The Rationality of Affective Polarization

Brian Kogelmann

Paul Wheatley

Abstract: While philosophers have extensively examined whether belief polarization can be rational, they have largely ignored the rationality of affective polarization. Affective polarization is the tendency for partisans to view opposing partisans negatively and copartisans positively. This paper investigates the rationality of affective polarization. We evaluate this phenomenon by relying on two prominent philosophical theories of rational emotions. The first theory holds that emotions are (cognitively) rational when they are fitting responses to objects or events in the world, while the second theory maintains that emotions are (instrumentally) rational when they help agents achieve their goals. We argue that affective polarization fails to meet the standards of rationality under either framework. Affective polarization is thus irrational.

Keywords: polarization, partisanship, political epistemology, political parties, political anger.

1. Introduction

Turn on the news and it is common to hear that the United States is more polarized now than any other time in its history except for the Civil War. Social scientists have carefully studied polarization, hoping to shed light on its causes, consequences, and potential solutions. Philosophers are beginning to study polarization as well (Kogelmann and Manor, 2025, pp. 729-730). Some ask what polarization means for normative theories of liberalism and democracy (Vallier, 2021). Others – and this is the literature we're interested in – examine the extent to

which polarization is *rational* (Pallavicini et al., 2018; Singer et al., 2019; Joshi, 2020; Dorst, 2023; Dellsén, 2024; Kopecky, 2024; Westfall, 2024). This literature concentrates on belief polarization, which is the tendency for individuals' political beliefs to become more extreme and correlated over time. Can belief polarization result from rational epistemic processes, or does it necessarily involve epistemic irrationality?

Polarization manifests not only in our beliefs but also in our affect. Affective polarization refers to 'the tendency of people identifying as Republicans or Democrats to view opposing partisans negatively and copartisans positively', although the definition need not be couched in such parochial U.S.-centric terms (Iyengar and Westwood, 2015, p. 691). We can more generally define affective polarization as the tendency of people identifying with a political party to view opposing partisans negatively and copartisans positively. Affective polarization involves 'prejudice, anger, and activism on behalf of that prejudice and anger' (Mason, 2018, p. 4). When we are affectively polarized, we see 'each other only as two teams fighting for a trophy' (Mason, 2018, p. 4). This affective dimension to polarization is well-documented empirically. Affective polarization can occur even when belief polarization does not. When this happens 'we act like we disagree more than we really do' (Mason, 2018, p. 4).

Social scientists spend just as much time studying affective polarization as they do belief polarization. And yet, philosophers have paid little attention to it. Specifically, no philosopher has probed the *rationality* of affective polarization as they have belief polarization. This paper fills this gap. We ask: are the actual patterns of affective polarization we see in the United States and similar countries rational? They are not, we conclude. In most cases, it is irrational to harbor antagonistic emotions towards opposing partisans.

Here is the structure of the paper. The next section overviews the empirical literature on affective polarization and refines our guiding question in response to it (§2). After this, we consider the two main theories of rational emotions from the philosophical literature. The first view says that emotions are (cognitively) rational when they are fitting responses to objects or events in the world (§3). The second view says that emotions are (instrumentally) rational when they help agents achieve their goals (§4). Neither account of rational emotions can vindicate the actual patterns of affective polarization we find in the world. We are left to conclude that affective polarization is irrational.

2. Affective Polarization

Evidence demonstrating the United States is affectively polarized falls into three buckets: survey self-reports, implicit measures, and behavioral measures (Iyengar et al., 2019, pp. 131-134). Let's take a brief look at all three, starting with survey self-reports. These studies survey partisans, asking them about their attitudes towards opposing partisans with a variety of different kinds of questions. For example, in a 2010 survey 49 percent of Republicans and 33 percent of Democrats reported that they would be unhappy if their child married someone from the opposing party (Iyengar et al., 2012, p. 418). Another survey asks partisans to associate certain traits (e.g., 'intelligent' or 'selfish') with members of their own and opposing parties. Unsurprisingly, they are overwhelmingly more likely to associate negative traits with the opposing party (Iyengar et al., 2012, p. 420). Most alarmingly, a November 2017 survey found that just over 60 percent of partisans believed opposing partisans are 'a serious threat to the United States and its people'; 40 percent said opposing partisans are 'evil'; and 20 percent said

members of the opposing party 'lack the traits to be considered fully human – they behave like animals' (Kalmoe and Mason, 2022, pp. 44-45).

There is always a fear that partisans do not fill out surveys truthfully. To address this, the second bucket of evidence looks at implicit measures of affective polarization. One way to do this is by administering an implicit bias test for partisanship. An implicit bias test asks people to associate different groups (e.g., white or black, Republican or Democrat) with different positive or negative attributes as quickly as they can; the quicker the reaction time, the more deeply ingrained the association, hence an implicit, rather than explicit, bias. Researchers who perform partisan implicit bias tests find it 'is ingrained; approximately 70% of Democrats and Republicans show a bias in favor of their party' (Iyengar et al., 2019, p. 133). Another study demonstrating implicit partisan bias asked participants to award scholarships to high school students; among factors such as GPA, the fictional student's partisan allegiance was on their resume. The main finding is that 'even when the candidate from the opposing party was more qualified, partisans opted to award the scholarship money to the copartisan candidate' (Iyengar and Westwood, 2015, p. 699).

The final bucket of evidence looks at actual behavior. In one experiment, participants are told they will play a game against an opponent whose partisan allegiance they are made aware of. When you lose the game, participants are told, your opponent will be hit with a loud blast of sound. Participants can then choose how loud the sound will be for their opponent, should their opponent lose. Researchers found that 'partisans were significantly more likely to assign the maximum "painfully loud" option to the opposing party rather than to their own' (Kalmoe and Mason, 2022, p. 101). In another experiment, after performing a series of tasks, participants are given two options for their compensation. The first option is a lump sum of money with no

strings attached; the second option is a lump sum of money that is double the amount of the first option, but if this second option is chosen, then money is also donated to the opposing political party. The finding is that 'three-fourths of respondents were willing to give up a doubling of their bonus payment simply to avoid making a donation to the other party' (McConnell et al., 2018, p. 15).

