

Interthematic Polarization

Finnur Dellsén

Forthcoming in *American Philosophical Quarterly*
(please cite official version when available)

Abstract

In recent epistemology, *belief polarization* is generally defined as a process by which a disagreement on a single proposition becomes more extreme over time. Outside of the philosophical literature, however, ‘polarization’ is often used for a different epistemic phenomenon, viz. the process by which people’s beliefs on unrelated topics become increasingly correlated over time. This paper argues that the latter type of polarization, here labeled *interthematic polarization*, is often rational from each individual’s point of view. This suggests that belief polarization is not necessarily a failure of individual rationality, but instead a failure of the social structures within which we live our epistemic lives.

Keywords: polarization; disagreement; social epistemology.

1 Two Types of Belief Polarization

Broadly understood, *belief polarization* occurs when the beliefs or attitudes of two or more parties become increasingly dissimilar from each other over time. In the recent philosophical literature, however, belief polarization is generally defined more narrowly as the process by which a disagreement on a single proposition P becomes more extreme over time, e.g. with one party becoming more confident that P is true while another becomes more

confident that P is false.¹ This type of polarization is especially puzzling in cases where the parties obtain the same evidence regarding P , or when the parties are actively discussing and deliberating about P . Naturally, then, philosophers have wondered whether belief polarization in this sense could be epistemically rational in some recognizable sense of the term (see, e.g., Kelly, 2008; Dorst, 2019; McWilliams, 2019; Singer et al., 2019; Pallavicini et al., 2021).

Outside of the philosophical literature, however, ‘polarization’ is often used to refer to a different change in people’s beliefs.² Consider situations in which two parties, who previously only disagreed on some proposition P (or some set of propositions $\{P_1, \dots, P_m\}$), subsequently come to disagree also on an unrelated proposition Q (or set of such propositions $\{Q_1, \dots, Q_n\}$), where two propositions count as ‘unrelated’ just in case learning that one is true would not in itself make the other more plausible.³ In this type of case, the two parties will not necessarily end up with more extreme views about initially-disagreed-upon proposition P (or set $\{P_1, \dots, P_m\}$). Rather, the total set of their opinions – about the P s and the Q s collectively — will become increasingly similar to those within their party, and increasingly dissimilar to those on the other side. Put differently, the opinions in the population as a whole, regarding the P s and Q s, will become increasingly correlated. To put a name on this species of belief polarization, let us call it *interthematic polarization*. We may then call the other kind of belief polarization, more familiar from the philosophical literature, *monothematic polarization*.

¹This characterization of belief polarization is also often used in the psychological literature on polarization, beginning with Lord et al. (1979).

²This is especially true of the recent social scientific literature on ‘polarization’ (e.g. Baldassarri and Gelman, 2008; Pew Research Center, 2014; Abramowitz, 2018) and public discussions of political polarization (e.g. Nivola, 2005; Klein, 2020).

³In a probabilistic (Bayesian) framework, where p is one’s personal probability function, this notion of unrelatedness between two propositions P and Q can be identified with their (unconditional) independence: $p(P \wedge Q) = p(P)p(Q)$. Since it follows from this that $p(P|Q) = p(P)$ and $p(Q|P) = p(Q)$, it intuitively means that whether a given P_i from the first theme is true tells us nothing in itself about whether a given Q_j from the second is true, or *vice versa*. (Note that learning that P or Q is true is not identical to learning that some particular individual claims that P or Q is true. The latter is the type of learning situation that I consider below.)

As noted, the philosophical question of whether monothematic polarization might be rational has already been discussed at some length. Interthematic polarization, by contrast, has received much less attention.⁴ And yet it should be clear that much of the public discussion and concern about belief polarization is not merely or even primarily about monothematic polarization. For example, claims to the effect that the American public has become more polarized in recent years should at least in part be understood in terms of increasing interthematic polarization.⁵ That is, the concern is not merely that Americans are adopting more extreme views on a particular topic, e.g. income inequality; rather, the concern is at least in part that their views on this topic are increasingly becoming strongly correlated with their views on orthogonal topics, e.g. gun control (see, e.g., Baldassarri and Gelman, 2008; Lewandowsky et al., 2013; Abramowitz, 2018).

The remainder of this paper addresses what I take to be the key *epistemological* question about interthematic polarization: *Is it rational?*⁶ I will argue that, from each individual's point of view, the answer is often 'yes'. Roughly, this is because the fact that another group or individual disagrees with you on one issue is often a reason for you to disbelieve claims made by that group or individual on another issue, even when the two issues are unrelated. As we'll see, this argument does not apply to all cases of interthematic polarization, but it applies to many. (As I develop the argument, we'll learn more about what types of cases it does and doesn't apply to.) I conclude by suggesting that this has important upshots for how to explain and

⁴I am aware of only two cases in which phenomena similar to interthematic polarization have been discussed in the philosophical literature so far: in Bramson et al.'s (2017, 129-130) definition of 'belief convergence', and in Weatherall and O'Connor's (2020) computer simulation model of 'epistemic factionalization' (for further discussion of the latter, see section 3). I prefer 'interthematic polarization' to either of the aforementioned terms because it (a) highlights the similarities and differences between it and *monothematic* polarization, and (b) lines up with common usage of the unqualified term 'polarization'.

