

**Draft of Date 2/21/24.**


*Please do not cite or circulate without permission.*

## **How Beliefs Persist Amid Controversy: The Paths to Persistence Model**

Kerem Oktar and Tania Lombrozo

Department of Psychology, Princeton University

### **Author Note**

Kerem Oktar  <https://orcid.org/0000-0002-0118-5065>

Word Count: 12867

We have no known conflicts of interest to disclose.

A summary of the framework was presented at the 2022 meeting of the Cognitive Science Society and appears in the conference proceedings. The framework was also accepted for presentation at the 2022 meeting of the Society for Philosophy and Psychology. We are grateful to Thomas L. Griffiths, Molly Crockett, Kevin Dorst, Eldar Shafir, Alin Coman, Corey Cusimano, Sally Xie, Thalia Vrantzidis, Thiago Varela, and members of the Concepts and Cognition Lab for valuable feedback.

Correspondence concerning this article should be addressed to Kerem Oktar, Dept. of Psychology, Princeton University, Princeton, NJ 08540. Email: [oktar@princeton.edu](mailto:oktar@princeton.edu)

### **Abstract**

From abortion to vaccination, we frequently know that millions disagree with us on controversial issues, yet we remain firmly rooted in our convictions. What enables this capacity to sustain controversial beliefs? To answer this question, we connect insights from psychology, epistemology, political science, and probability theory to develop the Paths to Persistence (PTP) model. The PTP considers four drivers of persistence, formally parsing the unique contributions of the epistemic, meta-epistemic, non-epistemic, and bounded paths to persistence. We explain how each path can individually drive persistence, and then introduce a rational analysis that integrates these considerations into a formal model grounded in meta-reasoning. The resulting analysis has theoretical, empirical, and normative implications for our understanding of the psychology of controversy.

*Keywords:* disagreement, controversy, persistence, belief, meta-reasoning

**Draft of Date 2/21/24.**

*Please do not cite or circulate without permission.*

## **How Beliefs Persist Amid Controversy: The Paths to Persistence Model**

When, how, and why do people persist in their beliefs amid controversy? Why don't the dissenting opinions of millions give us pause about whether God exists, whether vaccinations should be mandated, or whether abortion is immoral? Why does dissent so rarely make us question or update our beliefs instead? Our aim in the current paper is to address this puzzle of persistence: the widespread tendency for people to remain anchored to their beliefs amid large-scale disagreement.

Despite pertinent work in psychology (e.g., Minson et al., 2023), philosophy (e.g., Frances, 2014), political science (e.g., Erikson & Tedin, 2019), economics (e.g., Golman et al., 2016), linguistics (e.g., Angouri & Locher, 2012), business (e.g., Boothby et al., 2023), legal theory (e.g., Reynolds, 2020), and sociology (e.g., Wagner-Pacifi & Hall, 2012), recent reviews highlight major gaps in our understanding of disagreement and persistence. For instance, compared to other topics in philosophy, the study of disagreement is “a mere infant” that has focused almost exclusively on disagreement among peers (vs. groups; Frances & Matheson, 2019). In the sociological literature on large-scale opinion dynamics, “basic empirical questions about how to underpin model assumptions [e.g., about how individuals respond to evidence from disagreement] remain unanswered” (Flache et al., 2017). Similarly, relevant work in political science rests on “a rather shaky foundation; there are legitimate differences of opinion—sometimes explicit, often implicit—about what disagreement is” (Klofstad et al., 2013). Underlying this cross-disciplinary uncertainty is a dearth of communication: Studies of disagreement are highly siloed across disciplines, in part due to the absence of a comprehensive model of persistence that can bridge across literatures.

Such gaps in our understanding of disagreement are especially worrying in light of

persistent political, scientific, and moral divisions within the U.S (Jones, 2021; Newport, 2023) and in democracies across the globe (Carothers & O’Donohue, 2019; Levitsky & Ziblatt, 2018). Severe societal disagreements carry drastic consequences for individuals (e.g., partisan discrimination; Iyengar & Westwood, 2015) and states (e.g., loss of trust in democratic institutions; Hetherington & Rudolph, 2015). Developing a principled understanding of persistence can facilitate the design of interventions aimed towards mitigating such harmful consequences.

Even this brief summary reveals that there is a lack of clarity, and much at stake—practically, scientifically, and philosophically—when it comes to our understanding of the psychology of belief persistence. In this paper, we present a framework that aims to further this understanding by distilling findings across disciplines into candidate explanations for persistence, and organizing these explanations into a coherent psychological model.

## Overview

This paper is structured in four parts. In the first part, we define disagreement and situate persistence as one of several possible responses to it. We then consider a Bayesian analysis of belief revision in the face of disagreement. This analysis clarifies how disagreement differs from other forms of contrary evidence, and highlights when simple Bayesian updating can (and when it cannot) explain persistence.

In Part 2, we present a broad taxonomy of drivers of persistence called the ‘Paths to Persistence’ (PTP) model. This model allows us to explain many cases of persistence, including those that our Bayesian analysis cannot account for. This taxonomy is comprised of four primary paths, each of which offers a conceptually distinct basis for persistence supported by prior research.

In Part 3, we build on prior sections to offer an integrated, formal account of persistence. This account leverages meta-reasoning as a mathematical framework to map the paths discussed in Part 2 to persistence, while allowing for interactions across paths.

In the final section, we use our integrated account to address the puzzle of why people typically persist in their contentious beliefs, and identify conditions under which disagreement may cause them to question their views instead. We discuss key theoretical, practical, and normative implications of our model—from the design of belief-change interventions to explanations of societal opinion dynamics—and conclude with open questions about societal disagreement.

## Part 1: Defining and Responding to Disagreement

### *What is Disagreement?*

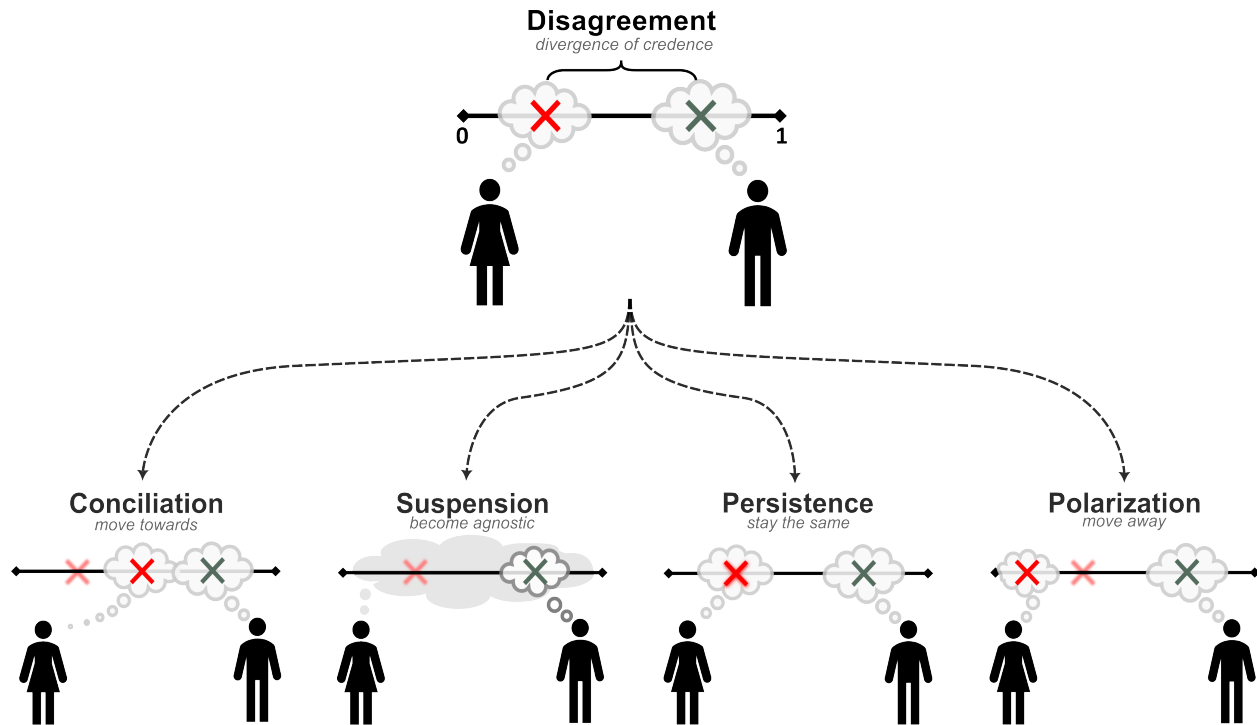
To define disagreement, we turn to philosophy, where epistemologists conceptualize disagreement as the state that obtains when two or more parties have differing beliefs about a proposition (Frances & Matheson, 2019). In Bayesian epistemology, for instance, beliefs are conceptualized as subjective probability assignments called ‘credences’ (Bovens & Hartmann, 2004): If I believe in climate change, that means I assign a high credence to it—say, 80% probability of climate change being real.

