The Common Ground – Rational Polarization

Marcello Di Bello - ASU - Spring 2023 - Week #6

The topic of today's class is polarization and whether rational or irrational mechanisms give rise to it.

Polarization occurs when people belonging to different communities, walks of life, socio-economic demographics, etc. endorse widely different opinions about certain matters of fact, and their confidence in these different opinions is also quite high.

What explains this phenomenon? Does it occur because of a failure of processing evidence correctly or a failure to gather evidence by one group? We discuss today the possibility of 'rational polarization' by reading Kevin Dorst's paper.[1] Another paper we will keep in the background to inform the discussion is by Thomas Kelly.[2]

[1] Kevin Dorst, Rational Polarization Forthcoming in *The Philosophical Review*.

[2] Kelly (2008), Disagreement, Dogmatism, and Belief Polarization, *The Journal of Philosophy*, 105(10):611-633.

Doorst Rational Polarization

The Standard Story (§1)

The standard story about why people polarize about an issue—say whether guns increase safety—is this:

> predictable polarization is due to epistemic irrationality—the fact that people's beliefs are insufficiently constrained by evidence. Instead, people glom onto the beliefs of their peers, confirm and entrench those beliefs, and become wildly overconfident in them. Combined with the informational traps of the modern internet, we have a simple explanation of the rise of polarization. (p. 2)

One part of this story is empirical, but the other is normative: polarization happens because of irrationality. Kevin Dorst shows that another story is plausible: people's rationality paired with ambiguity in the evidence they process can explain polarization. Hence, polarization can be rational.

(Im)possibility Theorems (§3)

Here is a formal way to sketch the idea (pp 9-10):[3]

> Theorem 3.1. Given No Ambiguity, P values $\tilde{P}$ iff P obeys Reflection toward $\tilde{P}$.
>
> Theorem 3.2 (Informal). Whenever $\tilde{P}$ is ambiguous but valued by some P, Reflection fails.

These two theorems (roughly) say this. Whenever we process evidence in a non-ambiguous manner—that is, we are always sure how

[3] P and $\tilde{P}$ are probability functions representing one's degrees of belief, where P is the prior and $\tilde{P}$ the updated one. The expression "P values $\tilde{P}$" means that $\tilde{P}$ is a rational update of P, for example, one that improves expected accuracy. Reflection is defined on p. 5, roughly, it says that your prior (rational) degree of beliefs in any q must equal your expected updated (rational) degree of belief in any q. If Reflection holds, an individual's rational degrees of belief cannot expectedly polarize.

to respond to the evidence and form beliefs—then processing evidence rationally rules out polarization. However, when the evidence we process is ambiguous but we still process it rationally, then we do tend to polarize. So, ambiguity tracks polarization closely.

Ambiguity (§4)

What does it mean for evidence to be ambiguous? Consider a word-completion task: P_G_ER. Suppose you cannot find the completion. That's your evidence. What does this evidence tell you? That either the prompt is non-completable or completable but you could not find how to complete it.[4] Contrast with this completion task: P_A_ET. Here you find a completion: planet. This evidence shows the prompt was completable. There is an asymmetry between evidence for completable words and evidence for non-completable words. The latter is ambiguous evidence, the former is not.[5]

Now, compare these two individuals. Here is Haley:

> She's wondering whether a fair coin landed heads. I'll show her a word-search determined by the outcome: if heads, it'll be completable; if tails, it'll be uncompletable. Thus her credence in heads equals her credence it's completable. She'll have 7 seconds, then she'll write down her credence. She knows all of this. .... Intuitively: it's easier for her to assess her evidence when the string is completable (when the coin lands heads) than when not. So if heads, her credence should (on average) increase a lot; if tails it should (on average) decrease a bit; and the average of 'increase a lot' and 'decrease a bit' is 'increase a bit' (p. 11)[6]

And here is Thomas:

> Like Haley, he's about to see a word-search, determined by the (same) coin toss. But while she'll see a completable string iff heads, he'll see a completable string if tails. By parallel reasoning, Thomas's opinion should expectably polarize in the opposite direction: it'll be easier for him to assess his evidence if tails than if heads, so his average posterior rational credence in heads should be lower than 50%. (p. 14)

Because of the different setup they are in, Haley will (on average) increase her confidence that the coin is heads, while Thomas will (on average) do the opposite. So they will split apart, rationally.[7]

Predictable polarization (§5)

The argument so far has shown that Haley and Thomas will *expectedly* polarize, but does not show they will *predictably* polarize.[8] We can get predictable polarization by iteration:

[4] You are unsure about how you should respond to the evidence that you did not find a word. You doubt you might have missed something and the prompt is actually completable.

[5] There is *ambiguity-asymmetry* between completable and uncompletable tasks.

[6] The bottom line is that Haley should expect to increase her confidence in heads. If the coin lands heads, she can get clear evidence of completion and thus increase her confidence in heads a lot (on average). If the coin lands tails, her evidence will be ambiguous. So she should decrease her confidence in heads by only a little.

[7] Crucially, both Thomas and Haley update their beliefs in a way that increases expected accuracy. So they are behaving rationally.

[8] What is the difference?

