

## Chapter 12

# The Politics of Irrationality

### Abstract

Recent politics has experienced a pathological rise of enmity, fall of trust, and rise of intolerance. I argue that there is theoretical and empirical reason to think that a widespread belief in Reasonable Convergence has exacerbated these trends, convincing us that our fellow citizens are unreasonable. Accepting that we are Reasonably Polarized can help. I argue that it should soften our enmity, strengthen our trust, redouble our commitment to meeting our fellow citizens where they are, and commit us to a duty of persuasion.

Our politics has become pathological.

You can see it in the *rise of enmity*, aka ‘affective polarization’—the fact political opponents increasingly dislike each other. Between 1994 and 2022, the proportion of partisans with a ‘very unfavorable’ view of the opposing party *tripled*—from 21% to 62% for Republicans, and from 17% to 54% for Democrats (Figure 12.1).

You can see it in the *rise of intolerance*—the fact that people are less committed to democratic norms, and less open to engaging with those on the other side. From 2016 to 2022, the rate at which Democrats and Republicans said that the other side was more ‘close-minded’, ‘dishonest’, ‘immoral’, and ‘unintelligent’ each rose to majority positions.

You can see it in the *fall of trust*—the fact that people now have less trust in the judgment of institutions and their fellow citizens. From 1958 to 2019, the percentage of Americans who said they trusted the government to do the right thing most or all of the time fell from 73% to 17%.

In this chapter, I’ll defend three claims. Empirically, these trends are exacerbated by belief in Reasonable Convergence (§12.2). Normatively, if we accept that we are Reasonably Polarized, we should soften our attitudes toward our political opponents (§12.3–12.4). And politically, the ever-present possibility of reasonable polarization thwarts attempts to tame politics by containing debate to ‘reasonable’ positions—and instead imposes a *duty of persuasion* on citizens of democracies (§12.5).

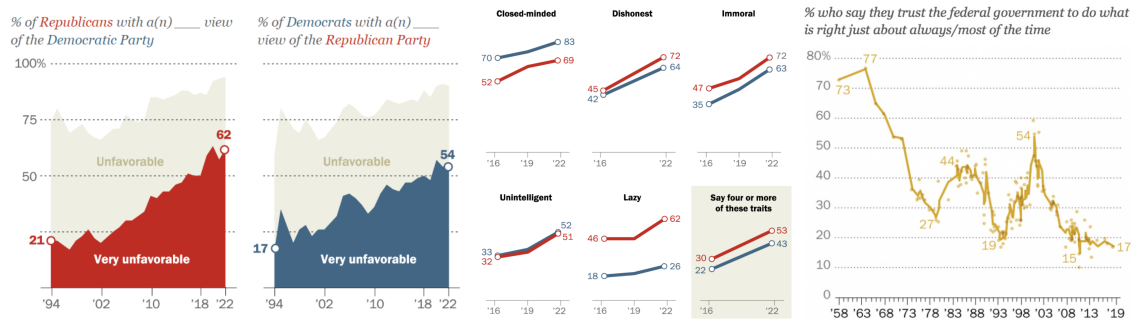


Figure 12.1: *Left:* Rise of enmity [CITE]. *Middle:* Rise of intolerance [CITE]. *Right:* Fall of trust [CITE].

## 12.1 Our Political Pathologies

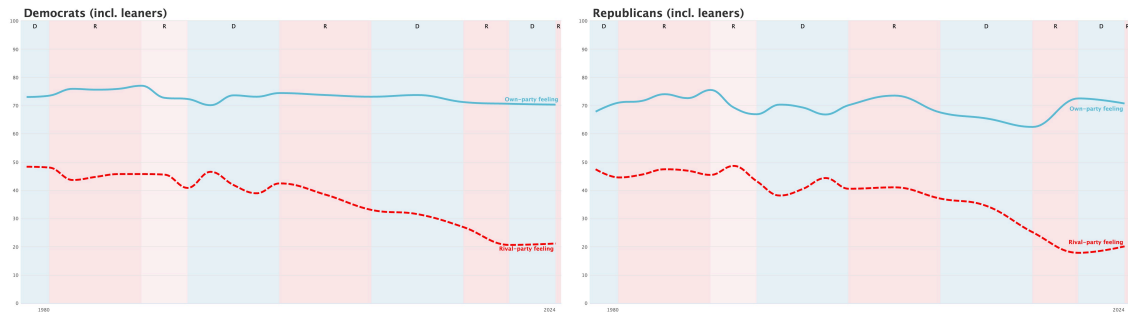
To begin, I need to show that we really have seen a rise of enmity and intolerance, and a fall of trust.

First, the rise of enmity. For several years in political science there was a debate over whether the American public was increasingly polarized on policy issues, or had instead merely *sorted* into different political camps (Fiorina et al. 2004, 2008; Abramowitz 2010; Fiorina 2016). The generally accepted conclusion is that, at the very least, Americans have become increasingly *affectively* polarized over politics—they increasingly dislike, distrust, and blame the other side (Iyengar et al. 2012, 2019). There are many ways we can measure this.

One is with ‘feeling thermometers’. The American National Election Studies (ANES) has for years asked people how ‘warm or cold’ they feel toward various groups. From 2000 to 2024, the average Democrat feeling toward the Republican party fell from 42.4 to 21.1; for Republicans, the fall was from 40.5 to 20.1 (Figure 12.2).

Another is with how willing you would be to have members of the other party in your life. In 1960, 2008, and 2010, Democrats and Republicans were asked how they would feel if their child married someone from the opposite party. In 1960, around 5% of participants said they would be ‘displeased’. In 2008, 27% of Republicans and 20% of Democrats said they would be somewhat or very upset, and in 2010, 49% of Republicans and 33% of Democrats said they would be somewhat or very unhappy (Iyengar et al. 2019). In a 2024 YouGov poll, the numbers were 40% for Republicans and 45% for Democrats [CITE].

A third is with trait ratings. We can present partisans with positive and negative traits (eg intelligent vs. selfish), and ask them whether they’d ascribe such traits to members of their own party and members of the other party. Adding these up and taking the difference between them two gives us an in-out party discrepancy. In 1960 the average discrepancy was +0.12 for positive traits and  $-0.07$  for negative traits, meaning the mean participant ascribed 0.12 more positive traits and 0.07 fewer negative traits to their in-party than to the out-party. For instance, 33% of participants said their own party was ‘intelligent’, while 27% said the opposing party was. But by 2008, the average discrepancy had grown to +2.24 for positive traits and  $-1.70$  for negative traits—for example, 62% of participants said their own party was ‘intelligent’, and only 14% said



**Figure 12.2: Growing enmity.** Feeling thermometers for in-party vs. out-party, 1980–2024.

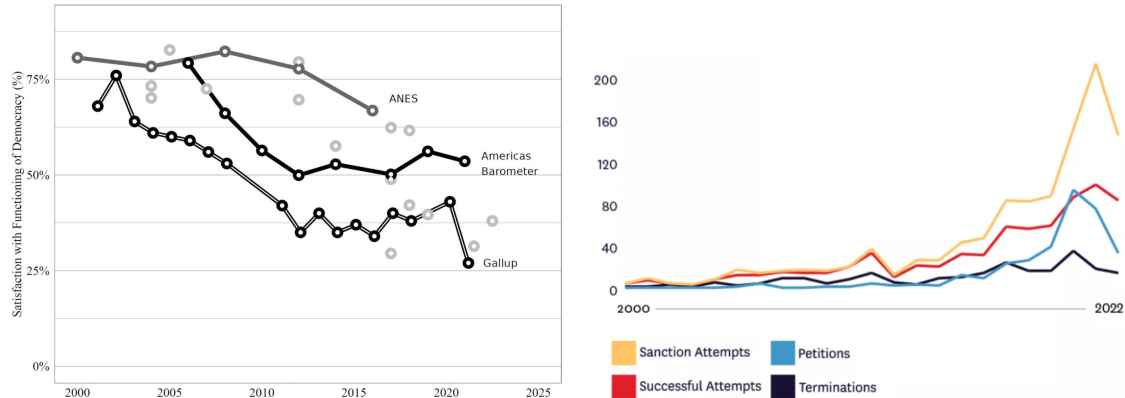
the other party was.

We see growth in these trends more recently in the middle chart from Figure 12.1, which shows that from 2016–2022 there was a sharp increase in the percentage of partisans who were willing to label members of the other party as ‘close-minded’, ‘dishonest’, ‘immoral’, ‘unintelligent’, and ‘lazy’.

These trait ratings show both a rise in enmity, and a rise in intolerance for those on the other side. But we can measure this rise of intolerance more directly. First, we can look at support for democracy. There are many ways to measure this, but it seems to be falling in the U.S. For example, pulling data from a variety of different sources, Foa and Mounk 2025 document falling satisfaction with the functioning of democracy from 2000 to the early 2020s—depending on how you measure it, the fall was from about 65–80% to 25–50% of Americans being satisfied with democracy. (Figure 12.3). Meanwhile, according to the General Social Survey [CITE], the proportion of Americans who are ‘somewhat’ or ‘very’ proud of the way democracy works fell from 89% in 2004 to 67% in 2024, with the proportion who were ‘very’ proud halving from 34% to 16%.

Second, we can look at support for anti-democratic norms and actions. Although it is difficult to get extended time-series data for these questions, Holliday et al. 2024 found that a nontrivial number of Americans on both sides supported censoring partisan media (around 20%), having more loyalty to their party than the Constitution in contested elections (around 17%) ignoring out-party judicial decisions (around 13%), and reducing out-party polling stations (around 9%). It is easy to overstate support for these actions, and indeed the paper replicated a well-known finding that people have ‘inaccurate meta-perceptions’ about the other side: people on average thought that closer to 55%–60% of the opposing party would support each of those actions, inflating the true number by 3–5 times. And the paper also found that support for outright political violence, including assault, arson or murder, remained extremely low—around 2–3%.

