

Was it Polarization or Propaganda?

C. Thi Nguyen

*(This is a pre-print. Please cite the final version, which appeared in
Journal of Philosophical Research 46: 173-191 [2021]:
<https://doi.org/10.5840/jpr20211022183>)*

ABSTRACT

According to some, the current political fracture is best described as political polarization – where extremism and political separation infest an entire whole population. Political polarization accounts often point to the psychological phenomenon of belief polarization – where being in a like-minded groups tends to boost confidence. The political polarization story is an essentially symmetrical one, where both sides are subject to the same basic dividing forces and cognitive biases, and are approximately as blameworthy. On a very different account, what's going on is best described propaganda – where a discrete set of bad actors have manipulated some part of the media environment. The propaganda story is usually told as a highly asymmetrical story, where only some media consumers are under the spell of the propagandists. Which is right? I consider two analyses of the 2016 American election, and suggest that the propaganda account has better empirical support. I also offer a diagnosis of the appeal of the polarization story. Those who accept a polarization account are often political centrists, who accuse those at the political extremes of motivated reasoning – of believing what they find comfortable. Such centrists also tend to treat political extremism as the product of the irrational belief polarization, arising from living in like-minded groups. But, I argue, these arguments are too quick. First, we can't dismiss a group as irrational merely because they are likeminded. The existence of like-minded group can be explained in terms of irrational belief polarization, but it can also be explained by rational convergence on the truth. Second, belief polarization is not always irrational, such as when its emotional effects are used to repair impaired self-confidence. Third, political centrists are also subject to similar debunking argument. When we accept a polarization account, we get to feel the comfort of being "above it all". Political centrists are just as plausibly subject to the irrational effects of living in like-minded groups. Belief polarization isn't just for extremists.

KEYWORDS: polarization, propaganda, social epistemology, political philosophy, post-truth, echo chambers, fake news

I would like to stage a fight between two different accounts of the current political landscape — which has been sometimes called the “post-truth” era, the infodemic, or the end of democracy.

On one account, what’s going on is best described as *systemic polarization*. In this account, our once-peaceful society has come to be profoundly riven into polarized camps. Extremism and political separation are the core problems, and the fix is something like re-connection, intermingling, and friendship across party lines. The sound of this polarization story, in ordinary life, is somebody issuing a plea for civility, “in these divisive times”.

On a very different account, what’s going on is best described as *propaganda*. In this account, some people have, on purpose, filled some part of the media environment with false or misleading information, for political purposes. Under the propaganda story, the problem is the intentional manipulation of the media environment by bad actors. The solution lies in fighting or controlling those bad actors. The sound of the propaganda story, in ordinary life, is somebody saying something, “Fox News is all lies and bullshit,” and making gestures in the direction of Steve Bannon, Rupert Murdoch, and the Koch brothers — or the equivalent story from the right, pointed at the media left.

Systemic polarization is an essentially *symmetrical* story. The problem arises through a widespread social dynamic, where like-minded individuals clump together, start talking, and end up boosting each others’ confidence beyond what is reasonable. Conservatives and progressives are both subject to this polarization. The social forces at play — social mobility, online media bubbles, algorithmic filtering — are pervasive and their effect near-universal. The symmetry may not be perfect, but symmetry is the basic underlying shape. Both sides are approximately as vulnerable and approximately as blameworthy. The propaganda story, on the other hand, is usually told *asymmetrically*, where one side has come to be in the grip of this propaganda machine, and the other side is fighting against it. (The propaganda story could be told about both sides, but symmetry is not central to its basic logic.)

Of course, polarization and propaganda surely both occur to some degree. The real question is: which are the dominant factors, and in what proportion are they responsible for the current political landscape?

To make a first stab at this question, I want to stage a battle between two contemporary representatives — two recent books which, in my estimation, make particularly compelling, rich, and empirically sensitive cases for their respective stories. In the polarization corner, we have the political philosopher Robert Talisse’s recent book, *Overdoing Democracy*. In the propaganda corner, we have *Network Propaganda*, by Yochai Benkler, Robert Faris, and Hal Roberts — a team of researchers whose expertise lies in the empirical study of political and online life, especially the network topology of social media and professional media. Both books take as their starting point the political landscape around the 2016 American election of Donald Trump. And in the opening pages of both, you can hear the shock and intellectual disorientation in the writers, as they struggle to come to terms with what, exactly, has happened to American democracy.

I. TWO STORIES

Robert Talisse offers, in *Overdoing Democracy*, a sophisticated version of the polarization account, which links political philosophy with empirical work on the behavior of like-minded groups. Talisse’s account takes, as its starting point, Bill Bishop’s well-known analysis from *The Big Sort* (2009). According to Bishop, social mobility creates the opportunity for people to exert greater control over their location and community. This creates large-scale geographical sorting, as conservatives cluster with other conservatives, and liberals with liberals. This self-sorting aids the creation of like-minded enclaves, and decreases encounters with one’s opposites. To Bishop’s account we can add a new technological flourish. Social media, by enabling even more frictionless self-sorting, accelerates that process.

In Talisse’s telling, the effects of geographic mobility and frictionless technologies are significantly worsened by our tendency to, as he puts it, “overdo democracy”. What this means: once upon a time, there were plenty of public spaces where people of different political inclinations mingled. In those spaces, they could form a sort of connection. Liberals and conservatives might not actually befriend each other, but they could come to see each other as good parents, at the Little League game; or as charitable, while working together at the soup kitchen. In many cases, we wouldn’t even know what one another’s political orientations were. The point is, that politics would be besides the point. In such contexts it is easy to form, in Talisse’s terms, “civic friendship” — a relationship of respect for one another as fellow citizens, deserving of an equal role in democratic governance. And this form of respect is required for the basic functioning of democracy, since in democracy,

we need to accept the authority of rules that we disagree with, when our side is outvoted.