This naturally leads to a probabilistic interpretation of disagreement, whereby discrepancies in people’s credences (i.e., subjective probabilities) about the truth of a proposition can characterize how much they disagree. For example, we can define disagreement as a state in which two parties, A and B, do not hold the same credence regarding a proposition  $S$  (i.e.,  $P_A(S) \neq P_B(S)$ ). Using another definition, disagreement can be any state where the difference between the two parties’ credences exceeds some threshold,  $\tau$  (i.e.,  $|P_A(S) - P_B(S)| \geq \tau$ ).

Which of these definitions best corresponds to people’s judgments of disagreement is an open empirical question. We therefore broadly define disagreement as divergence in credences, without committing to a particular measure of divergence. For instance, individuals who meaningfully differ in the probability they assign to the proposition that climate change is real can be said to disagree. Generalizing to group settings, large-scale disagreement can be defined in terms of divergence over the set of every agent’s credences.

*How Could We Respond to Disagreement?*

In principle, an individual can respond to disagreement in one of four ways: conciliation, suspension, persistence, and polarization. As illustrated in Figure 1, conciliation involves moving one's credences towards the disagreeing other; suspension involves withholding judgment on the issue; persistence involves remaining steadfast in one's prior credences; and polarization involves moving one's credences away from the disagreeing other. How an individual responds to disagreement depends on many factors,

**Figure 1***Four Possible Responses to Peer Disagreement*

*Note.* Responses to divergence shown through changes in the left figure's credences.

including the proposition in question, their goals, and their informational and cognitive resources. In this first section, we follow the epistemological literature in assuming that the subject of disagreement is a matter of fact (vs. subjective opinion), that the individual's goal is to have the most accurate beliefs, and that they have the information and capacity

to update their beliefs (Matheson, 2015). Despite sharing these assumptions, epistemologists disagree about the appropriate response to peer disagreement, with some advocating for persistence (Kelly, 2005), others for conciliation (Christensen, 2007), and others for suspension (Feldman, 2007). Recent work additionally argues that polarization can be a rational response to some contradictory evidence as well (Dorst, 2023; Jern et al., 2014). One insight emerging from this literature is that disagreement offers what philosophers call higher-order evidence.

Unlike first-order evidence—which bears directly on the truth of propositions—disagreement can offer evidence about evidential relations (Christensen, 2010; Feldman, 2009; Kelly, 2005). Suppose you try to mentally calculate  $32 \times 47$  and find 429. After performing the calculation, you note that you are extremely fatigued and thus prone to making mistakes. You should grow less confident in your calculation, not because fatigue bears on mathematical truth, but because the mental evidence you generated is less reliably related to mathematical truths than it would be normally. Similarly, learning that there exists disagreement regarding some issue does not offer direct (first-order) evidence regarding that issue, but it does offer higher-order evidence that the mechanisms generating beliefs may not be as reliable as we thought, and so we might have reason to reduce our own confidence or change our beliefs.

Importantly, this literature from epistemology focuses on cases of ‘peer disagreement,’ and thus does not consider two factors key to the psychology of disagreement: First, disagreements can involve asymmetries in the evidence or expertise available to each side; and second, people disagree not just with individuals, but also with groups. In the next section, we propose a formal, Bayesian framing that can take these considerations into account, and that specifies how one should respond to disagreement from a purely epistemic standpoint.

### *A Bayesian Response to Disagreement*

Consider  $S$ , a proposition to which an agent assigns some prior probability (denoted  $P(S)$ ). If the agent obtains evidence of disagreement,  $D$  (e.g., an informant asserting a divergent credence with regard to  $S$ ), Bayes' rule prescribes the following belief update:

$$P(S|D) = \frac{P(D|S)P(S)}{P(D|S)P(S) + P(D|\neg S)P(\neg S)}. \quad (1)$$

Any change in the updated belief (denoted  $P(S|D)$ ) from the prior belief reflects learning from disagreement, and persistence can be cast as the absence of learning. Note that this inference depends on two considerations: the strength of the prior belief, and evaluations of the likelihood terms ( $P(D|S)$  and  $P(D|\neg S)$ ), which intuitively capture the reliability of the informant concerning the truth (or falsity) of  $S$ . For example, a doctor is generally reliable concerning the cause of some symptom (and so will have a large impact on belief revision); a five-year-old generally less so (and so will have a correspondingly smaller impact on belief revision). More precisely, a doctor would be likely to correctly identify the cause of a symptom, so the probability that they would tell you that you have a disease would be high when you actually have a disease, and low when you do not (so  $P(D|S)$  is high and  $P(D|\neg S)$  is low); a child would be uninformative (so  $P(D|S)$  and  $P(D|\neg S)$  would be relatively similar, leading to a small update). Past work has developed Bayesian models of such inferences, and found that people can use their previous knowledge to jointly infer who is reliable, and how much to update one's beliefs, upon receiving testimony (Alister et al., 2023; Shafto et al., 2012).

Next, consider what happens when we encounter evidence of disagreement from multiple sources. This requires an expanded version of our update equation. Instead of simply denoting disagreement as  $D$ , we can break it up into the credences of the disagreeing individuals in a group of  $N$  people, denoted  $C_1, C_2, \dots, C_N$ :

$$P(S|C_1, \dots, C_N) = \frac{P(C_1|C_2, \dots, C_N, S)(\dots)P(C_N|S)P(S)}{P(C_1, \dots, C_N|S)P(S) + P(C_1, \dots, C_N|\neg S)P(\neg S)}. \quad (2)$$

Where  $(\dots)$  is used to abridge the remainder of the chain rule expansion. Now, in



addition to reliability, we also have to consider how informative people’s credences are of each other—in other words, how dependent they are as informants (i.e., we condition  $C_1$  on  $C_2, \dots, C_N$ ). For instance, if you encounter ten independent sources who disagree with you about the safety of GMOs, that should be more persuasive than ten sources who all derive their opinions from the same news source. Research has investigated such inferences of shared information across informants using Bayesian models as well (Enke & Zimmermann, 2019; Whalen et al., 2018).

### *Can a Simple Bayesian Analysis Explain Persistence?*

Though our simple Bayesian analysis is useful for highlighting the distinctive computations that disagreement entails, it can only explain persistence under two specific conditions: First, if prior beliefs are extreme (i.e., definitely false,  $P(S) = 0$ , or definitely true,  $P(S) = 1$ ), and second, if the evidence is seen as entirely uninformative (i.e., if  $P(D|S) = 1$  &  $P(D|\neg S) = 1$ ). Otherwise, there would be some update.<sup>1</sup> Indeed, differing priors provably erode to convergence for Bayesian agents that perfectly communicate their posteriors under a variety of conditions (as described by the ‘no-disagreement’ theorem; Aumann, 1976).

The conditions that lead to persistence are highly restrictive: Extreme priors imply that beliefs will never be revised—not just in the presence of disagreement, but in the face of any evidence whatsoever. Uninformative likelihoods, on the other hand, imply that an agent could have any prior belief about  $S$ , and would never learn anything from the evidence in question.

Is it plausible that real-world cases of persistence can be characterized in terms of these conditions? While the conditions may apply on occasion, they are likely the exception rather than the rule. Consider a controversy, such as whether the U.S government should conduct raids on workplaces to detain illegal immigrants. People seem

---

<sup>1</sup> Though we could think of small updates as persistence, Part 3 clarifies why a no-update definition is distinctively useful in accounting for the phenomenon of persistent societal dissent.

to persist in such cases while retaining some uncertainty in their views, and without assuming that *all* disagreeing others are *entirely* uninformative (Kalla & Broockman, 2020). Capturing such real-world cases of persistence seems to require something beyond the simple Bayesian analysis we have offered in this section.

Despite their limitations, however, the epistemological and Bayesian analyses are useful in highlighting three ways in which psychological responses to disagreement plausibly differ from responses to more canonical forms of contrary evidence (e.g., scientific data refuting beliefs about the health benefits of alcohol; Kappes et al., 2020; Nickerson, 1998). First, responding to disagreement involves distinct evaluations of reliability. Whereas evaluating first-order evidence (e.g., how diagnostic a medical test is of some illness) requires expertise in the subject matter, evaluating higher-order evidence from disagreement requires judging the relative epistemic standing of disagreeing others (e.g., how diagnostic a physician’s opinion is of some illness, Shanteau, 2015; see also Harris et al., 2018; Plunkett et al., 2020). Second, whereas evaluating dependence across pieces of first-order evidence requires pattern detection or causal inference (Penn & Povinelli, 2007), evaluating dependence amid disagreement requires inferring social and informational relations across informants, for which people utilize distinct cognitive strategies (Son et al., 2021; see also Desai et al., 2022; Yousif et al., 2019). Finally, and as mentioned already, disagreement offers higher-order evidence. Higher-order evidence can have a broader disconfirmatory reach than first-order evidence (Whiting, 2020). Consider the previous example of trying to do mental math while tired. Fatigue should not only lower your confidence in your calculation but also other judgments made during that time. Similarly, higher-order evidence from disagreement can call into question your expertise in a domain, or even your overall capacity for reasoning, undermining self-trust. Evidence from disagreement is therefore cognitively risky (e.g., consider gaslighting: Spear, 2019, or conservative responses to unfamiliar advisors: Soll & Larrick, 2009).

