



ARTICLE

The Polarization Paradox

Brandon Carey 

Department of Philosophy, California State University, Sacramento, Sacramento, CA, USA
Email: brandon.carey@csus.edu

(Received 1 August 2024; revised 29 September 2025; accepted 5 October 2025)

Abstract

The Dogmatism Paradox begins with the claim that I know some proposition p and uses apparently good reasoning to draw a seemingly irrational dogmatic conclusion: that I should therefore dismiss any new evidence against p . A standard solution to this paradox is to note that when I acquire new evidence against p , that evidence defeats my justification for believing p . As a result, I no longer know that p , and so the reasoning used to generate the paradox begins with a false premise.

By appealing to recent work in social epistemology by Endre Begby and C. Thi Nguyen, I develop a new, stronger version of the Dogmatism Paradox that is immune to this standard solution. This version of the Dogmatism Paradox has significant consequences for contemporary polarized political disagreements, in which subjects on both sides of the disagreement have reason to distrust new evidence against their beliefs. So, unless the paradox can be solved, many political disagreements in sufficiently polarized communities will turn out to be rationally intractable in the sense that agreement can only be reached by one side irrationally revising their beliefs.

Keywords: Polarization; dogmatism; the dogmatism paradox; echo chamber; evidential preemption

The Dogmatism Paradox begins with the claim that I know some proposition p and uses apparently good reasoning to draw a seemingly irrational dogmatic conclusion: that I should therefore disregard any new evidence against p . A standard solution to this paradox, the “defeat solution,” relies on the idea that when I acquire new evidence against p , that evidence defeats my existing evidence for believing p . As a result, I no longer know that p , and so the reasoning used to generate the paradox begins with a false premise.

However, recent work in social epistemology by Endre Begby and C. Thi Nguyen demonstrates that subjects can sometimes rationally mistrust any future sources that offer evidence against propositions they believe, and a new version of the paradox arises for these “evidentially polarized” subjects: if I have good reason to believe that p and further good reason to believe that any sources offering evidence against p are untrustworthy, I can again use apparently good reasoning to draw the worryingly dogmatic conclusion that I should disregard any new evidence against p . The defeat solution to the original paradox is ineffective against this version, as any new evidence

against p that would ordinarily defeat my justification for p is itself defeated by my pre-existing good reasons to distrust sources offering evidence against p .

This new paradox has significant consequences for contemporary polarized political disagreements, in which subjects on both sides of the disagreement have reason to think that the sources of evidence used by the other side are not trustworthy. In addition to the practical problem of how to persuade those who disagree with us about basic facts about the world, the paradox raises an epistemic problem: if the reasoning in the paradox is sound, then it would be irrational for either party in such a disagreement to revise their beliefs, no matter what evidence is presented by the other side. So, unless the paradox can be solved, many disagreements in sufficiently polarized communities will turn out to be rationally intractable in the sense that agreement can only be reached by one side irrationally revising their beliefs.

1. The Dogmatism Paradox

The Dogmatism Paradox was first proposed by Saul Kripke (2011), but its most elegant presentation comes from Gilbert Harman (1973), who notes that we can reason as follows for any proposition h that we know:

“If I know that h is true, I know that any evidence against h is evidence against something that is true; so I know that such evidence is misleading. But I should disregard evidence that I know is misleading. So, once I know that h is true, I am in a position to disregard any future evidence that seems to tell against h .” This is paradoxical, because I am never in a position simply to disregard any future evidence even though I do know a great many different things. (p.148)

Though Kripke and Harman both formulate the puzzle in terms of knowledge, it is inessential to the puzzle that I know the initial proposition – we can derive the same dogmatic conclusion merely from the premise that it is epistemically rational for me to believe that h is true. I assume here *evidentialism* about epistemic rationality, according to which it is epistemically rational to have exactly the doxastic attitudes that currently fit one’s evidence.¹ There are many incompatible theories of what constitutes someone’s evidence, but what I have in mind is roughly what Earl Conee (2001) calls a person’s “data”: “the nonderivative indications that the person has of the truth value of the proposition” (p.102). On this conception, one’s evidence consists of the basic indicators one has of the truth values of various propositions, which I take to include perceptual experiences, memories, instances of testimony, and so on. One’s evidence may also include less obvious considerations that bear on the truth value of the target proposition, such as one’s awareness of different hypotheses that can account for one’s other data (see Thomas Kelly (2008) on this point), so long as that awareness serves as an indicator of the proposition’s truth that is not entirely derived from other items of evidence. However, in my view, a subject’s evidence does not include data that they could have had, but do not in fact have. So, for example, what Emily McWilliams (2021) calls a “motivated defeater,” a piece of defeating evidence that a subject could have acquired by further reflection on their other data had they been differently motivated, is not included in a subject’s evidence, nor is, for example, evidence that the subject could easily have acquired through more responsible inquiry. The general picture, then, is that what