Social mobility erodes the space for civic friendship, but the true nail in the coffin, for Talisse, is saturated politicization. Once we attach political meanings to all kinds of acts, we will radically accelerate our sorting into like-minded groups. Once hunting gets coded as “conservative” and cleaning up parks as “liberal”, and so on for all of the other activities in our lives, then we lose our last spaces for running into, and forming some kind of relationship with, our political opponents (2019, 71-94). Even when we do physically mix on the street, our way of dressing has become so saturated with political signals that we can enforce social sorting, even when we have to occupy the same physical spaces. Talisse points to, for example, your MSNBC tote bag, or your camouflage t-shirt, as forms of such political signaling (91).

Once we sort ourselves so thoroughly, along political lines, we will tend to lose our capacity to be civic friends with the other side. And such respect is required for successful governance. Talisse summarises his position:

The past thirty years have seen the gradual decline of traditional sites of activity in which people cooperated together without regard for political affiliation. Specifically, our workplaces, neighborhoods, places of worship, households, and shared public spaces have become both more politically homogenous *and* more politically intoned. Although these spaces — workplaces in particular — used to serve as venues where citizens who might be politically divided could nonetheless work with each other and through those interactions come to regard one another as, say, a *dependable and skilled* coworker, *good neighbor*, or a *responsible parent*, they are progressively becoming settings in which individuals interact against the background of their salient political homogeneity. So our day-to-day social interactions are increasingly likely to put us in contact only with others who share our politics, and more and more of what we do together is regarded by us as also an expression of our politics. Our conception of a good neighbor, dependable coworker, and responsible parent are now likely to be infused with our political allegiances such that we gradually come to regard our political rivals as incapable of embodying these roles. (22-23)

Talisse’s basic slogan here is: the attempt to make everything political actually undermines the goods of politics.

Why does this occur? Once sorted into like-minded enclaves, we get a very specific effect. Talisse here leans heavily on empirical research into *belief polarization*.¹ Belief polarization “besets individuals who talk only or mainly to others who share their fundamental commitments,” and leads them to “embrace a more extreme version of their initial position” (97). So the formation of like-mindedness enclaves leads to increased moral and political certainty — which, combined with the lack of exposure to the other side, makes it easier to demonize the other side, and condemn them as malicious actors or moral imbeciles. And Talisse argues that the existence of belief polarization, as a psychological phenomenon, has been empirically, and robustly, demonstrated.

The current political rift, Talisse suggests, arises from the profound mutual disrespect between the two sides which is, in turn, rooted in this basic fracture in the social landscape. The fix, says Talisse, is to restore civic friendship — to find our way back to respecting the other side. Cass Sunstein, one of the major proponents of this sort of polarization picture, thought that, since the problem was structural, we needed a structural solution. It wasn’t just individual people making choices, but vast social changes in mobility and communications technology. So Sunstein suggests that we need public efforts to build more public spaces — to bring back the public square of old, in some newfangled technological form; to subsidize non-profit media channels, that will increase exposure to the other side (Sunstein 2007, 190-211). Talisse, on the other hand, presents a more personal solution (2019, 158-165). Even if we create, Sunstein style, public forums for political exchange between partisan opponents, those forums will still be *political*, with all the self-sorting and dismissal that implies (139). We, as individuals, need to make our own individual efforts to nurture *civic friendships* — where we can respect our opponents and their right to participate in shared governance, even if we disagree with them (150).

First, says Talisse, we need to change our view of our rivals. We need to see our political opponents as sincerely holding their values — we need to have “democratic sympathy” (147). And that involves, among other things, seeing that belief polarization affects everybody. The whole problem, says Talisse, is that we tend to think that it’s only the other side that’s in the grip of an irrational effect. We act as if belief polarization affects people asymmetrically (159). But this belief in asymmetry, says Talisse, is itself the result of belief polarization. The first step is to recognize that the effect is symmetrical — and that “our own views are likely products

¹ Talisse notes that this effect is usually referred to as “group polarization” in the literature, but he thinks this terminology muddies the waters (97, fn 4). I follow Talisse’s usage in this discussion.

of belief polarization" (160). And then, once we have repaired our inward dismissive tendencies towards the other side, we should engage in non-political cooperative projects with them. Pick up litter together, he suggests, or teach somebody to read at the library, or join a bowling league (163). Pick projects where you don't even know your fellows' political affiliations. This is how we undo total political saturation, and restore the possibility of civic friendship.

We find a very different story in *Network Propaganda* (2018). According to Benkler, Faris, and Roberts, a host of factors went into creating the political landscape around the election of Donald Trump. Many factors contributed something, they say, but, the dominant causal factor was propaganda. Propaganda, for them, is the intentional spread of false or misleading information, for the sake of political power. And, in their analysis, the most important sources of propaganda isn't the fancy new stuff, like Russian troll farms or bots. It's the old familiar stuff: the propaganda in mainstream news outlets. In Talisse's story, the prime movers are social mobility and pervasive political saturation. In Benkler et. al.'s story, the prime movers are Fox News, Breitbart, and their funders and allies among the political elite.

At the center is an effect they call the "propaganda feedback loop". In the propaganda loop, some media outlets stop trying to present truths, and stop trying to check their fellow outlets for the truth of their claims. Instead, those media outlets seek to give their followers partisan news that confirms the followers' world-view, instead of challenging it. These outlets will tend to draw followers who "seek confirmation more than truth" (78). Those followers then grow more used to the experience of constant confirmation, and grow more intolerant of any challenges to their belief system. As a result, they will decrease their trust in other media outlets, which don't conform to the loop's world-view. At the same time, the loop will attract political elites willing to align themselves with the loop's prevailing belief system, in exchange for the attention and approval of those followers. Once in the loop, departures from the party line are punished by lowered attention or disdain from the rest of the loop. So media outlets and political elites in the loop are highly incentivized to keep to the loop's party line. The result, say Benkler et. al., is a "self-reinforcing feedback loop that disciplines those who try to step off it with lower attention or votes, and gradually over time increases the costs to everyone of introducing news that is not identity confirming, or challenges the partisan narratives and frames" (79).²

² This propaganda loop I take to be a specific type of echo chamber, as I understand the term. Elsewhere, I have argued that the modern conversation tends to con-

