

# Why Rational People Polarize

Kevin Dorst

U.S. politics is beset by increasing polarization. Ideological clustering is common; partisan antipathy is increasing; extremity is becoming the norm (Dimock et al. 2014). This poses a serious collective problem. Why is it happening? There are two common strands of explanation.

The first is psychological: people exhibit a number of “reasoning biases” that predictably lead them to strengthen their initial opinions on a given subject matter (Kahneman et al. 1982; Fine 2005). They tend to interpret conflicting evidence as supporting their opinions (Lord et al. 1979); to seek out arguments that confirm their prior beliefs (Nickerson 1998); to become more confident of the opinions shared by their subgroups (Myers and Lamm 1976); and so on.

The second strand of explanation is sociological: the modern information age has made it easier for people to fall into informational traps. They are now able to use social media to curate their interlocutors and wind up in “echo chambers” (Sunstein 2017; Nguyen 2018); to customize their web browsers to construct a “Daily Me” (Sunstein 2009, 2017); to uncritically consume exciting (but often fake) news that supports their views (Vosoughi et al. 2018; Lazer et al. 2018; Robson 2018); and so on.

So we have two strands of explanation for the rise of American polarization. We need both. The psychological strand on its own is not enough: in its reliance on fully general reasoning tendencies, it cannot explain what has *changed*, leading to the recent rise of polarization. But neither is the sociological strand enough: informational traps are only dangerous for those susceptible to them. Imagine a group of people who were completely impartial in searching for new information, in weighing conflicting studies, in assessing the opinions of their peers, etc. The modern internet wouldn't *force* them to end up in echo chambers or filter bubbles—in fact, with its unlimited access to information, it would free them to form opinions based on ever more diverse and impartial bodies of evidence. We should not expect impartial reasoners to polarize, even when placed in the modern information age.

In short, I agree with the standard story: the above-mentioned “reasoning biases” are an important component of the polarization process—they are the tendencies which make us susceptible to modern informational traps (Sunstein 2009, 2017; Lazer et al. 2018; Robson 2018). What I disagree with is the claim that these “reasoning biases” are *biases*. Rather: I'll argue that these tendencies—to interpret conflicting evidence as confirmatory; to search for confirming arguments; and to react to discussion by becoming more extreme

—may be the result of fully rational processes. In particular, they can all arise when rational people are sensitive to the fact that some forms of evidence are systematically more *ambiguous* than others. Now, I can't establish that the way people actually exhibit these tendencies is rational. But I can establish that it *could* be—that our explanation of the rise of polarization doesn't *require* individual irrationality.

Why does this matter? Because it changes the nature of the collective problem we face. Step back. *Whenever* we observe a collectively bad outcome for a group of interacting people, there are two quite different kinds of explanations we can give: we can locate the problem in the *choices* of the individuals or the *structure* of their interaction. To illustrate, contrast two cases:

**Case 1:** We face a heart disease epidemic—we would all be better off if we all ate better.

**Case 2:** Our fisheries are being depleted—we would all be better off if we all fished less.

The bad outcome in Case 1 is due to individual failings: each of us would have a better outcome if we *unilaterally* ate better—we are irrational to overeat. Thus the collectively bad outcome “filters up” from suboptimal choices of the individuals. Call this an **individual problem**. In contrast, the bad outcome in Case 2 is due to structural failings: none of us could get a better outcome by unilaterally changing our decision—we are *rational* to overfish. The problem is that we are caught in a tragedy of the commons: although the total fish population would be larger (and so fishers more productive) if everyone fished less, the best way for an individual to maximize their catch is to overfish (Hardin 1968). Thus the collectively bad outcome “filters down” from the suboptimal structure of the group. Call this a **structural problem**.

To put it bluntly: individual problems arise because people are *dumb*; structural problems arise because people are *smart*. You solve an individual problem by getting people to make better choices—for example, by educating them about nutrition. You solve a structural problem by changing the choices they face—for example, imposing fines for overfishing.

Back to politics. As commonly articulated, the standard story represents polarization as an *individual* problem: it implies that polarization arises, ultimately, from irrational informational choices on behalf of individuals. It therefore predicts that the solution will require teaching people to make better ones. Educate them; inform them; improve them.

If I'm right, this is a mistake. The informational choices that drive polarization are not irrational or suboptimal. Like the tragedy of the commons, polarization is a structural

problem. It arises because people are exquisitely sensitive to the informational problems they face—because they are *smart*. It doesn't require educating them; it requires altering the choices they face.