¹Maria Lasonen-Aarnio (2014) takes this to be at least a necessary condition on epistemic rationality in her discussion of this paradox. In my view, it is also sufficient for determining which is the rational attitude, though whether one holds that attitude based on the evidence that makes it rational is a further question.

attitude it is rational for you to have toward a proposition at a time is determined by the information you have to go on at that time in deciding what to believe about the world. The truth value of the proposition, whether you believe it, the causal relationship between that truth value and your belief, what you would have believed or what evidence you would have had in other nearby possible worlds, and any number of other factors that we might think are relevant in determining whether you know a proposition are thus irrelevant to the epistemic rationality of believing that proposition.

With this picture of epistemic rationality in mind, we can generate the same sort of paradox in any case where it is epistemically rational for a subject to believe some proposition. For example, sitting in the airport, I have good evidence from my memory of checking my itinerary this morning that my flight is at 5 pm. So, I reason:

1. My flight is at 5 pm.
2. If my flight is at 5 pm, any evidence that it is not is misleading evidence.
3. I should disregard misleading evidence.
4. So, I should disregard any evidence that my flight is not at 5 pm.

Believing 1–3 fits my evidence. 1 is straightforwardly supported by the memory of my itinerary. If, as is common in discussions of this puzzle, we take “misleading evidence” to just mean evidence against a true proposition, then 2 is supported by a brief reflection that, since my flight is at 5 pm, any evidence that it is not is evidence against a truth and therefore misleading. On the assumption that my goal is to believe the truth, misleading evidence will, by definition, thwart my efforts to achieve that goal, which gives me good reason to accept 3. And 1–3 collectively support that evidence that my flight is not at 5 pm is evidence of a type that I should disregard, given that I am trying to believe what’s true, giving me good reason to accept 4. But 4 recommends the same kind of irrationally dogmatic attitude that Harman inferred from knowing *h*. This attitude seems irrational in part because, although I have good evidence for 1, that evidence is intuitively *defeasible*, such that it can be undermined or overridden by other evidence. For example, if I hear an announcement in the airport that my flight is at 7 pm, I have good reason based on 4 to think that I should disregard this evidence. However, I am seemingly not in an epistemic position to rationally disregard this evidence (even if it is in fact misleading) – my memory of my itinerary is fairly good evidence, but it does not seem so good as to trump any possible counterevidence. Instead, I should heed the announcement and give up my belief that my flight is at 5 pm. And yet, prior to the announcement, it is epistemically rational for me to believe that my flight is at 5 pm, and each step in the reasoning from 1 to 4 seems to preserve that rationality. Since I can reason similarly for any proposition that it is epistemically rational for me to believe, the paradox threatens similarly implausible dogmatism for all of my beliefs.

1.1. The defeat solution

Fortunately, the idea of defeating evidence that grounds the intuition that 4 is irrational also grounds a straightforward solution to the puzzle. Harman’s initial presentation of this solution is nearly as elegant as his explanation of the paradox:

Since I now know that [*h* is true], I now know that any evidence that appears to indicate something else is misleading. That does not warrant me in simply disregarding any further evidence, since getting that further evidence can change what I know. In particular, after I get such further evidence I may no longer know

that it is misleading. For having the new evidence can make it true that I no longer know that [h is true]; if I no longer know that, I no longer know that the new evidence is misleading. (Harman, pp.148–149)

The key insight here is that new evidence can change what I know, even if I previously knew that evidence to be misleading. As before, it seems to me that the reason for this has to do with the epistemic rationality of my belief, rather than any other necessary condition on knowledge. In the airport case, prior to the announcement, believing 1 fits my evidence. After the announcement, I now have new data: testimony from a seemingly reliable and authoritative source that my flight is at 7 pm. This evidence serves as a *rebutting defeater* for my previous evidence from my memory of my itinerary that my flight is at 5 pm by exerting evidential force in favor of an incompatible proposition such that believing 1 no longer fits my evidence. Depending on the details, perhaps I should instead believe that my flight is at 7 pm, or perhaps I should merely suspend judgment and go try to find more evidence. It can also become irrational for me to maintain a belief when I encounter an *undercutting defeater* for my evidence for that belief that undermines my reason for believing without supporting any incompatible proposition. If, for example, I received a message from my assistant apologizing for sending me the wrong itinerary this morning, that would undermine my reason for believing 1 and thereby my reason for accepting 4 (even if my assistant is mistaken and the original itinerary was correct). In either case, since my reason for accepting 4 depends on my evidence for 1, which has now been defeated, I no longer have a good reason to accept 4. So, encountering evidence against my initial belief defeats my reason for disregarding such evidence, providing a solution to the puzzle. Even if it is epistemically rational, prior to encountering evidence against my beliefs, for me to believe that such evidence is misleading and should be disregarded, once I acquire such evidence, it defeats the very evidence that would make me rational in disregarding it. Since that evidence is defeated, I no longer have a good reason to disregard this new evidence. So, we can accept that my reasoning in the airport case is good while still accounting for the intuition that I should not disregard new evidence that I encounter about the time of my flight, since encountering that evidence will defeat my reason for disregarding it.