At the center of the propaganda loop is a shift in the evidential interest in communication. Communication becomes more about establishing agreement and shared identity, than about finding truth.³ Importantly, the propaganda loop is not the inevitable outcome of some systemic social or technological force. It is a specific condition that afflicts some media networks, but not others. In fact, several chapters of *Network Propaganda* are spent carefully documenting the asymmetry of the existence of propaganda loops in US politics. At the center of their analysis is a careful, empirical study of the network topology of the media consumption environment around the 2016 election, based in part on their algorithmic analyses of large data-sets of Twitter and Facebook activity. To summarize their results: the basic network topology of the right-wing media ecosystem was extremely different from the network topology of the rest of the media ecosystem. The right-wing network centered around Fox News and Breitbart and excluded any mainstream media sources — and any sources that conformed to standard norms of objective journalism. False and misleading claims could — and did — circulate in the right-wing media, getting amplified without criticism from anywhere else in the trusted network. The left/mainstream network also included many left-wing media sources that delivered false information and failed to conform to the norms of objective journalism. But it also included, in the same network, left and centrist news sources that did conform to norms of objective journalism. And most readers in the left network read widely across all those sources — and treated them as beholden to, and in conversation with, one another.

According to Benkler et. al., the propaganda feedback loop is not inevitable, and not pervasive. In fact, they say, at the time of the 2016 election, the mainstream/left media network operated with an entirely different dynamic: what they call a “reality-check dynamic” (77-8). In the reality-check dynamic, media outlets are incentivized to check up on each other. For one thing, media outlets are rewarded with attention and support for discovering errors in other outlets inside

flate two very distinct social structures. *Epistemic bubbles* are structures where members don't *hear* the other side. *Echo chambers* are structures where members don't *trust* the other side (Nguyen 2020). My analysis is based on Jamieson and Cappella's (2008) groundbreaking empirical analysis of the echo chamber around Rush Limbaugh and Fox News, and I use the term in the way they used it — a usage that I think has been mostly forgotten of late.

³ I offer an extended discussion of several such shifts evidential interest in (Nguyen and Williams 2020; Nguyen forthcoming a, forthcoming b). Many of these cases seem to involve dropping one's epistemic standards in exchange for pleasure or a sense of security.

the same network. Such aggressive fact-checking is incentivized — catching another media outlet's mistake counts as "getting the scoop". As Benkler et. al. carefully document, in the period around the 2016 election, the network dynamics of the right-wing media ecosystem were entirely different from the network dynamics of the mainstream/left ecosystem. On the right, false reporting was promulgated and largely went unchecked. On the left and center, false reporting was often initially promulgated — and then quickly discovered and criticized by a different media outlet. Usually, these disagreeing fact-checks were also widely promulgated. The false story quickly died out — and the reporters of those falsehoods were typically disciplined or fired (85-99). As Benkler et. al. put it, in the propaganda loop dynamic, media outlets are incentivized to remain loyal to the party line and police it; while in the reality-check dynamic, media outlets are incentivized to seek the truth and police failures in truth-seeking (80).

Reporters and politicians on both sides often lie. Benkler et. al. report that their preferred fact-checker, PolitiFact, recorded approximately the same percent of "mostly false" claims from Republican notable Mitch McConnell (43 %) and Democratic notable Nancy Pelosi (41 %), and a similar symmetry between Bill O'Reilly (53%) and Rachel Maddow (48%). Notably, the right-wing media ecosystem had much higher heights of falsity — Rush Limbaugh weighed in at 81%. It seems plausible that these higher heights are enabled by the very loop dynamics described by Benkler et. al. What matters most, however, are the "network dynamics" in response to those falsehoods (83). The right-wing media ecosystem showed no tendency to correct errors or punish false reporting. The mainstream/left media ecosystem did often deliver false or misleading news stories, but also quickly subjected those stories to fact-checking. Errors were quickly discovered, which almost always led to retractions, wide-spread reporting of retractions, firing, and other standard performances of the apparatus of journalistic objectivity. In short, the mainstream/left-wing ecosystem was a large and unified network that contained actors who conformed to norms of objective journalism and were incentivized to find errors in their brethren. And their audience treated these various actors as in conversation and beholden to each other. These dynamics tend to keep the whole system mostly honest.

This happens as a result, they suggest, of a differential incentive structure. Inside the propaganda loop, there is no incentive for error-correction; all the incentives are towards punishing defectors from the party line. Outside the propaganda loop, there is actually plenty of incentive for error-correction. Among other things, once a propaganda loop becomes established on one side, such that its media sources

become typically unreliable, then the other side is incentivized to hold strongly to norms of journalistic integrity. Once the right-wing media environment went all-in on propaganda, it was to the advantage of the left to embrace and uphold those norms of journalistic integrity — partly because those norms provide plenty of ammunition against the right! This dynamic is why we often see asymmetry in propaganda dynamics between political opponents, say Benkler et. al. Once one side enters a propaganda loop, the other side has very good reason to uphold norms of journalistic integrity (81).

Network Propaganda includes some significant consideration of other agents in the 2016 election — including right-wing trolls, Russian hackers, and algorithmic online effects. But the researchers conclude that these other agents were far less causally important than the propaganda loop in large-scale, institutional media. And they specifically reject the framing of “polarization”.

...The behavior of the right-wing media ecosystem represents a radicalization of roughly a third of the American media system.... To speak of “polarization” is to assume symmetry... No fact emerges more clearly from our analysis of how four million political stories were linked, tweeted, and shared over a three-year period than that there is no symmetry in the architecture and dynamics of communication within the right-wing media ecosystem and outside of it. (14)

They also offer an argument against the Talisse-style systemic polarization story. If the systemic polarization story is right, they say, then we should see similar effects across both sides of the political spectrum, and across different geographical areas and communities. But we don't. In fact, we see demonstrably asymmetric network dynamics, as Benkler et. al. have documented. And we see great differences in polarization in different regions — despite similar degrees of social mobility and exposure to technology. This is best explained in terms of long-term, asymmetric political changes, like the racial politics of the GOP's Southern Strategy (295-339).