Still there are warning signs about political violence. The American Values Survey found that between 2021 and 2025, the percentage of people who ‘completely disagreed’ with the claim that ‘Because things have gotten so far off track, true American patriots may have to resort to violence in order to save our country’ fell from 59% to 49%. And according to Lilliana Mason [CITE], the rate of people who say anything other than ‘never’ to the question of whether its acceptable to use political violence has risen from around 7% in 2017 to around 15–20% in the summer of 2024.



**Figure 12.3: Rising intolerance.** *Left:* Satisfaction with democracy in the U.S. from 2000–2022 (Foa and Mounk 2025). *Right:* Rates of attempts to sanction academics in the U.S., 2000–2022.

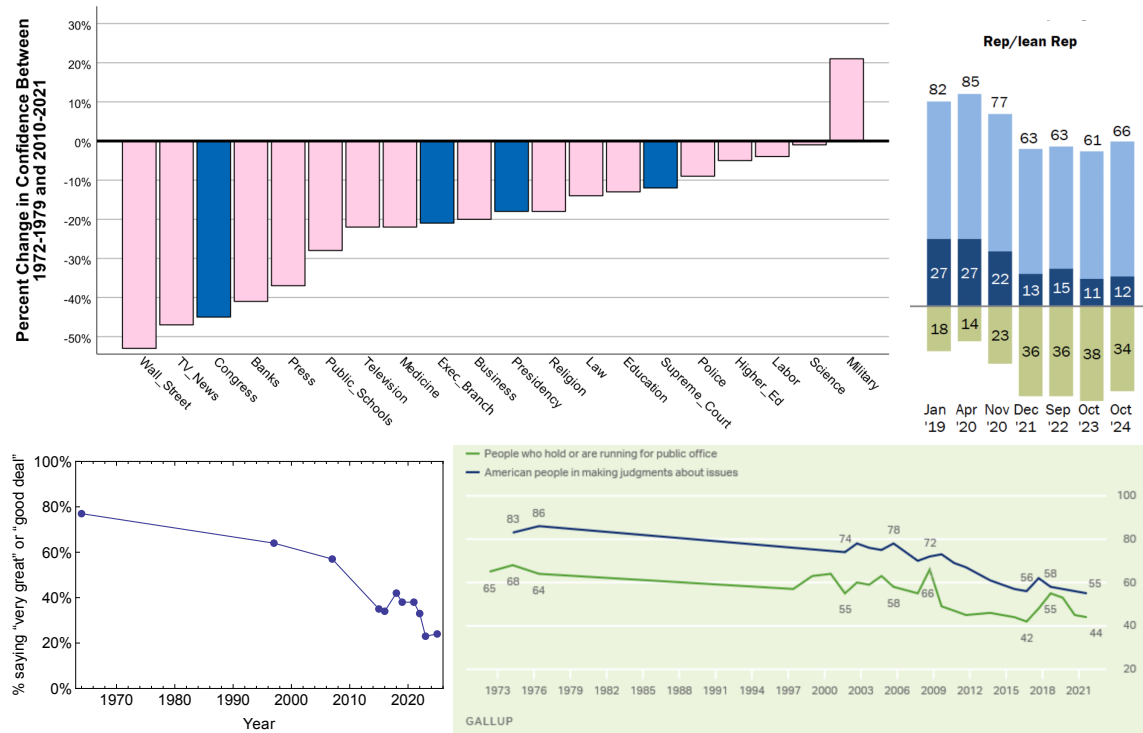
A third way to measure the rise of intolerance is by looking at support for free speech. According to a Knight Foundation report [CITE], the percentage of college students who think that freedom of speech is ‘secure’ or ‘very secure’ in the country fell from 73% in 2016 to 43% in 2024, while those who said freedom of the press was secure fell from 81% to 50%. According to a 2022 report from The Foundation for Individual Rights and Expression (FIRE), the rate of attempted and successful public sanctions campaigns against scholars has risen substantially over the last two decades, from around a handful of sanction attempts per year in the 2000s to almost 200 per year in 2020–2022. Finally, we can look at willingness to allow an unwelcome person, like a racist, to give a speech using data from the GSS. In 2002, 63% of people said a racist speaker should be allowed to speak at a public event, while 37% said they should not be. By 2024, the majority had flipped: only 46% said they should be allowed to speak, while 54% said they shouldn’t be.

What about the fall of trust? Trust in both (1) institutions and (2) the judgment our fellow citizens has fallen dramatically over the last few decades.

Institutions first. As we saw above, trust in the federal government to do the right thing is at historic lows. This is part of a general trend: Brady and Kent 2022 document that trust in virtually all public institutions—other than the military and science—has fallen dramatically between the 1970s and the 2010s. For example, trust in TV news and Congress have each fallen by about 45%, while trust in business and the presidency have each fallen by around 20% (Figure 12.4). In this data, trust in science is steady, but in the wake of the Covid-19 pandemic, even that has taken a hit. Pew reports that between January 2019 and October 2024, the proportion of Americans with ‘a great deal’ or ‘a fair amount’ of confidence in scientists to act in the best interests of the public fell from 86% to 76% [CITE]. While Democratic trust in science slipped only a few percentage points, Republicans went from 82% down to 66%, and the proportion with ‘not too much’ or ‘none at all’ doubled from 18% to 34%.

What about trust in other people? Many measures of ‘generalized trust’ are fairly ambiguous in what they’re asking about, usually calling to mind whether you trust strangers not to steal your bike or whatnot. So measured, generalized trust has fallen slightly—from 46% in 1972 to 34% in

## 12.1. OUR POLITICAL PATHOLOGIES



**Figure 12.4: Falling trust.** *Top Left:* Change in trust in various institutions in the U.S. between the 1970s and the 2010s (Brady and Kent 2022). *Top Right:* Rates at which Republicans report ‘a great deal’ or ‘a fair amount’ of trust (blue bars) in scientists to act in the best interests of the public, 2019–2024 (Pew [CITE]). *Bottom Left:* How much trust and confidence do you have in the wisdom of the American people when it comes to making political decisions? [CITE] *Bottom Right:* Gallup.

2024—with most of this fall happening before the 90s. [CITE]

But when we focus on confidence in people’s *judgments*, especially around political matters, the verdict is stark and clear. The percentage of people who say they have ‘trust and confidence in the wisdom of the American people when it comes to making political decisions’ has fallen from 77% in 1964 to 24% in 2025, with most of the fall-off happening after 2000 (Figure 12.4). Other measures show similar trends. Gallup asked people how much trust they have in the political judgment of people running for office and the American people in general. Between 1976 and 2021 trust in the former fell from 64% to 44%, while trust in the latter fell from 86% to 55%, with most of the fall-off being after 2006. Public perception matches these trends: 71% of people think that Americans have less trust in each other than 20 years ago [CITE]. Moreover, there seems to be a generational gap: only 29% of those who are 65 or older agree that ‘most people *can’t* be trusted’, while 43% of those 50–54 do, 52% of those 30–49 do, and fully 60% of those 18–29 do [CITE]. While this might just be an age effect—perhaps younger people just trust others less—it might also be a leading indicator, suggesting that that the current social and political environment is worse than the past.

So it’s not just the vibes. In the last few decades, we really have seen sharp rises in enmity, tolerance, and distrust.

## 12.2 Reasonable Convergence Makes It Worse

Recall the foil of this book:

**Reasonable Convergence:** When evidence is plentiful, if people take reasonable steps to figure out the truth, they'll usually succeed.

I've claimed that most people implicitly accept Reasonable Convergence. In this section, I'll argue that belief in Reasonable Convergence exacerbates our political pathologies.

### 12.2.1 The Theoretical Argument

First, the theoretical argument. The idea is simple. Suppose we believe that when evidence is plentiful, if people take reasonable steps to figure out the truth, they'll usually succeed. We all know that, on many political topics—from the causes of climate change to the rates of immigration to the constitutionality of presidential actions—the evidence is plentiful. Despite this fact, people patently have *not* converged to each other, meaning at least one side hasn't converged to the truth. Accepting Reasonable Convergence, we conclude that people—especially those on the other political side—*haven't* been taking reasonable steps to figure out the truth. We think, 'They must be being unreasonable; why else would they vote and act and believe as they do?' And if people are being unreasonable and not caring about the truth—especially on important topics they *should* care about—they are (1) blameworthy, (2) not worth listening to, and (3) untrustworthy. Let's look at each of these in turn.

**(1) The rise of enmity.** Why would belief in Reasonable Convergence be linked to our increasing willingness to blame the other side? Because blame and other reactive attitudes are highly sensitive to epistemic rationality (Rosen 2008).

Suppose Dr. Delila prescribes me penicillin, and I have a serious reaction leading to hospitalization. Should I blame her? It depends. If she had checked all the latest medical advice, had thought through it carefully, and had good reason to think that I would be fine, then she isn't blameworthy. But if she misread the evidence, failed to check for an interaction effect that she should've, or otherwise *should've known* that it was a big risk, then I'll rightfully be furious. Generally, when an action has a bad outcome, we want to know whether or not there were good epistemic reasons to do it. At a first pass: if there were, then the actor is blameless; but if there weren't, they are blameworthy. Epistemic rationality provides moral excuses. (We'll consider exceptions below.)

Likewise when the action in question is a political action. Imagine that you're convinced that Trump's second term is having terrible consequences. Now consider a Trump voter. If you think they *should've known better*, you'll be angry and upset with them, and likely conclude that they are either unintelligent or immoral, pursuing the wrong values. But if, on the other hand, you think that they are *reasonable* to have the opinions and goals they have—even if they're wrong—you will perhaps be frustrated and disappointed but blame them much less. And, of course, if you believe in Reasonable Convergence, you *will* believe that they should've known better. So belief in Reasonable Convergence should exacerbate the rise of enmity.

**(2) The rise of intolerance.** For similar reasons, belief in Reasonable Convergence should make you less tolerant of their political judgments. Of course, no matter what you believe about

rationality, we're holding fixed that you believe they are *wrong* in their political judgments—they voted for the wrong person. But it's one thing to get it wrong reasonably, due to sensitivity to a large swath of facts that genuinely pull in the direction that they chose, but which it turns out are outweighed overall. It's quite another to get it wrong due to rampant bias and irrationality.