2. The polarization paradox

My purpose here is not to evaluate the defeat solution to the Dogmatism Paradox, but rather to introduce a new paradox that arises for subjects with a certain kind of additional evidence. Let's say that a subject is **evidentially polarized** in favor of a proposition p just in case it is epistemically rational on that subject's evidence to believe both:

- (i) p, and
- (ii) any source offering evidence against p is epistemically untrustworthy about p.

"Epistemically untrustworthy" is an umbrella term meant to capture the many ways in which a source of evidence can be undeserving of our trust on a particular subject: sources might be significantly unreliable, uninformed, misinformed, deceptive, unconcerned with the truth, etc. So, an epistemically polarized subject has good reason to believe p and good reason to distrust any sources that offer evidence against p. When we have reason to think that a source is epistemically untrustworthy about a proposition, it seems plausible that we should disregard the evidence offered by that source about that

proposition, in the sense that we should not take that evidence as a direct reason to believe or disbelieve the proposition.² So, evidentially polarized subjects can reason as follows:

1. Any source offering evidence against *p* is epistemically untrustworthy about *p*.
2. I should disregard evidence against *p* from sources that are epistemically untrustworthy about *p*.
3. So, I should disregard any evidence I encounter against *p*.

But, as in the original Dogmatism Paradox, 3 seems irrationally dogmatic. In disregarding any future evidence against *p*, the subject treats their evidence for *p* as indefeasible, but, outside of a few special cases, our evidence for our beliefs always seems in principle defeasible.³

However, this Polarization Paradox is not as broad as the original Dogmatism Paradox, since it only applies to evidentially polarized subjects. And, though the first condition of evidential polarization is easily and commonly satisfied, the second condition looks quite strong. Under what conditions could I be rational in believing that *any* source offering evidence against one of my beliefs is untrustworthy? If I cannot, then we can solve this paradox before it even gets going by simply denying that it is possible for a subject to be evidentially polarized. However, recent work in social epistemology suggests that evidential polarization of this kind is not only possible but commonplace.

2.1. Evidential preemption

Endre Begby (2021) uses “evidential preemption” to name a kind of testimony in which a speaker testifies that *p* but also warns the audience that they will encounter (misleading) testimony for some proposition *q* incompatible with *p* using phrasing like:

My opponents will tell you that *q*; but I say *p*.

The idea is that this kind of testimony can give you a good reason to accept *p* while also warning you that you should expect other sources to offer you apparently good, but actually misleading evidence for *q*. The truth is that *p*, I might say, but you’ll hear some untrustworthy sources out there who will tell you *q* – be sure to ignore them. The intended effect of this kind of testimony is to provide a preemptive undercutting defeater for evidence that would otherwise count against *p*:

If that evidence were subsequently made available to me without forewarning, I would be rationally required to see it as cancelling out my reasons for believing that *p*. But now that the evidence has been preempted, the subsequent disclosure is neutralized and thereby rendered epistemically moot. (Begby 2021, p. 519)

As Begby points out, although many instances of evidential preemption, such as those that occur in conspiracy groups or partisan political discourse, seem obviously epistemically vicious, there are also cases where evidential preemption is a reasonable

²Disregarding a source’s evidence about *p* in this way is compatible with still taking that evidence as a reason to believe/disbelieve other propositions. For example, I may take a source’s testimony that *p* as a reason to believe that the source testified that *p* without taking it as a reason bearing on the truth of *p*. As we’ll see later, evidence against *p* that is disregarded in this way can also *indirectly support* *p* by reinforcing one’s trust in other sources.

