

21. Kelly 2008: Is Confirmation Bias Reasonable?

Kevin Dorst
kevindorst@pitt.edu

PHIL 1555, Rationality

Problem with Nielsen and Stuart 2018: the *direction* of polarization is predictable.

E.g. what happens when people with different opinions are both presented with mixed evidence?

They tend to strengthen prior beliefs.

Confirmation Bias: People are inclined to interpret mixed evidence in a way that favors their prior beliefs.

“Biased assimilation” of evidence

Example: we disagree over *Deterrent*; both presented with two studies, one favoring and one disfavoring.

Deterrent = capital punishment has a deterrent effect

S_1 : States A and B are next to each other; A has capital punishment, B does not; A has lower murder rate.

S_2 : States C and D are next to each other; C has capital punishment, D does not; they have the *same* murder rate.

Result? I increase my confidence in *Deterrent*; you decrease yours.

Why think irrational?

- Dogmatism?
 - No. Rather, *selective scrutiny*.
 - Reasonable because of limited cognitive resources. Scientists.
- Failures of *commutativity*?
 - Kelly: properly understood, this is *not* a commutativity failure—so not necessarily irrational.

Key Epistemological Fact: Given evidence E , how confident you should be in H depends on what other alternative explanations of the evidence (H' , H'' , ...) you are aware of.

Obviously right? Additivity.

Obviously wrong? Update by conditioning.

Doesn't change hypothesis space.

Subtly right? Bayesian model too simplified. Cognitively limited agents must generate hypotheses¹ on the fly, on an as-needed basis.

¹ Or, perhaps, *label* them, generating evidence about them.

Subtly wrong?

→ If we *don't* know we do this, can't explain how we can predict our own polarization.

→ If we *do* know we do this, shouldn't we “price in” the effects of selective scrutiny in how we adjust our credences?

Reflection: $cr(A) = \mathbb{E}_{cr}(cr^+(q))$

Q: What do people think?