If you believe in Reasonable Convergence, you'll see bias and irrationality as the cause of your disagreement—you will think that half of your fellow citizens are disconnected from reality, and not even tracking the basic facts on the ground. So *of course* you'll be less enthusiastic about democracy and will be less inclined to allow people you strongly disagree with to speak—you are confident that you don't have anything important to learn from them, and worry that other people will be fooled by their rhetorical tricks, or whatever it was that fooled *them*.

**(3) The fall of trust.** Likewise for distrusting the judgment of our fellow citizens. If you believe in Reasonable Convergence, then you'll conclude that your political opponents' judgments are not sensitive to the facts, and are instead driven by bias or motivated reasoning. This will not only make you less tolerant of their views, but also make you trust their judgment far less—you will think you have little to learn from them, and come to think that not only did they make a mistake in this case, but that they are likely to make mistakes in many more cases.

Belief in Reasonable Convergence also exacerbates our distrust of institutions. When institutions (or people within them) don't agree with each other, Reasonable Convergence predicts that either (i) evidence is not plentiful, or (ii) they are not being reasonable. Either way, seeing disagreement should lead to a loss of trust and deference. Meanwhile, suppose the institution *does* have a broad consensus—say, on climate change. Suppose (for whatever reason) you don't end up deferring to them, i.e. you are not confident that they have converged to the truth. If you thought they might be reasonably polarized, then you could still think that they have plenty of evidence that you could learn from and that they are reasoning about it well. But if instead you believe in Reasonable Convergence, you'll conclude that either (i) they don't have plentiful evidence, or (ii) they're not using it reasonably. Your distrust of the institution will run deeper.

Those are the theoretical reasons to think that belief in Reasonable Convergence should exacerbate our trends. But do we find empirical evidence that it does?

### 12.2.2 The Empirical Argument

We do. To get such evidence, we must find some psychological measure that is a proxy for belief in Reasonable Convergence, and then see how it relates to affective polarization and distrust. Of course, there is no perfect measure—and I am not in the business of designing a new one. So this is an imperfect test. But still, we can find several psychometric scales that plausibly capture aspects of belief in Reasonable Convergence.

First, there is what sometimes gets called the 'objectivity illusion' or 'naive realism', which is intended to measure the tendency to think that *we* see the world objectively, while those who disagree with us must be either uninformed, irrational, or biased (Ross and Ward 1996). For instance, Schwalbe et al. 2020 measured how much Trump and Clinton supporters in 2016 thought that supporters on each side were moved by normative vs. non-normative considerations. The former included 'careful consideration', 'knowledge of facts and history', and 'interest in what's best

I'm thinking it makes sense to put fall of trust before rise of intolerance, since the latter is more of a consequence

for the country’; the latter included ‘self-interest’, ‘biased media coverage’, ‘misinformation’, and ‘propaganda’. If I’m right that belief in Reasonable Convergence exacerbates affective polarization, we should expect this measure to be positively correlated with it.

A second, popular psychological measure is ‘intellectual humility’, which is usually defined as ‘recognizing that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one’s own limitations in obtaining and evaluating relevant information’ (Leary et al. 2017, 793). This is measured using agreement with a series of items like ‘I question my own opinions, positions, and viewpoints because they could be wrong’, ‘I recognize the value in opinions that are different from my own’, and ‘I like finding out new information that differs from what I already think is true’. This measure doesn’t fit Reasonable Convergence perfectly, since it is also measuring a general degree of confidence in your own opinions. But since belief in Reasonable Convergence leads people to be less inclined to worry about counter-evidence and to be less interested in the opinions of those who disagree with them, it should be a decent proxy for belief in Reasonable Convergence. If I’m right that belief in Reasonable Convergence exacerbates affective polarization, we should expect intellectual humility to be negatively correlated with affective polarization.

We find what we expect. Schwalbe et al. 2020 found that their measure of objectivity bias was robustly correlated with affective polarization. For instance, it predicts how upset subjects would be if their child married someone of the opposite party ( $r = 0.33$ ), even controlling for strength of candidate support. It also predicts the degree to which they rated an opposing blog author as evil ( $r = 0.34$ ), as well as political adversaries generally as evil ( $r = 0.2$ ), and whether they thought their political adversaries were likely to commit acts of terrorism ( $r = 0.2$ ). All of these effects remained significant after controlling for strength of support for their candidate.

Relatedly, some studies find that that interventions warning people about naive realism led people to judge others less harshly. For example, Nasie et al. 2014 found that it made hawkish participants in the Israel-Palestine conflict more open to narratives from the other side, and Cuadrado et al. 2022 found that it increased the rate at which subjects rated Moroccan women—a stereotyped and discriminated-against group—as moral, sociable, and competent. That’s some (limited) causal evidence that reducing belief in Reasonable Convergence can reduce enmity, intolerance, and distrust.

The verdicts are similar for intellectual humility. For instance, Stanley et al. 2020 found that intellectual humility was negatively correlated with both belief in the poor intellectual capabilities of opponents ( $r$  ranging from  $-0.25$  to  $-0.37$  across topics), and belief in the immorality of opponents ( $r$  ranging from  $-0.25$  to  $-0.49$ ), even after controlling for political leaning and the importance of party affiliation to your identity. They also found that those who with more intellectual humility were more willing to befriend political opponents.

Similarly, Knöchelmann and Cohrs 2024 measured intellectual humility about specific topics, affective polarization (with feeling thermometers and trait differences), and approach vs. avoidance behavior, such as willingness to talk with or befriend a political opponent. They found that intellectual humility predicted less affective polarization as measured by feeling thermometers ( $r = -0.33$ ) and trait differences ( $r = -0.17$ ), and that subjects with higher intellectual humility were much more willing to approach and much less inclined to avoid political opponents. Likewise, McMurtrie

et al. 2024 develop a measure of affective polarization that combined social distance (e.g. caring about which side your children marry), aversion (e.g. how much you dislike the other side), and incivility (e.g. how willing you are to help someone from the other side). They find that intellectual humility is negatively correlated with all three ( $r$  ranging from  $-0.22$  to  $-0.33$ ).

In short: there is good theoretical and empirical reason to think that belief in Reasonable Convergence exacerbates the enmity, intolerance, and distrust that pervade our politics.

## 12.3 Forgoing Blame

The thesis of this book is that Reasonable Convergence is wrong—that instead that we are:

**Reasonably Polarized:** People reasonably approximate Bayesian solutions to the problems they face. Under clarity, this leads to convergence—but under ambiguity, it leads to bias and polarization.

I now want to argue that this thesis has political implications. If we accept or lend significant credence to it, we should soften our attitudes toward our opponents (§12.3), increase our trust in them (§12.4), and owe them a duty of tolerance and persuasion (§12.5).

### 12.3.1 Epistemic Rationality Excuses

First, enmity. If we accept that we are Reasonably Polarized—and, therefore, that our political opponents are epistemically rational to believe what they do—then we should reduce our blame and dislike. We should become less affectively polarized.

Blame, dislike, feeling ‘cold’ toward our opponents, as well as thinking that they are ‘close-minded’, ‘dishonest’, and ‘immoral’, are *reactive attitudes* in the philosophers’ sense. They are feelings that we only have toward other moral agents—you can’t blame or ‘feel cold’ toward a toaster, even if it isn’t working.

Philosophers distinguish between the ‘objective’ and ‘participant’ stances we can take toward other people (Strawson 1963). The objective stance is the one we take toward engineering or mechanical issues—viewing the person as a problem to be solved or avoided, much as you’d view a stalled-out car or a broken elevator. The participant stance is the one we adopt when we treat another as a *person*, who has responsibility for her actions, autonomy in her choices, and to whom we owe and expect mutual respect. We are in the participant stance whenever we feel the ‘reactive attitudes’ of love and appreciation, anger and blame, or gratitude and resentment. You cannot feel these things toward a toaster. You can feel them toward a Republican.

There are many philosophical questions about such attitudes, but let’s focus on their relationship to epistemic rationality. As we’ve seen, at a first pass it seems like the relationship is very tight. Recall Dr. Delilia who prescribed me penicillin, leading to an interaction with another medication and to me getting hospitalized. She did something that had a bad outcome for me. Should I blame her? Obviously it depends.

If it turns out she intended to harm me, then she’s blameworthy. But let’s stipulate that she had well-meaning, morally permissible goals—say, helping me get better while expanding her medical practice. Does it follow that she is blameless?

Obviously not. If she *should've* known better—or, more generally, had reason to expect this was not the best course of action—then she is (negligently) blameworthy for pursuing it, even if she believed she was taking sensible steps. To a first approximation: if she should've, given her evidence, expected the penicillin to help and not harm me, then she's blameless; but if she should not have expected that, she's blameworthy. So, at a first pass, it looks like when someone's goals are morally permissible and their actions sensible given their beliefs, then they are blameworthy for bad outcomes if and only if they weren't epistemically rational in their beliefs. If that were true, there would be an extremely tight connection between epistemic rationality and blame.

But it's not true. Epistemic rationality, understood as rationally responding to your evidence, does not necessarily insulate you from apt blame.

Suppose that Dr. Delila failed to satisfy some *zetetic* norm—an epistemic norm of inquiry, i.e. to gather evidence (Friedman 2020, 2024). Perhaps she neglected to look at my chart to see what other medications I was taking, or knew them but failed to check their possible interactions. Then even if she was perfectly epistemically rational *given the evidence she had*, I should feel anger and blame toward her—she should've gathered the evidence she had reason to expect was relevant.

Of course, it's important that the evidence is something she had reason to expect was there. Suppose she had every reason to expect that all would be well, but unbeknownst to her a paper was published last week showing a previously-unknown type of interaction effect. Then although there *was* available evidence that would've changed her mind—all she needed to do was type in the relevant search term to find it—since she had no reason to expect it to be there, she had no zetetic duty to gather such evidence.