Affective polarization is a genuine phenomenon. A few more points of clarification are needed. First, what attitudes do the affectively polarized harbor towards opposing partisans? Researchers are not precise here, using a variety of terms to describe the relevant sentiments: 'dislike', 'distrust', 'animosity', 'anger', 'prejudice', 'hostility', 'negative', and more (Iyengar et al., 2019, p. 130; Mason, 2018, p. 4; Iyengar and Westwood, 2015, p. 691). Rather than assume a specific emotion to characterize the affectively polarized, we assume they harbor what philosopher Berit Brogaard (2020, p. 12) calls the 'antagonistic emotions', which include blame, anger, resentment, indignation, disrespect, disgust, contempt, and more. Antagonistic emotions are negatively valenced, target individuals or groups, and typically arise in response to perceived offenses of the targeted individuals or groups. Going forward, we say affectively polarized partisans 'dislike' opposing partisans as shorthand for saying they exhibit some kind of antagonistic emotional response towards them.

Who specifically do partisans dislike when they are affectively polarized? In one empirical paper, researchers try to answer this by asking survey respondents about their attitudes towards elites of the opposing party (i.e., elected officials) as well as ordinary voters. It turns out that partisans dislike both elites and ordinary voters from the opposing party, but they dislike elites just a little bit more (Druckman and Levendusky, 2019, p. 119). For the purpose of this paper, we focus strictly on whether it is rational for partisans to dislike ordinary rank-and-file

voters of the opposing party. We do this because it is the harder case; if it is ever rational to dislike an anonymous Republican voter, then surely it is rational to dislike Republican politicians. To refine this paper's guiding question: is it ever rational to dislike rank-and-file members of the opposing political party, where 'rank-and-file members' are defined as those whose political activity consists of regularly voting for the opposing party and regularly consuming partisan news media, but not much more than that?

Furthermore, the rank-and-file opposing partisans we are concerned with are *anonymous*. They are people whom you have never met and likely never will meet. We are not asking whether it's rational to cut off relations with your friends or parents because they support a different political party than you. Jason Brennan (2023) already has an excellent paper on that question. We are asking what we think is a more difficult question to answer in the affirmative: is it rational to dislike rank-and-file members of the opposing political party whom you have never met and likely will never meet? If you are a Democrat, is it ever rational to dislike a person you do not know personally simply because they are a Republican?

As a final clarifying point, we are interested in whether the *actual patterns* of affective polarization we see are rational. We are not interested in logical possibilities. So long as one accepts that emotions and attitudes are subject to rational scrutiny (more on this below), then it is trivial that there are *some* set of circumstances in which it is rational to dislike rank-and-file members of the opposing political party whom you have never met and likely will never meet. A more interesting question is whether the affective polarization that increasingly characterizes our politics is rational. So, to reform our guiding question one last time: is it rational for Democrats (Republicans) to dislike rank-and-file Republicans (Democrats) whom they have never met and likely will never meet?

3. Cognitive Theories of Rational Emotions

While some might think emotions are paradigmatically irrational, many philosophers disagree. Emotions can be rational in at least two different senses, the *cognitive* and the *instrumental* (Solomon, 1973, p. 33; Sousa, 1987, pp. 163-164; Elster, 1994, p. 36; Greenspan, 2000, pp. 478-479; Scarantino and Sousa, 2018, §10). We discuss the instrumental theory of rational emotions in the next section. This section asks: is it rational in the cognitive sense for Democrats (Republicans) to dislike rank-and-file Republicans (Democrats) whom they have never met and likely will never meet?

Emotions are rational in the cognitive sense when they are appropriate or fitting responses to objects or events in the world. C.D. Broad offers a nice statement of this view. He writes:

> Now some kinds of emotional quality are *fitting* and others are *unfitting* to a given kind of epistemological object. It is appropriate to cognise what one takes to be a *threatening* object with some degree of *fear*. It is inappropriate to cognise what one takes to be a fellow man in *undeserved pain* or *distress* with *satisfaction* or with *amusement* (Broad, 1954, p. 209).

If the affectively polarized are to be rational in the cognitive sense, then their antagonistic emotions must be a fitting response to objects or events in the world. When are antagonistic emotions fitting? There are two answers in the literature.

First, it is fitting to harbor antagonistic emotions towards an individual or group when the individual or group displays bad character. Michelle Mason advances this view. According to

her, contempt is fitting when it 'is directed at a person as a response to his violation of an interpersonal ideal of the person' (Mason, 2003, p. 250). Contempt is person-focused rather than action-focused, in that it takes as its object an individual's character, rather than what she does (Mason, 2003, p. 246). You can have contempt for someone who does you no harm. You might have contempt for a movie star who is an adulterer, a politician who displays a disregard for the truth, or an athlete who shows poor sportsmanship. On this view, it is fitting and thus rational to have some kind of antagonistic emotional response towards an individual or group because of *who they are*.

Second, it is fitting to feel antagonistic emotions towards an individual or group when that individual or group culpably harms you. This is the view developed by P.F. Strawson in 'Freedom and Resentment'. Strawson's (1974, p. 10) concern in this essay is with the *reactive attitudes*, which are the 'natural human reactions to the good or ill will or indifference of others towards us, as displayed by in *their* attitudes and actions'. It is fitting for Alice to harbor an antagonistic reactive attitude (such as resentment) towards Betty if Betty harms Alice due to ill will or indifference. If Betty trips and accidentally steps on Alice's hand, then it is not fitting for Alice to resent Betty because, though Betty harms her, she displays no ill will or indifference. But if Betty malevolently steps on Alice's hand, or does so because she is not paying attention, then resentment is fitting (Strawson, 1976, p. 5). On this view, it is fitting and thus rational to have some kind of antagonistic emotional response towards an individual or group because of *what they do*.