So: which is right? On the face of it, *Network Propaganda* has one major mark in its favor. Talisse's account largely relies on small-scale laboratory studies about belief polarization. We know, certainly, that belief polarization is a psychological phenomenon. There is no direct evidence that the phenomenon of belief polarization is the primary factor at play in causing the large-scale political effects being observed. We are supposed to connect the dots between the small-scale studies showing that there is such an effect, and the large-scale conclusion that the effect provides the best explanation of the observed social discord. And we would have

good reason to so connect the dots, if there were no other good explanations for that discord. But there are: *Network Propaganda* provides at least one. What's more, *Network Propaganda*, does provide at least some large-scale evidence of the probable impact of propaganda effects. The spread of intentionally-created propaganda across information networks can be directly observed and measured. Of course, this may be because media misinformation is easier to document, while showing belief polarization on a mass scale is significantly harder. So we need to proceed with some significant caution, as we examine the prospects of these two competing accounts.

II. TWO POLARIZATIONS

But first: there are very different notions of polarization in play here. Since the terminology seems highly variable between different fields of scholarship, let's adopt Talisse's usage. According to Talisse, we should distinguish between *political polarization* and *belief polarization*. Political polarization, says Talisse, is a measure of the *distance* between political opponents. The further apart the various elements — the party planks, the commitments of the member of the party — the greater the political polarization.⁴ Belief polarization, on the other hand, is a specific phenomenon that happens inside a particular group — one where like-minded enclaves increase the extremity of their belief (2019, 96-100).

Belief polarization and political polarization are entirely different phenomena, conceptually and mechanically. Belief polarization is an *effect* — it is a *dynamic psychological phenomenon that besets like-minded enclaves and makes them more extreme*. Political polarization is a *state of a populace* — a measure of how far apart the basic clumps are. In Talisse's accounting, these two distinct phenomena are causally linked: symmetrical, system-wide belief polarization *leads to* that wide gap, which is political polarization. But we should note: belief polarization is only one path to political polarization.

Let's take some time here, because confusing these forms of polarization can lead to some deadly misreadings of the data. First: these two forms of polarization are conceptually distinct; and as real-world phenomena, they can come apart. There can be political polarization without belief polarization. Here's a (perhaps

⁴ Talisse provides a useful discussion of the dimensions of political polarization, including polarization of party platyforms, degree of ideological purity in the parties, and the degree of antipathy towards members of the opposing group (98-9).

excessively) obvious example: pre-war Nazi Germany was a highly politically polarized environment, under this very thin definition of “political polarization”. But I think that it’s pretty obvious that this political polarization wasn’t the result of the specific process called belief polarization — at least, not symmetrically. Perhaps the phenomenon of belief polarization helps explain why the Nazis were so Nazi. But what’s the best explanation of the extremely anti-Nazi stance of the German resistance? Very plausibly, it’s that the Nazis were evil, morally corrupt — and the German resistance was responding accurately and sensitively to the moral stench. The case of pre-war Germany is one of political polarization, but the primary cause of that political polarization is not simply systemic and symmetrical belief polarization.

Similarly, you can have belief polarization without political polarization. Here’s a simple example: imagine a population of all like-minded centrists, who all believe in the goods of moderation and civility. The belief polarization effect could just as easily take hold of them, pushing their degree of confidence in their beliefs up beyond their evidence. Is it even possible to be too committed to moderation? Sure it is: one might, for example, become excessively confident in one’s moderate beliefs. The tenor and content of one’s moderate beliefs might also change. One might become instantly dismissive of any instance of moral outrage; one might knee-jerkly demand moderation in all circumstances, even when it is inappropriate, or condemn with vitriol any passionate political extreme. Imagine a society-wide belief polarization in favor of centrism in a time of wide-spread slavery, racism, or misogyny. (Just imagine!) Such a belief polarization would produce the opposite of political polarization — but that is not obviously always for the best.

Here, we also need to be wary of different senses of “extreme”. In political polarization, the notion of an “extreme” view is one of distance from the center — towards the far right or the far left. But in the discussion of belief polarization, groups tend towards increasing the extremity of the content of their beliefs, by *increasing their degree of confidence* and by increasing *the extremity of their belief*. So it’s possible, in this latter sense, for belief polarization to affect centrists — who will then become excessively confident in their politically moderate beliefs, or acquire beliefs with contents more aggressively dismissive of political radicals. This sounds like terminological nit-picking, but confusions across these senses of “polarization” and “extreme” are common in the contemporary cultural discourse.⁵

⁵ The effect shown in empirical studies of belief polarization is simply that like-minded groups tend to increase their confidence and “extremity” of their beliefs, whatever that belief. Certainly this seems like it should apply to like-minded centrist groups.

Let's extract some key lessons. It is quite tempting, for some, to go from the mere observation there is a considerable distance between political opponents — political polarization — and conclude that the cause was some irrational and symmetrical effect. But the mechanism of belief polarization is only one route to political polarization. There are other routes, including ones that do not symmetrically debunk the rationality of both sides. It's safe to assume that political polarization is real. So, in order for Talisse to make the case that belief polarization is the best explanation, and that civic friendships are the cure, he owes us an account of why belief polarization is the right origin story for political polarization, as opposed to one of the others. And note that Benkler et. al.'s description — of the reality-check dynamic and its response to the propaganda loop dynamic — also explains the observation of political polarization, but in a way that does not symmetrically discredit the rationality of both sides.

III. IDENTIFYING BELIEF POLARIZATION

Interestingly, I can find no direct empirical studies that inquire into the possibility of centrist belief polarization. The studies that Talisse describes, at least, are ones in which belief polarization is tested by sorting people into two groups — approximately, one liberal and one conservative, and noting that each group becomes more liberal and conservative (103-4). See, in particular, Sunstein's (2009, 1-20) discussion of his studies in belief polarization, on which Talisse relies as a touchstone. In these studies, no test group of like-minded centrists was created or measured. This suggests that, perhaps, the experimental design of these polarizations contains a presumption about what counts as "polarization towards an extreme", and are, as experimental designs, worrisome in that regard. Talisse's discussion seems to reflect this omission — the examples are of groups of feminists becoming more ardent feminists, racists become more committed to their racism (109). There is simply no discussion of the question of centrist polarization, nor the possibility of more extreme centrist beliefs. I worry that this represents a systemic issue, and presumption, in the literature on polarization. In particular, I worry that the term "polarization", and the associated image, have confused the issue. The documented effect — of like-minded groups increasing their confidence and extremity of their beliefs — is not necessarily a matter of movement towards poles on a single dimension. But the term and image "polarization" perhaps steers us in that linear direction. But the underlying effect — the increase in confidence and intensity of belief in like-minded groups — could potentially involve movement away from the poles, towards the center. In any case, the issue is surely worth further consideration. (For a useful discussion of how the conversation on polarization often collapses very different phenomena, see Bramson et. al. (2017).)