So much for motivation; here are the three reasoning tendencies I will focus on:

**Biased assimilation:** People tend to interpret conflicting evidence as supporting their prior beliefs (Lord et al. 1979)

**Confirmation bias:** People tend to seek out new arguments that confirm their prior beliefs (Nickerson 1998).

**Group polarization:** When people discuss their opinions in groups, their opinions tend to become more extreme in the direction of the group's initial inclination (Myers and Lamm 1976).

How could these tendencies be rational? The answer I'll give relies on three epistemological facts.

**Fact 1:** Some types of evidence are *more ambiguous* than others: it is harder (even for rational people) to know the rational way to react to them.

To illustrate, compare two cases. The question is whether the coin I just flipped landed heads. First case: all you know is that it's a fair coin. It's obvious that this evidence should lead you to be 50-50 on whether the coin will land heads—the evidence is unambiguous. Second case: you know it's a fair coin; after flipping it, with a slight grin I say, "You might want to think about heads." This *might* be evidence that it landed heads. But maybe the grin indicates that I'm trying to trick you, and it's actually evidence that it landed tails. But *maybe* I'm thinking that you'll think that, and the grin indicates that I'm trying to *double-trick* you—in which case it's actually evidence for heads! But *maybe...* You see where this is going. In short, it's far from obvious whether my comment should lead you to be more or less than 50-50 confident that the coin landed heads—in other words, this evidence is ambiguous.

Why does evidential ambiguity matter? Because of the second epistemological fact:

**Fact 2:** The more ambiguous a piece of evidence is, the *weaker* it is: if you should be quite unsure how you should react to a piece of evidence, then that evidence shouldn't lead to a radical shift in your opinion.

For example, in response to my comment that "You might want to think about heads," it might be reasonable to be more confident than not that it landed heads—but it is certainly not reasonable to be very confident it landed heads. Contrast this with a third case: I

flatly say “It landed heads.” In this third case, you *should* be very confident it landed heads. The reason for the difference in strength of evidence is due to the fact that in the former case the evidence is ambiguous, while in the latter case it’s not.

Why are Facts 1 and 2 important? Because rational people can use them to guide their choices:

**Fact 3:** Some information-gathering strategies will *predictably* yield more ambiguous (hence weaker) evidence than others.

For example, suppose you can choose whether (1) I make a comment about the coin in person—with all the facial and inflectional subtleties that involves—or (2) I simply send either the word ‘heads’ or ‘tails’ over text message. You can expect that the evidence you’d receive from (1) will be more ambiguous—and so, by Fact 2, weaker—than the evidence you’d receive from (2). So it could be rational to prefer (2) to (1).

Philosophers and economists have given various models of ambiguous evidence (Levi 1974; Joyce 2005; Dorst 2018), but they all agree on Facts 1–3. Here’s why these facts are important: *the ambiguity of a piece of evidence is correlated with whether it confirms or disconfirms your prior beliefs*. That is, there are fairly general settings in which rational people can expect that receiving evidence that (probably) *confirms* their opinion will result in less evidential ambiguity than receiving evidence that (probably) *disconfirms* their opinions. Since the former will (other things equal) be stronger, it can thereby be rational to choose to receive it—even though doing so predictably leads to strengthening prior beliefs.<sup>1</sup> I’ll first explain two general reasoning scenarios in which this dynamic occurs, and then discuss how our analysis of them can rationalize biased assimilation, confirmation bias, and group polarization.

First reasoning scenario: *exploring explanations*. You believe something—say, that lowering the corporate tax doesn’t help the economy. You are presented with conflicting pieces of evidence (studies, news items, etc.)  $E_1$  and  $E_2$ —the former tells in favor of this belief, the latter tells against it. For example, suppose  $E_1$  is a study of five countries that lowered the corporate tax but didn’t see significant changes in economic growth; meanwhile  $E_2$  is a study of five countries that lowered the corporate tax and *did* see a rise in economic growth. You only have time to scrutinize one of these studies more closely—to examine it for methodological flaws, mistaken assumptions, and to (more generally) explore *alternative explanations* that may remove the study’s support for its conclusion.

---

<sup>1</sup> This can be given a rigorous formulation in a Bayesian setting—Salow (2017) shows how. He offers a strategy to avoid the result, but it can be shown that this strategy works only if it rules out evidential ambiguity entirely (Dorst 2018).