Putting it all together, affective polarization is rational in the cognitive sense if opposing partisans either display bad character or culpably harm members of the other party. Is either true? This is what we now turn our attention to.

3.1 *Who they are*.

One view in the literature says antagonistic emotions towards an individual or group are fitting when that individual or group displays bad character. For this view to say affective polarization is rational, it must be that opposing partisans display bad character. Do they?

An obvious answer says opposing partisans display bad character when they embrace a morally odious ideology. Ideology should be understood broadly, as a set of loosely connected beliefs about how to address social, political, and economic questions (Hinich and Munger, 1996, pp. 9-13). Liberalism and conservatism are examples. Depending on its content, embracing an ideology can be indicative of bad character, which makes it fitting to dislike someone who holds such an ideology. An easy example is someone who embraces a racist ideology that says some races are superior to others and should receive special social, economic, and political privileges as a result. An individual who embraces this ideology has bad character, so it is fitting to dislike them.

To explain our actual patterns of affective polarization we need less extreme examples. Perhaps it is fitting for Democrats to dislike Republicans because Republicans believe we should mass deport illegal immigrants, which will likely result in breaking up families, which some think is unnecessarily cruel. Perhaps it is fitting for Republicans to dislike Democrats because Democrats believe we should give transgender children medical interventions such as surgery or hormone replacement therapy, which some think is tantamount to child abuse. Hence, affective polarization is rational because opposing partisans demonstrate bad character by embracing morally odious ideologies.

The problem with this view is that most rank-and-file partisans don't actually embrace morally odious ideologies, because most rank-and-file partisans don't actually embrace any ideology at all. This provocative thesis was first advanced by Philip Converse (2006) in what is now a classic paper, 'The Nature of Belief Systems in Mass Publics'. It is common to think ideology drives political participation: Alice votes Republican because she has a conservative ideology, while Betty votes Democratic because she is a liberal. While this is true for political elites – individuals who either work in or take great personal interest in politics – ordinary voters are different. According to Converse (2006, pp. 51-52), 'large portions of an electorate do not have meaningful beliefs, even on issues that have formed the basis for political controversy among elites for substantial periods of time'. Most voters are 'remarkably innocent' of ideology (Converse, 2006, p. 66).

Several pieces of evidence support this conclusion. Most compelling is the lack of stability in people's political beliefs. Converse conducted multiple interviews with the same individuals over time and so was able to examine the extent to which their political beliefs remain stable. Little stability was found: 'only about thirteen people out of twenty manage to locate themselves even on the same *side* of the controversy in successive interrogations, when ten out of twenty could have done so by chance alone' (Converse, 2006, p. 45). The findings have been replicated (Zaller, 2011, pp. 30-32). For instance, voters interviewed in January and June of 1980 were asked what the U.S.'s stance towards the Soviet Union should be. Researchers found that 'altogether, only 50 percent of respondents took the same position in June that they had taken in January' (Zaller, 2011, p. 30). If individuals have genuine ideologies, their political beliefs would be more stable.

Another piece of evidence supporting the claim that most people don't have a political ideology is that responses to political questions are incredibly sensitive to how they are worded or phrased. For example, if you tell partisans that a leader of their party endorses policy *x*, they are far more likely to say they endorse *x* as well; if you tell them a leader of their party does not endorse *x*, they too will probably reject it (Cohen, 2003; Barber and Pope, 2019; Lewis and Lewis, 2023, pp. 7-9). Voters follow their leaders (Lenz, 2012; Holcombe, 2023). As Gabriel S. Lenz (2012, p. 19) summarizes the evidence: 'instead of politicians following the will of the people, the people seem to be following the will of politicians'. Like instability, this finding suggests that most voters don't have genuine ideologies, for if 'citizens have preexisting "true attitudes" that they merely reveal to the inquiring pollster, such "endorsement effects" should not occur' (Zaller, 2011, p. 33). In general, the evidence for ideological innocence among the electorate is significant and compelling (Kinder and Kalmoe, 2017, ch. 2).

We thus cannot say that actual patterns of affective polarization are rational because opposing partisans display bad character by embracing odious ideologies. Most opposing partisans probably don't actually believe the ideologies the elites in their parties embrace. Most voters have no ideology at all.

In response, it might be argued that mindlessly parroting a morally odious ideology reflects bad character, even if one does not embrace it. Consider an analogous case. Imagine a young man, Charlie, who hangs out with a group of friends who embrace a racist ideology. Charlie finds the racist ideology abhorrent but has no other options for companionship. Though Charlie rejects the racist ideology, he goes along with it to fit in. He frequently uses racial slurs even though he secretly thinks this is wrong. It seems fitting to have contempt for Charlie, for he demonstrates moral cowardice by not standing up to what he knows is wrong. Likewise, though

the empirical evidence suggests most partisans don't have ideologies, the literature does suggest they will typically parrot their party's ideology when it is made salient to them. So long as the ideology they parrot is morally odious, this is worthy of contempt.

This argument violates what we call the *hypocrisy constraint*. The hypocrisy constraint says it is fitting for you to dislike someone for falling short of a character ideal only if you are not guilty of the same transgression. Contempt is fitting, writes Mason (2003, p. 253), when the 'attitude is directed by an agent who does not possess a similar fault or, if he does, is committed to regarding himself in the relevant circumstances as likewise contemptible in virtue of it'. To see why the hypocrisy constraint is plausible, let's continue the example from above. We said it is fitting to have contempt for Charlie because he parrots the morally odious ideology of his friend group. This response is not fitting for everyone, however. Consider Doreen, who hangs out with a group who believes that assassinating the CEOs of healthcare companies is morally justified. Doreen doesn't believe this, but she goes along with it because she has no other options for companionship. Though contempt for Charlie might be fitting in general, it is not a fitting attitude for Doreen specifically. The hypocrisy constraint explains why: she falls short of the same character ideal.