Let's say that your action is *fully epistemically justified* iff you (1) satisfied all the genuine zetetic requirements on you to gather evidence, (2) responded rationally to the evidence you received, and (3) in light of that evidence you expected your choice to be a permissible course of action (relative to your permissible goals). It's natural to think that if your action was fully epistemically justified, then it can't be blameworthy—if bad things happen, you have an iron-clad excuse. ('I couldn't have known; I did my best with the limited information we had'.) Call it:

**Epistemic Rationality Excuses:** If your action was fully epistemically justified, then it is not apt for blame or resentment.

Below we will consider some edge-case exceptions to this principle, so it's better to think of it as a heuristic. But in the main, this is surely right (cf. Rosen 2008).

What happens when we combine the hypothesis that we are Reasonably Polarized with the heuristic that Epistemic Rationality Excuses? Well, consider one of your political opponents—someone who you vehemently disagree with.

Let's imagine that you firmly believe that Trump has violated the Constitution in his second term, and that many of his policies are cruel, arbitrary, ineffective, and pursuing reprehensible goals. Meet Max, who's a died-in-the-wool MAGA supporter: he firmly believes that Trump has followed the Constitution, that his immigration and tariff policies are justified and necessary to reduce crime and increase opportunity for Americans, etc.<sup>1</sup>

---

<sup>1</sup>If you believe the opposite, imagine Max is a Marxist who believes Trump should be in jail and who, despite abhorring violence, has sympathy for the reasoning that led a radicalized young man to murder Charlie Kirk.

If you believe in Reasonable Convergence, then you believe that Max should know better. After all, you think, Trump *has* violated the Constitution, and that his immigration and tariff policies are *not* effective, necessary, nor helpful for reducing crime and increasing opportunity for Americans. And it's not some secret—there is plentiful evidence to this effect, broadcast from every (left-leaning) media outlet, seen in the uncontroversial economic data, and so on. So Max either has impermissible goals (he wants to harm people for no reason), or he is badly off the rails in his beliefs about what will lead to good goals. Either way, *of course* you blame him for his actions and words; *of course* you think he is immoral, close-minded, and unintelligent.

But suppose instead that you believe we are Reasonably Polarized. Then you have good reason to think that Max—or, at least, most of your political opponents, including many people like him—is fully epistemically justified in his actions. After all, he *says* his goals are to reduce crime and increase opportunity for Americans. Those are perfectly permissible political goals. And he *thinks* the actions Trump is taking are and will serve those goals—albeit with tradeoffs and some bad consequences, but no more so than many other politically necessary actions.

If he is Reasonably Polarized, he is *fully epistemically justified* to think those things. It's immediate that he is responding rationally to his evidence—he is (approximating) Bayesian conditioning given his reasonable but ambiguous priors. Less obvious, but also true, is that he has also been zetetically rational. Or, at least, failures of zetetic rationality do not explain his beliefs—even if he had been perfectly zetetically diligent, he would still think what he does. This is true even if Max is quite selective in his media habits and patterns of cognitive search—perhaps he only watches Fox news and Ben Shapiro, and whenever he sees liberal media he (internally and externally) scrutinizes it relentlessly.

Two reasons. First, much of his information environment is not up to him—his friends, family, and work are constantly presenting him with arguments on one side of the issue. That would likely suffice for his beliefs being what they were, even if he scrutinized in a very different way. Second, and more importantly, recall that on my theory, polarization results from a *diachronic tragedy*. Although balancing your search for evidence between both sides is *at the outset* expected to wind you up with more accurate beliefs, *at each stage* the step that is most-expected to improve (short- and long-run) accuracy is to search for evidence selectively. When he wakes up and sees news headlines about the latest government shutdown, he has every reason to expect that Ben Shapiro will give him an accurate read of the situation and that The Bulwark will give a skewed and unrepresentative picture. Wanting to figure out what to make of the situation, of course he tunes in to Shapiro. He rationally thinks that the sources he listens to are the most balanced and reliable on offer—and that although listening to others might persuade him (Martingale fails, after all), they are likely to make him less accurate.

In short: if Max is Reasonably Polarized, then he's just like the Dr. Delilia when she has no reason to suspect that there'll be an interaction effect or that she'll find important information by searching for it. Even though we (now) know that she chose poorly, we do not blame her for doing as she did. Likewise for Max: even though we think (perhaps know) that he chose poorly, you should think that Max's actions—to vote for Trump, to defend him on social media, to donate to his campaign, and so on—are fully epistemically justified. If Epistemic Rationality Excuses—at least in cases like Max's—then Max is not blameworthy or apt for resentment, *even if* his actions

Hm, this is partially redundant with the theoretical argument from earlier. Need to streamline.

have terrible consequences.

More is true. You don't need to be *convinced* that we (or Max) is Reasonably Polarized; all you need is to take it seriously, lending it significant credence.

Go back to Delila. This time imagine, while sitting in the hospital bed, we're *unsure* whether she acted negligently or had done her due diligence and gotten unlucky. Suppose we learn that a warning about the interaction effect had gone out the day before she prescribed penicillin, but that it only went out to 70% of doctors. We're unsure whether she received-and-ignored the warning, or she didn't receive it. If the former, she should've known better; if the latter, she was fully epistemically justified. Let's stipulate that we know we can't find out the answer—Delila has, by coincidence left the country on a years-long trip. So we're stuck with 70%-credence that she was negligent, and 30%-credence that she was fully epistemically justified. Do we blame her?

It seems not. As Buchak 2014 points out, blaming someone seems to require *outright belief* that they did something wrong, in a sense that's relatively strong. Knowing that 70% of crimes in situation *X* are done by men, and that the only two people who could've done it are Jim and Jane, we cannot yet blame Jim. Mere statistical evidence—at least when it's not overwhelming—doesn't suffice for apt blame.

The same is true if the evidence is not statistical but there's significant uncertainty. Change the case. Again, Delila is off on a years-long trip. This time, no warning went out, but the interaction risk is a new and relatively-unknown finding. You're trying to get a sense of whether Delila should've known, and you ask three friends who are doctors. Two of them tell you that yes, she should've thought of the interaction; one of them tells you no, *they* never would've thought of that. You're genuinely unsure who is right—perhaps leaning toward thinking she should have, but being far from certain. No new evidence is forthcoming. Do you blame her? Again, it seems to me not—or at least, not unreservedly. 'I am furious at her for what she did, though of course there's a significant chance it was exactly what she should've done' has a Moorean ring to it. So long as you lend significant credence to the possibility that someone's action was fully epistemically justified (again, modulo the riders below), blame and resentment seem inapt.

In short: even if you haven't been convinced that we're Reasonably Polarized, if you lend it significant credence, it should substantially soften your reactive attitudes toward your political opponents.

It's not just about normative beliefs about your opponents. If we are Reasonably Polarized, then it follows that although people's political beliefs are biased, they are not biased in the way you probably thought. They do not ignore evidence—Arguments Work. They do not simply nod along to bad arguments—Quality Matters. They are not simply avoiding information they don't like— Selective Search is largely driven by expected accuracy. Indeed, they are gathering and processing information in a way that makes them *locally* more accurate about the relevant evidence. What is happening is that ambiguity-asymmetries are making them more sensitive to evidence pointing one way than the other, skewing their overall interpretation of the weight of evidence, and leading to polarization or radicalization. We will return to these empirical trends shortly, when we turn to trust. The point for now is that if you come to think we are Reasonably Polarized, that will change both your normative *and descriptive* beliefs about your political opponents—you will come to think that they are fully epistemically justified, and come to think that their beliefs are driven by genuine

Make sure to add this trend in Ch. 8

information and evidence supporting them, even though they've gotten the overall balance wrong.

### 12.3.2 The Benefits of Epistemic Empathy

This is a verdict I find not only plausible, but promising. It gives me a way to think about those I disagree with strongly—whether in politics, office quarrels, or personal relationships—that is helpful.

I think this'll be obvious in personal relationships. Seeing our friends and loved ones with epistemic empathy can help us to see where they're coming from, to be self-aware and self-critical, and to maintain and repair relationships. It supports a principle of charity: those who disagree with you probably are onto something, even if they're wrong about the big picture. We all hope that others will see *us* this way; and surely most of us will admit that our discussions with our loved ones go better when we see *them* this way.

More controversially, I think it is also a helpful approach to politics. One of our core pathologies is the high level of affective polarization—the stark degrees of enmity and dislike we feel for our political opponents, and which they feel for us. Now, it's important not to over-state this. Our 'meta-perceptions' are inaccurate: people vastly *over*-estimate how negatively their political opponents view them. For example, Moore-Berg et al. 2020 found that indeed Republicans were more inclined to 'see the humanity in' Republicans than Democrats: 84 vs. 62 on a 100-point scale, or a gap of 22 points. But they also found that Democrats *way* over-estimated that gap: their average estimate was that Republicans would rate Republicans at 91 and Democrats as 36, i.e. with a gap of 55 points, fully 2.5 times bigger than the gap actually is. The numbers are similar for Democrats' views of Republicans, and Republicans' 'meta-perceptions' about Democrats' views of Republicans. We over-estimate affective polarization (see Yudkin et al. 2019; Druckman et al. 2022; Stone 2023).

Still, as we've seen, we *are* quite affectively polarized. No one likes it when people on the other side feel this negatively toward them. But it also doesn't feel good to feel this way about others. When we are in the grip of the assumption of Reasonable Convergence, we are baffled by our political opponents—'How can they believe and say and do those things?' We think they are deeply confused or duped, and blameworthy for it. We think they must be self-interested and blind to what's happening around them. We worry that perhaps they are moral monsters, totally unlike us in their values and reasoning. We may start to think that they are incorrigible—a "basket of deplorables", in Hilary Clinton's infamous phrase—and so it is useless to try to understand them or persuade them. We may conclude that they are pawns of the powerful, duped by those who would use them for their own political ends—lost souls that "cling to guns or religion or antipathy toward people who aren't like them", in Barack Obama's infamous phrase. We may conclude that they hate America and everything that is good and just. There are too many Trump quotes to mention, but here's one: "The radicals on the left are the problem, and they're vicious and they're horrible and they're politically savvy."