These three features of disagreement—the nature of reliability, dependence, and

higher-order evidence—introduce additional reasons for why accounts of belief revision developed for first-order evidence (including simple Bayesian models) will not readily capture real-world cases of our target phenomenon: belief persistence amid disagreement.

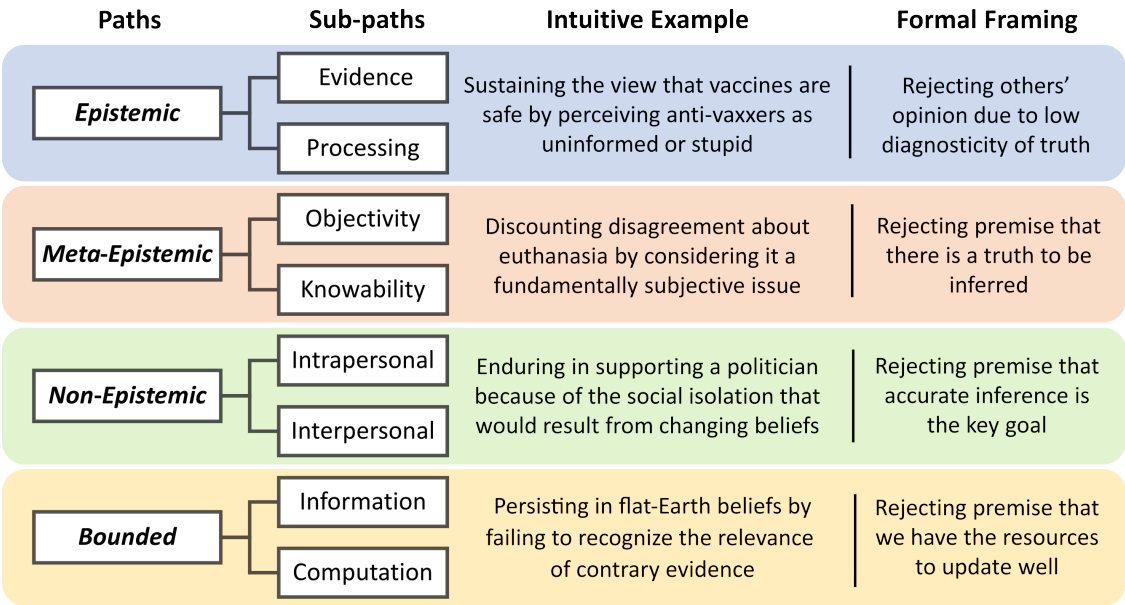
### ***Summary***

In this section, we defined disagreement as divergent credences, outlined four responses to disagreement (conciliation, suspension, persistence, and polarization), and introduced a Bayesian analysis of updating beliefs given disagreement. Our discussion highlights four insights: Disagreement (i) is a form of higher-order evidence that requires assessing (ii) the reliability and (iii) dependence of informants; and (iv) straightforward Bayesian updating is unlikely to capture all cases of persistence.

## **Part 2: The Paths to Persistence Model**

Why does our Bayesian analysis fail to capture some cases of persistence, such as the politics of immigrant prosecution? It fails in part because it is limited by three strong assumptions we borrowed from the epistemology of disagreement: That the subject of disagreement is an objective fact, that the individual’s goal is maximizing accuracy, and that the individual has the informational and cognitive resources to support accurate inferences on the basis of others’ beliefs. These assumptions do not generally hold in real-life cases of disagreement, and a psychologically realistic model of persistence should be able to accommodate persistence in cases where these assumptions are violated. In this section, we present four primary ‘paths’ that offer theoretically distinct—and potentially interacting—routes to maintaining beliefs in the face of controversy (see Figure 2). Only the first of these, the epistemic path, preserves the assumptions that motivated our simple Bayesian analysis. Below we offer a brief introduction to all paths before considering each in more detail. The epistemic path captures how belief persistence can result from considerations such as the quality of evidence or competence attributed to those who disagree. For instance, an individual might persist in her belief that the moon landing was staged despite disagreement if she believes that everyone else is entirely misled by

**Figure 2**  
*A Taxonomy of Four Paths to Persistence*



*Note.* Each branch in the taxonomy represents a distinct explanation for belief persistence. Importantly, the paths are not mutually exclusive—a given instance of belief persistence can involve multiple paths acting simultaneously. The formal framing for Bounded refers to both informational and cognitive resources necessary to perform accurate inference, as described below.

unreliable sources.

The meta-epistemic path captures how people can persist if they don't see an issue as having an underlying 'truth' for people to converge on. For instance, an individual might persist in her belief that euthanasia is morally permissible because she regards this proposition as fundamentally subjective. Within our Bayesian analysis, this can be seen as a rejection of a core premise: that the disagreement concerns some proposition about which there is some 'truth' that can be discerned.

The non-epistemic path captures how beliefs can persist for reasons other than their presumed truth, such as their personal or social value. For instance, a fervent supporter of a politician accused of crimes might not be swayed by disagreeing strangers. This may occur because the belief is held out of loyalty or social pressures, not because the believer

aims to maximize the probability of being correct. Within our Bayesian analysis, this can be seen as a rejection of the goal of maximizing accuracy.

Finally, the bounded path captures the possibility that belief persistence may originate from a failure to recognize disagreement or fully process its implications. For instance, an anti-vaxxer may see a poll on the news that indicates societal dissent, but switch to a different channel without reasoning about its implications. This can be seen as a rejection of the premise that the update described by our Bayesian analysis will be computed accurately.

In the rest of this section, we describe these intertwined paths individually; giving concrete examples for each, reviewing relevant evidence, and identifying open questions.

### **The Epistemic Path to Belief Persistence**

*Theo buys organic produce for his family because he believes that genetically modified foods are less healthy for human consumption. Theo is aware of the GMO controversy and has spent time carefully researching it. He is not bothered by disagreement over this issue because he considers himself to be a smart, informed consumer, unlike those who disagree—they either do not care enough to seek the facts, or are dumb enough to be misled by the same corporate lies.*

Theo's case illustrates how epistemic explanations can sustain controversial beliefs. He is aware of the deep disagreement over GMOs (Pew Research Center, 2016), but thinks that disagreeing others are less reliable than he is at tracking scientific truths. He thus does not update his views based on their credences. We organize our discussion of such epistemic explanations around inferences of others' *evidence* and *processing*.

### ***Epistemic: Inferences of Inferior Evidence***

As Theo's case illustrates, attributions of inferior evidence can push us towards persistence. Research in social and developmental psychology has shown that people readily attribute evidential inferiority when evaluating disagreement.

The literature on naïve realism (people's tendency to assume that their own

perceptions reflect reality as it is) has documented that people often judge disagreeing others as ignorant (Robinson et al., 1995; Ross & Ward, 1996), with larger disagreements leading to inferences that others have correspondingly worse evidence (Pronin et al., 2004).<sup>2</sup> Children also judge disagreeing individuals as uninformed and ignorant in some domains of disagreement (e.g., in moral disagreements, but not cultural ones, Wainryb et al., 2001; see also Aboody et al., 2022). Such inferences are not necessarily unfounded—selectivity in learning is a basic component of our social reasoning toolkit, and undergirds epistemic vigilance (Koenig & Harris, 2005; Sperber et al., 2010). Accordingly, novices conciliate towards the opinions of trusted experts because they are thought to possess more (and better) evidence in their areas of expertise (Kruglanski et al., 2005).

An important nuance is that perceiving others’ evidence as merely inferior need not, on its own, be sufficient to drive persistence. If others have inferior evidence that is at least somewhat diagnostic, and if others’ evidence is not subsumed by one’s own, then some (minor) belief revision could still be appropriate. Others’ evidence therefore has to be perceived as entirely undiagnostic, as incompatible with one’s own evidence, or as a subset of one’s own evidence to trigger persistence.

Much work has documented attributions of inferior evidence to disagreeing individuals. To our knowledge, there has been no direct work on whether people make biased evidential inferences about disagreeing *groups* (though people are overconfident about many judgments; Moore & Healy, 2008). Note that attributing evidential inferiority to a group requires taking a strong stance: That the disagreeing group, as a whole, has access to worse evidence compared to oneself (or the agreeing group). One approach to justifying these attributions is through perceptions of dependency. If disagreeing others are

---

<sup>2</sup> We note that the vast majority of the empirical evidence in this paper comes from studies conducted in the U.S, and should not be assumed to generalize to all people (Henrich et al., 2010). For the sake of conciseness and readability, we will use the term ‘people’ when referring to the results of studies, but readers should keep in mind that extensive cross-cultural work is needed to investigate the generalizability of these results.

perceived as receiving their information from the same source, for example, their informativeness would be reduced to that one source. Given that  $\sim 87\%$  of both Republicans and Democrats perceive each other to be “brainwashed” (Yudkin et al., 2019), that people consider others to be more easily persuaded by mass media than themselves (Duck & Mullin, 1995; Sun et al., 2008), and that outgroups are perceived to be highly homogenous (Quattrone & Jones, 1980; Rubin & Badea, 2012), inferences of dependency can plausibly justify attributions of inferior evidence to entire groups.