The hypocrisy constraint clearly says affective polarization is irrational if the ideologies that both political parties parrot are morally odious. Suppose, however, that rank-and-file Republicans parrot a morally odious ideology, but the ideology rank-and-file Democrats parrot is not similarly objectionable. Is it rational for Democrats to dislike Republicans in these circumstances? We are skeptical. This is because the only thing preventing Democrats from parroting a morally odious ideology is not superior character, but luck. We have every reason to

believe they *would* parrot a morally odious ideology if the leaders of their party embraced one. The tendency for voters to follow their leaders is bipartisan.

That most Democrats would parrot a morally odious ideology if their leaders embraced one makes all the difference. Consider an example. The hypocrisy constraint clearly says it is not fitting for an adulterer to have contempt for the politician who cheats on his spouse. But what about the individual who would like to and often tries to have an affair, but is not attractive enough to find a suitable partner? The hypocrisy constraint also says it is not fitting for this person to have contempt for the unfaithful politician, for they too would be an adulterer were circumstances different. It is not superior character that stops them from cheating, but luck. Likewise, it is not superior character that prevents Democrats from parroting a morally odious ideology, but luck. If their leaders adopted an objectionable policy platform, most would follow. The hypocrisy constraint still applies.

The following might be offered in response: 'Though I tend to parrot whatever my political party says, I chose to join a party that either will not or is far less likely to adopt a morally odious ideology. Opposing partisans have not done this, which is why it is fitting for me to have contempt for them'. On this view, opposing partisans' character flaw is not that they follow their leaders to bad places, but that they chose bad leaders. An analogy will be helpful. Edgar has a choice of friend groups. He can hang with the jocks, goths, or nerds. Hanging with the jocks is most desirable, but they engage in morally vicious conduct. Joining the goths or nerds is less desirable, but they are not vicious. Edgar joins the jocks and ends up acting viciously. The complaint against him is not that he caved to peer pressure – most of us probably would were we in his shoes – but that he chose to join a group he knew might lead him astray.

The problem with opposing partisans is not that they follow their leaders to bad places; the problem is they join a party with bad leaders.

This argument assumes partisan identity is a conscious choice. For most, it's not. The biggest determinant of partisan identity is your parents' partisan identity (Campbell, et al. 1980, p. 147; Lewis-Beck et al., 2011, p. 139). Also relevant is social identity (e.g., race, ethnicity, religion, etc.) (Achen and Bartels, 2016, p. 307). Given the massive impact these factors have on party choice, the hypocrisy constraint becomes relevant again. To see why, let's continue the case from above. Both of Edgar's parents and all his siblings are jocks; he is raised in a culture that reveres sports. It is thus unsurprising he becomes a jock. This, unfortunately, leads him astray. Both of Felicia's parents and all her siblings are nerds; she is raised in a culture that reveres math, science, and technology. It is unsurprising that she becomes a nerd. This friend group, luckily, does not lead her astray.

Felicia has contempt for Edgar because he chose to join a friend group that acts viciously. Her contempt is not fitting, in our view. Like the person who does not have an affair only because he is unattractive, Felicia avoids Edgar's mistakes only due to circumstances beyond her control. If she were born into a different family and culture, she would probably be a jock and make the same mistakes. Likewise, it's not fitting for Democrats to dislike Republicans for joining a party that may lead them astray, for they too would probably join the Republican Party were they born into different circumstances.

3.2 *What they do*.

A second view in the literature says antagonistic emotions towards an individual or group are fitting when that individual or group culpably harms you. For this account of the rational emotions to vindicate the patterns of affective polarization we see in the real world, it must be that Democrats (Republicans) harm Republicans (Democrats) due to ill will or indifference. Do they?

Before answering this, we must get clear on terminology. What is meant by 'harm'? It cannot mean any reduction in welfare. It is in the nature of politics – especially democratic politics, where policies are determined by majorities – that there are winners and losers. Few policies are Pareto efficient in that they make everyone better off, or at the least no one worse off. Even policies that strike many as overwhelmingly good (such as basic unemployment insurance) will make *some* people worse off (e.g., the rich who see their taxes increase or future generations who must pay the debt taken on to finance the initiative). If it is rational to dislike an opposing partisan whenever they culpably harm you in this sense, then it will *always* be rational to dislike opposing partisans, but this seems overly capacious. Harm must be interpreted more narrowly.

We follow the classic definition provided by Joel Feinberg (1984, p. 215), which says 'the term "harm" as it is used in the harm principle refers to those states of set-back interest that are the consequence of wrongful acts or omissions by others'. A wrongful act or omission is one that violates a right; person *A* wrongfully sets back *B*'s interests when '*B*'s set-back interest is one that he has a *right* to have respected' (Feinberg, 1984, p. 108). Thus, to harm another is to set back their interests in violation of their rights. While raising taxes on the wealthy to pay for unemployment insurance might set back the wealthy's interests, most theories of rights will not call this a rights violation. So, for the current account of rational emotions to vindicate the

patterns of affective polarization we see in the real world, it must be that Democrats (Republicans) set back the interests of Republicans (Democrats) in violation of their rights due to ill will or indifference. Do they?

Perhaps an opposing partisan violates rights through their elected representative's policy decisions. A Democrat might be harmed by a Republican because the Republican voted for Trump who won and then implemented policies that prevent her from receiving reproductive healthcare (suppose, for the sake of argument, there is a right to reproductive healthcare). A Republican might be harmed by a Democrat because the Democrat voted for Harris who won and then significantly restricted access to firearms (suppose, for the sake of argument, there is a right to bear arms). The basic idea, then, is that opposing partisans cause harm by how they vote, making them fitting targets of resentment.