⁵Indeed, Baldassarri and Gelman (2008, 410-411) suggest that American public opinion can be said to be polarized only in this latter sense of the term.

⁶I will not be directly addressing various empirical questions about interthematic polarization, such as the extent to which it is occurring in modern societies and what psychological mechanisms might cause or facilitate it. For discussions of such empirical questions, see, e.g., Baldassarri and Gelman (2008), Lewandowsky et al. (2013), Abramowitz (2018) and references therein.

counteract interthematic polarization.

2 Interthematic Polarization Through Rational Distrust

In what follows, I assume that what it is rational for an individual to believe depends in part on what other agents tell them is the case, i.e. on testimony. Not every piece of testimony in favor of P provides equally strong support for P , of course. Whether, and the extent to which, someone's testimony that P provides a reason to believe P depends, among other things, on the credibility of the testifier. Other things being equal, if you have some reason to distrust someone's testimony regarding P , their testimony in favor of P provides less support, or perhaps no support at all, for P . Conversely, if you have some reason to think that someone is especially trustworthy regarding P , their testimony in favor of P provides more support for P than it otherwise would have, other things being equal. (Of course, it is a substantive issue what sort of factors give us reason to trust or distrust another person regarding some proposition; we'll get to that shortly.)

To explain why interthematic polarization can be rational from each individual's point of view, I will consider how each individual in a group could rationally reason so as to make the group as a whole (increasingly) interthematically polarized. Roughly, I will argue that it would be rational for each individual to endorse an argument for placing less confidence in the testimony of those with whom they already disagree about something else as compared to those with whom they agree on that other issue. (This is not to say that every, most, or even any, of the individuals in the group will in fact reason in the way I describe. But at least in one sense of the term 'rational' — and it is that sense in which I am interested here — the fact that someone *could* reason their way to a particular conclusion, via some argument that it would be rational for them to endorse, means that the conclusion is indeed rational for them.)⁷

⁷This sense of the term 'rational' is roughly what is often called *propositional justification*, as distinguished from *doxastic justification*. I use the term 'rational' instead of 'justified' because the latter is usually applied to certain doxastic states, viz. beliefs, whereas I am most directly concerned with *changes* in doxastic states.

Let us consider a maximally simple case in which an agent, S_I , both encounters an agent that *disagrees* with them about P , S_D , and also encounters another agent that *agrees* with them about P , S_A .⁸ Suppose that S_D and S_A provide S_I with conflicting testimony on Q , where P and Q are unrelated. Now consider the following argument, made from S_I 's first-person perspective, concerning how much confidence to place on S_D 's claims about Q :

- (1) Since my (S_I 's) belief about P is true, S_D 's contrary belief about P is false.
- (2) If S_D 's belief about P is false, then it's because S_D 's reasoning regarding P is defective, and/or because S_D 's evidence regarding P is misleading.
- (3a) If S_D 's reasoning regarding P is defective, then there is some (additional) reason to believe that S_D 's reasoning regarding Q is also defective.
- (3b) If S_D 's evidence regarding P is misleading, then there is some (additional) reason to believe that S_D 's evidence regarding Q is also misleading.
- (4) So, there is some (additional) reason to believe that S_D 's reasoning regarding Q is defective, and/or that S_D 's evidence regarding Q is misleading. [From (1)-(3b).]
- (5) So, other things being equal, it's rational for me to place less confidence in S_D 's testimony about Q .⁹ [From (4).]

S_I clearly cannot run the same argument regarding S_A 's claims about Q , since S_A agrees with S_I about P . If rationally endorsable by S_I , this argument would therefore make it rational for S_I to place more confidence in

⁸There are two salient ways to conceive of agreement and disagreement, depending on whether we are working with full beliefs or degrees of confidence. In a full-belief framework, agreement (disagreement) about P simply amounts to having the same (contrary) belief regarding P . In a degree-of-belief framework, by contrast, agreement about P amounts to having identical (i.e. equal) degrees of confidence in P , and disagreement thus consists in having different (i.e. unequal) degrees of confidence in P .

⁹I.e., less than if S_D hadn't disagreed with me about P — or less than the confidence it's rational to place in the testimony of an otherwise identical agent who doesn't disagree with me about P , such as S_A .

S_A 's claims about Q than in S_D 's, other things being equal. This, in turn, would move S_I closer to S_A 's attitude towards Q , and away from S_D 's, thus increasing the extent to which doxastic attitudes on P and Q are correlated among S_I , S_D , and S_A . The upshot, in other words, is interthematic polarization.

The question, then, is whether it would be rational for S_I to reason in accordance with the above argument. Would it be rational for S_I to believe each of the premises (1), (2), (3a), and (3b), and make both inferences to (4) and (5)? Let us consider each of these steps in turn.