***Epistemic: Inferences of Inferior Processing***

People are sensitive to asymmetries in processing as well. For instance, children learn from their parents not just because adults have better evidence about the world, but also because adults are more competent—that is, more likely to make the right inference given the same evidence (Harris, 2012). Adults, on the other hand, persist on the basis of their greater knowledge and competence. We can break down such inferences into two varieties: attributions of intellectual inferiority and attributions of bias.

**Inferences of Intellectual Inferiority.** Anecdotally, people often denigrate disagreeing groups, labelling them “childish, stupid people” or claiming that they are “ignorant, stupid, or insane” (quotes from opinion pieces on controversies; Cunningham, 2021; Dawkins, 1989). There is little work in psychology on such inferences. As Hartman et al. (2022) point out in a recent paper, “the only investigations (...) [of political attributions of] unintelligence were conducted by polling organizations.” Their results echo the findings of these polls: Partisans are likely to view each other as unintelligent (a third agree with such attributions; Pew Research Center, 2019). Developmental studies suggest that children often make domain-dependent attributions of ‘unintelligence’ in response to disagreement as well (Wainryb et al., 2004).

**Inferences of Bias.** Much research in social psychology has shown that people consider disagreeing others to be more biased (Kennedy & Pronin, 2008), influenced by self-interest (Reeder et al., 2005), unfair (Frantz, 2006), and influenced by group pressures

(Cohen, 2003) than themselves. While these uncharitable inferences are not identical, they function similarly in the context of disagreement: If an informant is perceived to be biased, their beliefs should carry less epistemic weight.

One way people may justify the intellectual inferiority (or greater bias) of entire groups is through essentialist stereotypes (Gelman, 2004). For instance, ethnographies show that some flat-earthers justify their epistemic superiority over disagreeing others by placing them into a social category ('sheeple' or 'globies') and stereotyping them as intellectually inferior, "blind and uncritically obedient" (Toseland, 2019). Similarly, opposing partisans hold many stereotypes of one another, likely including unintelligence (Judd & Park, 1993; Rothschild et al., 2019).

### **The Meta-Epistemic Path to Belief Persistence**

*Brandon loves eating meat, and believes that it is morally okay to do so. He is aware that many vegetarians disagree with him—and he respects their personal preference. Yet their views do not influence his: To Brandon, there are no right or wrong answers to moral questions, just subjective opinions. And even if there were some universal moral code that establishes the 'truth' about the morality of eating meat, he's convinced that no one knows what it is, anyway.*

Brandon persists in his view about the morality of eating meat based on issue-level, meta-epistemic inferences—such as the impossibility of identifying shared moral 'truths.' This impossibility can result from either the truth of a statement being fundamentally agent-relative (i.e., subjective), inaccessible (i.e., unknowable), or both. These explanations are meta-epistemic because they determine whether epistemic considerations are relevant to a given question: If there is no shared truth to be established about an issue, then others' opinions are irrelevant to their truth.

### ***Meta-epistemic: Inferences of Subjectivity***

Whereas beliefs and decisions in some domains are perceived as objective (e.g., medicine and mathematics), other domains are seen as subjective (e.g., fashion and



romance; Kuhn et al., 2000). There is a deep connection between such subjectivity and disagreement captured by the Latin adage *de gustibus, non est disputandum* ('in matters of taste, there can be no disputes'). This is because aggregate 'truths' about subjective issues are ill-defined—there is no such thing as the best song for everyone, for instance, but there may be a best treatment for an illness (Kivy, 2015). To the extent that an issue is considered subjective, others' opinions thus become epistemically irrelevant to our beliefs (Egan, 2010). Note how perceptions play a key role here: Meta-epistemic persistence does not require a statement to in fact be subjective—it merely requires people to think that it is. Past research has shown that such perceptions of subjectivity are consequential, leading to more positively-valenced judgments in cases of moral disagreement (Sarkissian et al., 2011).

But what underlies perceptions of subjectivity? One factor is the presence of disagreement itself. Goodwin and Darley (2012) found that presenting participants with evidence that many others disagree with them decreased perceptions of the objectivity of moral claims (e.g., whether downloading a TV program in violation of copyright laws is immoral). Similarly, greater perceived consensus regarding the moral status of a claim predicted greater perceived objectivity (for similar findings about non-moral claims, see Ayars & Nichols, 2020; Heiphetz & Young, 2017). Disagreement can thus lead to inferences of subjectivity, which in turn allow individuals to persist amid said disagreement, resulting in entrenched cleavages of opinion.

In contrast to epistemic persistence, subjectivity-based persistence does not depend on negative inferences about the disagreeing party. Perhaps for this reason, 'ice-breakers'—introductory activities that establish rapport—often rely on sharing of preferences (Chlup & Collins, 2010). However, not all meta-epistemically justified beliefs are merely ice-breaker material.

*Meta-epistemic: Inferences of Unknowability*

Does hell exist? Some domains, such as religion, raise important questions that many expect to be beyond human understanding; others, such as science, raise questions that we expect to have discoverable answers (e.g., whether the moon causes tides; Davoodi & Lombrozo, 2022b; Liquin et al., 2020). If people expect the truth of a statement to be fundamentally unknowable, they may persist in their beliefs amid disagreement without assuming that others have weaker epistemic standing (since no one’s judgment on the issue is informative).

Recent work has found that people have systematic beliefs about what is knowable, and by what means. For instance, Gottlieb and Lombrozo (2018) found that participants judged some psychological phenomena (such as conscious experience and belief in God) as more likely than other phenomena (such as depth perception) to fall beyond the scope of what science can explain. Kominsky et al. (2016) observed that children and adults gravitate towards informants who show ‘virtuous ignorance’—that is, acknowledging ignorance about unknowable matters, such as the number of blades of grass in New York (see also Heiphetz et al., 2021; Johnson et al., 2016).

More direct evidence pertaining to unknowability judgments comes from recent studies on paradoxical knowledge, where people recognize something as unknowable, but claim to know it nonetheless. Gollwitzer and Oettingen (2019) found that paradoxical knowledge is commonplace across domains (with more than 90% of participants endorsing at least one claim similar to the following participant-generated example: “I know that there is no God... I know this, even though it is unknowable”), is particularly prevalent for goal-relevant beliefs, and is associated with a willingness to join and adhere to extreme groups. Relatedly, research on conspiracy theories has identified widespread incoherence in conspiratorial beliefs that is accompanied by paradoxical inferences of unknowability (e.g., that climate change cannot be predicted, but that we are heading into an ice age; Lewandowsky et al., 2018, see also; Wood et al., 2012); and people can inject

unknowability into their construal of key political and religious issues when facts threaten pre-existing worldviews (Friesen et al., 2015).

In sum, persistence can result from meta-epistemic inferences: If there is no accessible truth about an issue, either because it is subjective or unknowable, there is no reason to update one’s beliefs due to disagreement.

### **The Non-Epistemic Path to Belief Persistence**

*Matt works at a rifle store in Texas and often discusses gun laws with his family. He shares their belief that gun laws in the U.S. are too restrictive and owns an impressive collection of munitions at home. Moreover, his belief in the unrestricted right to bear arms grounds much of his understanding of what it means to be an American, a Republican, and a proud Texan.*

Matt’s case illustrates the non-epistemic route to persistence: Changing his views could cost him his job and alienate him from his loved ones, in addition to jeopardizing his larger worldview and sense of self. We cluster these non-epistemic values of beliefs into two categories: inter- and intrapersonal.

### ***Non-epistemic: Interpersonal Drivers of Belief***

Beliefs play a profound role in our social lives. Having the wrong beliefs in the wrong context can get you shunned, exiled, or executed (Poliakov, 2003). Historically, clashes between groups with different sets of beliefs have driven much animosity, war, and bloodshed (Golman et al., 2016)—and even today, much armed conflict in the world arises over differences in beliefs (Svensson, 2013).

Beliefs are consequential in part due to their social function as signals of group affiliation (Golman, 2023; Golman et al., 2016). Signaling the right affiliations by curating group-congruent beliefs can allow people to reap the benefits of social integration (Thoits, 2011), while avoiding the costs of social exclusion (Roberts et al., 2021). Accordingly, people form beliefs on novel issues that align with those of their in-group (Kahan, 2010), and infer that out-groups have beliefs that differ from their own (Dion, 2003).

Foundational studies in social psychology, such as Sherif’s experiments in group conflict (1956), demonstrate the strength of these pressures: even groups that are formed randomly and arbitrarily can generate prejudice and discrimination.