Arguably, expressing these attitudes is poor political strategy. But unarguably, it feels bad to think these things about your fellow citizens. And it will likely leave you puzzled, if you manage to get out of your immediate social circle and spend some time with people from the other side, doing something that has nothing to do with politics. Those on the other side are not in some

distant place—even in the bluest city or reddest town, there are Republicans and Democrats. They are neighbors and friends and family members. You will find that they too coach Little League, and help strangers on the street, and worry about the country. You will be left wondering how agreeable, bright, caring people could be so politically off the rails and morally bankrupt.

Believing we are Reasonably Polarized offers a better way. Those people *are* bright and caring. They are reasonable and well-meaning. They are also wrong—but not for any negligence or failing on their part. There but for a twist of fate go we. I find that thinking my political opponents are Reasonably Polarized helps me to disagree with them without blaming or hating them—and to approach them with curiosity rather than scorn. And I find that that is a much more useful way to engage with them. I have no argument that you must also. What I can and have argued is that you are making a mistake if you view your political opponents as evil or unreasonable or blameworthy for their views. And what I can do is invite you to try viewing them a different way, and see if you too find it politically helpful.

But is it doable? We all have immediate reactive attitudes to what others say and do, including when they express beliefs that we are inclined to think as ‘crazy’. I believe that we are Reasonably Polarized about as strongly as anyone. But when my neighbor says in passing that there won’t be elections in the U.S. next year—or when my (Moroccan) Lyft driver says the town has changed because all of these sleazy, dishonest Indians have moved in with high-paying jobs but no interest in giving back, or when my friend makes light of Charlie Kirk’s murder—my immediate thought is ‘What??’, and my immediate reaction is blame. Believing we’re Reasonably Polarized, in my experience, does not stop you from making split-second judgments of people and their beliefs.

But it does help you with what comes next. I find the belief that people are Reasonably Polarized to be a kind of mindfulness practice. As any meditator will tell you, meditation won’t stop unpleasant thoughts from popping into your mind—but it will help you respond to them, gaining distance or letting them pass or being less reactive. Likewise, in my experience, with believing that most of those ‘crazy’ people out there are reasonable. Instead of reacting to their crazy beliefs with scorn or distance—or, at best, changing the subject—it helps me to get curious. What exactly did they mean? What experiences or arguments led them to believe it?

This practice is hard. But I find that it works. I find that I am more curious and less judgmental. I find that they notice as much, and open up. I find that I learn and listen more, and have more genuine interactions with people who are genuinely other people, in Bernard Crick’s memorable phrase (Crick 2005). I find that the neighbor is a perennial pessimist, that the Lyft driver feels disrespected by his clients, and the friend had been frightened by Kirk’s tactics for years.

This curiosity is consistent with assigning zero credence to the possibility of them being right. After all, to think that people are Reasonably Polarized is to deny Reasonable Convergence—to think that reasonable belief is only loosely connected to the truth. It is to think that reasonable people go off the rails all the time, but to recognize that they had reasons and evidence that led them there. You can have curiosity about those reasons, while retaining confidence that the destination is wrong.

### 12.3.3 When To Blame

But is it desirable? Sometimes, you might think, we *should* feel anger or blame toward our political opponents. Does my view forbid it?

There are a couple ways anger and blame might be valuable. First, it could be necessary: sometimes getting angry is the only way to get you (or someone) to do something that they should. If you realize that Delila is about to give me a medication that has a high risk of serious harm, you should do something to stop her—even if (you know that) she’s being epistemically rational. If options are limited or she is recalcitrant, you should use more forceful methods, including violence if necessary. If the only way to get yourself to use such methods is by getting angry, then you should get angry. Likewise, of course, in politics. My view doesn’t say otherwise.

But to say that anger and blame are a useful or necessary means to achieving a worthwhile end, is not to say the anger or blame is *apt* in the sense of accurately representing its target. Compare: if the only way you can get yourself to take action against Delila is to get yourself to believe you are strong enough to lift a car, you should get yourself to believe that. But the fact that the belief is necessary doesn’t make it true. Sometimes we should blame or hate our (political) opponents because we must do so in order to take necessary action; it does not follow that they are blameworthy or hate-worthy. Indeed, if they are Reasonably Polarized and so fully epistemically justified, they are not.

But there’s another justification you might find for righteous anger. Consider opposing politicians or public figures who, you (rationally and, let’s stipulate, truly) believe are taking actions with terrible consequences. Right now, Trump is orchestrating mass deportations of immigrants and RFK is deconstructing our public health infrastructure. If you’re like me, you’re convinced that these actions have horrific consequences. Am I committed to saying you cannot aptly blame Trump or RFK or your least-favorite politician?

I don’t think so. First of all, the view that we are Reasonably Polarized is not the view that every person at all times is being reasonable. Sometimes people are not, and perhaps we have special reason to think Trump or RFK are not—in which case we have reason to think they are not epistemically justified, and thereby reason to think they are blameworthy.

But I don’t want to rely on that. Let’s stipulate for the sake of argument that we know our given politician—say, RFK—*is* Reasonably Polarized. Must we not blame him, then? Does full epistemic justification *always* excuse?

No. As Rosen 2008 points out, sometimes we can incur zetetic obligations to inquire over and above what epistemic norms demand. Suppose Dr. Delila has done her due diligence and has every reason to think that the penicillin won’t harm me. Being a hypochondriac, I ask her if she will double-check and do one more round of research. Let’s stipulate that she has no epistemic duty to do so, but she promises me she will. Then she gets busy, fails to do the extra research, and prescribes the medication anyways. Let’s further stipulate that if she *had* done the extra research, she would’ve found the warning about the interaction effect. Then when I am harmed, I can blame her—*despite* the fact that she was fully epistemically justified. Sometimes we can incur moral duties to inquire over and above our epistemic duties; and when we do, that expands the remit of what we ‘should’ know, in the sense relevant for apt blame.

Can we use this to explain why blaming RFK is apt? Sometimes, perhaps we can. But again, I don't want to rely on it, since it won't always hold. Granted, becoming a politician—especially one running a public health agency—incurrs moral duties to inquire into the relevant questions fully and with an open mind. But it does not and cannot impose a duty to converge to the truth, nor necessarily to search for evidence 'equally' in both directions. (Think of climate scientists.) At best it requires you to diligently inquire in a way that you expect to make you most accurate. But, as we've seen, inquiring in a way that maximizes expected accuracy (of your short- or long-run beliefs) is no guarantee that you will converge. So we can stipulate that we are in a case where RFK has done that, and failed to converge to the truth. If this explanation were right, then we could not blame him in such a case.

Here, I think, is a better tack. There is something right about thinking about promises to inquire. We have *normative powers*: there are things we can do which change our obligations to and expectations of each other. This includes promises and vows and (in the right circumstances) commands, but it also includes softer variants of these like assurance ('Don't worry, I'll be there') and pressure ('I'd love it if you could come') [CITES].

Suppose Dr. Delila sees that I am worried, and assures me I'll be fine, 'Don't worry, I prescribe this all the time. Trust me, it won't harm you.' Suppose that she was fully epistemically justified in giving this assurance—she had every reason to think that I was worrying unnecessarily, and that such assurance would put me at ease and avoid adverse psycho-somatic reactions. But then I get hospitalized due to an interaction effect—Delila was wrong. In this case, *despite* Delila being fulling epistemically justified at each step, it seems to me that I can rightfully blame her.

Why? Well, why did I *want* her assurance to begin with? Obviously to calm my nerves and make me feel more better. But why would it have that effect? Think of the economics of medical assurances. Suppose that doctors knew that if they were fully epistemically justified, then no matter what they said they would not be blameworthy if things go awry—in other words, suppose Epistemic Rationality (*always*) Excuses. Then their assurance wouldn't mean much—they would be incentivized to give their assurance no matter how confident they were, so long as they were justified in believing they were taking the best course of action. And if they would be incentivized to give their assurance no matter what, pretty soon their assurance would be meaningless, like the signs on every deli proclaiming 'world's best bagels'.

What I *want* when talking to Delila about my worries is for her to be able to give me a *costly* signal of her confidence—one that she would only give if she were genuinely confident. But the only way to guarantee that is if there is some risk of blowback or blame if the signal is sent in error. In other words, we need a practice whereby doctors can put themselves genuinely on the hook by what they say, so that if things go awry, then *even if they were epistemically justified*, they are blameworthy. Contrast how you feel if Delila says (1) 'Look, there are no guarantees in medicine; but this treatment almost always works' vs. (2) 'Trust me, you'll be fine'. (1) is much less reassuring, precisely because Delila has *not* put herself out on a limb—she has only told you the medical facts, and can't be blamed if things go awry. (2) is meaningful precisely because it removes her excuse—something she would only be willing to do if she was genuinely confident that I would be fine.

Likewise, perhaps, for politicians. We're all familiar with the observation that presidents are

blamed for bad economies and praised for good ones, despite the fact that the economy is hardly under the president's control. This is often seen as irrational, but there's an alternative interpretation: politicians are giving us their assurances. In proposing their policies, they are putting themselves on the hook so that if things go bad, then *even if they are fully epistemically justified*, they are blameworthy.

The reasons why we would have this practice are similar: think of the economics of political assurances. In a complex, chaotic, and ambiguous world, all sorts of beliefs about the impacts of policies can be epistemically justified. If such epistemic justification were always excusing, politicians would be off the moral hook if things go badly. But that would be a bad system—it would prevent politicians from sending us a *costly* signal of their commitment to their policies, since they would pretty much always have an excuse. (Or, at least, we wouldn't be confident they didn't—so we couldn't justifiably blame them.)

To solve this problem we might develop a practice whereby people—including politicians—can make *moral gambles*. They could give public assurances, thereby giving up their epistemic-rationality-based excuse if things go poorly. They are on the hook, even if they were justified and the outcome was not in their control. On the other hand, if things go well then they are set up for praise—again, even though things weren't under their control.