Why is the mechanism so important here? Systemic belief polarization is, in Talisse's hands (and the hands of many similarly-minded accountings) a symmetrically discrediting story, because it tells a story of symmetrical irrationality. Belief polarization involves an irrational change in belief. As Talisse puts it, it is not a matter of explicit reasoning, but an unconscious effect. It *happens to* people, rather than being the product of actual thought (2019, 105). It, says Talisse, "typically produces a shift in belief regardless of one's evidence. The phenomenon often causes us to hold beliefs that are not adequately supported by the evidence that we have" (106). Or, as he puts it later:

...Belief polarization produces extremity shifts in our belief contents and in our overall commitment to our viewpoints, but it does not provide us with correspondingly better arguments, reasons, or evidence. To the contrary, the phenomenon leaves us generally in a cognitively *worse* position than we are likely to have been in prior to the effect. Once belief-polarized, we believe more confidently things that we are less able to support with reasons. (123)

Consider the mechanisms of belief polarization. First, suggests Talisse, like-minded enclaves expose us to too much of one kind of argument in favor of one side, and a scarcity of arguments on the other. This, he says, is plausible, but cannot be the whole story. Second, like-minded enclaves give us a reason to conform to the group beliefs, so as to appear authentic. This, he says, explains some of the data, but also has its limitations. Talisse's preferred view is that the simple experience of corroboration *feels good* to us — and that these good feelings make us more committed (111-4). But whichever of these mechanisms you prefer — or perhaps you prefer a stew of all three — the discrediting force is apparent. Manipulated exposure to the arguments, authenticity-signaling, and hedonism — all of these are bad reasons to adjust one's beliefs. And their result is not only bad beliefs, but bad attitudes towards one's opponents.

From the standpoint of that intensified outlook, opposing views and countervailing considerations are bound to appear distorted, feeble, ill-founded, and strenuous. To employ an image that hopefully is not too farfetched, the belief polarization phenomenon operates almost like the side-view mirror on an automobile; it consolidates our first-personal perspective and makes everything else appear distant and disfigured... Once we are sufficiently belief-polarized, those who espouse views that differ from ours will strike us as more and more alien, benighted, incoherent, and perhaps even unintelligible. (118)

But when should we think that belief polarization is the right explanation for what's going on? We know that belief polarization happens; it's well documented. But is it the prime mover in any case where we find increasing degrees of belief in a like-minded group? And was it the prime mover in the case of the 2016 election?

It is tempting to look at a like-minded group increasing their degree of confidence and conclude that their high confidence arises from an irrational process of belief polarization. And if that were actually a dependable inference, we would be possessed of an incredibly powerful sword of discrediting. Any confidence that we found in a like-minded group could be instantly debunked and blamed on an irrational effect. Notice that this form of debunking involves a transition from some external features of a group's shared belief to a conclusion about their rationality. We see a like-minded group increase their confidence; we attribute that rise of confidence to the irrational effect of belief polarization; so we conclude that the group's degree of confidence is irrational.

But, again: there are other, alternative explanations for these external appearances. Let's give these external appearances a name that does not intrinsically include a presumption of irrationality. First, let's call any group with a shared belief, which they hold with a high degree of confidence, a *confident enclave*. Identifying a confident enclave is something you can do at a particular moment in time. Furthermore, we can notice that some groups went through a process of collective bolstering — that the individual people once held a belief with a lower degree of confidence, but then gathered in a like-minded enclave and subsequently increased their degree of confidence. Let's call that *enclave boosting*. Enclave boosting happens whenever an enclave of like-minded individuals increases their confidence in a shared belief.⁶

However, we can't infer from the existence of a confident enclave, that their confidence necessarily arose as the result of the irrational process of belief polarization. Here's an example of a confident enclave: the group of trained medical

⁶ Note that you could use the term "belief polarization" in two ways. First, you could take belief polarization to be equivalent to what I call a "confident enclave". In that case, belief polarization does not necessarily mean irrationality. Second, you could take belief polarization to always involve some irrational process. This is not a substantive difference, but merely a semantic difference — but we should be careful to avoid translation errors between the different senses of belief polarization. I have elected to employ the latter usage, because it seems more in line with Talisse's account, this his language changes slightly in the book. In the first introduction, belief polarization is described as typically leading to irrationality (103-5), but later in the text, the language is stronger (123).

doctors, who share a high degree of belief in the efficacy of vaccines. The existence, in a group, of like-minded confidence can have some very different sources. It might originate in belief-polarization-style effects, or it might originate because the group shares a methodology. And if that methodology is sound (and if it was executed independently), then the existence of a high degree of confidence across the group is actually a reason to increase one's trust in that group's shared belief. The convergence of independent scientists is a marker of epistemic success. In other words: Talisse, like others who study belief polarization, warns us that belief polarization tends to create more "uniform" beliefs in a population (108). But there are multiple potential explanations for uniformity: one is an irrational force of homogenization, and another is convergence on the singular set of true beliefs.

Fine, we might say, the mere observation of confidence across an enclave isn't enough. What about when we observe, over time, the process of enclave boosting? Can we conclude that the *process* was irrational? Again, I don't think we can do this so easily. First, as with the medical case, learning a reliable methodology may be an entrance requirement for the group. Doctors often have similar beliefs, because the process of becoming a doctor involves a similar kind of training — and insofar as that training involves learning a reliable methodology, the similarity of beliefs across medical doctors is not, by itself, discrediting.