Alternatively, people may privately conciliate in response to encountering disagreement with the out-group, but choose not to express divergent beliefs to their in-group (Noelle-Neumann, 1977). That 62% of Americans today say they have political beliefs they are afraid to share (The Cato Institute, 2020), and recent evidence that partisans “parrot the party line, but do not vote it” (Lenz, 2013) support this idea. However, such dissonant beliefs may erode over time (Harmon-Jones & Mills, 2019), in part due to the difficulty of sustained deception (Hippel & Trivers, 2011; Schwardmann & Van der Weele, 2019), and in part due to a preference for expressing authentic beliefs (Brown et al., 2022; Erickson, 1995; Oktar & Lombrozo, 2022).

In sum, beliefs have important social consequences—such as exclusion and prejudice—that drive people to maintain group-consistent beliefs. Holding a belief for such interpersonal reasons, rather than having the primary goal of holding the most accurate belief, can support persistence.

### ***Non-epistemic: Intrapersonal Drivers of Belief***

Dissent can induce uncertainty and ambiguity, which complicate decision-making. Accordingly, people may persist amid dissent to preserve decision-promoting beliefs (Kagan, 1972; Kruglanski, 2004): I may have to decide whether to vaccinate on a given date, and it may be inefficient for me to debate the pros and cons endlessly, as opposed to committing to a course of action. Relatedly, self-esteem facilitates decision-making and the pursuit of long-term goals (Bandura, 2010). Given that disagreement can lower confidence (Pool et al., 1998), people may also persist in their beliefs to protect their self-esteem (Cohen et al., 2000): If I am a staunch pro-vaccine advocate, doubting my stance on vaccines could lead me to doubt my capacity to form robust beliefs on key issues, reducing my self-esteem.

Beyond providing value by guiding decisions, mounting evidence suggests that

beliefs are a source of value in and of themselves—that is, beliefs are not merely a means to an end (in service of decision-making or signaling), but also directly confer utility (Bénabou & Tirole, 2016; Bromberg-Martin & Sharot, 2020). For instance, beliefs have affective consequences: religious belief can buffer against existential anxiety (Norenzayan, 2013), and just-world beliefs promote a sense of safety and happiness (Hafer & Sutton, 2016); similar examples abound (Abramson et al., 1989; Altay et al., 2023; Davoodi & Lombrozo, 2022a; Hafer & Sutton, 2016; Molnar & Loewenstein, 2020). When beliefs are held for these reasons, epistemic evidence from disagreement may not be relevant. Relatedly, theories of persuasion underscore the functional role of emotions in guiding attitude change: For instance, if messages can associate negative affect with one’s existing beliefs, attitude change is more likely—though the relationship between affect and attitude change is moderated by the extent of elaboration (Petty & Briñol, 2015).

In sum, both inter- and intrapersonal benefits of belief can drive persistence. Such ‘returns’ provided by a belief can consciously or subconsciously guide people’s likelihood of persisting in that belief, in line with the literature on motivated reasoning (Cusimano & Lombrozo, 2023; Epley & Gilovich, 2016; Kunda, 1990), and with influential work on the functional approach to the study of attitudes (Katz, 1960; Shavitt, 1989).

### **The Bounded Path to Persistence**

*Lisa believes that many vaccines cause infertility. She has no real expertise in biology, and could not really articulate a plausible mechanism that would explain how vaccines might cause infertility—but she is unaware of how shallow her understanding is. She was recently channel-surfing when a poll on vaccination beliefs flickered on her television. She skipped it without paying much attention; there were better things to watch and she was already convinced that many people agree with her.*

Lisa’s case illustrates how constraints on our understanding, reasoning, and information can drive persistence. Lisa (falsely) believes that many support her views, and when she encounters evidence to the contrary, she prioritizes other tasks or information

instead of revising her views based on the disagreement. Moreover, even if Lisa tried to update her beliefs about vaccination, her lack of understanding would pose challenges for how she ought to revise her views. We can categorize such limitations as constraints on our information and computation.

***Bounded Persistence: Informational Constraints***

Appropriately responding to societal disagreement first requires accurately representing the presence and properties of disagreement. However, there is a large body of work that suggests that people frequently “operate within a ‘false’ social world” (Fields & Schuman, 1976) and drastically misestimate the content and distribution of others’ views and beliefs. At times, people can even mistake the minority belief for the majority, and vice versa (Shamir & Shamir, 1997). Such erroneous inferences of collective belief can shape behavior by miscalibrating descriptive and prescriptive norms (Miller & Prentice, 1994). For instance, underestimating scientific consensus can drive rejection of anthropogenic climate change (Lewandowsky et al., 2022).

Importantly, our inferences of others’ views are not merely erroneous, but also systematically biased: Ross et al. (1977)’s work on the false-consensus effect showed that people typically think their own beliefs are more common than alternatives. If someone believes that it is okay to litter, for instance, they are likely to believe that most other people think littering is fine, too (see also Mullen et al., 1985). Relatedly, people tend to overweight the informativeness of local samples in drawing inferences about global environments (Broomell, 2020)—so the perception of local agreement can lead to the inference that there is less global disagreement than there really is. People may be especially prone to underestimating the extent of disagreement with their own position—perhaps to the point of entirely discounting it.

The argument here is that people may persist due to systematically biased inferences about others’ beliefs that understates disagreement. However, people can also exaggerate disagreements. For example, Americans think that more than half of opposing

party members hold extreme views, when in reality less than a third do (Westfall et al., 2015; Yudkin et al., 2019). Such exaggerations likely occur for controversies that loom large in public discourse and media, such as global warming (McCombs & Valenzuela, 2020). Thus, though unawareness due to informational constraints can account for some cases, persistence on highly publicized issues may stem from other limitations.

### ***Bounded Persistence: Computational Constraints***

When Lisa briefly saw the vaccination poll on her television, did she accurately update all of her relevant beliefs in light of this information? Likely not. Lisa, like all of us, faces significant constraints on her cognition, and may persist in her belief by failing to update her relevant beliefs due to these constraints—even if she correctly infers how many Americans disagree.

**Limited Cognitive Resources.** Human cognition is shaped by significant constraints on time, computation, and communication (Griffiths, 2020). An important consequence of these limitations is that few beliefs enjoy extensive elaboration—most information is processed in a qualitatively shallow way, ‘going in one ear and out the other’ ( Craik & Lockhart, 1972; Pennycook & Rand, 2019), as attention and computation have to be deployed strategically (Lieder & Griffiths, 2020). Moreover, people are often motivated to avoid information that would require extensive elaboration to evaluate (Evans & Stanovich, 2013). Combined with the increased taxation of attentional resources in modern society (Hemp, 2009), the need to strategically deploy limited computation can precipitate persistence.

**Narrow Understanding.** An adjacent possibility is that updating may fail in a broader sense; not due to attentional limitations or shallow processing, but rather due to a lack of coherent conceptual structure supporting belief in the first place (Chater, 2018). For example, non-physicists may believe that gravity exists but find themselves grasping at straws if asked to defend or explain it (as seen in the illusion of explanatory depth; Rozenblit & Keil, 2002). Relatedly, theories of communal knowledge representation

emphasize that most lay beliefs are stored as propositional pointers to sources of further information, such as experts (Rabb et al., 2019). Such shallowness poses a challenge for computations on beliefs, as generalizing inferences from disagreement ideally involves direct access to a comprehensive representation of the relevant beliefs.

**Messy Beliefs.** Mental representations may be redundant or inconsistent (Bendana & Mandelbaum, 2021; Sommer et al., 2023), rather than perfectly integrated (e.g., in a Bayesian network). We are often unaware of these inconsistencies—we believe that we have robust beliefs about politics, for instance, but our beliefs can be swayed by irrelevant factors.

For example, in a seminal paper on public opinion measurement, Hyman and Sheatsley (1950) found that 37% of Americans supported Communist reporters being able to freely visit the U.S and report on their experiences, but that the fraction nearly doubled to 73% when the question was preceded by a question asking whether Russia should let American reporters in (Klein et al., 2014) and similar order and framing effects have been found for other issues (though there is much variation in the size of framing effects; Tourangeau & Rasinski, 1988). Accordingly, public opinion scholars have long considered stable and consistent views to be the exception rather than the rule (Converse, 1964). As Zaller summarized, “for most people, most of the time, there is no need to reconcile or even to recognize their contradictory reactions to events and issues (...) individuals do not typically possess ‘true attitudes’ on issues (...) but a series of partially independent and often inconsistent ones” (1992). Updating messy beliefs may be particularly difficult if it entails the recall and reconciliation of inconsistent fragments to construct coherent beliefs prior to updating.

### **Interactions Across Paths**

We have thus far considered the four paths independently. In reality, there is good reason to think that they often interact; dynamically reinforcing or substituting for each other in generating persistence—and one of the primary affordances of the PTP framework



is its ability to generate predictions about what such interactions might look like. Here, we identify three pairwise interactions that emerge from the first three paths, and highlight relevant empirical evidence (see Figure 3 for illustrations).