This is a practice it would make sense to have: politicians, public figures, and other involved parties can put themselves on the hook for the political and economics policies they stand for, giving up their epistemic excuses. It would allow them to show that they genuinely believe in their policies. But it would only work if they *are* blameworthy if their policies have terrible consequences—and that if we rationally believe as much, we can justifiably blame them. Perhaps that is why anger and blame are often apt toward the opposing side's politicians and public figures—they have taken a moral gamble, and (we believe or know) lost.

But notice that this aptness of blaming politicians does not extend to the vast majority of the private citizens who support them. Those citizens haven't given us their assurance—all they've done is given a politician their vote. In short: believing we're Reasonably Polarized should soften our least-productive forms of affective polarization—hating our fellow citizens—while still allowing us to blame politicians who we disagree with.

## 12.4 Rebuilding Trust

Trust in both individuals and in institutions has fallen dramatically over the past decades. If we come to believe that we are Reasonably Polarized, then there is a sense in which this lack of trust is apt, but also senses in which it's not—and things we can do to rebuild it.

Start with trust in individuals' judgments. Are we justified in distrusting our fellow citizens? In some sense, yes: to believe that people are Reasonably Polarized is to believe that they are biased, and will likely fail to converge to the truth even when the evidence is plentiful. It is to believe that the connection between rationality and truth is far weaker than we usually assume. So if trust in the wisdom of the American people means trusting that the majority of them will usually get the correct answer on hard political questions, then I do not advocate trust in the American people.

But there is another sense of trusting people’s judgment that we might have, perhaps conveyed by that word ‘wisdom’. A wise person’s advice doesn’t always turn out for the best, but it is sensitive to the right reasons. Distinguish trust in *judgment* from trust in *reasoning*: trusting people to get the right conclusion, vs. trusting them to be sensitive to the right reasons. Believing Reasonable Convergence, we have believed that these stand or fall together—that if you were sensitive to the right reasons, then you would get to the truth. As a result, we have concluded from people’s failure to converge to the truth that they are not sensitive to the right reasons—perhaps not even in the *business* of truth-seeking, instead simply engaging in tribal cheerleading [CITE].

If I am right, then this is wrong. If we are Reasonably Polarized, then most people *are* sensitive to the right reasons—recall that Arguments Work, Quality Matters, and that Selective Search is largely driven by expected accuracy. This matters. If people are not even sensitive to the right reasons, then they have nothing to contribute to the discussion, and no basis for their beliefs. They are *simply* biased or pigheaded or self-interested or evil. What could we learn from them? No wonder we don’t want to hear from them.

But if people *are* sensitive to the right reasons, then—even if they’re polarized, radicalized, and wrong—they know things that we can learn from. There are *genuine reasons* driving their beliefs, even if they’ve over-weighted those reasons or under-weighted the opposing ones. This suggests that we should be more open to hearing the reasons motivating even our radical political opponents.

Moreover, although convergence is not to be *expected* under ambiguity, it is still *possible*. One strand of reaction to the literature on bias is to infer that many existing standards of assessment—especially when they have elements of subjective judgment—are simply tools for the powerful. Perhaps sometimes they are. But if we are Reasonably Polarized, we cannot infer from the relentless ambiguity of a subject-matter that there is no fact of the matter, nor that no one has found it. If people have searched for evidence in the right balance of directions, and have assessed enough of it, they might well have converged to the truth.

What about institutional trust? Again, if we are Reasonably Polarized then the verdict is mixed. Even institutions run by smart and well-meaning people can easily get things wrong and go off the rails when they investigating a relentlessly ambiguous topic. The fact that all the experts agree is far from a guarantee of truth—especially if they are all searching for evidence in the same direction. All the institutional skeptics out there are right about something: some of our institutions are not as trustworthy as they claim to be.

It’s important to acknowledge these genuine reasons for skepticism. But it’s also important to not overstate them: well-run institutions often *do* achieve feats of convergence. This is most robust when they achieve clarity about a subject-matter, as we’ve seen in our discussions of science (Ch. 9). If we accept that we are Reasonably Polarized, then our patterns of deference to institutions should be highly sensitive to whether or not they’ve achieved clarity about their subject-matter—and, if they have not, how they’ve searched for evidence.

As we’ve seen, even under ambiguity, if institutions retain the right sort of intellectual diversity, with a ‘Foxing’ strategy of arguing different sides of an issue—*in utramque partem*, as Roman rhetoricians used to say—then they can achieve feats of convergence to the truth even under ambiguity. We’ve also seen that once you achieve clarity, direction of search matters much less for where your beliefs end up. Therefore, whether or not you have achieved clarity, the hypothesis that we are

Reasonably Polarized provides a straightforward argument in favor of free discussion, open debate, and the adversarial model of inquiry used in the courts and much of academia. The most reliable way for the truth to come out, especially under ambiguity, is for smart, honest people to make the strongest case they can on opposing sides of the issues. And people know that. So one way to rebuild trust in institutions is to publicly re-entrench these norms.

## 12.5 Saving Persuasion

Turn, finally, to intolerance. There is a long history of attempts to constrain political and ethical debates to a certain set of ‘reasonable’ opinions that share much common ground. This is clear in philosophy. You see it in Rawlsian conceptions of ‘public reason’ and the thought that political power can only be legitimately exercised in ways that all ‘reasonable citizens’ can be expected to endorse [CITE]. You see it in Scanlonian contractualism, with its link between what is wrong and what systems of norms someone could ‘reasonably’ reject [CITE].

Others?

But it is also clear in contemporary politics. These days, the internet and partisan media regularly confront us with viewpoints that we think should be outside the bounds of permissible political discourse. In the face of such an assault, there is a constant desire to want to draw lines between ‘us’—the good, sensible, well-meaning folks over here—and ‘them’, the deplorables with unacceptable views. There is both social and moral pressure to avoid engaging with those views, despite the fact that they are held by a substantial portion of our fellow citizens. This pressure helps give rise to ‘cancel culture’ and other attempts to police the bounds of discussion. Here’s Ta-Nehisi Coates, in an interview with Ezra Klein about “Bridging Gaps vs. Drawing Lines”:

For me, the bigger question is where are the lines? You know what I mean? I think there’s no problem with saying, listen, you can’t hurl epitaphs at people. You’re out if you do that. I’m sorry. . . If you think it is okay to dehumanize people, then conversation between you and I is probably not possible. [CITE]

Call this—in both its philosophical and political guises—the attempt to *tame politics*. Any such attempt presupposes that we can draw a line between the ‘reasonable’ and ‘unreasonable’ political opinions that is both (1) justifiable and (2) substantive. It must be justifiable in the sense that we are right to enforce it, and those with views outside the line are wrong to object to our enforcement. And it must be substantive in the sense that it constrains the views people can endorse and defend so that, at the very least, public discourse is more constrained than it is currently. As an explicit minimal condition, I’ll assume ‘substance’ requires all permitted views to endorse the basic tenets of a liberal democracy, including the rule of law and the equality of persons.

I believe that if we accept that we are Reasonably Polarized, then we must conclude that we cannot tame politics. There is no way to draw a line around the ‘reasonable’ opinions that is both justifiable and substantive. This is both a critique of current attempts to tame political discussion, and an objection to political philosophies that presuppose that we can. And it suggests a different way forward—that we might ground our obligations to our fellow citizens in a duty of persuasion (Garsten 2006).

### 12.5.1 When To Persuade?

Start with a simple case. Suppose we have a community garden which we've all been working on. It is time for the harvest, and we're having a town meeting about how to divvy up the proceeds—who gets how many carrots, and pumpkins, and turnips? The rules of the meeting are that everyone in town gets to say their preferred proposal, then there's open discussion, and then a vote. Jane proposes everyone should get equal shares. Fred proposes that people should rank their favorites, so that people with different preferences can get more of what they want. Then Max stands up and says that he and his family should get *all* the proceeds.

Huh? We're confused; why would he be proposing that?

Suppose we know that Max is being unreasonable—that he has impermissible values (he thinks his family is entitled to more, despite doing no more work), or that he has irrational beliefs (on no evidence, he thinks that he and his family did all the work). Then clearly we are under no obligation to hear him out or address his concerns. We are obliged to let him speak for his allotted two minutes, and we may feel some pressure to briefly explain why his proposal is unfounded and unfair. But we needn't genuinely listen to his reasons (we know they are bad), and we don't need to spend time in discussion arguing with or trying to persuade him away from them. The reasonable ones among us can justifiably (and truly) say that it is an unreasonable proposal that we don't need to take seriously, and go about our business deciding amongst the reasonable proposals.

Of course, if it turns out that unreasonable people like Max make up a substantial proportion of the town, then shutting them down or labelling them as unreasonable might not be good political strategy—it might make them angry or more vociferous or less willing to follow the rules. Still, we would be correct to view their proposal as carrying no genuine weight in our deliberation. We would view them from the objective stance (Strawson 1963), as obstacles to be overcome and managed—not agents to be listened to and reasoned with.

But suppose, instead, that we have reason to think that Max *is* being reasonable. He genuinely does have permissible values, and is fully epistemically justified in his beliefs; or at least, suppose we should lend significant credence to this possibility. Perhaps we have reason to think that he has been misled by his family, who claimed to him that *they* had been doing all the gardening and that none of the other townsfolk have helped. Perhaps he has reason to think that the original agreement was that he would fund the entire garden and reap all the produce, and that the rest of the town had volunteered to help just for fun. Perhaps he has reason to think that other people have been skimming produce off the garden already, and taken more than their fair share, so that his family is the only honest one who hasn't received anything yet. Perhaps something else.