The reasoning from the paragraph above is actually quite contentious (Brennan, 2011; Maskivker, 2019). Whether opposing partisans cause harm with their votes depends on one's theory of causation. If causation is understood in the Lewisian sense – where event *c* causes event *e* if and only if both *c* and *e* occur, and if *c* had not occurred, *e* would not have occurred – then voting almost never causes harm, because an individual's vote is almost never causally efficacious. However, a different theory of causation can be adopted, such that voting does cause harm. To cause harm perhaps one must only *contribute* to it, even if one's contribution isn't strictly causally necessary (Goldman, 1999, pp. 205-206). To be as ecumenical as possible, we grant that an individual vote can cause harm.

This, however, is not sufficient to make resentment fitting, according to Strawson's view. For resentment to be fitting the harm must result from ill will or indifference. So, do opposing

partisans act from ill will or indifference when they vote for a candidate who subsequently violates rights?

In some cases, the answer is obviously no. Suppose a candidate runs for office on a perfectly normal, non-rights-violating platform. Moreover, there is nothing about the candidate's past behavior or actions on the campaign trail that suggest he will violate rights. After getting elected, the politician violates rights. The opposing partisan harms in this case by voting for the politician, but she does not do so culpably, for she had no reason to think the politician would violate rights once in office.

What about cases in which the candidate clearly says, as a part of his platform, that he will violate rights? Suppose women have a right to reasonable abortion access and that a Republican politician campaigns on banning all abortions. Suppose the politician is elected and, in conjunction with other Republican legislators, bans all abortions. While this might seem to be a clear case in which a Republican voter culpably causes harm, in fact things are more complicated, for one can vote for such a candidate with neither ill will nor indifference. There are a few ways this might happen.

Note first that what politicians say on the campaign trail must always be taken with a grain of salt. Politicians lie, pander, exaggerate, grandstand, and bullshit (Gibbons, 2024). There are strong incentives for them to do so (Gibbons, 2023). They might exaggerate problems to increase their perceived importance, promise unrealistic solutions to differentiate themselves from other candidates, or take extreme positions during primaries to appeal to the party base. Hence, the fact that a politician campaigns on doing *x* does not imply that they actually want to do *x*, and most voters know this. It is not unreasonable for a voter to think: 'this politician is campaigning on rights-violating policy *y*, but I don't think they actually want to do that; it's just

rhetoric'. Because the politician has other attractive policies the voter does believe they sincerely embrace, they go ahead and vote for him. If this is why an opposing partisan votes for a politician who credibly promises a rights-violating policy, then they do not act from ill will or indifference.

Suppose, however, that the opposing partisan knows for certain that the politician she votes for is truthfully disclosing his intention to ban all abortions, and votes for him anyways. Even so, it's still possible for someone to vote for such a candidate without being driven by ill will or indifference. The voter can be opposed to the politician's sincerely held rights-violating policy, but reasonably doubt he can implement it.

The road from electing a politician to implementing policy is long and flush with obstacles. A single politician cannot implement any policy on his own; he will have to muster a majority in the House of Representatives and a supermajority in the Senate. Political parties will play a role, disciplining politicians to ensure they either support or oppose certain policies. Special interest groups and lobbyists will exert influence. Even if the elected politician gets the policy he wants, the judiciary may strike it down. Given these complicating factors, it's not unreasonable for a voter to say: 'I'm voting for this politician for reason *x*; I don't like his rights-violating policy *y*, but I don't think he can actually implement *y*'. The voter might have reason to think other members of the politician's party will stop him, or perhaps the judiciary. If this is why an opposing partisan votes for a politician who credibly promises a rights-violating policy, then she does not act from ill will or indifference.

Of course, some opposing partisans *do* vote for politicians who campaign to violate rights due to ill will or indifference. Returning to our running example, some vote for the politician who campaigns on banning all abortions *precisely because* he wants to ban all abortions. We

take this to be a case of ill will. Or, they might vote for the candidate for another reason but simply do not care about his abortion policy; that is a price worth paying for other desirable policies (e.g., tax cuts). This is indifference.

If an opposing partisan votes for a candidate who campaigns on violating rights and then does so, and if we have evidence that the opposing partisan voted for the candidate due to ill will or indifference, then it is rational to dislike them. This, however, does not address our guiding question, for our question is whether it is rational for Democrats (Republicans) to dislike rank-and-file Republicans (Democrats) *whom they have never met and likely will never meet*? Given the anonymity, one cannot have evidence that any particular opposing partisan acts from ill will or indifference. For example, suppose a Democratic administration just violated your rights and you pass a car with a bumper sticker supporting the Democrats. In this case, you simply cannot know whether the Democrat supports the current administration due to ill will or indifference (thus making her culpable for harming you), or whether she voted for the administration because she did not believe what the candidate said or doubted his efficacy (thus rendering her not culpable). Due to this uncertainty, it is irrational to dislike the Democrat you pass on the highway.

4. Instrumental Theories of Rational Emotions

In the prior section we argued that the patterns of affective polarization we find in the real world are not rational in the cognitive sense; many opposing partisans do not display bad character nor do they culpably harm. But there's a different sense in which emotions can be rational: the instrumental sense (Solomon, 1973, p. 33; Sousa, 1987, pp. 163-164; Elster, 1994, p. 36; Greenspan, 2000, pp. 478-479; Scarantino and Sousa, 2018, §10). This view is primarily

defended by social scientists, such as Robert H. Frank (1988, ch. 3), Thomas Schelling (2008, p. 37), and Francis Fukuyama (2011, p. 39). Typically, we talk about *actions* being instrumentally rational when they help agents achieve their goals. For instance, it's rational to exercise and eat plenty of protein if your goal is to lose weight. Likewise, emotions are instrumentally rational when they help agents achieve their goals.