### ***Epistemic and Meta-epistemic***

Paths can moderate each other in motivating persistence. Perhaps the most salient predicted interaction in the PTP is that between the epistemic and the meta-epistemic: the meta-epistemic path is named as such *because* subjectivity and knowability determine whether epistemic considerations are relevant. This predicts the following interaction: For objective and knowable issues, others' reliability should have a large effect on persistence; for subjective or unknowable issues, this effect should be attenuated or eliminated. Consistent with this prediction, Cheek et al. (2021) find evidence for an interaction along these lines: highlighting the subjectivity of judgments in a domain (e.g., art)—hence enabling meta-epistemic persistence—can reduce the extent to which participants judge those who disagree with them as biased (an epistemic inference). Similarly, Wainryb et al. (2001) find that children are less likely to judge disagreeing others as uninformed or unintelligent in more subjective domains (e.g., cultural disagreements). These examples involve pairwise disagreements (rather than group-level or societal disagreements), but they nonetheless illustrate the predicted interaction and offer tentative support for the more global claim.

### ***Epistemic and Non-epistemic***

Paths can also be synergistic. For instance, Mercier and Sperber (2017) argue that reasoning is primarily a vehicle for social engagement, enabling justification in the interest of persuasion and power (see also Haidt, 2001; Tetlock, 2002). From this standpoint, reasons galvanize one's pursuit of practically or socially convenient beliefs. This highlights a second predicted interaction in the PTP: Because epistemic reasons can provide convincing arguments for sustaining belief, we may expect persistence to be especially likely when both epistemic and non-epistemic paths are available (i.e., it is both practically useful and

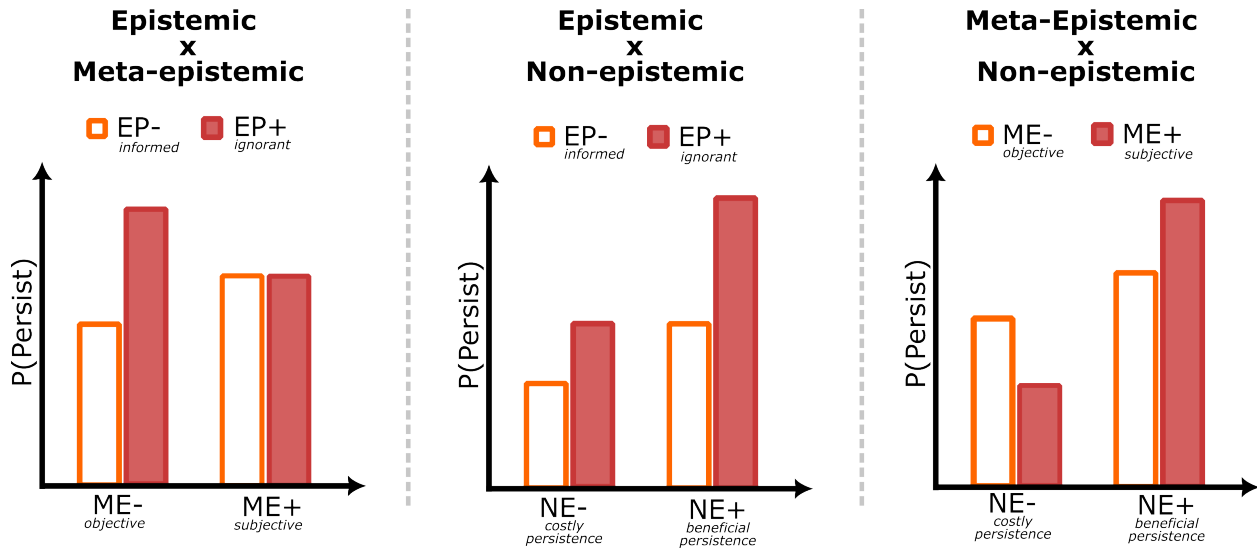
justifiable to persist). On the flip-side, questioning may be especially likely when neither path is available. For instance, van Prooijen and Sparks (2014) find that anti-climate change attitudes changed substantially more following an epistemic intervention when it was coupled with a self-affirmation exercise that presumably attenuated the non-epistemic consequences of view change. The PTP predicts that this interactive relationship should generalize to epistemic interventions that pertain to perceptions of disagreeing others.

### ***Meta-epistemic and Non-epistemic***

Finally, interactions can cause reversals. For example, though subjectivity is often a path to persistence, it can also enable questioning when it is socially desirable to question and conform (e.g., consider the rapid adoption of aesthetic fads; Bikhchandani et al., 1998). This prediction falls out of a similar logic to the previous one: It is harder to appear competent if you frequently change your mind in arguments about ostensibly objective issues; but subjectivity offers a route to conciliating without social costs. Accordingly, early work comparing conformity on objective vs. subjective statements in small-group discussions with *peers* found that subjectivity led to higher *conformity* (i.e., lower likelihood of persistence; Blake et al., 1957). Similar work on *anonymous*, online interactions, however, finds that subjectivity drives *persistence* (Wijenayake et al., 2020), plausibly due to the absence of interpersonal consequences to dissent. The PTP framework thus leads to the prediction that whether meta-epistemic factors increase or decrease persistence depends on the non-epistemic context.

### ***More Complex Dependencies***

Beyond the relatively simple pairwise interactions we discussed here, there are likely more complex dependencies across paths that guide persistence. Most obviously, higher-order interactions (e.g., Epistemic x Meta-epistemic x Non-epistemic) may have an influence, and considerations from the bounded path could influence whether and how other paths are pursued. Considering reasoning about controversy as a sequential process reveals more subtle dependencies as well: Initial evaluations of one path may influence subsequent

**Figure 3***Illustration of Plausible Pairwise Interactions in the PTP*

*Note.* The figure illustrates three plausible interaction patterns discussed in the text. EP denotes Epistemic, ME denotes Meta-epistemic, and NE denotes non-epistemic; (+) indicates that the path is available, (-) indicates that it is not available. Whether and how the paths interact is a largely open empirical question; here we present predicted relationships.

evaluations of other paths, resulting in complex, time-dependent interactions. Further complicating the picture are possible interactions across specific sub-paths (e.g., inferences of bias x unknowability). The pairwise interactions described above are thus meant to illustrate the possibility of rich dependencies, rather than offer a comprehensive analysis.

### ***Ambiguous Cases and Assumptions***

The examples presented so far were selected to support unambiguous classification within a single path, but many real-life instances of persistence are likely to be underspecified or hard to sort. For example, you may persist upon disagreeing with a disliked other for no reason beyond your dislike—we can impute various non-epistemic causes to understand such persistence, but it is plausible that it may be an unreflective response or otherwise indecipherable (but see Minson & Dorison, 2022).

Moreover, disagreement can sometimes persist without people persisting *per se*. This may happen due to biased informational networks (e.g., information silos) that drive persistent opinion dynamics at the societal level (Cinelli et al., 2021; Dinas, 2014), or in highly insulated communities where dissent may be masked, such as cults (Singer & Lalich, 1995). We do not focus on external network effects here, as the paths to persistence outline individual psychological mechanisms.

Similarly, dissent may persist amid referential ambiguity, as studied in the pragmatics of disagreement (Angouri & Locher, 2012; Shields, 2021). For example, if a Republican believes that the morality of abortion boils down to whether fetuses have souls, and thinks that Democrats consider the key question to be about women’s rights instead, he may ignore dissent on abortion because he takes Democrats to be engaging with a different question. Persistence due to communicative, pragmatic inferences such as these fall outside the scope of our present analysis.

Finally, beliefs are typically sensitive to a mixture of first-order evidence (in the form of reasons, evidence, and arguments) and higher-order evidence from disagreement (Hedden & Dorst, 2022). Foundational work has analyzed phenomena such as persuasion and polarization by highlighting the contributions of these two components (Isenberg, 1986; Myers & Lamm, 1976; Petty & Cacioppo, 1986). Here, we focus on clarifying the psychological mechanisms underlying the latter.

We take these points—network effects, pragmatics, and mixtures of evidence—to constitute fruitful next steps for research on persistence.

And with these clarifications, our exposition of the four paths is complete.

### **Part 3: Putting the Paths Together**

As Theo, Brandon, Matt, and Lisa’s cases illustrate, each path can offer a sufficient basis for persistence. This gives us some insight into the prevalence of persistence: For most important real-world controversies, it is exceedingly likely that at least one path will be available to support persistence. For instance, part of the reason why anti-vaxxers

persist in their beliefs about the harms of vaccines is that they take themselves, as parents, to have greater expertise in matters relating to their own children (Powell et al., 2023).

Work on naive realism has shown that such perceptions of epistemic superiority are extremely common (Pronin et al., 2004); similarly, the other paths outline factors that are likely to be quite common across many real-world cases.