³Perhaps, for example, I can reasonably disregard any evidence against the proposition that I exist, but the list of propositions for which this attitude seems reasonable is short and uninteresting.

way of warning people about misinformation that they might otherwise trust. We might do this by telling them not to trust such information or by just advising them to avoid it altogether. Charles Peirce describes the latter kind of case in his “The Fixation of Belief”:

I remember once being entreated not to read a certain newspaper lest it might change my opinion upon free-trade. “Lest I might be entrapped by its fallacies and misstatements,” was the form of expression. “You are not,” my friend said, “a special student of political economy. You might, therefore, easily be deceived by fallacious arguments upon the subject.” (1877, p.7)

This kind of advice seems like a common and reasonable part of our social epistemic lives, the sort of thing we teach in critical thinking and media literacy courses to help students protect themselves from bad information. Some sources are untrustworthy, so you should avoid or ignore them. In fact, being a rational epistemic agent often involves appreciating this kind of preemptive defeat:

If I know that I am reading *The Onion*, I am rational to disregard what might otherwise seem like a credible news story.

If I recall taking a powerful hallucinogen an hour ago, I am rational to disregard my visual evidence indicating that the walls are melting.

Knowing that “gut” impressions during job interviews are swayed by various kinds of bias but also notoriously difficult to ignore, I might advocate that our hiring committee not conduct interviews at all.

In all of these cases, epistemic rationality requires that I disregard or avoid evidence that I would otherwise use as a basis for belief, because I have pre-existing evidence that that new evidence will be misleading or otherwise defective. The reason for disregarding the new evidence can be tied to the specific source (*The Onion*), a period of time (until the hallucinogen wears off), or a kind of evidence (“gut” impressions). Importantly, this kind of rational disregard for new evidence can also be tied to what proposition the evidence supports. For example, suppose I receive the following message from the university’s IT department:

Our support team will never ask you to divulge your password. Persons or messages asking you to reveal your login or password are illegitimate even if they might seem to be legitimate. Be suspicious – ignore or delete such requests.

If I subsequently receive evidence that seems to indicate that a member of the support team is requesting my password, it seems rational to disregard that evidence, regardless of when, where, or from whom it comes. Furthermore, it seems rational for me to believe that the source of that evidence is untrustworthy, even if it otherwise seems trustworthy, simply because that source is trying to convince me that the support team is asking for my password. If that’s right, though, then I am evidentially polarized in favor of the proposition that the support team is not asking for my password: it is rational for me to believe that proposition and to believe that any source trying to convince me otherwise is untrustworthy. Generalizing from this case, the familiar practice of evidential preemption can cause subjects to become evidentially polarized in a range of ordinary cases.

2.2. Echo chambers

Echo chambers provide another good example of seemingly rational evidential polarization. Paraphrasing C. Thi Nguyen (2020), an echo chamber is an epistemic community in which sources offering evidence against the community's core beliefs are actively discredited. Cults and conspiracy theory groups are common examples, in which members are taught not only some claims about the world (e.g., that US astronauts never walked on the moon) but also that sources that offer evidence against these claims are untrustworthy for some reason (e.g., because they are part of a conspiracy designed to mislead the general public). Interestingly, though, echo chambers as defined are not obviously bad epistemic communities.⁴ This is because sometimes we have good reason to think that we are in an epistemic environment containing many untrustworthy sources that will offer misleading evidence against our beliefs, and in that kind of situation, discrediting those sources is a good epistemic practice. For example, we say in the Round Earth echo chamber that:

The Earth is not flat, but there are a number of people out there who will try to convince you that it is. They may even present you with arguments or experiments that seem to demonstrate that the Earth is flat. But these arguments and experiments are always misleading, and the people making them are either sincerely confused or maliciously deceptive. So, you should not trust what they have to say on this topic.

Sources that offer evidence against the belief that the Earth is not flat are discredited here, but there is no obvious epistemic sin involved. This seems like good advice for people heading into an epistemic environment that, for whatever reason, contains a bunch of misleading evidence about the shape of the Earth. In fact, we might think that we have an obligation to inoculate the uninformed against this kind of misinformation when we know they are likely to encounter it. Otherwise, we are knowingly letting them be fooled by purveyors of misinformation. Here again, the reason for distrusting the source is tied to the proposition that the source offers evidence for: the reason given for disregarding the evidence offered by Flat-Earthers is that it is evidence for the proposition that the Earth is flat. So, this kind of echo chamber seems to be another example in which subjects are evidentially polarized, with good reason to think any source offering evidence against some well-supported proposition (e.g., that the Earth is not flat) is untrustworthy about that proposition.