In each of these cases, we are sure that Max is wrong: he has mistaken beliefs about the history of the garden, the rules with which it was set up, and what other people have done. But if (we have reason to believe) that his mistaken beliefs are *fully epistemically justified*, then clearly we are *not* morally justified in simply ignoring or dismissing his proposal without explanation. Other constraints may prevent us from thoroughly engaging with him—perhaps time is short and everyone else agrees that he's in the wrong, so we table his proposal and go ahead. But, in this case, we clearly owe him something. If time permits, we owe it to hear him out, to listen to what reasons he has for thinking that his family deserves all the produce. Likewise, we have a duty to try to explain

to him where he's gone wrong—why his justified beliefs are mistaken, why we aren't going to accept his proposal, and why he should accept ours instead. Helpfully, since we believe he's reasonable, we have reason to think he will be sensitive to the reasons we offer, and might well be persuadable. In short: even if we know Max is wrong, when we have reason to think that he's reasonable—with morally permissible values and epistemically rational beliefs—we owe him a *duty of persuasion*: a duty to listen to where he's coming from, and to try to convince him where he's gone wrong, and bring him around to our side of the issue.

I think this simple case is clarifying—it tells us something important about politics. Suppose a member of our political community, our fellow citizen, disagrees strongly with us. Suppose he's broadly conforming to the norms of political discussion—he is not being violent or dangerous, although he may well be aggressive and confrontational, as political discussions often are. We believe or know he's mistaken. How should we treat him? It depends.

Suppose we have reason to believe that he is being unreasonable: that he has morally impermissible values, epistemically unjustified beliefs, or otherwise should know better. Then we are justified in adopting the objective stance, treating him as an obstacle to be overcome but not a person to be addressed. If it is strategically viable, we are justified in ostracizing or shunning or ignoring him and his proposals. If circumstances don't demand otherwise, it is not our job to educate unreasonable people who have all the tools at their disposal to educate themselves. Political tolerance need not extend to sociopaths or fools.

But suppose we have reason to think otherwise. Suppose we should think that he is being morally and epistemically reasonable, engaging with good faith. Perhaps he is rude or agitated or mean toward us, or people we care about. But suppose we believe that he genuinely has reason to be—he is fully epistemically justified in thinking that *we* are engaging in bad faith, or have impermissible values, or are being epistemically unreasonable and should know better. Then, I think, we owe him a duty of persuasion. If time and space allow, we should hear him out—hear out the genuine reasons he has for his beliefs and proposals, to see where he's coming from. That will allow us to put our beliefs about the correct (incompatible) proposals in proper perspective, and to better articulate to him why he is wrong and why we are right. Perhaps we know we won't be able to convince him, but since we believe he is reasonable we believe he will be sensitive to our reasons—and perhaps, with time, will come around to them. If our fellow citizens are Reasonably Polarized, we owe it to them to try to persuade them.

As always, our obligations are shaped by political reality and strategy. Perhaps we know our attempt to persuade will backfire. Perhaps we know that we are at risk of being bamboozled by his reasons, tricked into mistakenly giving up our (correct) position. Perhaps the issues are important and pressing enough that we don't have time to engage with him, and know that we can successfully stonewall him and push him out of the political decision-making. In such cases, all things considered, we should not engage and should work around him. But when we must, there are moral traces [CITE]—we should feel regret at having to do so, pressure to apologize when and if it is doable, and plan to try to persuade him in another setting.

I think these conclusions are common sense: other things equal, if people are being reasonable then they deserve to be heard out and reasoned with; if not, they can justifiably be ignored. I suspect most of us already accepted these claims, at least implicitly. But when combined with the

hypothesis I've defended at length—that in our political debates, office dramas, and couple quarrels, we are Reasonably Polarized—they have radical consequences.

### 12.5.2 Political Discourse

This should be obvious for our current political discourse. In our age of a rabbit-holed internet and partisan politics, we are constantly barraged by political viewpoints that we intuitively think are crazy, evil, or beyond the pale. The attitude voiced by Ta-Nehisi Coates is common: we are inclined to think, at the very least, that productive conversation between us and the people who hold those beliefs is not possible. (Many of us are inclined to think much stronger things.) Substantial numbers of us have lost friends or cut off family members because of political disagreements. In 2021, 15% of people reported ending a friendship because of politics, and 11% said they stopped talking to a family member over it [CITE]. In 2020 fully 45% of people said they had stopped talking politics with a friend because of something they said [CITE]. Currently the numbers are worse for Democrats than Republicans: in 2024, 18% of Democrats and 10% of Republicans reported *in the last year* cutting off a friendship because of politics; 8% of Democrats and 4% of Republicans reported doing so for a family member, and 5% of Democrats and 1% of Republicans reported doing so for a spouse or partner [CITE]. Similarly, a 2022 study found that 24% of 'Very Liberal' respondents had stopped talking to a family member because of politics, compared to 9% of 'Very Conservative' respondents [CITE].

These responses make sense if we believe in Reasonable Convergence. The evidence about these basic political matters—that our side of the debate is reasonable, deserves respect and dignity, and should be met with curiosity rather than vitriol—is plentiful. If, given such evidence, reasonable attempts to get to the truth would succeed, then a surprising number of our fellow citizens, 'friends', and family members are not being reasonable. They should know better, and should do better. We might find in ourselves the grace and moral fortitude to engage with them and try to persuade them, but we don't *owe* it to them—and we are perfectly justified in cutting them off, if that is what we feel is best.

But if (we have reason to think that) they are Reasonably Polarized, things change. It is of course true that our side of the debate is reasonable, deserves respect and dignity, and ideally should be met with curiosity. It is of course true that the evidence in favor of these basic political facts is plentiful. But Reasonable Convergence is false: it does not follow that our fellow citizens, friends, and family will realize as much if they are being reasonable. Indeed, I've argued at length that in the informational and social environments in which we now live, they would likely *not* realize as much even if they were being reasonable. For they are inundated with arguments to the effect that our side is *not* reasonable, does *not* deserve respect, *has* acted in bad faith, *should* know better, and so on. These arguments are conducted under conditions of ambiguity, wherein even though Quality Matters—and the conclusion of the argument is false, and there are plenty of good arguments to the contrary—reasonable people will quite often be convinced by them.

So the people voicing the crazy and evil views you see are likely, in the main, being reasonable. And when people are being reasonable—even if they are badly mistaken—we owe it to them to listen, and try to bring them around to our side. We owe them a duty of persuasion.

This is not to say that every person is well-meaning and reasonable—obviously there are genuinely bad actors and genuinely unhinged and unjustified people. Obviously we are justified in cutting off such people, if it's the best strategy for us. But estimates for the rate at which people are genuinely sociopathic or so mentally ill that they are disconnected from reality are both around 1% [CITE.] There is evolutionary reason to think that the rates of such illnesses in the population would be relatively stable and low. And there is absolutely no reason to think that changing social or media conditions would, in a matter of years, skyrocket the proportion of genuinely unhinged people to the order of 40 or 50%. Let alone that our mothers, fathers, brothers, sisters, children, and friends—people that we have lived with for years—would suddenly become unreasonable people. Far more likely is that they found themselves in a rabbit hole of one form or another, where the balance of evidence and arguments was strongly skewed, so that the same reasonable mechanisms they've always used to form their beliefs have led them off the rails. They are probably Reasonably Polarized. And *if* they are, we should be listening, and learning, and persuading—bridging gaps, not drawing lines.

citation?

In short: a belief in Reasonable Convergence plays a crucial role in justifying and sustaining the intolerance that pervades our politics. And coming to believe—or even take seriously—the hypothesis that we are Reasonably Polarized should change that.

### 12.5.3 Political Philosophy

But I think the conclusion cuts deeper than contemporary politics. There is a storied tradition in political and moral philosophy of trying to derive the rules of what we owe to each other from what 'reasonable' people would endorse, or be unable to reject. Such proposals require people to use 'public reason' or in other ways engage with each other using a constrained type of reasoning. If I am right, then this tradition has overlooked the important and inevitable possibility that reasonable people will end up deeply polarized over empirical and (thereby) normative matters. And it suggests that a different proposal, rooted in persuasion, is a better norm for democratic politics.

Political philosophers tend to accept the inevitability of normative disagreements—between people with different 'comprehensive doctrines' about morality and the good life, in Rawls's terminology (Rawls 1971, 1987, 2005, 2006). Some moral philosophers do likewise—for example, in giving a contractualist theory of morality, Scanlon does not want to assume that the 'reasonable' people who would decide on the rules all agree on how to live (Scanlon 2000).

Such theories tend to worry much less about empirical disagreements. Yet they are a problem, especially for contractualist views (cf. de Kenessey 2025). Our political and moral disagreements are never pure normative disagreements. For almost every political or moral question that is debated in real life, there are large empirical disagreements about the causes, status, and consequences of different courses of action. And virtually all agree that if the empirical answers swung one way hard enough, that could flip the correct answer.

This is obvious in extreme cases. I assume you believe that the best or perhaps only permissible political system is a liberal one that accords equal and substantial rights to all individuals, including freedom of expression, religion, and association. But here is an empirical hypothesis: *in any society with access to the internet, if it has freedom of expression it will break apart into bloody civil war*

*and widespread massacres within decades; while if freedom of expression is curtailed in the (say) Chinese form of authoritarian capitalism, society will be stable and people will flourish.* Now, I don't think this or anything like this empirical hypothesis is true. But presumably we all agree that if it (or something close) *were* true, then the best political systems would *not* include freedom of expression. When the consequences get bad enough, what is morally and politically acceptable shifts. And what those consequences are is an empirical question.

Likewise with non-consequentialist considerations. We have agreed that we have a duty to listen to our fellow citizens when they are well-meaning, engaged in good-faith deliberation, and have rationally permissible beliefs given their evidence. But we are *not* obliged to engage with them if we have (sufficient) reason to believe they are acting in bad faith, are pursuing immoral ends, or have unhinged and unreasonable beliefs. Yet these, yet again, are empirical hypotheses. So whether we are obliged to work with and listen to our fellow citizens—or instead stonewall and circumvent them—depends on (rational) empirical beliefs about *their* beliefs and motives.

This much should be uncontroversial. It is not yet a problem in itself, for if Reasonable Convergence is true, and we *are* all well-meaning and epistemically reasonable, then there will be plentiful evidence that we are—so we should all agree that we are acting in reasonable good faith, and thereby agree that we are obliged to engage collaboratively with each other.