An example will illustrate (Frank, 1988, 48). Greg grows wheat and Hypatia raises cattle on adjacent plots of land. If Hypatia's cattle wander over to Greg's field, they will cause $1,000 worth of damage. Hypatia can prevent this if she builds a fence, which costs her $200. Hypatia knows that, were her cattle to wander over to and damage Greg's field, it would cost Greg $2,000 to bring a civil suit against her. If all parties are rational, then Hypatia won't build the fence, for she knows that Greg won't pursue a civil case, as the damages he would receive are less than the cost of bringing the case. But suppose Greg has a temper. If something upsets him, he's known for being the kind of guy who pursues vengeance at great personal cost. Hypatia will now build the fence, for she knows that if her cattle wander over to his field, he will pursue a civil case even if it costs him more than he wins. Because the cost of the fence is less than the cost of paying damages, Hypatia will build the fence because she knows an intemperate Greg will take her to court. Greg's anger is instrumentally rational, because it helps him achieve his goal: Hypatia builds the fence, and his crop is safe.

It's worth noting that emotions that are instrumentally rational can (but need not) be cognitively *irrational*. Imagine a man who becomes hysterically upset whenever he has a minor disagreement with his partner. His reaction is cognitively irrational, since hysteria is not an appropriate or fitting response to a trivial conflict. Yet suppose that his hysteria consistently leads his partner to give in, allowing him to get what he wants. In that case, the emotion is

instrumentally rational because it effectively helps him achieve his goals, even though it is cognitively irrational.

For affective polarization to be rational in the instrumental sense, it must be that rank-and-file Democrats (Republicans) are more likely to achieve their goals precisely because they dislike rank-and-file Republicans (Democrats) whom they have never met and likely will never meet. Is this true? Drawing on recent work in the political philosophy of anger, we develop an argument suggesting that rank-and-file Democrats (Republicans) might indeed advance their goals by disliking rank-and-file Republicans (Democrats). We then show that this argument fails. It fails because it considers only the potential benefits of affective polarization and ignores its costs; an instrumental justification, however, must take both into account.

Let us turn to the growing literature in philosophy on political anger. Many point out that political anger can be motivating (Srinivasan, 2018, p. 126; Cherry, 2021, ch. 3; Cherry, 2022, p. 5; Lepoutre, 2023, p. 45). In particular, it can help oppressed individuals fight injustice. Audre Lorde (2007, p. 127) offers the best description of this view: 'every woman has a well-stocked arsenal of anger potentially useful against those oppressions … which brought that anger into being. Focused with precision it can become a powerful source of energy serving progress and change'. For instance, suppose it is 1900 and women cannot vote. Irene wants to vote. Since Irene cannot affect change through ordinary political means – women cannot vote, after all – she must pursue change through more subversive means, such as protest and civil disobedience.

Social scientists have long pointed out that engaging in protest and civil disobedience is irrational (Kogelmann, 2025). If all women simultaneously protested, then politicians would probably give in and grant them the right to vote. But Irene can only control her actions, not the actions of others. If enough others protest, Irene would personally be better off staying home and

free riding on their efforts – she would gain the right to vote without facing any personal risks or costs. If enough others don't protest, she would also be better off not protesting, as she would face personal risks without achieving any gains. So regardless of what others do, Irene's rational choice is to not protest. Everyone reasons the same way. In equilibrium, women never achieve suffrage.

But Irene is angry at her unjust treatment, and this can alter her decision calculus. Marching in the streets is a powerful way for Irene to express her emotions. It may be cathartic, like a pressure valve releasing pent-up frustration. Protesting also allows Irene to stand in solidarity with others who are angry. These benefits to protesting change Irene's decision calculus. Absent anger, when deciding to protest, Irene weighs the benefits of protesting (i.e., the miniscule chance her marching in the streets makes a difference) against the costs (i.e., the moderate chance she is punished and the time and effort it takes). With anger, the benefits of protesting change. They now include expression, catharsis, solidarity, and more. These added benefits may tip the scales, such that protest is now rational.

This argument plausibly explains why it is instrumentally rational for *political activists* to be affectively polarized. Anger towards Republicans (Democrats) gives Democrats (Republicans) a reason to march in the streets and engage in other costly political activities that are otherwise irrational. Our question, though, is whether it is rational for rank-and-file partisans to be affectively polarized. These are individuals whose political activity consists of regularly voting for their party and consuming partisan media, but not much else. Empirical work shows that many are affectively polarized. We cannot explain the rationality of their emotions by saying it gets them out in the streets, for they almost never engage in protests or similar activities. What is anger motivating them to do?

Deciding whether to protest is not the only collective action problem we confront in politics. A more common dilemma the rank-and-file face is the decision of whether to vote. The paradox of voting points out that the costs of voting typically outweigh its benefits, which makes voting irrational (Brennan, 2020, §1). The costs of voting are the time and effort it takes to register to vote, drive to the polling place, stand in line, etc. The benefits of voting are the benefits you receive if your preferred candidate wins multiplied by the probability that your vote changes the outcome of the election. Because the chances of you changing the outcome are very small, voting's benefits are very small. In most cases, the costs of voting exceed its benefits, so it is irrational to vote.

Anger changes this. Suppose Irene and her fellow women have won the right to vote, but now want to challenge the legal prohibition on women owning property. There are two candidates in an upcoming election, one who wants to grant women the right to own property and one who does not. Irene has a strong preference here, but her vote won't matter, so she cannot bring herself to go to the polls. Now suppose Irene is angry. By voting against the candidate who wants to prevent her from owning property, Irene expresses contempt for him, doing so may be therapeutic for her, and she engages in an act of solidarity with other angry women. Irene's anger instrumentally serves her desire to vote. In support of this argument, evidence suggests the affectively polarized are more likely to vote than those who are not affectively polarized (Iyengar and Krupenkin, 2018).