2.3. No defeat solution

If, as these examples suggest, evidential polarization is not only possible but in fact fairly common, then we either need to accept the dogmatic conclusion of the Polarization Paradox for the beliefs of quite a few subjects in seemingly ordinary circumstances, or we need to find some fault in the reasoning used to generate that conclusion. However, the reasoning in this paradox does not share the same flaw as the reasoning in the Dogmatism Paradox, and so the defeat solution proves ineffective here. The flaw in reasoning revealed by the defeat solution involves failing to account for the defeating effect of new evidence, but for evidentially polarized subjects, that effect is preempted –

⁴Nguyen's full account of echo chambers includes some additional claims about the way they function that may guarantee that echo chambers are epistemically bad in at least some respects. Those faults are not entailed by the thinner definition used here, though. See Westfall (2024) for more discussion of "good" and "bad" echo chambers.

they already have an undercutting defeater for any incoming evidence against *p*, since, on their evidence, the fact that a source offers evidence against *p* is a reason to think that source is epistemically untrustworthy about *p*. As a result, any evidence that would otherwise serve as a defeater for the subject's evidence for *p* will not succeed in defeating the subject's existing evidence for *p*. So, such evidence will not undermine the subject's reason for believing *p*, meaning that, even after the new evidence is acquired, the subject will remain evidentially polarized toward *p*.

For example, suppose, as before, that I recall that my itinerary said my flight was at 5 pm and also that I was told the following by the airline agent at the ticket counter:

Our automated announcement system is broken today. You'll probably hear it say that your flight is at some other time, but don't worry – it's definitely at 5 pm.

When I subsequently hear an announcement that my flight is at 7 pm, this does not seem to defeat my evidence that the announcement is misleading (and so should be disregarded) in the same way it did in the original case. This is because, in the original case, my only evidence that the announcement is misleading comes from my memory of my itinerary. Now, though, my reason for thinking that the announcement is misleading is not just my evidence that my flight is at 5 pm – the ticket agent has given me additional reason to think that I will hear misleading announcements about my flight time. This undercuts the force of the evidence I gain from the announcement, preventing it from having the defeating effect that the defeat solution relies on. Since evidentially polarized subjects are, by definition, rational to believe that any evidence against *p* comes from an untrustworthy source, they have a similar undercutting defeater for any evidence they encounter against *p*. So, while the defeat solution may solve the Dogmatism Paradox, it cannot solve the Polarization Paradox.

However, we might think that a modified defeat solution can solve the Polarization Paradox by denying that evidentially polarized subjects are entitled to entirely disregard evidence from sources that they rationally distrust. Perhaps evidence against *p* from sources that we rationally believe to be untrustworthy about *p* ought to be given considerably reduced weight in determining the rational attitude toward *p*, but it should still be given *some* weight.⁵ That is, perhaps Begby is mistaken in thinking that rationally preempted evidence is “epistemically moot” – instead, it merely has much less evidential force than it would have had it not been preempted. If this is right, then polarized subjects would not be rational in accepting premise 2 of the paradox, as new evidence against *p* should not be disregarded but merely downgraded. These subjects' evidence in favor of *p* would then no longer be in principle indefeasible, since sufficiently strong cumulative evidence against *p* from a sufficient number of independent sources could eventually outweigh it, even if each individual piece of counterevidence is given very little weight.

There are reasons, though, to think that merely giving less weight to evidence against *p* is not the rational policy for evidentially polarized subjects. First, in the kinds of cases that Begby and Nguyen use to illustrate evidential preemption and echo chambers, the overall structure of the subjects' evidence includes what Nguyen calls a “disagreement-reinforcement mechanism,” through which evidence presented against *p* indirectly strengthens the subjects' overall reason for believing *p* by reinforcing the credibility of their sources of evidence for *p*. If my sources have told me that *p* and also that there are many misleading sources who will try to convince me that *p* is false, then when

⁵Thanks to an anonymous referee for this journal for pressing this objection and the one in the next section, as a result of which this paper is considerably stronger and more interesting.

I subsequently encounter sources who offer evidence against *p*, “they are strengthening rather than weakening the evidential support for my belief that *p*. They are strengthening it, of course, not by providing direct evidence for *p*, but rather by providing evidence for the credibility of my testimonial source for believing *p*” (Begby 2021, p. 539). When I encounter yet another YouTube video presenting evidence that the Earth is flat, for example, this gives me no reason to doubt what I’ve been told by my fellow members of the Round Earth echo chamber. On the contrary, it is just one more piece of evidence that they are credible sources, since one of the things they told me is that I would see many such videos. In light of this additional evidence of their credibility, I ought to (slightly) increase my confidence in my beliefs that are based on their testimony, including that the Earth is not flat. In this way, encountering evidence that would otherwise count in favor of the proposition that the Earth is flat instead counts, weakly and indirectly, against it by reinforcing my trust in its critics. Given this, it seems implausible that I should count each new Flat Earth video I encounter as any reason at all to think that the Earth is flat, which is to say I should disregard these videos as evidence against my belief that the Earth is not flat.