(1) states that since S_I 's belief that P is true, S_D 's contrary belief is false. This might seem trivially rational for S_I , since any rational person would arguably have to regard their own beliefs as true, and contrary beliefs as false. There is, however, an important objection to this step that highlights a potential restriction of the range of cases to which the above argument is applicable. The objection is that once our agent S_I becomes aware that they disagree with S_D , S_I should suspend judgment about P (i.e. cease to believe P). After all, S_I might have no reason to think they are any more reliable than S_D regarding P , so perhaps this is the type of peer disagreement that calls for both agents to suspend belief (see, e.g., Feldman, 2006; Christensen, 2007). If so, then as soon as S_I learns of S_D 's contrary belief on P , S_I is no longer in a position to rationally appeal to the argument above. So, according to this objection, the argument would not even get off the ground for S_I .

As I have indicated, I do concede that this objection might apply in *some* cases, thus potentially restricting the range of cases to which the argument is applicable. But this restriction will at most be very limited, since it only applies if S_I , as a result of recognizing that they disagree with S_D on P , rationally ends up with a doxastic attitude regarding P that is *identical* to S_D 's. If there is any difference between S_I and S_D 's attitudes post-conciliation, e.g. in that S_I ends up with a slightly higher degree of confidence than S_D

¹⁰See, e.g., Joyce (1998) and Pettigrew (2016) for influential measures of accuracy for degrees of confidence.

(which may be compatible with S_I and S_D both suspending belief in the binary sense of ‘belief’), then an argument to the same effect can be made by replacing ‘belief about P ’ with ‘degree of confidence in P ’ and ‘true’/‘false’ with ‘more accurate’/‘less accurate’.¹⁰ So the argument should be restricted only if, or in so far as, S_I is rational in adopting a doxastic attitude towards P that is identical to the one adopted by S_D .

Now, what it’s rational for S_I to do upon recognizing that they disagree with S_D will depend on which epistemological theory of disagreement is correct. It is only on the most radically conciliatory theory of disagreement, the Equal Weight View (e.g., Elga, 2007; Matheson, 2015), that epistemic peers would be rationally required to adopt identical doxastic attitudes upon recognizing that they disagree. So assume, for the sake of the objection, that the Equal Weight View is correct. Even so, S_I and S_D should end up with identical doxastic attitudes towards P only if S_I and S_D are (and/or regard each other as) epistemic peers in the strictest sense of the term. However, although strict epistemic peers are useful hypothetical agents for the purposes of idealized thought experiments, they are notoriously hard to find in the wild (see, e.g., Elga, 2007; Frances, 2010; King, 2011). It is thus a quite minor limitation of the argument if it is restricted to cases in which S_D is not (and/or is not regarded by S_I as) S_I ’s strict epistemic peer.

Furthermore, even if S_I were to adopt the conciliatory attitude that is recommended by the Equal Weight View in response to recognizing S_D as their (strict) epistemic peer, S_I and S_D might still very well end up disagreeing on P . After all, the *other* agent involved in the disagreement, S_D , might still fail to conciliate, either at all or to the extent required by the Equal Weight View. (And let’s face it: most people do not conciliate much, if at all, in cases of peer disagreement.) Indeed, S_D ’s failure to (sufficiently) conciliate upon learning¹¹ about S_I ’s belief about P might rationally indicate to S_I that S_D is not their epistemic peer — because, by S_I ’s lights, a genuine epistemic

¹¹I am assuming here that S_D becomes aware of their disagreement with S_I . However, it’s clearly also possible for S_D to remain unaware of their disagreement with S_I , as when S_I reads about S_D opinions on P and Q but not vice versa. In such a case, S_D would presumably not be rationally required to conciliate towards S_I , even according to the Equal Weight View.

peer would be fully rational and thus conciliate — which in turn would make it rational for S_I to revise their conciliatory response towards more steadfastness, thus further widening the gap between S_I and S_D 's doxastic attitudes towards P .

To summarize, even on the Equal Weight View, there will be plenty of cases in which it would be rational for S_I to believe (1) (or its functional equivalent in terms of degrees of confidence), viz. all cases in which either (i) S_D fails to be (and/or to be regarded by S_I as) S_I 's strict epistemic peer, or (ii) S_D fails to fully comply with the Equal Weight View such as to end up with an identical doxastic attitude to that adopted by S_I after conciliating. In either case, it will be rational for S_I to have a doxastic attitude towards P that differs from S_D 's, thus making it rational for S_I to believe (1) (or its functional equivalent in terms of degrees of confidence).

Consider next whether it would be rational for S_I to believe (2). This premise asserts that S_D 's (false) belief about P is due to defective reasoning and/or misleading evidence. Premise (2) thus distinguishes, in an admittedly course-grained way, two possible causes of false beliefs. It is hard to see what else could cause an agent to believe falsely.¹² Note that (2) does not assert that these two explanations for the other agent's false belief are equally plausible, just that one of them is correct. Indeed, I will shortly suggest that the second explanation is often considerably more plausible in most realistic cases of interthematic polarization.