But what if Reasonable Convergence is false? If we can be Reasonably Polarized over what the best immigration policy is or what happened in the 2020 election, then obviously we can be Reasonably Polarized *about whether we are Reasonably Polarized*. That is: even if we are all well-meaning and epistemically rational, we may well be rationally convinced that people on the other side are *not* well-meaning and rational. And if we are (reasonably) convinced of that, we are under no obligation to engage with them constructively, take their perspective, or even to provide them reasons for our actions. We are not obliged to use 'public reason' to reach our conclusions, because (we justifiably think) we are not engaging with people who are reasonable and engaged in good faith. We are justified in breaking the political norms and rules, precisely because we are (reasonably, but wrongly) confident that *they* are doing so.

And that is the problem. Contractualist views of politics or morality say that the rules we are required to follow are (something like) those that we would all agree to—or not be able to reject—if we were each reasonable and committed to coming up with a fair system of cooperation. But suppose such reasonable people should be expected to polarize over the empirical consequences of any proposed system, the historical causes of the present state of affairs, and even whether others are engaging in good faith and committed to coming up with a fair system. Then reasonable people committed to coming up with a fair system of cooperation would not be able to agree on such a system, and any such proposal *would* be able to be reasonably rejected by someone (namely, by someone who thinks the people proposing it are not reasonable and that the consequences would be dire). Rawls, Scanlon and company have presupposed that reasonable people would all agree on certain basic empirical premises, including that others are engaging reasonably. They have assumed a special case of Reasonable Convergence, failing to take seriously the depth of disagreement and polarization we should expect from even well-meaning and reasonable people.

In effect, such theorists have tried to offload some of their normative arguments to the epistemology: rather than having to argue simply that a given structure of society is best or justified, they

have thought that reasonable people would come to this conclusion themselves. If I am right, then they have been far too optimistic about what reasonable people would agree to. Very likely, if you start out with a group of fully epistemically rational people who all have morally permissible goals and are all committed to constructing a fair system of cooperation, they will break down into the polarized, distrustful, and intolerant sort of polity that *we* have become. And if they *do* all happen to agree on a system, they may well agree on an illiberal structure of society, for they may all be reasonably radicalized into thinking that that is the only stable or acceptable one.

That's what happens if 'reasonable' means, simply, everyone being epistemically rational with permissible goals and a commitment to coming up with a fair scheme of cooperation with others who are so-committed. What if, instead, Rawls and Scanlon want to build *more* into what it means to be 'reasonable'? Perhaps, for example, we stipulate that 'reasonable' people are all convinced that liberal democracy is the best basic structure of society?

The problem with such a stipulation is to justify it. How are we justified in labeling people who do *not* believe this as 'unreasonable', and excluding their points of view from the debate? Why would they be unjustified in resisting our exclusion, with anger or force? If we could assume that any such people would either have impermissible goals, or be epistemically irrational, then of course we would be justified in doing so—that's the conclusion we came to above, when considering the crazy proposals of our fellow community-gardeners. But, if I am right, we can assume no such thing.

Indeed, if I am right, the 'unreasonable' people we are excluding have permissible goals and full epistemic justification in believing that we are unfairly excluding them—on any standard theory of moral justification, they would be justified in resisting our exclusion. They would be *right* to say—as some people in public discourse do say—that us proponents of liberal democracies are ourselves exercising a sort of tyranny, a tyranny of policing the bounds of reasonable discussion. In other words, as Bryan Garsten (2006) argues, they would justifiably feel a form of the 'liberal alienation' that many actual people feel from attempts to tame public reason:

Many people, however, seem to feel coerced by the implicit claim that their own judgments are not reasonable. They believe that their commitments and convictions deserve as much respect as their capacity to revise those commitments. They may find a discourse of justification less respectful than the rhetorical alternative in which their opinions and sentiments—and their capacity for judgment—are directly engaged. (199)

That is the dilemma for contractualists and those who try to circumscribe the bounds of 'public reason'. They either (i) assume only that 'reasonable' citizens would be epistemically rational with morally permissible goals, or (ii) they build in more than that. If (i), then the inevitable possibility of reasonable polarization scotches their conclusions—they cannot derive substantive norms, such as commitment to the equality of persons and the rule of law, from what reasonable people would agree to. And if (ii), then they can't *justify* excluding people who fall outside the bounds that they have set for 'reasonable' points of view.

What does this mean for morality? That is a deep question, which it's not my place to try to answer here. But I do think the inevitable possibility of reasonable polarization is a problem that agreement- and ideal-observer based theories of morality need to contend with.

What does it mean for politics? How are we to engage with those who have radical disagreements with us, including those who think that *we* are not engaging in good faith or are wildly off the rails? I think we should follow the line of reasoning in Cicero, Lincoln, and Garsten, and entrench a duty of persuasion.

Cicero argued that orators have a duty to argue both sides (*in utramque partem*) since inevitably both sides are holding onto partial truths. [CITE]

Lincoln (1842) argued that the best strategy is to meet people where they are:

To have expected them to do otherwise than they did—to have expected them not to meet denunciation with denunciation, crimination with crimination, and anathema with anathema, was to expect a reversal of human nature, which is God’s decree, and never can be reversed. When the conduct of men is designed to be influenced, persuasion, kind, unassuming persuasion, should ever be adopted. . . . If you would win a man to your cause, first convince him that you are his sincere friend. . . . On the contrary, assume to dictate to his judgment, or to command his action, or to mark him as one to be shunned and despised, and he will retreat within himself, close all the avenues to his head and his heart... [CITE]

And, beyond strategy, he argued that that to do otherwise (in the context of the temperance movement) was unjust:

But I have said that denunciations against dram-sellers and dram-drinkers are *unjust* as well as impolitic. . . . If, then, what I have been saying be true, is it wonderful, that some should think and act now as *all* thought and acted twenty years ago? And is it just to assail, contemn, or despise them, for doing so? . . .

Another error. . . was, the position that all habitual drunkards were utterly incorrigible, and therefore, must be turned adrift, and damned without remedy. . . . There is in this something so repugnant to humanity, so uncharitable, so cold-blooded and feelingless, that it never did, nor ever can enlist the enthusiasm of a popular cause. . . . It looked so fiendishly selfish, so like throwing fathers and brothers overboard, to lighten the boat for our security—that the noble minded shrank from the manifest meanness of the thing.

A *duty of persuasion* to our fellow citizens would require three things of us. First, it would require a willingness to meet them where they are—to see what they believe, and listen to their reasons for it, so we can better understand what moves them, and how we might change their mind. As Garsten puts it:

The effort to persuade requires us to engage with others wherever they stand and to begin our argument there, as opposed to simply asserting that they would adopt our opinion if they were more reasonable. (2006, 3)

Second, it would require a firm belief in the correctness of our own views, so that we can listen genuinely to those who disagree with us but still believe that we are in the right, so that we can begin to craft an argument or narrative that will lead them to our position:

[T]rying to persuade others requires us to step outside our particular perspectives without asking us to leave our particular commitments behind. (2006, 210)

And third, it would require a belief—perhaps a faith—in the power of argument and reasoning to change people’s mind’s. Here’s Garsten describing Lyndon B. Johnson’s 1960 tour of the South, using his own Southern roots as a basis to convince people to vote for him and JFK:

He [got people to think] by not giving in to the view that southern opinion was irredeemably prejudiced and dogmatic but instead indulging the hope that there could be found, even within opinions characterized by great fear and prejudice, deliberative pathways and a capacity for situated judgment. He maintained a fundamentally democratic faith in the possibility of persuasion. (2006, 194)

So what we would need is (1) a willingness to listen, (2) a belief in ourselves, and (3) a faith in the possibility of persuasion. If I am right that we are Reasonably Polarized, then the story I have told buttresses each of these needs.

It buttresses (1) a willingness to listen, since it tells us that those ‘crazy’ people who disagree with us are not simply pig-headed or misinformed or irrational. They are being reasonable, with genuine reasons for their beliefs—they are sensitive to partial truths and real considerations. We have much to learn from them.

It buttresses (2) a belief in ourselves, because it warrants a type of steadfastness in the face of disagreement. *We* have been rational in forming our beliefs, and noting our disagreements with our reasonable opponents or our past selves should not undermine them. Those disagreements are exactly what we expected to find.

And it buttresses (3) a faith in the possibility of persuasion, because the type of rationality I have described is neither arbitrary nor permissive. Rational people respond to reasons—they are (approximately) Bayesians, treating things as evidence for things that make them likely. So Quality Matters and Arguments Work: if you provide people with genuinely good reasons for your position, they will—with reasonable skepticism, and perhaps painfully slowly—begin to take it more seriously. Especially if you acknowledge the rationality and insight of where *they* are coming from.

In short: if we come to accept that we are Reasonably Polarized, that should change both our normative and descriptive beliefs about our political opponents. It should show why we cannot hope to tame politics—we cannot justifiably keep political debates within a safe bound of ‘public reason’. Garsten argues for a similar conclusion, but even he frames it pessimistically:

We might be driven to a politics of persuasion in large part because there are no good alternatives, because efforts to avoid it tend to produce increasingly dogmatic forms of rhetoric and increasingly intransigent forms of conflict. [CITE]

But believing that we are Reasonably Polarized should make us more optimistic than that. It tells us that even those ‘crazy’ people out there are sensitive to reasons—and that they will be sensitive to *our* reasons, if presented the right light. There is nothing for it but to meet our fellow citizens where they are. But where they are is not nearly so far gone as we tend to fear.

## 12.6 Conclusion

Our politics has experienced a pathological rise of enmity, fall of trust, and rise of intolerance. There are many causes to these trends, but I have argued that an important one is a widespread and implicit belief in Reasonable Convergence—for it implies that our political opponents are not merely wrong, but unreasonable.

If we come to accept that we are Reasonably Polarized, that should help ameliorate these trends. It should soften our attitudes toward our opponents, revealing them as more justified and less blameworthy. It should raise our estimation of our opponents, revealing them as more sensitive to reasons and onto genuine considerations. And it should increase our commitment to our opponents, revealing that we cannot justify excluding them and we can succeed in persuading them. Politics cannot be tamed. But it also has not been lost.