Anger can change an individual's cost-benefit analysis, thus incentivizing her to vote when she otherwise would not. This, however, is not yet sufficient to make voting instrumentally rational. To see why, consider a case. Jane has the goal of earning a law degree and practicing law. She can go to two law schools. They provide the exact same quality of education and exact

same career opportunities after graduation. They are identical in all respects except for price. One law school costs $100,000, the other costs $50,000. Suppose Jane chooses the more expensive option. Even though she achieves her goal of becoming a lawyer, she acts irrationally. This is because instrumental rationality is not only about achieving your goals, but also doing so with the most effective (i.e., least costly) means possible.

Return to Irene, whose affective polarization leads her to overcome the paradox of voting and cast a vote for her preferred candidate. The question is: are there more effective ways to get Irene to vote? Yes. Note first that being affectively polarized harbors significant costs. It is bad for us to be angry. Anger makes us epistemically irrational. It makes us worse reasoners. Summarizing several empirical findings, Glen Pettigrove (2012, p. 362) writes that 'even a very modest degree of anger, such as can be generated in a lab setting, has a marked influence on our reasoning even when we are reasoning about cases that are completely unrelated to what provoked our anger'. Anger also impacts our epistemic rationality by changing how we assess risk: 'in a series of studies … when angry, people assess risks to be lower than they do in a neutral emotional state' (Pettigrove, 2012, p. 362). Anger can make us epistemically irrational by altering perception. Angry people 'are more likely to interpret unpleasant events as being due to the actions or intentions of responsible agents (as opposed to causes that are not agent related)' (Pettigrove, 2012, p. 363). They also 'appear to give more weight to stereotypes than do sad or neutral subjects' (Pettigrove, 2012, p. 363). Beyond its impact on our epistemic rationality, if one

simply reflects on the phenomenology of anger, it is an unpleasant emotional state. All things equal, it is best to avoid anger.[1]

Pointing out that anger is bad for us does not yet show it is irrational, for if anger is the *only* way to resolve the paradox of voting, and if voting is a sufficiently important goal, then anger may still be rational. But in fact, anger is *not* the only solution to the paradox of voting. There are many proposals for how to alter a voter's decision calculus to make it rational for her to vote. For instance, if a voter thinks she has a duty to vote and cares about discharging this duty, then it is rational for her to vote (Mackie, 2014). Instead of a voter wanting to express contempt for the candidate of the opposing party, she may want to express support for the candidate of her own party (Brennan and Lomasky, 1993). Expressing support can make it rational for her to vote. Perhaps the voter doesn't just care about the outcome of the election but also wants to feel like she contributes to the outcome (Tuck, 2008). This too can make it rational for her to vote.

There are many ways to solve the paradox of voting other than being angry. From a purely rational perspective, all of them are preferable to affective polarization, because they achieve the same goal (i.e., they get you to vote) but don't involve the immense personal costs

[1] One might add that anger can also be *socially* harmful; it can erode the fabric of society. This is certainly true. But this fact enters into an individual's cost–benefit analysis only if the individual actually cares about the health of the social fabric. If societal well-being does not appear in an individual's utility function, then the social harm of anger plays no role in determining the instrumental rationality of that individual's anger.

affective polarization involves. Because of this, we conclude that affective polarization is not instrumentally rational.

References

Christopher H. Achen and Larry M. Bartels, *Democracy for Realists* (Princeton: Princeton University Press, 2016).

Michael Barber and Jeremy C. Pope, 'Does Party Trump Ideology? Disentangling Party and Ideology in America', *American Political Science Review* 113 (2019), 38-54.

Geoffrey Brennan and Loren Lomasky, *Democracy and Decision* (Cambridge: Cambridge University Press, 1993).

Jason Brennan, *The Ethics of Voting* (Princeton: Princeton University Press, 2011).

Jason Brennan, 'The Ethics and Rationality of Voting,' in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (2020), <https://plato.stanford.edu/archives/win2020/entries/voting/>.

Jason Brennan, 'Friendship and Blackballing for Bad Beliefs', *Philosophy* 98 (2023), 191-214.

C.D. Broad, 'Emotion and Sentiment', *Journal of Aesthetics and Art Criticism* 13 (1954), 203-214.

Berit Brogaard, *Hatred* (Oxford: Oxford University Press, 2020).

Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes, *The American Voter* (Chicago: University of Chicago Press, 1980).

Myisha Cherry, *The Case for Rage* (Oxford: Oxford University Press, 2021).

Myisha Cherry, 'Political Anger', *Philosophy Compass* 17 (2022), e12811.

Geoffrey L. Cohen, 'Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs', *Journal of Personality and Social Psychology* 85 (2003), 808-822.

Philip E. Converse, 'The Nature of Belief Systems in Mass Publics', *Critical Review* 18 (2006), 1-74.

Finnur Dellsén, 'Interthematic Polarization', *American Philosophical Quarterly* 61 (2024), 45-58.

Kevin Dorst, 'Rational Polarization', *Philosophical Review* 132 (2023), 355-458.

James N. Druckman and Matthew S. Levendusky, 'What Do We Measure When We Measure Affective Polarization?', *Public Opinion Quarterly* 83 (2019), 114-122.

Jon Elster, 'Rationality, Emotions, and Social Norms', *Synthese* 98 (1994), 21-49.

Joel Feinberg, *The Moral Limits of the Criminal Law* (Oxford: Oxford University Press, 1984).

Robert H. Frank, *Passions within Reason* (New York: W.W. Norton & Company, 1988).

Francis Fukuyama, *The Origins of Political Order* (New York: Farrar, Straus and Giroux, 2011).