Premises (3a) and (3b) are structurally similar. They both claim that if someone's (false) beliefs about P are due to a particular type of error (defective reasoning or misleading evidence), then that's some reason to think that they are more likely to have beliefs about Q that are affected by the same type of error. Specifically, (3a) effectively says that agent who is known to reason defectively regarding P is more likely to reason defectively regarding Q . This seems rational to believe in a wide range of cases, in so far

¹²If I am overlooking some other source of false beliefs, then it could presumably simply be listed as item (c) in premise (2) and then accordingly be addressed in a separate premise in an analogous fashion to what I have done for defective reasoning and misleading evidence in premises (3a) and (3b).

as it would be rational to believe that the quality of another's reasoning is normally not completely domain-specific.

To be sure, there are plausible exceptions to this general rule. For example, most of us realize that defective reasoning about especially difficult mathematical problems may not be correlated with defective reasoning about the reasons for other people's behavior. (In those cases, premise (3a) fails to be plausible, and so the argument relies entirely on the plausibility of premise (3b). If that premise fails as well, then the argument is not applicable at all to the relevant cases.) More commonly, however, an agent will have some reason to think that those who reason defectively about one issue, e.g. because they confuse correlation with causation or because they commit the gambler's fallacy, will be more likely to similarly reason defectively about another issue. So it seems that it would often (though certainly not always) be perfectly rational for S_I to believe premise (3a).

One might object that the above rationale for (3a) only applies to S_I if they have some insight into what caused the relevant agent – S_D , in this case – to reason defectively. After all, for all S_I knows, S_D might normally be an excellent reasoner who just had a bad day, was momentarily distracted, or otherwise formed their belief about P in a way that has no bearing on how they formed their belief about Q . Wouldn't S_I have to rule out this type of explanation of S_D 's defective reasoning regarding P in order for there to be any reason for S_I to believe that S_D 's reasoning regarding Q is defective, as per (3a)?¹³

Not really, no. To be sure, if it were rational for S_I to be *absolutely certain* that S_D just had a bad day, etc., then it admittedly wouldn't be rational for S_I to believe (3a). But such situations are rare at best, since one should rarely, if ever, be absolutely certain about such things. In the more common type of situation in which S_I cannot rationally rule out the possibility that S_D did *not* just have a bad day, etc., it remains true that S_D is, from S_I 's point of view, more likely (perhaps only slightly more likely, but more likely nonetheless) to reason defectively regarding Q given that they reasoned de-

¹³Thanks to an anonymous reviewer for raising this objection.

fectively regarding P . Roughly, this is because the overall probability of S_D 's reasoning defectively regarding Q is determined, in part, by the probability of S_D 's reasoning defectively regarding Q conditional on the possibility that S_D did *not* just have a bad day, etc. Even if this possibility is unlikely by S_I 's lights, it will positively affect the probability that S_D reasoned defectively regarding Q as long as its probability is not zero.¹⁴ Of course, *how much more likely* S_D is to have reasoned defectively regarding Q will depend, among other things, on how (un)likely bad-day-type explanations are by S_I 's lights. But as long as it's rational for S_I not to absolutely rule out alternative explanations for S_D 's defective reasoning regarding P , S_I is indeed rational in finding it at least *somewhat* more likely that S_D reasoned defectively regarding Q as well.

Next consider premise (3b). In general, possessing misleading evidence is often a plausible explanation for false beliefs — perhaps more plausible, indeed, than defective reasoning. Much of the evidence we possess, especially regarding the types of issues on which polarization is common, comes from sources that are viewed as suspicious by the other side, such as politicized media outlets. In any case, (3b) effectively asserts that an agent who has misleading evidence regarding one theme is more likely to have misleading evidence regarding another theme. For example, suppose S_D has misleading evidence about whether climate change is anthropogenic. It is then quite reasonable, in many if not most circumstances, to suppose it more likely (perhaps only slightly more likely, but more likely nonetheless) that S_D also has misleading evidence regarding, say, whether the MMR vaccine causes early onset autism. After all, it seems plausible that S_D 's sources of evidence regarding climate change would overlap with their sources of evidence regarding the MMR vaccine, so if their evidence regarding the former is misleading then their evidence regarding the latter is more likely to be so as well.

Now, it is quite clear that the extent to which premise (3b) is plausi-

¹⁴It is a straightforward theorem of the probability calculus that raising the conditional probability $P(A|B)$ necessarily raises the unconditional probability $P(A)$ provided that $P(B)$ is non-zero and that other terms are held constant.

ble in a specific case of interthematic polarization will depend heavily on the propositions in question, P and Q . Premise (3b) will be most plausible whenever S_I has reasons to believe that S_D 's evidence for P and Q would come from at least partially overlapping sources, as in the case of climate change and the MMR vaccine. But, of course, there will also be cases in which S_I has reason to believe that S_D 's sources of evidence for their belief about P are completely distinct from their sources of evidence for their belief regarding Q — in which case (3b) is much less plausible. For example, even if I have reason to believe that you have misleading evidence regarding climate change, that gives me little reason to believe that you have misleading evidence regarding your own personal finances, since it is quite unlikely that the sources for your evidence regarding these issues overlap. (In those cases, premise (3b) fails to be plausible, and so the argument relies entirely on the plausibility of premise (3a). If that premise fails as well, then the argument is not applicable at all to the relevant cases.)