Second, enclave boosting may be the result of a genuine and epistemically appropriate exchange of reasons. This, again, happens in the sciences. One group of scientists concludes that Vitamin D prevents cancer based on population-level epidemiological studies; another arrives at the same conclusion from laboratory tests on lab rats; another does it through the computer modeling of human cells. When these various groups gather, they each learn new and additional reasons to confirm the beliefs they already had. Different people can come to the same conclusion by different pathways, and when they learn of each others' pathways, they have good reason to increase their confidence. And the reason that they gathered may be that they have, say, similar interests and high standard of reasoning. This may also help to explain why, before gathering, they had already developed some inclination towards the same belief. Insofar as they are following some truth-tracking methodology, they may have already started down the right path. But, in gathering, they exchange their various separate reasons, and their confidence increases — for perfectly legitimate reasons. Both their initial like-mindedness, and the ensuing enclave boosting, can be explained in terms of the convergence of good epistemic actors on the truth.

So enclave boosting has at least two possible explanations: it could be from a

rational effect of information-exchange, or it could be from some raw emotional boost in mutual confidence that arises from the mere fact of being surrounded by like-minded folk. But suppose we discover, for a given case of enclave boosting, that there is no exchange of reasons. Suppose there was only an emotional boost to the enclave's general degree of confidence, merely from the creation of a like-minded community. Even so, the process is still not necessarily irrational. Some deployments of community-based emotional support can rationally increase confidence.

Suppose you have low confidence in your observations and beliefs, and you come to think that this low confidence arises, in part, from political circumstances. The ambient culture disrespects you and your intellectual capacities, perhaps because of your race or gender, and you have acquired the habit of excessive self-distrust (Jones 2012; Medina 2013, 27-8). But this self-distrust doesn't vanish as soon you realize its irrational origin. As Karen Jones puts it, self-trust is not purely reasons-responsive; it is an emotional disposition acquired through a social circumstance. And such emotional dispositions can be recalcitrant. Merely realizing that you have been so exposed to a lifetime of disrespect and that your self-confidence is too low, may not be enough, in and of itself, to repair your self-confidence in your reasoning abilities. You may need a long soak in a curative bath — another lengthy exposure to a supportive community, to get the right emotional counter-balance. So if you have good reason to think that your self-trust has been unwarrantedly depressed, then you have good reason to consciously seek out a like-minded enclave, to help the process of repairing your dispositions and emotional self-outlook, and set them to a more appropriate level. And that emotional boost, if deployed in an appropriate response to the circumstances, is not a distorting one, but a reparative one — a cure for a long-standing social ill. Sometimes, toxin and cure need to come in a similar emotional key.

In short: the support of like-minded individuals is an effect that increases confidence. If that effect is deployed on people with an appropriate degree of confidence, and ends up boosting their degree of confidence to inappropriate degrees, then the effect drags us away from rationality. But if it is appropriately deployed to repair inadequate degrees of confidence, then the effect repairs rationality.⁷

⁷ Sunstein does make room for something like this effect, when he talks about the possibility of “good extremism”. He grants that good extremism is sometimes possible, but ends up quickly dismissing it (to my mind, too quickly). He is far more concerned about belief polarization carrying the group away from the evidence (2009, 153). Talisse does not seem to give serious consideration to the possibility that polarization can be

And this, I think, is precisely the idea behind various forms of safe spaces and similar strategies.⁸ The emotional effects of like-minded support may be harnessed to restore, rather than undermine, one's proper epistemic functioning. And the effect of this support can involve both the degree and content of one's beliefs. One may have the right belief, but be unable to bring oneself to be adequately confident in oneself and it. But perhaps more interestingly: inadequate confidence in one belief may prevent one from exploring down the line. If you are stuck over your worry about a particular belief, you may be unwilling or unable to follow the thought to where it leads. This is perhaps why, as Miranda Fricker points out, early feminist groups used "consciousness-raising circles" — like-minded groups, which functioned, not just to increase confidence, but to explore new conceptual terrain. She credits such consciousness-raising circles as creating the intellectual space which enabled women to first articulate the concept of "sexual harassment" (Fricker 2007, 148). Safe spaces can free one to follow a line of thought.

The point is not to show that belief polarization doesn't exist, or that we shouldn't be worried about it. Nor is the point that the left, or people who claim the banner of "safe spaces", are always good epistemic actors. The point is that various external signs — the existence of confident enclaves, or the process of enclave boosting — are insufficient to demonstrate that an irrational process of belief polarization has occurred. We need much more to show that belief polarization has happened, and much more to discredit a confident enclave, or a boosted one. We need to show, first, that the belief change is due to merely like-minded exposure, instead of as an appropriate response to the evidence. And second we need to show that such like-minded exposure led to overinflated confidence, rather than restoring appropriate confidence. There's more than one route to a confident enclave, and only some bear the mark of irrationality.

To put it another way: let's agree that like-minded enclaves increase degrees of confidence and extremity of belief. The claim that this change is essentially irrational depends on the presumption that the pre-enclave state was more likely to be rational, and that increasing confidence through mutual support is always a deviation from rationality. But that presumes that the level of confidence in the original, pre-enclave state was appropriate. To put it more bluntly: the framing of

harnessed for curative effects.

⁸ Anderson (2021) provides a far more detailed case for a similar point. DiPaolo (2019) provides a useful discussion of the epistemic justification of using such effects, as "norms of compensation".

belief polarization presumes that people basically have it right before they get together and get more extreme. But that presumes that independent thinkers who tend towards moderate thoughts are more likely to be right. If, on the other hand, you think that thinkers have been systematically led astray — by, for example, systematic racism and sexism — and that “independent” thought processes typically reiterate the biases and problems of the status-quo, then you have good reason to think that like-minded support, managed appropriately, could increase epistemic quality, rather than undermining it.

IV. SO MANY KINDS OF DEBUNKING!

Of course, one might worry that all of this is quite self-serving. Both the *Network Propaganda* view and the defense of “safe spaces” I’ve mounted weigh in favor of a progressive position. One might worry that this is the result of motivated reasoning — that I’m giving just the kind of self-serving argument that somebody on the left would give, to justify their beliefs. Perhaps I am drawn to such an argument precisely because I have already been part of a like-minded enclave, whose beliefs have already been boosted far past the truth. Perhaps I have been snowed by the pleasure of the confirmation of my pre-existing beliefs.