Adam Gibbons, 'Bad Language Makes Good Politics', *Inquiry* (Online First). Doi: https://doi.org/10.1080/0020174X.2023.2203164.

Adam Gibbons, 'Bullshit in Politics Pays', *Episteme* 21 (2004), 1002-1022.

Alvin I. Goldman, 'Why Citizens Should Vote: A Causal Responsibility Approach', *Social Philosophy & Policy* 16 (1999), 201-217.

Patricia Greenspan, ‘Emotional Strategies and Rationality’, *Ethics* 110 (2000), 469-487.

Melvin J. Hinich and Michael C. Munger, *Ideology and the Theory of Political Choice* (Ann Arbor: University of Michigan Press, 1996).

Randall G. Holcombe, *Following Their Leaders* (Cambridge: Cambridge University Press, 2023).

Shanto Iyengar and Masha Krupenkin, ‘The Strengthening of Partisan Affect’, *Political Psychology* 39 (2018), 201-218.

Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood, ‘The Origins and Consequences of Affective Polarization in the United States’, *Annual Review of Political Science* 22 (2019), 129-146.

Shanto Iyengar, Gaurav Sood, Yphtach Lelkes, ‘Affect, Not Ideology: A Social Identity Perspective on Polarization’, *Public Opinion Quarterly* 76 (2012), 405-461.

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

Hrishikesh Joshi, ‘What are the Chances You’re Right about Everything? An Epistemic Challenge for Modern Partisanship’, *Politics, Philosophy & Economics* 19 (2020), 36-61.

Nathan P. Kalmoe and Lilliana Mason, *Radical American Partisans* (Chicago: University of Chicago Press, 2022).

Donald R. Kinder and Nathan P. Kalmoe, *Neither Liberal nor Conservative* (Chicago: University of Chicago Press, 2017).

Brian Kogelmann, ‘Ideology vs. Collective Action’, *Erkenntnis* (Online First). Doi: https://doi.org/10.1007/s10670-025-00937-5.

Brian Kogelmann and Aylon Manor, ‘Political Epistemology’, in Gerald Gaus, Fred D’Agostino, and Ryan Muldoon (eds.), *The Routledge Companion to Social and Political Philosophy* (New York: Routledge, 2025), 726-736.

Felix Kopecky, ‘Argumentation-Induced Rational Issue Polarization’ *Philosophical Studies* 181 (2024), 83-107.

Gabriel S. Lenz, *Follow the Leader?* (Chicago: University of Chicago Press, 2012).

Maxime Lepoutre, ‘The Red Mist’, *Journal of Ethics and Social Philosophy* 24 (2023), 36-57.

Hyrum Lewis and Verland Lewis, *The Myth of Left and Right* (Oxford: Oxford University Press, 2023).

Michael S. Lewis-Beck, William G. Jacoby, Helmut Norpoth, and Herbert F. Weisberg, *The American Voter Revisited* (Ann Arbor: University of Michigan Press, 2011).

Audre Lorde, *Sister Outsider: Essays and Speeches* (Trumansburg: Crossing Press, 2007).

Gerry Mackie, ‘Why It’s Rational to Vote’, in Claudio López-Guerro and Julia Maskivker, *Rationality, Democracy, and Justice* (Cambridge: Cambridge University Press, 2014), 21-49.

Julia Maskivker, *The Duty to Vote* (Oxford: Oxford University Press, 2019).

Lilliana Mason, *Uncivil Agreement* (Chicago: University of Chicago Press, 2018).

Michelle Mason, ‘Contempt as a Moral Attitude’, *Ethics* 113 (2003), 234-272.

Christopher McConnell, Yotam Margalit, Neil Malhorta, and Matthew Levendusky, 'The Economic Consequences of Partisanship in a Polarized Era', *American Journal of Political Science* 62 (2018), 5-18.

Josefine Pallavicini, Bjørn Hallsson, Klemens Kappel, 'Polarization in Groups of Bayesian Agents' *Synthese* 198 (2018), 1-55.

Glenn Pettigrove, 'Meekness and "Moral" Anger', *Ethics* 122 (2012), 341-370.

Andrea Scarantino and Ronald de Sousa, 'Emotion', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (2021), <https://plato.stanford.edu/archives/sum2021/entries/emotion/>.

Thomas Schelling, *Arms and Influence* (New Haven: Yale University Press, 2008).

Daniel J. Singer, Aaron Bramson, Patrick Grim, Bennett Holman, Jiin Jung, Karen Kovaka, Anika Ranginani, and William J. Berger, 'Rational Social and Political Polarization', *Philosophical Studies* 176 (2019), 2243-2267.

Robert C. Solomon, 'Emotions and Choice', *The Review of Metaphysics* 27 (1973), 20-41.

Ronald de Sousa, *The Rationality of Emotions* (Cambridge: MIT Press, 1987).

Ami Srinivasan, 'The Aptness of Anger', *Journal of Political Philosophy* (2018), 123-144.

P.F. Strawson, *Freedom and Resentment and Other Essays* (New York: Harper & Row, 1974).

Richard Tuck, *Free Riding* (Cambridge: Harvard University Press, 2008).

Kevin Vallier, *Trust in a Polarized Age* (Oxford: Oxford University Press, 2021).

Mason Westfall, 'Polarization is Epistemically Innocuous', *Synthese* 204 (2024), 1-22.

John. R. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge: Cambridge University Press, 2011).

Brian Kogelmann (bkogel89@gmail.com) is an Associate Professor of Philosophy and Political Science at Purdue University. His work lies at the intersection of philosophy, politics, and economics (PPE). His recent books include *Secret Government: The Pathologies of Publicity* and *Political Meritocracy in the 21st Century*.

Paul Wheatley (pwheatle@purdue.edu) is a graduate student in Philosophy at Purdue University. He works in ethics and political philosophy and is writing a dissertation on how we should respond to extreme beliefs among those with whom we have close relationships.