Nevertheless, premise (3b) will be plausible even in a wide range of cases

¹⁵Note that the qualification that S_D must have been able to choose not to obtain the evidence from the source(s) in question is not always satisfied. Some sources of evidence, e.g. perception of one's immediate environment, are more or less impossible to avoid entirely. Accordingly, S_D 's having misleading evidence regarding climate change should not (at least not by the present argument) lead S_I to believe that S_D 's perceptual evidence about her immediate environment is also misleading.

¹⁶Here one might raise an objection similar to the one considered above in the discussion of premise (3a). Specifically, one might object that the above rationale for S_I to believe (3b) only applies if S_I has some insight into what caused the relevant agent (S_D , in this case) to have misleading evidence regarding P . After all, for all S_I knows, S_D might be someone who normally selects her sources of evidence carefully and judiciously, but was just unlucky or a bit sloppy in her selection of evidence on this particular occasion, i.e. regarding P . Wouldn't S_I have to rule out this type of explanation of S_D 's having misleading evidence regarding P in order for there to be any reason for S_I to believe that S_D 's evidence regarding Q is misleading, as per (3b)? For analogous reasons as before, the answer is 'No, not really'. Provided that S_I is not rationally required to be *absolutely certain* that S_D was just unlucky on this particular occasion, i.e. provided that S_I cannot rationally rule out the possibility that S_D was *not* just unlucky or a bit sloppy on this particular occasion, it remains true that S_D is, from S_I 's point of view, more likely (perhaps only slightly more likely, but more likely nonetheless) to have misleading evidence regarding Q if they have misleading evidence regarding P . Again, this is roughly because the overall probability of S_D 's having misleading evidence regarding Q is determined, in part, by the probability of S_D 's having misleading evidence regarding Q conditional on the possibility that S_D was *not* just unlucky or a bit sloppy in their selection of evidence regarding P (see also footnote 14).

in which there is no reason to think there would be any overlap between the sources of evidence regarding P and Q respectively. This is because S_D 's having misleading evidence regarding P would reflect badly on S_D themselves, epistemically speaking, in that it indicates that S_D is the sort of epistemic agent that would choose to trust a source that provides them with misleading evidence. That sort of epistemic agent is more likely to also trust another source that provides them with misleading evidence. And this holds even if the two sources are entirely distinct. So, in this slightly round-about way, it would arguably be rational to believe premise (3b) whenever it is at least somewhat plausible that S_D bases their beliefs about P and Q on any type of source from which they could have chosen not to obtain evidence.^{15,16}

This concludes my discussion of the four premises, (1)-(3b), all of which seem to me to be rational for S_I to believe in a wide range of cases. What about the steps from there to the conclusion, (5)? (4) is clearly implied by (1)-(3b). (5) does not follow logically from (4) alone, but it does follow given the epistemic truism that, other thing being equal, it is rational to place less confidence in the testimony of those that one has some (additional) reason to believe have reasoned defectively, and/or to have misleading evidence, regarding the relevant proposition. Thus I see no fault with this part of the argument either. I conclude that, in a wide range of cases, S_I would indeed be rational in placing less confidence in S_D 's claims about Q , given their disagreement with S_D about P , than in S_A 's claims about Q .¹⁷ As a result, S_I thus ends up with a doxastic attitude towards Q that is closer to S_A 's, and further away from S_D 's, for no other reason than the fact that they agree with S_A , and disagree with S_D , on an unrelated P . This is interthematic polarization.

Now, for any epistemic community that consists of multiple agents, and where each agents takes a stand on multiple propositions, there is an important further question of exactly *how* the community would interthemat-

¹⁷Indeed, by parity of reasoning, S_D would then also frequently be rational in placing less confidence in S_I 's claims about Q , thus further increasing the distance between S_I and S_D 's attitudes towards Q .

ically polarize if each agent reasoned in this way. For example, we may ask how dependent such polarization is on certain patterns of interaction between agents, and how many, and how large, groups of agents with highly correlated beliefs we would end up with. Interestingly, these sorts of questions have been addressed by political scientist Robert Axelrod (1997), who models communities of agents who influence each other's cultural traits more and less based on prior cultural similarity.¹⁸ In particular, an agent that already shares more cultural traits with their 'neighbor' (another agent in the model with which they could potentially interact) is more likely to adopt another of their neighbor's traits as well. This assumption in Axelrod's model corresponds roughly to what the above argument shows to be rational for each individual to do, viz. to place less confidence in someone's testimony regarding Q if they disagree with on another proposition P than if they had agreed with them on P .