For example, you might look closer at  $E_2$  to see what other economic changes were happening during the time studied that could account for the rise in growth.

Which strategy—scrutinizing  $E_1$  or  $E_2$ —is rational, given your background beliefs? Here's the crucial fact about this reasoning scenario: *successfully finding* an alternative explanation leads to evidence that is less ambiguous than *trying and failing* to find such an explanation. Start with an example. You are scrutinizing  $E_2$ . Case One: you find an alternative explanation—you realize that the study took place during a global economic boom. Upon discovering this, you should think, “*Of course* economic growth was rising—there's no reason to think it was the lowering of corporate taxes that did it.” In other words, when you successfully debunk the study, you wind up with unambiguous evidence—you know exactly how to react. Case Two: you think about the study for a while, but come up with no good alternative explanation of the rising economic growth. Upon failing to discover an explanation, you should think, “Maybe there's no other explanation—in which case this study *does* support cutting the corporate tax. But on the other hand, maybe there's a perfectly good explanation and I just haven't thought of it yet.” In other words, when you fail to debunk the study, you wind up with ambiguous evidence: you should be unsure whether (1) there is *no* available alternative explanation, or (2) there *is* one that you should have (but didn't) think of.

More generally, when you set out to debunk a piece of evidence, there are two possibilities: either (1) there is an available alternative explanation, or (2) there is none. If (1), then the evidence doesn't actually support its conclusion; if (2), then it does. Crucially, if you *find* an alternative explanation, then you know that you're in possibility (1)—you have unambiguous evidence. But if you fail to find an alternative explanation, you don't know whether that's because you're in case (2)—where there is none—or instead because you're in case (1)—where there is one, but you've (irrationally) failed to generate it. So if you fail to find an alternative explanation, you wind up with ambiguous evidence. By Fact 2, that means that trying and failing to find an alternative explanation provides you with *weaker* evidence than successfully finding one.

Where does this leave us? You can try to debunk either  $E_1$  or  $E_2$ . If your attempt is successful, that'll give you less ambiguous—hence stronger—evidence than if your attempt is unsuccessful. Since rational people try to get the strongest evidence possible, you should try to debunk the one that you think you're most likely to succeed at debunking. Which one is that? The one that tells *against* your belief, of course! (You have reason to think that evidence against your belief is more likely to be misleading than evidence in its favor—otherwise you wouldn't believe it!) Upshot: if you're sensitive to evidential ambiguity, then when presented with conflicting studies it is rational—other things equal—to try to debunk the one that conflicts with your prior beliefs.

That covers exploring explanations—our first general reasoning scenario where evidential ambiguity is predictably correlated with your prior beliefs. The second scenario is *assessing arguments*. Observation: when you are presented with (minimally reasonable) arguments for a claim, this tends to provide you with more evidence in favor of that claim than against it. This may seem obvious—that’s the point of arguments, after all. But the reason is actually quite subtle—it hinges on evidential ambiguity.

You might think that hearing an argument for a claim *C* *always* provides you with evidence in favor of *C*. But that’s not right—sometimes it provides evidence against *C*. Example: we know that a lawyer will try to construct the best argument possible for the claim that her client is innocent. Suppose that upon hearing her argument we can tell that it’s quite a bad one. Then it actually provides evidence that her client is *guilty*—for upon hearing it we should think, “That’s a bad argument. If he was innocent, she could probably construct a better one. So he’s probably not innocent.” In short, the mere fact that you’re hearing an argument in favor of a claim doesn’t mean that you’re getting evidence for it.

The real reason why arguments tend to provide evidence for their conclusions has to do with the differing levels of evidential ambiguity that they generate—and, in particular, with our (in)ability to *recognize* whether they are good or bad arguments. The issue is that bad arguments tend to masquerade as good ones—and thereby mislead you into thinking that they’re good ones—whereas good arguments rarely masquerade as bad ones. So when you come across a good argument, you’ll have relatively little reason to think that it’s a bad one; but when you come across a bad one, you’ll have relatively more (misleading) reason to think that it’s a good one.