Second, even in cases of evidential polarization without this kind of disagreement-reinforcement mechanism, merely giving less weight to evidence from a rationally distrusted source seems like an epistemic underreaction. If, for example, I rationally believe that my thermometer is broken and displays random values unrelated to the actual temperature, I should not merely decrease the weight I give to the evidence it provides, I should disregard it entirely. Even if the thermometer reads 80 degrees Fahrenheit and subsequent good evidence ultimately convinces me that that is the actual temperature, I should intuitively not count my broken thermometer’s reading among my reasons for this belief, given my background evidence about its malfunction. There is no disagreement-reinforcement mechanism at work – the reading gives me no indirect reason to think that the temperature is *not* 80 degrees – but the reading of a broken thermometer is no evidence at all about the temperature. So, the rational response to evidence from a source that we already rationally believe to be untrustworthy seems to be to disregard that evidence entirely. Similarly, if the editor of my favorite celebrity gossip site confesses that all of their stories are made up and I thereby come to rationally distrust the site, I should not merely downgrade those stories as items of evidence. I should instead disregard them and think that I no longer have any reason to believe what they report. Even if I later learn through some other credible source that one of the stories was in fact true, I should not count the initial story among my reasons for believing it, since I rationally believe that it was produced without regard for the truth. So, discovering that a source is untrustworthy seems to entirely undercut the force of the evidence it has provided, not merely to downgrade it.

So, since evidentially polarized subjects rationally believe that any source offering evidence against *p* is untrustworthy, they do seem entitled to disregard, and not merely assign less weight to, future items of evidence against *p*. As a result, these future items of evidence cannot come to cumulatively outweigh the subjects’ evidence in favor of *p* as the modified defeat solution suggests.

2.4. Accepting the conclusion

This immunity to defeat solutions might lead us to simply accept the conclusion of the Polarization Paradox. Although the conclusion of the paradox initially sounds irrationally dogmatic, evidentially polarized subjects always have good reasons to distrust sources offering evidence against their beliefs, just as I have good reasons to distrust announcements from a broken system, phishing e-mails asking for my

password, and the evidence offered by Flat-Earthers. As Kripke notes in discussing the original Dogmatism Paradox:

Sometimes the dogmatic strategy is a rational one. I myself have not read much defending astrology, necromancy, and the like . . . Even when confronted with specific alleged evidence, I have sometimes ignored it although I did not know how to refute it . . . One epistemological problem is to delineate the cases when the dogmatic attitude is justified. (p. 49)

So, perhaps we should think that the best response to the Polarization Paradox is to acknowledge that dogmatism is the rational attitude for evidentially polarized subjects, since they have additional evidence, beyond their evidence for *p*, that any evidence against *p* is misleading (or at least is not a reason to doubt *p*). The dogmatic conclusion of the original paradox seemed irrational at least in part because the subject lacked this kind of evidence, but since evidentially polarized subjects have it, they are rational in their dogmatism.

This response to the paradox would align with several recent arguments for the rationality of belief polarization, in which subjects either become more confident in their beliefs or adopt increasingly extreme beliefs in response to mixed or opposing evidence. Thomas Kelly (2008) has argued that belief polarization can be rational when subjects are exposed to a mixed body of evidence containing some evidence in favor of their beliefs and some against, since they are more likely to scrutinize evidence against their beliefs and uncover defeaters for that evidence. The resulting mix of some undefeated evidence for their beliefs and some defeated evidence against them then leads the subjects to rationally increase their confidence in those beliefs. Kevin Dorst (2023) has further argued that belief polarization resulting from such selective scrutiny is rational even when we can predict in advance that we will polarize and know that we are subjecting opposing evidence to additional scrutiny, due to the asymmetric impact of ambiguous and unambiguous evidence. Mason Westfall (2024) has also recently argued that belief polarization resulting from certain mechanisms, including selective scrutiny and the tendency to seek out confirming rather than disconfirming evidence, can be rational under the right conditions. If these arguments are successful in showing that these mechanisms can rationally polarize beliefs, then it is perhaps less obvious that the dogmatic attitude that results from accepting the conclusion of the Polarization Paradox is irrational.

3. Polarized disagreements

However, reflection on disagreements between evidentially polarized subjects makes accepting the conclusion of the paradox seem implausible. Suppose that Blue believes on the basis of several reports from seemingly trustworthy sources (B1) that Joe Biden legitimately won the 2020 US presidential election and (B2) that a number of political figures and news organizations are presenting various forms of misleading evidence that (B1) is false. Red, on the other hand, believes on the basis of several reports from seemingly trustworthy sources (R1) that Donald Trump legitimately won the 2020 US presidential election and (R2) that a number of political figures and news organizations are presenting misleading evidence that (R1) is false. Plausibly, we can fill in the details such that Blue and Red are evidentially polarized in favor of (B1) and (R1), respectively, since each has been told by sources they have reason to trust that any sources presenting evidence against their beliefs in these propositions are intentionally misleading them in,

for example, an effort to steal the election. So, if we accept the conclusion of the paradox, we must conclude that each would be rational in disregarding any evidence against their beliefs in (B1) and (R1), including any new evidence presented by the other against these beliefs.