It is thus interesting for our purposes to note some of the results from simulations of Axelrod's model. Over time, the model tends to yield a community that is split into some number of groups of agents that share no traits at all with any of their neighbors ('stable regions', in Axelrod's terminology). Predictably, the number of distinct such groups decreases as the agents in the community are assumed to be able to interact with a greater number of other agents (i.e. as the size of the 'neighborhood' increases). This fits the intuition one might have that more open channels of communication should yield greater diversity in the community. To be clear, however, there is still interthematic polarization in such diverse communities – it's just that the sorting of agents into groups with correlated traits is such that there is a greater number of such groups. A less predictable result from Axelrod's simulations (see also Axtell et al., 1996) is that very large communities of agents will tend to yield very few such groups (and much fewer than more moderately sized communities). For example, under cer-

¹⁸Naturally enough, Axelrod does not address the epistemological question that I focus on in this paper, viz., whether this process might be epistemically rational from each individual's point of view. Rather, Axelrod is concerned with explanatory questions of how certain patterns of cultural differences could arise given that individual agents tend to behave in certain ways.

tain assumptions the average number of such groups is down to only about two in communities of 10,000 agents (which are the largest communities in Axelrod’s simulations). This may provide some insight into actual patterns of interthematic polarization, e.g. in the context of American politics.

3 Separating the Rationality of Monothematic and Interthematic Polarization

In this section, I wish to briefly highlight that the argument of the previous section does not assume or imply that *monothematic* polarization is (ever) individually rational. Thus the argument avoids the highly controversial issue of whether monothematic polarization is ever, let alone generally, rational from each individual’s point of view (for discussion, see Kelly, 2005; Dorst, 2019; McWilliams, 2019; Pallavicini et al., 2021).

To see this clearly, note that the above argument does not imply that S_I would be rational in bringing their attitude towards P further away from S_D ’s as they discover that they disagree with S_D on P ;¹⁹ rather, it implies only that S_I would be rational in bringing their attitude on an unrelated Q further away from S_D ’s. This is an important difference between the route to polarization described above and a cruder route by which agents would simply move further away from any attitudes that are held by agents with whom they initially disagree on some proposition. For example, consider a mechanism by which an agent S_I , who notices that they disagree with S_D about P , is thereby inclined to move further away from S_D ’s attitudes generally, including their attitude on P itself.

Weatherall and O’Connor (2020) construct a formal model of the latter type of mechanism, and show through computer simulations that a form of interthematic polarization would emerge as a result. Specifically, Weatherall and O’Connor’s model assumes that agents update their degrees of confidence, on two distinct propositions B_1 and B_2 , by (mis)trusting

¹⁹Indeed, as noted in the discussion of premise (1) above, the argument is compatible with it being rational for S_I to ‘conciliate’ to a large extent, i.e. move *closer* to S_D ’s attitude on P — which, if anything, is a kind of monothematic *anti*-polarization.

evidence from other agents in proportion to the extent to which they already (dis)agree with these other agents on B_1 and B_2 . As Weatherall and O'Connor (2020, 9) acknowledge, this is simply a generalization of a simpler mechanism that applies to a single proposition B , and which leads to monothematic polarization on B (see O'Connor and Weatherall, 2018). Weatherall and O'Connor do not unqualifiedly claim that the modelled mechanism is rational from each individual's point of view – describing it instead as “arguably reasonable” (2020, 2) and “(semi) rational” (2020, 20). These qualified characterizations are indeed appropriate in my view, since it will at least be controversial whether a given agent may be rational in mistrusting the evidence provided by other agents regarding some proposition P solely on the grounds that they disagree with them about that very same proposition P .²⁰

By contrast, the argument that I have provided for the individual rationality of interthematic polarization in this paper does not make any such assumption. That is, the argument of the previous section is neutral on whether monothematic polarization is also rational from each individual's point of view. Indeed, the argument is even compatible with the contrary assumption that it is *never* (for any S_I , S_D , and P) rational for S_I to move further away from S_D 's attitude *on* P as they discover that they disagree on P ; rather, all that is required by that argument is that it would be rational for S_I to move further away from S_D 's attitude *on* Q (where P and Q are unrelated). Accordingly, this argument does not assume or imply that monothematic polarization is (ever) individually rational; let alone that interthematic polarization is rational only when monothematic polarization is rational. In short, the argument of this paper, in contrast to the simulation model of Weatherall and O'Connor, separates the rationality of the two forms of polarization, and shows that interthematic polarization would be rational even if (or when) monothematic polarization isn't.

²⁰To be fair, Weatherall and O'Connor are much less (if at all) concerned with arguing that the mechanism they model is epistemically rational than with showing in detail that, and how, this mechanism leads endogenously to (monothematic and interthematic) polarization under various conditions. So the central question of the current paper, ‘Is interthematic polarization rational?’, arguably lies outside of the main scope of Weatherall and O'Connor's paper.