Now consider what happens when you are presented with an argument for claim *C*. Either (1) it’s a good argument, or (2) it’s a bad one. If (1), then it provides evidence for *C*; if (2), then it provides evidence *against* *C*. The crucial asymmetry is your (in)ability to recognize whether you’re in case (1) or (2). If (1) it’s a good argument, you should be relatively sure that it’s good—thus you have relatively unambiguous evidence in favor of *C*. But if (2) it’s a bad argument, then since it will be masquerading as a good one, you should be relatively *unsure* that it’s bad—thus you have relatively *ambiguous* evidence against *C*. By Fact 2, ambiguous evidence is weaker than unambiguous evidence. So being presented with a good argument for *C* provides relatively strong evidence for *C*, whereas being presented with a bad argument provides relatively weak evidence against it. *This* is why, on the whole, hearing arguments for *C* tends to give you more evidence in favor of *C* than against it.

Where does this leave us? Suppose you face a choice between hearing one of two arguments,  $A_1$  or  $A_2$ —one *in favor* of your prior belief, one against it. If the one you

choose is good, being presented with it will provide less ambiguous (so stronger) evidence than if it's bad. Since you should try to get the strongest evidence possible, you should choose to be presented with the argument that you think is most likely to be good. Which one is that? The one that argues in favor of your prior belief, of course! (You have reason to think there are more good arguments for your belief than against it— otherwise you wouldn't believe it!) Upshot: if you're sensitive to evidential ambiguity and can choose between hearing two arguments, it is rational—other things equal—to listen to the one that supports your prior beliefs.

I've argued that there are two general reasoning scenarios—exploring explanations and assessing arguments—in which rational sensitivity to evidential ambiguity can lead you to prefer an information-search strategy that can be expected to confirm your prior beliefs. Suppose this is right. Then our psychological biases are rational.

First, *biased assimilation*—people's tendency to interpret conflicting evidence as supporting their prior beliefs. The empirical regularity appears to be due to the fact that people spend more energy trying to debunk the piece of evidence that tells against their prior opinions than the piece tells in favor of them (Lord et al. 1979; Kelly 2008). This should sound familiar: it is precisely this strategy which I argued is rational in the context of exploring explanations. Finding an explanation results in less ambiguity than failing to find one. A rational person who's interested in the truth should be trying to get strong evidence, and so to avoid ambiguity. That means—other things equal—they should search for alternative explanations in the place they expect to find them, focusing on debunking the evidence that conflicts with their prior beliefs. So rational people exhibit biased assimilation.

Second, *confirmation bias*—people's tendency to seek out new arguments that confirm their prior beliefs. That is, if people are given a choice of which slant of argument to hear (or—what comes to the same thing—which news source to watch), they will tend to choose the one that agrees with their opinion. This should sound familiar: it is precisely the strategy I argued is rational in the context of argument assessment. Hearing a good argument results in less ambiguity than hearing a bad one. A rational person who's interested in the truth should be trying to get strong evidence, and so to avoid ambiguity. That means—other things equal—they should prefer hearing arguments that they expect to be good, rather than those they expect to be bad. And, in general, they should expect arguments that agree with their beliefs to be better than those that disagree with them. So rational people exhibit confirmation bias.

Third, *group polarization*—the tendency for a group's opinion to become more extreme after discussion, in the same direction as the group's initial tendency. The empirical regularity appears to be due to two mechanisms: first, people discuss more arguments in

favor of the group's initial opinion than against it; and second, people are influenced by merely hearing of the opinions of others (Isenberg 1986; Baron et al. 1996). We can also use our results to explain why these effects could be rational. The first one is easy. We've seen that evidential ambiguity explains why being presented with arguments in favor of a claim C tends to provide you with evidence in its favor. Since groups that have an initial tendency to believe C will tend to know more arguments in favor of it than against it, sharing them will tend to provide more evidence for C than against it.

What about the effect of others' opinions? In general, learning that others believe some claim C provides evidence for C. However, in our political context it's common knowledge that loads of people agree with your views and loads of people disagree with them. So shouldn't these conflicting bits of evidence balance out? No—for we're back in our exploring-explanations scenario. You're presented with two pieces of conflicting evidence: these people all agree with your belief; but *these* people all *disagree* with it. When you have conflicting evidence and other things are equal, what's the rational thing to do? Spend more energy debunking the evidence that conflicts with your prior beliefs! It's therefore rational to spend time coming up with alternative explanations for why your political opponents disagree with you—and thus (upon coming up with such explanations) rational to give more weight to the political opinions of those who agree with you. So rational people also exhibit group polarization.