Talissee makes exactly such an accusation. He says that we tend to think belief polarization is asymmetrical and affects the other side, but not us. But this belief in asymmetry, says Talissee, is itself the result of belief polarization:

Indeed, it should be recognized that the tendency to disregard our own vulnerability to the phenomenon is itself a further manifestation of the polarization dynamic; the inclination to see belief polarization at work only in unlike others is yet another mechanism by which we are propelled into further extremity, hived into likeminded clusters, and insulated from the ideas and criticisms of anyone who is not like us. (159)

Of course, it’s true that any conclusion *could be* the result of belief polarization and other sorts of motivated reasoning. But notice that this sort of argument applies equally to all comers. Motivated reasoning isn’t just for extremists and radicals. Those worries apply just as well to those who might call for civility, preach for moderation, and disdain the extremes. Again: belief polarization can beset any enclave at any place on the political spectrum. And motivated reasoning can lead one

to prefer arguments for one's love of civility and moderation, just as well as motivated reasoning can bring one to prefer arguments for one's far-left (or far-right) position. We could even attempt to judo Talisse's suggestion, and argue that the temptation to accept a Talisse-style view of symmetrical belief polarization is itself a result of belief polarization — one arising in a body of like-minded centrists who would love to believe that the real problem was in all those irrational, polarized extremists.

In other words, we can certainly make an accusation of motivated reasoning, and supply a debunking argument, for somebody, say, on the left, who buys into the *Network Propaganda* view and is tempted by the epistemic safe-spaces argument I have offered. But there is a parallel accusation and debunking argument we can make for a Talisse-style view of systemic polarization. In fact, Benkler et. al. provide us with one:

As we have repeatedly seen... the prominent outlets on the left and center simply do not exhibit a parallel structure, content, or vehement outrage that we observe on the right. These facts are as inconvenient to academics seeking a nonpartisan, neutral diagnosis of what is happening to us as they are to professional journalists who are institutionally committed to describe the game in a nonpartisan way. Both communities have tended to focus on technology, we believe, because if technology is something that happens to all of us, no partisan finger pointing is required. But the facts we observe do not lend themselves to a natural, "both sides at fault" analysis. (292)

In other words, we can just as easily ascribe motivated reasoning to proponents of the systemic polarization story — such as the motivated reasoning of wishing to appear non-partisan and above the fray. Or, if we wished to be more radical: we could suggest that Talisse-type positions are anti-radical and so help justify the status quo, so anybody who benefits from the status quo has a motivated reason to accept such a position.

The point here is not to dismiss either position out of hand. The point is that the position of advocating for moderation, civility, and civic friendship does not somehow magically rise above the fray, and render itself, by its peaceable face, immune to debunking arguments and accusations of motivated reasoning. We can point out that *Network Propaganda* is comforting to liberals and leftists, but we should also point out that *Overdoing Democracy* is comforting to centrists, and anybody wary of the apparent extremism of those seeking radical change.

The dialectic here has a similar structure to another dialectic, in a nearby space

in the cultural discourse. It is common to hear accusations against moral crusaders of merely “virtue signaling” or “moral grandstanding”. The accusation is that all those people pouring out moral outrage and condemnation, say on Twitter, are just posturing, to gain the look of being moral. They are, in the words of Brandon Warmke and Justin Tosi (2016), using morality for social status. In everyday online life, such accusations are often leveled at leftist activists. And of course, they might be. But notice that the accusation of “virtue signaling” may be leveled at any form of displayed virtue, and any moral content. The calls for civility and moderation over extremism and outrage, are themselves just as plausibly forms of virtue signaling. In fact, the accusation of virtue signaling itself could be a kind of virtue signaling. It is only the particular kind of virtue that is signaled that varies — moderation and civility over, say, an interest in radical social change.

As the saying goes, debunking arguments are cheap. You can find them for pretty much any position. What do we do, if we can find a plausible debunking argument for all the involved positions? I think we have to fall back on the evidence. And there is evidence for the reality and causal power of asymmetric propaganda, and less evidence (so far) for the actuality and large-scale causal power of systemic belief polarization. Again, let me caution: this may simply be a matter of insufficient empirical research on a difficult and subtle matter. And we do have to be wary of the possibility that we are cherry-picking the evidence that supports our preferred position. But we also have to be wary of the possibility of dismissing the best-supported view, and all the evidence that supports it, with a too-quick leap to an accusation of motivated reasoning..

V. WHICH CAME FIRST, THE POLARIZED CHICKEN OR THE PROPAGANDIZED EGG?

It is important to figure out what the real causes are here, because we need to figure out the right fix. (And if it is a tangle of causes, we will need a tangle of fixes.) If Talisse is right, then his fix becomes paramount: we need to restore civic friendship and respect to restore peace to our fractured society. But if *Network Propaganda* is right, then Talisse’s suggestions will miss the target. If the cause of the fracture is intentional misinformation from media outlets, then the creation of civic spaces where we intermingle probably won’t help us very much. If the other side already believes you to be supporting a cabal of Satan-worshipping pedophiles (as is the actual core content of the QAnon conspiracy theory), then it doesn’t seem

like it will matter very much if our kids are in Little League together. The right response to the devil, and his followers, is constant vigilance, and not the open-hearted attitude that could lead to civic friendship and trust repair. Similarly, if Talisse is right, then cutting off one head of the propaganda hydra will be for naught; the underlying social dynamic will just grow another functional equivalent. But these two views may not be as far apart as it might first appear. In fact, there is a way to present each story such that it becomes the lynchpin of the other story.

Here's one way to tell the story. Let's call it the *polarization-first* story: the fracture of civic society, and the mutual disrespect that it created, laid the foundation for the effectiveness of that propaganda. All those efforts at propaganda efforts can't *take hold* until a certain amount of ambient civic disunity already exists. How can you believe your neighbor is a member of a Satanic death-cult if you still do stuff together and have mutual respect for each other? It's that civic distance that stops the inflow of evidence, that makes it possible for the cartoon pictures of the other to take hold.

