Henderson and Gebharter 2021, Bayesian Polarization?

Kevin Dorst 24.223 Rationality

Lord et al setup: conflicting studies; biased assimilation.

Polarization: parallel vs. contrary updating; convergent vs. divergent.

Could this be rational?

Too easy? Subjective Bayesians can disagree on any likelihoods you like, and therefore can update in opposite ways given the same evidence.

Some constraints from eg Principal Principle, but not nearly enough

Too hard? Subjective Bayesians who agree on *everything* can't disagree on anything given the same evidence.

Even if they only agree on *comparative likelihoods* (whether H is more to be expected given E or given $\neg E$), then can't diverge.

Is there any fruitful middle-ground?

H&G: have them agree on the "basic structure" of the evidential situation, where that means their beliefs can be represented with the same *Bayes net* (+ conditional probability values):

- A set of variables that can take different values.
- A *directed acyclic graph* between them, where an edge represents a direct (causal or) evidential dependency.
- A set of conditional probabilities specifying the probability of a variable taking a given value, given the values of its parents.

Propositions as *indicator variables*

A probability function P *factorizes over* the Bayes net iff it satisfies the conditional probabilities and the **Markov condition**: nodes are independent of their non-descendants, given their parents.

So all *roots* of the graph are independent

Eg H is whether you have covid and E is whether you test positive. Say: $H \to E$ where $P(E|H) = 0.8$ and $P(E|\overline{H}) = 0.2$.
$\to$ Any choice of $P(H)$ allowed.

Some shared Bayes nets permit divergent updating, others don't.

This one doesn't; always increase credence in H given E

Source reliability models:

- Single hypothesis H.
- A bunch of independent sources that can each either be reliable (R_i) or unreliable ($\overline{R_i}$).
- A bit of evidence E_i from each source R_i.
  If reliable (R_i), then E_i is perfect indicator of H.
  If unreliable ($\overline{R_i}$), E_i is just noise.

Their reliability is also independent of H. We'll come back to this.

$P(E_i|\overline{R_i}H) = P(E_i|\overline{R_i}\,\overline{H}) = a$.
Let's set $a = 0.5$

<u>Results:</u>
Single-source models: no polarization. Why? Suppose E.
We both agree that (1) if reliable, conclusive evidence for H; if not reliable, no evidence at all; and (2) R itself not evidence for H.

$\Rightarrow$ We both boost credence in H, proportional to how likely we thought it was that the source was reliable.

Two sources? Possible polarization!
We learn $E_1\overline{E_2}$. Alice takes the favoring E_1 to be stronger than the disfavoring E_2; Bob, vice versa. Eg:
$P_A(H) = 0.75$, $P_A(R_1) = 0.6$, $P_A(R_2) = 0.4$; and
$P_B(H) = 0.25$, $P_B(R_1) = 0.4$, $P_B(R_2) = 0.6$.
$\Rightarrow P_A(H|E_1\overline{E_2}) \approx 0.84$ and $P_B(H|E_1\overline{E_2}) \approx 0.16$.

Gets graph of credences-over-time right.

Worry 1: Requires differential priors on source reliability, which is implausible for the Lord et al setup.

Right, but seems plausible for many real-world settings[1].

[1] Unlike Jern et al.'s examples

Example. I'm a Republican, you're a Democrat, and we both watch Tucker and Rachel. Tucker says it's a great policy (E_1), Rachel says it's terrible ($\overline{E_2}$). I trust Tucker, you trust Rachel; so we diverge further.

Worry 2: Independence.
Two people can agree on a Bayes net only if they both treat the roots of the graph independently.
$\Rightarrow$ We agree on a source reliability model iff we think each source's reliability is independent of the others and of the hypothesis in question.

This can *sound* commonsensical: we might know that they are causally independent. **But that's a mistake.**

H&G use this reasoning to justify, saying the assumption "makes sense when the truth of the hypothesis does not influence the reliability of the source, nor does the reliability of the source influence the truth of the hypothesis" (10261).

Generally, known causal independence does *not* entail probabilistic independence. I can know that whether the coin lands heads on the first toss (h_1) is *causally* independent of the second toss (h_2), while still not "treating them independently".
$\rightarrow$ If it's an unknown bias, I *won't* treat them independently: h_1 is evidence for *heads-bias*, and therefore should boost my credence in h_2.

In the application of interest, we need each person to be such that learning source-1 *is* reliable has no affect on their credence in source-2 being reliable, nor on their credence in H. That's implausible in most real-world cases.

We all agree that *if* Rachel is reliable, then it's less likely that Tucker is.

What to make of this?