Politically polarized disagreements of this sort already present a practical problem. Since each side distrusts the sources of evidence trusted by the other, it's hard to see how they can reach any common ground or come to any agreement about the truth of (R1) and (B1). In this situation, it's not clear how either side could convince the other to change their beliefs, leaving this disagreement intractable in the sense that there is no way for Red and Blue to resolve their disagreement over who won the election. There is no philosophical mystery about this problem, though. Sometimes people trust bad sources and refuse to revise their beliefs in the face of evidence that should rationally compel them to do so. However, if we accept the conclusion of the Polarization Paradox, this practical problem is mirrored by an epistemic problem. Not only are Blue and Red unwilling to change their beliefs in response to new evidence the other might present, but, since they are evidentially polarized in favor of the propositions they believe, it would be irrational for them to do so. For, for example, Blue to change their mind and come to believe (R1) on the basis of a news report provided by Red, they would have to ignore their evidence for (B2) and trust a source that they have good reason to distrust. And of course, Red is in the same position. So, in addition to polarized disagreements like this being practically intractable, in the sense that neither side will in fact revise their beliefs in light of new evidence, they are also rationally intractable – neither side *should* revise their beliefs in light of new evidence.

This is not to say, of course, that Red's and Blue's beliefs are epistemically on par. Only one of (R1) and (B1) is true. Only Blue's or Red's beliefs are sensitive to the truth and based on testimony from credible sources. At most one of them has knowledge. But each has data that supports their beliefs in roughly the same way. Each relies on reports from sources that seem to them to be trustworthy for information both about the results of the election and about the trustworthiness of other sources. In this way, although one of them clearly has a set of beliefs that is epistemically deficient in many ways, neither of them is being epistemically irrational in disregarding evidence against their beliefs.⁶

This assessment of the disagreement seems implausible, however, for many of the same reasons that accepting the conclusion of the original Dogmatism Paradox seems implausible. First, it entails that the evidence Red and Blue have for (R1) and (B1), respectively, is *indefeasible*. No evidence provided by either party, regardless of its quality or quantity, can rationally overcome the other's existing evidence about the winner of the election. But although testimonial evidence from rationally trusted sources is quite good evidence, it does not seem like the sort of evidence that cannot in principle be defeated by other evidence.

Second, their dogmatism is *dismissive* in a way that parallels the dogmatic attitude in the original paradox. Just as all I need to know about a piece of evidence to disregard it in the Dogmatism Paradox is that it is evidence against something I believe, Red and Blue need not consider the merits of any piece of evidence offered by the other in order to rationally disregard it. Unlike rational belief polarization, this dogmatism cannot be accounted for by, for example, more closely scrutinizing opposing evidence and finding defeaters that neutralize its impact. Instead, if we accept the conclusion of the

⁶Of course, it may be that in actual cases of polarized political disagreement, subjects' beliefs are not at all epistemically rational. Maybe in most such disagreements, subjects are processing evidence poorly or not basing their beliefs on evidence at all. The point here is just that Red and Blue, as described, seem to be rational.

Polarization Paradox, Red and Blue are rationally entitled to pay almost no attention to opposing evidence. Once they know that it is evidence against their current beliefs, they are rational to disregard it and to distrust the source that presented it.

Third, if we imagine that Red and Blue each share all of their relevant evidence about the election, then accepting that their dogmatism remains rational is inconsistent with the Commutativity of Evidence principle:

The order in which the items of evidence a subject possesses were acquired is irrelevant to which attitudes are rational on that subject's total evidence.