> Haley knows the coin is fair but rationally estimates that the rational posterior is around 58%. So if we can repeat this with many independent fair coins and searches, since she's initially confident that around half the coins will land heads, she predicts that her average credence in $Heads_1, \ldots, Heads_n$ should rise to around 58%. Since they're independent, she can predict that she should become confident that around 58% landed heads and very confident that more than half landed heads. Since she's initially 50-50 in the latter, that's predictable polarization. (p. 17)

There is a question whether this iteration violates rationality.[9] But, if we assume that Haley can forget things and cares only about certain issues and not others—say she cares whether the majority of coins landed hands, but she does not care to remember the details of each completion task—then Haley would *predictably rationally* polarize about the proposition 'the majority of coins landed heads'.[10]

Dorst, borrowing an expression from Brian Hedden,[11] calls predictable polarization a form of *diachronic tragedy*:

> at each stage she expects the next step to make her more accurate and later ones to make her less so—despite knowing that once she takes the next step, she'll then expect the later ones to make her more accurate, and so will be willing to take them. This is the slippery slope to radicalization (p. 19)

This is all about word completion. How about the real world?

> For $Heads_i$ and $Tails_i$ substitute bits of evidence for and against the claim that guns increase safety. Going to a liberal university made me a Tailser—made me better at recognizing evidence against that claim. Living in a conservative town made Dan a Headser—made him better at recognizing evidence favoring that claim. Neither of us became worse at assessing evidence; we became better, in asymmetric ways. When we discuss individual facts (a school shooting; a case of self-defense), we often agree on which way they point. Yet since time and memory are limited, we are left disagreeing about high-level claims (guns increase safety) while being unable to share all the (rational) reasons we have for our differing opinions. (p. 20)[12]

Selective Scrunity (§6)

Ambiguity-asymmetries may arise in the empirical process that drive polarization. One such processes is confirmation bias. In particular, consider *biased assimilation*, the tendency to interpret mixed evidence as supporting your preferred hypothesis:

> Biased assimilation is driven by selective scrutiny: people spend more time looking for flaws with incongruent evidence than congruent evidence (p. 23)

[9] Spell out exactly what the problem might be.

[10] Theorem 5.1 (p. 19): "Haley can start out 50% confident of h, know that each update in a sequence is valuable with respect to how all the coins land (hence whether h), and yet predict with arbitrary confidence that the sequence will make her arbitrarily confident of h." (h stands for: the majority of coins landed heads)

[11] Hedden (2015), Options and Diachronic Tragedy, *Philosophy and Phenomenological Research*, 90(2):423-451

[12] Question: This picture suggests that rational polarization is partly driven by limited computational resources and memory. If we allow for memory failures, could Reflection fail and thus allow for polarization without ambiguity? In other words, what assumptions about memory are embedded in theorem 3.1, p. 9?

Suppose different groups engage in selective scrutiny intentionally.[13] Dorst shows that (1) it is rational to engage in selective scrutiny and (2) selective scrutiny leads to polarization.

First, selective scrutiny is rational in that it improves the accuracy of our beliefs.[14] This is a good reason to engage in selective scrutiny:

> Why do I tend to scrutinize incongruent studies over congruent ones? Because I expect doing so to make me more accurate, since it's more likely I'll be able to find a flaw, avoiding ambiguity. I may think it's more likely to contain a flaw—but even if I don't, I'll be more likely to find any flaws it contains. After all, part of being convinced of a claim is learning how to rebut arguments against it. (p. 24)

Second, groups that engage in selective scrutiny will polarize. The driving mechanism of the polarization is that one group is better at recognizing flaws in studies that favor a propositions q of interest, while the other better at recognizing flaws in studies against q:

> two groups of agents face a series of choices about which of two random studies to scrutinize. They start out 50% confident in a claim q, and at each stage they scrutinize in the way they expect to make their beliefs most accurate. But one group (red) is better at recognizing flaws in studies that tell against q, and the other (blue) is better at recognizing flaws in those that tell in favor of q. The result is polarization (Figure 6). (p. 25)

[13] Kelly (see below) argues that selective scrutiny should be done unintentionally for it to be rational. Dorst instead holds that selective scrutiny can be intentional and rational.

[14] See Figure 5, p. 25. Dorst wrtes: "I randomly generated models and compared P(Find | Flaw) to expected accuracy, finding a robust correlation (Figure 5, left). I then generated pairs of models in which you're (on average) more likely to find flaws that exist in the incongruent than the congruent study; expected accuracy quite often warrants scrutinizing the former (Figure 5, right)." (p. 25)

Kelly on Belief Polarization

Consider:

> *The Key Epistemological Fact*: For a given body of evidence and a given hypothesis that purports to explain that evidence, how confident one should be that the hypothesis is true on the basis of the evidence depends on the space of alternative hypotheses of which one is aware.

This fact justifies—normatively—why people scrutinize more forcefully claims that go against what they already believe compared to claims that agree with what they believe.[15] Then, polarization can be a rational phenomenon, or perhaps between a rational and an irrational phenomenon, so long as people are not maliciously or intentionally scrutinizing only claims that contradict their beliefs, but they do that as a result of the alternative hypotheses salient to them.

Kelly writes:

> a person who subjects apparent counterevidence to greater scrutiny … but is non-culpably ignorant of this, seems to constitute something of an intermediate case. On the one hand, he is unaware of the fact that a biasing factor played a role in his arriving at this body of total evidence. On the other hand, his agency is complicit in the fact that he now possesses a biased sample of evidence; the biasing mechanism is located in him. (end of section IV).

[15] How does that work exactly?