The paths can thus describe the mechanisms that generate persistence and partially explain its prevalence. However, this description lacks normative force: We can explain how people persist, but *should* they? When would it be rational not to persist? How should information across paths be integrated into such judgments?

To answer these questions, we need a normative account of responses to disagreement. One strategy to developing such an account, termed ‘rational analysis’ by Anderson (1990), first specifies the computational problem that an agent has to solve, and then generates the optimal solution to that problem (Oaksford & Chater, 2007). We take a similar approach here; leveraging what we learned about disagreement in Part 2 to identify the sort of computational problem posed by disagreement, and then providing the optimal solution to this problem.

## **A Rational Analysis of Disagreement**

We have already seen one attempt at formalizing how one should respond to disagreement. While developing the simple Bayesian model in Part 1, we viewed disagreement as posing a learning problem, and proposed that the laws of probability provide the optimal solution to this problem. Could we augment this simple analysis with more sophisticated Bayesian models of reliability (Shafto et al., 2012), dependence (Whalen et al., 2018), or other epistemic considerations (Hartmann et al., 2009; Hsiaw & Ing-Haw, 2022; Martini & Sprenger, 2017) to provide an adequate normative account?

Our discussion in Part 2 reveals that this approach would be inadequate. Beyond learning (as reflected in the epistemic path), our responses to disagreement need to reflect meta-epistemic commitments and non-epistemic goals as well. So another approach to

establishing a normative benchmark would be to integrate these factors into a joint model—one that incorporates sophisticated Bayesian models of epistemic considerations, but these other considerations as well. Rational models of sensorimotor control, for instance, posit a Bayesian learning module, a set of utilities, and an optimal controller that uses decision theory to integrate the world model and preferences into a set of physical actions (Körding & Wolpert, 2006, see also; Jara-Ettinger et al., 2016; Molinaro & Collins, 2023). Though we are not aware of Bayesian models of meta-epistemic inferences (and think this is a fruitful direction for future research), in principle such considerations could be integrated into a joint decision architecture as well. Would this complex architecture constitute an appropriate normative benchmark?

Not quite—as it would be missing the considerations outlined in the bounded path: the benefits of accurately assessing and responding to epistemic, meta-epistemic, and non-epistemic considerations need to be balanced against our cognitive constraints—which can be parsimoniously expressed through the computational costs of inference (Bhui et al., 2021; Lieder & Griffiths, 2020; Simon, 1990). Note that this balancing act requires reasoning about whether and when to reason. At the highest level, disagreement thus poses a *meta-reasoning* problem: we must strategically allocate our limited time and attention to evaluating the *right* disagreements. In other words, before we decide *how* to pursue a complete evaluation of epistemic, meta-epistemic, and non-epistemic considerations, we need to decide *whether* it makes sense to do so. The decision not to pursue this evaluation will automatically result in persistence. In this way, persistence differs from other responses to disagreement (i.e., suspension, conciliation, and polarization): it can result not only from a full evaluation of all paths, but also from the decision not to perform a full evaluation at all. A meta-reasoning framework is thus not only required to capture the constraints outlined in the bounded path, but also uniquely suited to offering an account of *persistence*.

The lens of meta-reasoning also has the virtue of offering a natural structure within which we can situate the considerations outlined by our other paths. We therefore leverage

the formal structure of a meta-reasoning model of decision-making (Lieder & Griffiths, 2017), and extend it to the context of judgment and disagreement. In describing this integrated account, we (i) first motivate the general form of the solution to the problem, then (ii) illustrate how the paths to persistence give us insight into the inputs to this solution, and (iii) finally show how the interactions between paths constrain the functional form of the solution.

### *Defining the Problem*

We begin our analysis with the set of possible disagreements that people can have in the world, notated  $\mathbb{D}$ , comprised of individual disagreements,  $D$ . Let  $\mathbb{C}$  refer to the set of contexts within which people could encounter each issue. Finally, let  $\mathbb{I}$  correspond to the set of possible responses to disagreement; in this case, people can either perform inference (i.e., fully evaluate the epistemic, meta-epistemic, and non-epistemic considerations at play within the bounds of their informational and computational limitations) or they can refrain from doing so, which would automatically result in persisting in their current belief.

Optimally responding to disagreements in each context requires choosing different response strategies for different issues—for example, seizing learning opportunities about matters of fact through careful inference, and persisting when the social costs of belief revision are prohibitive. Thus the problem of disagreement is learning an optimal mapping from disagreements to responses (i.e.,  $\mathcal{M} : \mathbb{D} \rightarrow \mathbb{I}$ ).

A simple approach to this problem would be to learn about each  $D$  independently, through brute force trial-and-error (as implemented in rote approaches to strategy selection, e.g., Shrager & Siegler, 1998). It is easy to see, however, that this would entail much error, since  $\mathbb{D}$  is extremely large and hence encountering each new  $D$  would entail a new period of potentially erroneous responses. Intuitively, arbitrarily persisting or thinking through issues to see what works best is highly inefficient, and this inefficiency is driven by the inability to generalize across disagreements. This could most easily be mitigated by learning the response that works best in a given context,  $C$ , instead (as implemented in

context-based approaches to strategy selection, e.g., Rieskamp & Otto, 2006). For instance, one could always think through scientific controversies, and refrain from doing so when it comes to normative issues (more generally, learning  $\mathcal{M} : \mathbb{C} \rightarrow \mathbb{I}$ ). Yet this strategy would entail incorrectly responding to each  $D$  that is atypical in a given  $C$ .

How can we achieve generalization across  $\mathbb{D}$  without sacrificing accuracy? One approach is to leverage the fact that there are many similarities across issues—both the causes of climate change and optimal tax rates might be seen as objective issues, for instance, whereas the best song of the year and the best candidate in an election might be seen as subjective. These similarities lie on latent dimensions, the full set of which we call  $\mathbb{L}$  (for example,  $L_{\text{Subjectivity}}$ ). The key idea is that regularities in these features across disagreements could serve as good cues to optimal responses (and hence, we can effectively *and* efficiently learn  $\mathcal{M} : \mathbb{L} \rightarrow \mathbb{I}$ ). Research in model-based reinforcement-learning has used this approach to develop agents that quickly generalize behavior to new stimuli (Dolan & Dayan, 2013; Sutton & Barto, 2018).

### ***What is the Right Mapping?***

To learn a good mapping from latent features to inference strategies, we crucially need to know what defines a good inference about a particular disagreement (i.e., we need an objective function). In artificial intelligence research, the *Value of Computation* (VoC) is a standard meta-reasoning objective that quantifies the value of a strategy (Russell & Wefald, 1991). It captures the insight that the value of a particular computation is derived from the utility of the actions it prescribes for a given problem, and the cost of engaging in that computation:

$$\text{VoC}(I, D) = \mathbb{E} \left[ \text{Utility}(I(D), D) - \text{Cost}(I, D) \right]. \quad (3)$$

The first term in the equation above captures the utility of all the actions that follow from employing inference  $I$  on issue  $D$ , and the second term is the cost of engaging in the computation. For instance, imagine that Kerem chose to persist in his views in response to



disagreement over whether vaccines are generally good for one’s well-being. The utility term would capture all of the positive and negative consequences of his persistence, from the survival benefits of taking necessary vaccines to the sadness of losing friends less keen on vaccines than he is. In AI research, the cost term corresponds to opportunity cost—inferences may require much time to compute (e.g., a complex Bayesian analysis of epistemic considerations), whereas persistence may entail much less computation (if any). In social settings, this cost could also include the consequences of choosing to deliberate about an issue at all (Oktar & Lombrozo, 2022; Tetlock, 2003).

Given this objective, we can define our problem intuitively as learning whether to evaluate or ignore different disagreements by tracking latent properties such as subjectivity. Formally, we can define the problem as:

$$\arg \max_{\mathcal{M}} \sum_{D \in \mathbb{D}} \text{VoC}(\mathcal{M}(L(D)), D) \cdot C(D). \quad (4)$$

In words, the goal is to learn a feature-to-response mapping,  $\mathcal{M}$ , that provides the highest expected VoC across disagreements,  $\mathbb{D}$ , weighed by the frequency with which these issues arise in a given context,  $C$ . Note that the mapping we choose here depends on the latent features for each disagreement,  $L$ , rather than the disagreements themselves—this allows us to easily generalize responses to novel disagreements.

**An Intuitive Example.** As an example, consider a toy case where we track one latent feature, subjectivity. Assume that, in terms of VoC, it is best to make inferences about objective issues, and to refrain from the extra computation associated with inference for subjective issues. There are four possible mappings to evaluate here (see Table 1):

By assumption, the best feature-to-response mapping is the first, and the fourth mapping is the worst. When it comes to the second and third options, however, there is uncertainty. Holding the costs of mistakes fixed (i.e., if persisting when one should have done otherwise is as bad as the reverse), the context determines which mapping is better: If objective issues are more common, then the second, ‘always infer’ mapping will be better. On the other hand, holding the frequency of issues fixed, the relative cost of making

**Table 1***Toy Example of Optimal Response Selection*

Mapping	Response to Objective	Response to Subjective	Optimality
1	Inference [✓]	No inference (persist) [✓]	Best
2	Inference [✓]	Inference [✗]	Middle
3	No inference (persist) [✗]	No inference (persist) [✓]	Middle
4	No inference (persist) [✗]	Inference [✗]	Worst

*Note.* The rows correspond to the four possible mappings, and the columns correspond to the values of the latent feature of importance. Checks mark optimal responses (as defined in the prompt), and crosses mark sub-optimal responses.

mistakes on objective vs. subjective issues would determine which mapping is better.