Kelly (2008) takes this principle to be at least one significant reason for the implausibility of the dogmatism licensed by the original paradox. Had I heard the announcement that my flight is at 7 pm before checking my itinerary, I could have dogmatically dismissed the evidence from my itinerary using parallel reasoning (my itinerary must be mistaken – I just heard that my flight is at 7!). So, the order in which I encounter these pieces of evidence affects what it is rational for me to believe about my departure time, which violates the Commutativity principle. As Kelly notes, this is especially implausible if I am aware of this violation, thinking, for example, that I should have disregarded the announcement and believed my flight was at 5 pm if I had checked my itinerary first, but since I heard the announcement first, I should instead think the flight is at 7 pm and disregard the itinerary. But if they share their evidence, Red and Blue also violate this principle. Had Blue acquired Red's evidence first, Blue would have been polarized, just as Red is, in favor of (R1). When subsequently encountering evidence for (B1), Blue would then have been rational in disregarding it, just as Red is. So, the order in which Blue acquired their evidence affects what it is rational for them to believe about the election results, violating the Commutativity principle. Even if Blue knows that they would be rational in believing (R1), just as Red does, had they merely encountered the evidence in a different order, accepting the conclusion of the Polarization Paradox requires that they are rational in continuing to believe (B1) while disregarding any evidence to the contrary. And this result seems counterintuitive in just the way that the parallel result does in the original Dogmatism Paradox.

Finally, accepting that both parties are rationally dogmatic in polarized disagreements is incompatible with an attitude that we might call "Evidential Optimism" – that disagreements between rational agents can in principle be resolved by gathering and evaluating evidence. Evidential Optimism licenses many common practices in philosophy and underlies much of the advice we offer in critical thinking courses. We teach students to make and evaluate arguments, to gather additional evidence when those arguments prove inconclusive, to avoid common errors in reasoning that might lead them to evaluate evidence irrationally, and so on. On the assumption that the evidence, when evaluated rationally, will settle the question of what they should believe, this is good advice. We may make mistakes, of course, and people will often refuse to accept a conclusion, even when they rationally ought to, but if everyone is being perfectly rational and all of the relevant evidence is available, they ought to reach the same conclusion. If they don't, then one party in the disagreement has made a mistake and is open to rational criticism. However, if we accept the conclusion of the Polarization Paradox, then disagreements between evidentially polarized subjects cannot be resolved in accordance with Evidential Optimism. Instead, these disagreements can only be resolved irrationally, which is a disappointing result for those who value epistemic rationality.

So, as with any paradox, we are forced with a choice. We can deny the initial premise by denying the possibility of evidentially polarized subjects, we can find some flaw in the reasoning that leads to the conclusion, or we can accept the conclusion that dogmatism is rational for evidentially polarized subjects. However, evidential preemption and echo chambers both seem to provide common examples of actual evidential polarization, and while there may be a flaw in the reasoning that generates the paradox, it cannot be the same flaw identified by the defeat solution to the Dogmatism Paradox. Meanwhile, the conclusion of the Polarization Paradox shares many of the counterintuitive features of the original paradox's conclusion and leaves us with the undesirable result that seemingly common polarized disagreements can only be resolved irrationally.⁷

References

- Begby, E. (2021). 'Evidential Preemption.' *Philosophy and Phenomenological Research* 102(3), 515–30.
- Conee, E. (2001). 'Heeding Misleading Evidence.' *Philosophical Studies* 103(2), 99–120.
- Dorst, K. (2023). 'Rational Polarization.' *Philosophical Review* 132(3), 355–458.
- Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.
- Kelly, T. (2008). 'Disagreement, Dogmatism, and Belief Polarization.' *Journal of Philosophy* 105(10), 611–33.
- Kripke, S. (2011). 'On Two Paradoxes of Knowledge.' In S.A. Kripke (ed), *Philosophical Troubles. Collected Papers Vol I*. New York, NY: Oxford University Press, 27–51.
- Lasonen-Aarnio, M. (2014). 'The Dogmatism Puzzle.' *Australasian Journal of Philosophy* 92(3), 417–32.
- McWilliams, E.C. (2021). 'Evidentialism and Belief Polarization.' *Synthese* 198(8), 7165–96.
- Nguyen, C.T. (2020). 'Echo Chambers and Epistemic Bubbles.' *Episteme* 17(2), 141–61.
- Peirce, C.S. (1877). 'The Fixation of Belief.' *Popular Science Monthly* 12(1), 1–15.
- Westfall, M. (2024). 'Polarization is Epistemically Innocuous.' *Synthese* 204(3), 1–22.

Brandon Carey is an Associate Professor of Philosophy at California State University, Sacramento, working mainly on questions in social epistemology about misinformation, polarization, and the transmission of knowledge. Email: brandon.carey@csus.edu

⁷I am grateful to the participants in the 2023 conference of the North American Society for Social Philosophy for their useful criticisms of an earlier version of this paper, to the students in my 2025 seminar on misinformation, polarization, and trust for extended discussion this puzzle, and to an anonymous referee for this journal for pressing the objections in the last two sections, as a result of which this paper is both more careful and more interesting.

Cite this article: Carey B. (2025). "The Polarization Paradox." *Episteme* 1–13. <https://doi.org/10.1017/epi.2025.10090>