4 Interthematic Polarization and Political Partisanship

In this penultimate section, I will address a concern about political partisanship that may be construed as an objection to my argument above.²¹ The concern is due to Joshi (2020), who raises an epistemic challenge to forming beliefs on orthogonal political issues in accordance with a particular party line, e.g. by adopting the official Republican stance on abortion, climate change, immigration, and gun control. In brief, Joshi argues that by adopting the party line on ‘orthogonal’ issues of this sort, one commits oneself to thinking that one’s own side is *reliable* on such orthogonal issues, and thus that the other side is *anti-reliable* on those issues. Joshi then goes on to suggest that such an assumption about the reliability of one’s own party and the anti-reliability of an opposing party would be epistemically problematic:

But what kind of belief forming method(s) would lead a group to get things consistently wrong on a set of orthogonal issues? [...] I want to suggest that it is implausible that a psychologically realistic method of forming beliefs can be reliably falsehood tracking with respect to such a set of orthogonal issues. Therefore, the partisan with respect to such orthogonal issues on either of the two prominent sides ought to moderate her political beliefs. (Joshi, 2020, 41)

It is hard not to agree with the sentiment that seems to be underlying Joshi’s argument here, viz. that blindly adopting the party line on unrelated issues is epistemically problematic. This might seem like a threat to the argument of section 2, since that argument suggests that it would at least sometimes be rational to trust supporters of one’s own party over those in an opposing party, even on topics that are unrelated to those that made one support the party in the first place, because those other supporters will share many of one’s other political beliefs. But then doesn’t the argument of section 2 also imply, as per Joshi’s line of reasoning above, that supporters of the other party are ‘getting things consistently wrong’ in a way that requires the existence of some implausible psychological mechanism to be present

²¹Thanks to an anonymous reviewer for bringing this concern to my attention.

in supporters of that party?

No, it does not. In fact, there are two quite separate reasons why it doesn't. The first reason should be apparent from the structure of the argument in section 2. Recall that premise (2) listed two possible causes for another person's false belief about a given proposition *P*. One of these was *defective reasoning* (the focus of premise (3a)); the other – noted then to be a considerably more common in most realistic cases – was *misleading evidence* (the focus of premise (3b)). Now, while the former (defective reasoning) might indeed need to be explained by positing a psychological mechanism of some sort, the latter (misleading evidence) clearly does not. For if the evidence from which you are reasoning is misleading, no special psychological mechanism is required to explain why you would end up with false beliefs. So one plausible explanation for why someone else might get a host of issues wrong is simply that they are repeatedly reasoning from misleading evidence, such as evidence from unreliable news sources. This requires no special psychological mechanism; only that the evidence possessed by those on one side of the partisan divide is more likely to be misleading.

The second reason why Joshi's objection to political partisanship fails to undermine the argument of section 2 concerns a subtle logical point about epistemic reliability. You can take someone to be epistemically reliable with regard to a series of propositions, in the sense that they are much more likely than not to be right about each one of these propositions, even while knowing full well that they will (or will almost certainly) be wrong about *some* of those propositions. Provided that you don't also know *which* of these propositions the person is wrong about, it might very well be rational to trust that person's testimony regarding *all* of the propositions. After all, you simply don't have the luxury of trusting them on only the proposition on which they happen to be correct. For example, suppose your local weather man is right 90% of the time about whether it will rain the next day. Since you don't know beforehand *when* he gets his predictions wrong, you would arguably still be rational in trusting him every day – even while knowing full well that on some days he will certainly be wrong.

To see the relevance of this point, note that Joshi's objection to political

partisanship rests on the idea that it requires the background assumption that one's own side is 'getting things consistently right' and that the other side is 'getting things consistently wrong'. Now, if this means that one must believe that supporters of one's own party are 100% reliable, and that those of the other side are 0% reliable, then this is clearly not implied by the argument of section 2. Rather, that argument implies, at most, that those with whom one agrees about another topic (such as fellow supporters of one's own party, perhaps) can rationally be believed to be somewhat more reliable than those with whom one disagrees about that topic. In light of the logical point above, it should be clear that this is fully compatible with believing that those with whom one agrees will (or will almost certainly) be wrong about some – indeed, perhaps many – of the issues on which one nevertheless chooses to trust them. Similarly, it is compatible with believing that those with whom one *disagrees* will (or will almost certainly) be right about some – indeed, perhaps many – of the issues on which one nevertheless chooses *not* to trust them.

In sum, then, Joshi's objection to political partisanship does not undermine the argument of section 2 for two separate reasons. First, there is a perfectly plausible explanation for why those with opposing views on *P* would more often be wrong about an unrelated *Q*, viz. that their evidence is more like to be misleading. This requires no special psychological mechanism that would be present only on one side of a polarized population. Second, the argument of section 2 does not require the agent in question to implicitly assume that those in whom they place their trust will always be right (or that those who they choose not to trust will always be wrong); on the contrary, it is fully compatible with believing that those that they trust will sometimes get things wrong (and that those that they distrust will sometimes get things right).

