Confirmation Bias Emerges from an Approximation to Bayesian Reasoning

Charlie Pilgrim (collaborators: Adam Sanborn, Eugene Malthouse, Thomas T Hills)

BIASR Model

1. **Source Reliability, R**
2. **Simultaneous updating, P(H,R|D)**
3. **Independence Approximation, P(H)P(R) ≈ P(H,R)**
4. **Sequential Updating**

> *Confirmation bias*, as the term is typically used in the psychological literature, connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand.[1]

1. Source Reliability

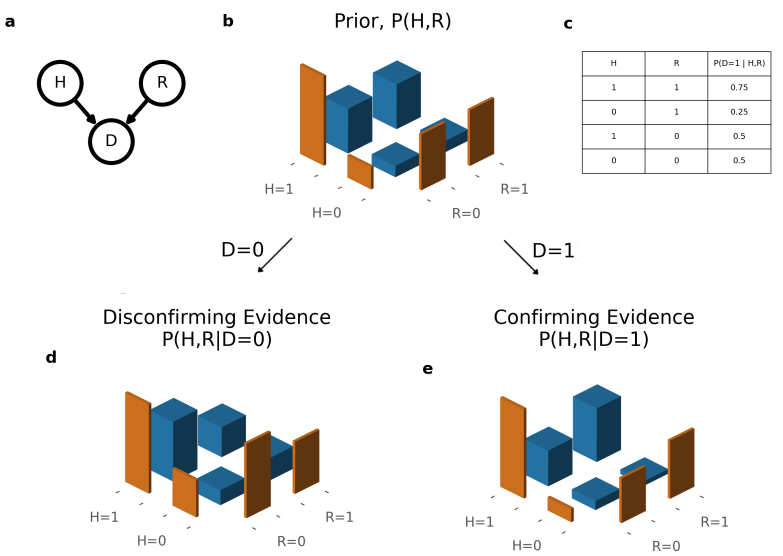
How reliable is Bob?

People track source reliability. [2,3]

2. Simultaneous updating

Someone tells you Elvis is outside. How reliable are they?

People update beliefs together. [4, 5]



H	R	P(D=1 | H,R)
1	1	0.75
0	1	0.25
1	0	0.5
0	0	0.5

Simultaneous updating of source reliability, R, and the central hypothesis, H. Blue (thick) bars show the joint belief distribution. Orange (thin) bars show the marginal belief distributions. a) The Bayesian network structure. b) Prior beliefs favour the central hypothesis, and are neutral about source reliability. The posterior following d) disconfirming and e) confirming evidence.

3. Independence Approximation

Upon receiving data, beliefs in H and R are no longer independent. [6]

Size of joint belief distribution:

2 beliefs -> 2^2=4

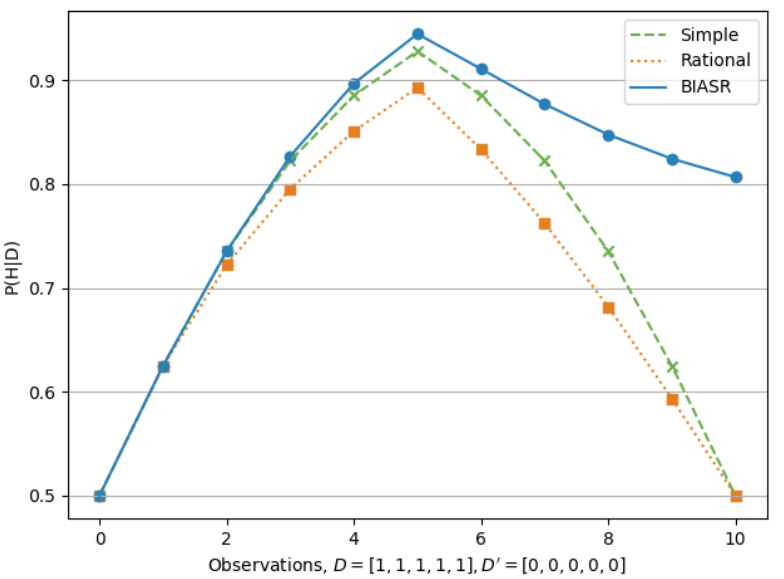
20 beliefs -> $2^{20} \approx 10^6$

300 beliefs -> $2^{300} \approx 10^{90}$

Unrealistic memory requirement.

4. Sequential Updating

Small errors magnify.



Data for, then data against, the central hypothesis are received from different sources given neutral initial priors in both the central hypothesis and source reliability. Under simple and rational models, the belief in the central hypothesis returns to the prior belief. With the BIASR model, biased assimilation dynamics mean that the data received earlier has a stronger effect on posterior beliefs than data received later.

	Simple Version of Bayes Theorem	Biased Evaluation Prior to Assimilation	Bayesian Updating Including Source Reliability	Bayesian Networks	BIASR. Bayesian updating with an Independence Approximation and Source Reliability
Biased Evaluation		✓	✓	✓	✓
Biased Assimilation		✓			✓
Attitude Polarisation				✓	✓
Belief Perseverance		✓			✓
Selection of Sources					✓

Comparison of information processing models of confirmation bias.

References

1. Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. Rev. general psychology 2, 175–220 (1998).
2. Lord, C. G., Ross, L. & Lepper, M. R. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. J. personality social psychology 37, 2098 (1979).
3. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. Am. journal political science 50, 755–769 (2006).
4. Gershman, S. J. How to never be wrong. Psychon. bulletin & review 26, 13–28 (2019).
5. Dunbar, K. How scientists really reason: Scientific reasoning in real-world laboratories. The nature insight 18, 365–395 (1995).
6. Pearl, J. Causality (Cambridge university press, 2009).