In short: there's reason to think that even fully rational people—those who are exquisitely sensitive to the potential ambiguity of their evidence—would tend to interpret conflicting evidence as confirmatory, tend to search for confirming arguments, and tend to react to group discussion by becoming more extreme. If that's right, then the fact that *you and I* exhibit these tendencies is no indictment. Rather, it may reveal just how exquisitely sensitive to our evidential situation we are.

Let's pull the strands together. We need to appeal to both new sociological trends and general psychological tendencies in order to explain the recent rise of political polarization. Many construe this rise as an *individual* problem—as a case where suboptimal individual choices explain the suboptimal collective outcome. I've argued that this may be a mistake. When we pay close attention to the subtle informational choices people face—in particular, the way in which they must navigate ambiguous evidence—we see that the relevant psychological tendencies may reveal *optimal* individual choices. If that's right, then we face a *structural* problem: the modern internet has created a social-informational situation in which optimal individual choices give rise to suboptimal collective outcomes.

If this is right, we should think of political polarization less as a health epidemic and more as a tragedy of the commons. That means two things. First, it means focusing less

on changing the choices people make, and more on changing the options they're given. And second, it means recognizing just how smart the "other side" is: that those who disagree with you may do so for reasons that are subtly—exquisitely—rational.

## References

- Baron, Robert S., Hoppe, Sieg I., Kao, Chuan Feng, Brunzman, Bethany, Linneweh, Barbara, and Rogers, Diane, 1996. 'Social corroboration and opinion extremity'. *Journal of Experimental Social Psychology*, 32(6): 537–560.
- Dimock, Michael, Doherty, Carroll, Kiley, Jocelyn, and Oates, Russ, 2014. 'Political polarization in the American public'. Pew Reserach Center.
- Dorst, Kevin, 2018. 'Higher-Order Uncertainty'. In Mattias Skipper Rasmussen and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, To appear. Oxford University Press.
- Fine, Cordelia, 2005. *A Mind of its Own: How Your Brain Distorts and Deceives*. W. W. Norton & Company.
- Hardin, Garrett, 1968. 'The Tragedy of the Commons'. *Science*, 162(3859): 1243–1248.
- Isenberg, Daniel J., 1986. 'Group Polarization. A Critical Review and Meta-Analysis'. *Journal of Personality and Social Psychology*, 50(6): 1141–1151.
- Joyce, James M, 2005. 'How Probabilities Reflect Evidence'. *Philosophical Perspectives*, 19: 153–178.
- Kahneman, Daniel, Slovic, Paul, and Tversky, Amos, eds., 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Kelly, Thomas, 2008. 'Disagreement, Dogmatism, and Belief Polarization'. *The Journal of Philosophy*, 105(10): 611–633.
- Lazer, David, Baum, Matthew, Benkler, Jochai, Berinsky, Adam, Greenhill, Kelly, Metzger, Miriam, Nyhan, Brendan, Pennycook, G., Rothschild, David, Sunstein, Cass, Thorson, Emily, Watts, Duncan, and Zittrain, Jonathan, 2018. 'The science of fake news'. *Science*. 359 (6380): 1096–1096
- Levi, Isaac, 1974. 'On indeterminate probabilities'. *The Journal of Philosophy*, 71(13): 391–418.
- Lord, Charles G., Ross, Lee, and Lepper, Mark R., 1979. 'Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence'. *Journal of Personality and Social Psychology*, 37(11): 2098–2109.
- Myers, David G. and Lamm, Helmut, 1976. 'The group polarization phenomenon'. *Psychological Bulletin*, 83(4): 602–627.
- Nguyen, C. Thi, 2018, April 9. 'Escape the echo chamber'. *Aeon*.

Nickerson, Raymond S., 1998. 'Confirmation bias: A ubiquitous phenomenon in many guises.' *Review of General Psychology*, 2(2): 175–220.

Robson, David, 2018, April 17. 'The myth of the online echo chamber'. *BBC*.

Salow, Bernhard, 2017. 'The Externalist's Guide to Fishing for Compliments'. *Mind*. To appear.

Sunstein, Cass R., 2009. *Republic.com 2.0*. Princeton University Press.

Sunstein, Cass R., 2017. *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Vosoughi, Soroush, Roy, Deb, and Aral, Sinan, 2018. 'The spread of true and false news online'. *Science*, 359(6380): 1146–1151.