5 Conclusion and Upshots

I have argued that individual agents would often be rational in forming beliefs in ways that constitute *interthematic polarization*, where agents who

disagree on one issue also come to disagree on an unrelated issue. In brief, this is because it is often rational for an individual to distrust those with which they disagree, even regarding issues unrelated to the original disagreement. In this sense, interthematic polarization is often individually rational. This is emphatically not to say that we should embrace or be unconcerned with interthematic polarization as a social phenomenon, seeing as it evidently leads to a breakdown of communication between people who might otherwise learn from one another. Rather, this suggests that interthematic polarization exemplifies a way in which individual and social rationality can come apart: What's epistemically good for individuals may not be epistemically good for the communities they comprise (cf. Mayo-Wilson et al., 2011).

Two further upshots are worth mentioning. First, although the argument outlined above is primarily meant to address the normative question of whether interthematic polarization is rational from each individual's perspective, it may also help to address the descriptive question of why interthematic polarization in fact occurs. Admittedly, much of the actual causes of interthematic polarization may well be a- or irrational ones, e.g. in-group bias, social cascades, and various other socio-psychological mechanisms. However, if interthematic polarization is also at least somewhat rational from each individual's perspective, then that may partially explain its prevalence and robustness (especially among seemingly rational agents). Second, the fact that interthematic polarization is often rational from each individual's perspective strongly suggests that measures to counteract this type of polarization should focus on social institutions, such as a society's media landscape, rather than on individual agents and their allegedly irrational tendencies.²²

²²I am very grateful to Glenn Anderau, Jonathan Matheson, James Norton, and several anonymous reviewers for written comments on drafts of this article. An earlier version of this paper was presented at the Social (Distance) Epistemology Network in 2020 and at the 2020 Gender Equality Forum in Reykjavik, Iceland.

References

- Abramowitz, A. I. (2018). *The Great Alignment: Race, Party Transformation, and the Rise of Donald Trump*. Yale University Press, New Haven.
- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *The Journal of Conflict Resolution*, 41:203–226.
- Axtell, R., Axelrod, R., Epstein, J., and Cohen, M. D. (1996). Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory*, 1:123–141.
- Baldassarri, D. and Gelman, A. (2008). Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114:408–446.
- Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., and Holman, B. (2017). Understanding polarization: Meanings, Measures, and Model Evaluation. *Philosophy of Science*, 84:115–159.
- Christensen, D. (2007). The Epistemology of Disagreement: The Good News. *Philosophical Review*, 116:187–218.
- Dorst, K. (2019). Why Rational People Polarize. *The Phenomenal World*, <https://phenomenalworld.org/analysis/why-rational-people-polarize>.
- Elga, A. (2007). Reflection and Disagreement. *Nous*, 41:478–502.
- Feldman, R. (2006). Epistemological Puzzles About Disagreement. In Hetherington, S., editor, *Epistemology Futures*, pages 516–236. Oxford University Press, Oxford.
- Frances, B. (2010). The Reflective Epistemic Renegade. *Philosophy and Phenomenological Research*, 81:419–463.
- Joshi, H. (2020). What are the chances you’re right about everything? An epistemic challenge for modern partisanship. *Politics, Philosophy and Economics*, 19:36–61.

- Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65:575–603.
- Kelly, T. (2005). The Epistemic Significance of Disagreement. In Gendler, T. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, vol. 1, pages 167–196. Oxford University Press, Oxford.
- Kelly, T. (2008). Disagreement, Dogmatism, and Belief Polarization. *Journal of Philosophy*, 105(10):611–633.
- King, N. (2011). Disagreement: What’s the Problem? or a Good Peer is Hard to Find. *Philosophy and Phenomenological Research*, 85:249–272.
- Klein, E. (2020). *Why We’re Polarized*. Simon & Schuster, New York.
- Lewandowsky, S., Oberauer, K., and Gignac, G. (2013). Nasa faked the moon landing—therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, 24:622–633.
- Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37:2098–2109.
- Matheson, J. (2015). *The Epistemic Significance of Disagreement*. Palgrave Macmillan, Basingstoke.
- Mayo-Wilson, C., Zollman, K. J., and Danks, D. (2011). The Independence Thesis: When Individual and Social Epistemology Diverge. *Philosophy of Science*, 78:653–677.
- McWilliams, E. C. (2019). Evidentialism and belief polarization. *Synthese*, 198:7165–7196.
- Nivola, P. S. (2005). *Thinking about Political Polarization*. The Brookings Institution, Washington, DC.
- O’Connor, C. and Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, 8:855–875.
- Pallavicini, J., Hallsson, B., and Kappel, K. (2021). Polarization in groups of

- Bayesian agents. *Synthese*, 198:1–55.
- Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press, Oxford.
- Pew Research Center (2014). *Political Polarization in the American Public*. Pew Research Center, Washington, DC.
- Singer, D. J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., Ranginani, A., and Berger, W. J. (2019). Rational social and political polarization. *Philosophical Studies*, 176:2243–2267.
- Weatherall, J. O. and O'Connor, C. (2020). Endogenous epistemic factionalization. *Synthese*, 198:6179–6200.