Here's another way to tell the story: the *propaganda-first* story. Propaganda lays the groundwork for civic fracture. In fact, the propaganda-first account becomes even easier to tell, once you accept the particulars of Talisse's account of polarization. For Talisse, the civic fracture comes in key part from our having developed pervasive beliefs about what activities count as conservative and progressive. The destruction of civic spaces can be driven by the manipulation of our beliefs about the political valence of various activities, which, the propaganda-firster could claim, are fostered by propaganda. It's propaganda that teaches people that, say, only liberals pick up litter — and that's what's driving the political saturation.

I don't think we can find an answer in the current state of the evidence, at least as it is presented in these two books. The work on belief polarization has tended towards small-scale studies of particular groups and computer modeling. Talisse provides plenty of evidence that belief polarization is *a* real phenomenon, mostly citing small-scale studies of group behavior. But he never, as far as I can tell, provides evidence that it is the dominant force responsible for present-day political polarization. In my survey of his cited empirical research, I have seen little to show that belief polarization is actually the primary causal force at play in political polarization. Nor are there significant attempts to compare belief polarization to other, more rational effects that lead to polarization.

On the other hand, *Network Propaganda* consists of large-scale studies of network connectivity. Benkler et. al.'s research concerns the structure of news media consumption, analyzed using trawls through data about what news outlets tend to

be viewed by which groups. There isn't much work on the ground level psychological forces that lead individuals to enter into these network topologies. Though, again, there is Benkler et. al.'s argument that the current political landscape is unlikely to be the result of systemic belief polarization, because of the deep asymmetry in the network topology and the deep unevenness of the effect across different geographic terrains. There is space for a polarization-first retrenching of Benkler et. al.'s analysis, though it would have to do significant work to explain those documented asymmetries.

What conclusions can be drawn? Talisse's presents an enormous degree of confidence in his conclusion about civic fracture as the original sin of the current discord. I find this confidence unwarranted by the quality of the presented evidence. I am, given the current state of the evidence, more swayed by Benkler et. al.'s data about the deep asymmetry of polarization and the incompatibility of those observations with systemic polarization. Notably, Talisse barely discusses propaganda and misinformation, and provides no reasons to prefer his systemic polarization account. This seems particularly striking, since claims about propaganda and misinformation were widespread and highly salient in the 2016 election.

Here is a particularly telling piece of data marshaled by Benkler et. al.: animosity is actually asymmetric. The portrait Talisse tells is one of equivalent animosity — which is exactly what we should expect from an explanation in terms of systemic and symmetrical belief polarization. But, the data show that this animosity is actually moderately asymmetric, at least in US politics. Democrats and Republicans certainly have increasing animosity towards each other, but Republican negative affective response has increased more. And though negative affect was high across both Democrats and Republicans, among the group of strongly partisan Democrats and Republicans, "outparty animus is significantly higher among Republicans."⁹

As Benkler et. al. note, the symmetrical picture of polarization has to combat our "lived experience" of asymmetry — of, as they note, constant lying from Donald Trump, and constant media support for those lies (292). An account like Talisse's owes us a note of explanation here. Talisse could give us an account of why that lived experience is basically false — either committing to showing that the right-wing media ecosystem is not full of lies, or that the center-left media ecosystem is as full of lies. I take it that Talisse does not want to go so far. But in that case, Talisse

⁹ This comes from Benkler et. al. (307-8) summarizing a set of studies, most notably Iyengar and Westwood (2015).

owes us an account of why his story of symmetrical belief polarization is worth treating as the dominant story, when the story of asymmetric propaganda fits so nicely with much of the empirical data, as well as the basic observations of our lived experience.

But that is all given the current state of the evidence. And I think the most important thing, going forward, is to gather more. The phenomenon of the political moment is so strange and vast and tidal; I urge we resist the temptation to leap on the nearest comfortable story, and instead seek further evidence. I do, however, want to emphasize one last thought: Talisse's position is not a neutral position. There is no particular reason to treat it as the right opening assumption, and to place the burden of proof wholly on departures towards more asymmetric explanations. The moderate, anti-conflict position is not somehow free of the possibility of debunking, discrediting, or accusations of motivated reasoning. Civility is not a default. And leaping to accept, with total confidence, a story of symmetrical irrationality, without sufficient evidence — when there are other competing accounts, which themselves have significant evidential support — itself bears the mark of motivated reasoning.¹⁰

BIBLIOGRAPHY

Anderson, Derek. 2021. "An Epistemological Conception of Safe Spaces," *Social Epistemology*. DOI: 10.1080/02691728.2020.1855485

Aaron Bramson, Patrick Grim, Daniel J. Singer, William J. Berger, Graham Sack, Steven Fisher, Carissa Flocken, and Bennett Holman. 2017. "Understanding Polarization: Meanings, Measures, and Model Evaluation," *Philosophy of Science* 84 (1): 115-159

DiPaolo, Joshua. 2019. "Second Best Epistemology: Fallibility and Normativity," *Philosophical Studies* 176: 2043-2066.

Jamieson, Kathleen Hall and Frank Cappella. 2008. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University

¹⁰ Special thanks to Mark Alfano, Melissa Hughs, Geoff Pynn, Aaron Rabinowitz, and Cailin O'Connor Weatherall, for their contributions to this paper. Thanks to my students at the University of Utah 2021 Social and Political Epistemology class, who helped chew over these texts with me. And a special note of thanks to Joshua DiPaolo for all his help in thinking these matters through. His past and in-progress work on these issues have contributed enormously to my understanding of polarization and debunking.

Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power & the Ethics of Knowing*. Oxford: Oxford University Press.

Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization," *American Journal of Political Science*, 59(3): 690-707.

Jones, Karen. 2012. "The Politics of Intellectual Self-trust." *Social Epistemology*. 26(2): 237-151.

Medina, Jose. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.

Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17(2): 141-161.

—. Forthcoming a. "The Seductions of Clarity," *Royal Institute of Philosophy Supplements*.

—. Forthcoming b. "How Twitter Gamifies Communication," *Applied Epistemology*, ed. Jennifer Lakey. Oxford: Oxford University Press.

Nguyen, C. Thi and Bekka Williams. 2020. "Moral Outrage Porn," *Journal of Ethics and Social Philosophy* 18 (2): 147-172.

Warmke, Brandon and Justin Tosi. 2016. "Moral Grandstanding," *Philosophy & Public Affairs* 44 (3): 197-217.