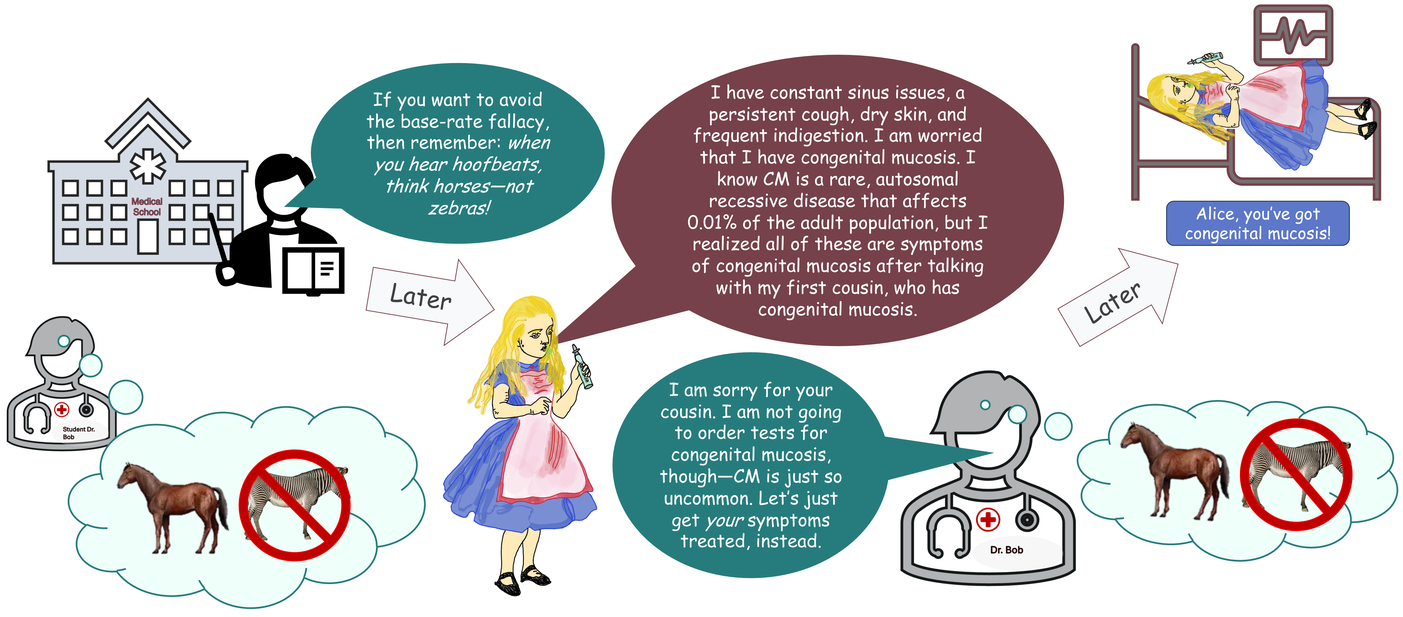
$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\text{Base Rate}}{1 - \text{Base Rate}} \times \frac{\Pr(E|H)}{\Pr(E|\neg H)}$$

Fixation

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\text{Base Rate}}{1 - \text{Base Rate}} \times \frac{\Pr(E|H)}{\Pr(E|\neg H)}$$

Decision-Theoretic Base-Rate-Fallacy Fallacy

Occurs when agents decide not to seek more evidence solely because of the presence of a low base rate.



What it isn't:

- Inverse fallacy (Koehler, 1996)
- Inverse base-rate effect (Don et al, 2021)
- All cases of conservatism effect (Phillips and Edwards, 1966; Howe et al., 2022)

Competing Explanations:

- Unproblematic Bayesian search
- Failure to use the correct base-rate / failure to update on *all* evidence
- Proper updating without reaching action threshold

Further Questions

- Testing for whether and how often the BRFF occurs: experimental design ideas?
- Are there other competing explanations not considered here?
- Can you think of other cases of the BRFF?

Acknowledgements: I am grateful to Tania Lombrozo, Jonah Schupbach, Matt Haber, Tom Griffiths, and Gesiel B. da Silva for feedback on this project.

References

Don, H. J., Worthy, D. A., and Livesey, E. J. (2021). Hearing Hooves, Thinking Zebras: A Review of the Inverse Base-Rate Effect. *Psychonomic Bulletin & Review*, 28(4):1142–1163.

Howe, P. D. L., Perfors, A., Walker, B., Kashima, Y., and Fay, N. (2022). Base Rate Neglect and Conservatism in Probabilistic Reasoning: Insights from Eliciting Full Distributions. *Judgment and Decision Making*, 17(5):962—-987.

Koehler, J. J. (1996). The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges. *Behavioral and Brain Sciences*, 19(1):1–17.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a Simple Probability Inference Task. *Journal of Experimental Psychology*, 72(3):346–354.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm.* Chapman & Hall.

Titelbaum, M. G. (2022). *Fundamentals of Bayesian Epistemology 1: Introducing Credences.* Oxford University Press.

Tversky, A. and Kahneman, D. (1982). Evidential Impact of Base Rates. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment Under Uncertainty: Heuristics and Biases*, pp. 153–160. Cambridge University Press.

My website