Thus, both VoC and context determine the relative optimality of mappings.

### Paths to Persistence in a Meta-reasoning Setting

What are the latent features that track important similarities across disagreements? We propose that our paths, in organizing relevant psychological dimensions, provide a natural starting point for models of persistence. More specifically, the considerations captured by the first three paths (epistemic, meta-epistemic, and non-epistemic) could serve as latent features (that factor into the response utility term in the VoC computation). We offer an example of such latent features below. The bounded path, on the other hand, captures the computational cost. With all paths in place, this integrated meta-reasoning model thus defines a space within which we can embed possible path-to-response mappings, where each point in feature-space corresponds to response, in this case whether or not to engage in inference (see Figure 4).

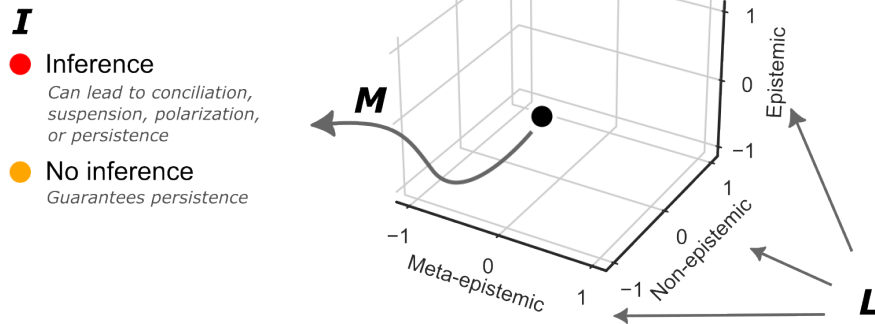
For example, consider intelligent people disagreeing with you about something trivial (i.e., with low non-epistemic value) and objective—this is a point in latent feature space, and may be associated with inference. Regions of such points define a mapping,  $\mathcal{M}$ ,

and the bounded path scales the region for each response with its computational cost. Note that the bounded path plays a distinct role here: More computational constraint translates to more persistence across all mappings that can be realized in the PTP model, as persistence entails less computation than other responses (conciliation, suspension, or polarization). We therefore focus on the other paths next, as they can generate persistence in more complex ways.

**Figure 4**

*The Paths to Persistence Model*

**$M$**  is a mapping from latent features ( **$L$** ) to responses ( **$I$** )



*Note.* The PTP suggests three latent dimensions that people map to inferences; illustrated here as normalized axes. Each dimension encapsulates many ‘sub-paths.’

### ***Three Plausible Mappings***

Our discussion of the paths in Part 2 stressed how each path can independently drive persistence—Theo, Brandon, and Matt each persisted in their views based on singular considerations. This intuitively translates to what we can call the ‘or mapping’ of persistence, whereby people will only engage in inference (and thereby potentially respond to disagreement through conciliation, suspension, or polarization) if no paths to persistence

are available; otherwise, they will persist:

$$\mathcal{M}_{\text{OR}} : \mathbb{I} = \begin{cases} \text{Infer,} & \text{if } \prod_{i=1}^3 (L_i < \alpha) = 1 \\ \text{Persist,} & \text{otherwise.} \end{cases} \quad (5)$$

Where  $L_{1,2,3}$  correspond to measures of the epistemic, meta-epistemic, and non-epistemic paths, respectively; and  $\alpha$  is a threshold that determines when these measures are sufficiently low to prompt full computation. The or mapping is simple to implement algorithmically. Whenever we encounter a disagreement, we can pick a feature at random, make a judgment about whether it justifies inference, and so forth.

The simplicity of the or mapping limits its capacity to account for trade-offs and variation in importance across paths. Consider Theo, whose belief in the unhealthiness of GMOs was grounded in epistemic inferences about others' ignorance and stupidity. It seems plausible that changes in the other paths could impact his likelihood of questioning his views. For example, if organic produce became more expensive, and Theo's girlfriend became irritated with his GMO obsession, these non-epistemic reasons could lead him to reconsider his views, even if he still thinks he has better evidence than those who disagree with him. The simplest way to flexibly allow for such trade-offs is a 'linear mapping':

$$\mathcal{M}_{\text{Linear}} : \mathbb{I} = \begin{cases} \text{Infer,} & \text{if } (\sum_{i=1}^3 w_i L_i) < \alpha \\ \text{Persist,} & \text{otherwise.} \end{cases} \quad (6)$$

Where  $w_i$  are weights that determine how influential each path is in driving mappings, and  $\alpha$  is now a threshold on the sum of weighted paths (i.e., a linear decision boundary).

Implementing the linear mapping requires more computation—in particular, estimating weights and summing distinct qualities (e.g., instrumental and epistemic values). Past research suggests that people can take these sums when evaluating beliefs (Sharot et al., 2023; van Lieshout et al., 2020).

The linear mapping is more flexible than the or mapping, but is still highly constrained. Importantly, it cannot account for the cross-path interactions we discussed

before. Incorporating such interactions into a ‘joint mapping’ requires appending the linear mapping with the interaction terms:

$$\mathcal{M}_{\text{Joint}} : \mathbb{I} = \begin{cases} \text{Infer,} & \text{if } \sum_{i=1}^3 w_i L_i + \sum_{i=1}^2 \sum_{j=i+1}^3 w_{ij} L_i L_j < \alpha \\ \text{Persist,} & \text{otherwise.} \end{cases} \quad (7)$$

These examples illustrate the diverse set of plausible mappings from paths to persistence. Which mapping is optimal ultimately depends on the informativeness of latent features, the relative utility of thinking through different kinds of disagreements, and the frequency with which those disagreements arise—as specified by the VoC computation in Equation 4.

A nuance worth noting is that inferring can itself lead to persistence. However, this will only occur when the features used to compute the meta-reasoning problem in some way mischaracterize the case at hand, such that engaging in full computation results in an outcome that mismatches the expected one. When an individual is perfectly calibrated, we should expect persistence to occur when  $\mathcal{M}$  prescribes persistence, and not otherwise.

Another important nuance relates to how the considerations outlined in the bounded path correspond to computational cost. Most cases that fall within the bounded path relate to constraints on time or effort, and so correspond to the kinds of considerations that have been modeled in prior work on meta-reasoning (Lieder & Griffiths, 2017; Russell & Wefald, 1991). Some cases, however, may involve more nuanced ‘costs’—such as basic constraints that arise from limitations in understanding that would prevent an individual from relating evidence of disagreement to some messy web of beliefs. We expect that many of these cases will also come down to resource limitations (for example, with enough time and effort some messy webs of belief could be tidied), but the more important observation for our model is that these more subtle manifestations of the bounded path will function like more traditional computational costs in the sense that they trade off against other reasons to engage in computation.

The analyses presented in this section raise interesting questions, such as how people could learn these mappings, whether there can be variation across people in

mappings, and which sets of features and responses would be optimal to consider. These are important questions for future research, but are out of the scope of the current paper. We focus on key implications of the PTP model next.

#### **Part 4: The Theoretical, Practical, and Normative Implications of PTP**

Consider your beliefs about the following:

- Is abortion morally acceptable?
- Do we burn fossil fuels more than we should?
- Should we mandate vaccinations in epidemics?

With policy following public opinion in democracies, answers to life-and-death questions like these end up shaping our societies. And we often have answers to them—answers that we confidently sustain in the face of controversy. You may have found yourself holding beliefs about the issues above with confidence, for example, even when hundreds of millions disagree.

In this paper, we have developed a model of how such belief persistence operates from a psychological perspective. We first formally defined disagreement and persistence. Then, in Part 2, we offered a comprehensive taxonomy of explanations for persistence in the form of four ‘paths.’ In a nutshell, if people (i) perceive disagreeing others to be epistemically inferior, (ii) reject the possibility of evaluating the shared ‘truth’ of an issue, (iii) are swayed by the costs and benefits of holding particular beliefs, or (iv) fail to recognize or process disagreement appropriately, they may persist in their belief amid disagreement. In Part 3, we leveraged these insights to argue that disagreement poses a meta-reasoning problem, and developed a rational analysis of persistence that integrates the four paths into a computational model.

In this final section, we first return to our guiding puzzle. We then consider key theoretical, practical, and normative insights revealed by the Paths to Persistence model, and discuss fruitful directions for future research.

## Resolving the Puzzle of Persistence

Why is persistence so common, and when can disagreement make us question our views instead? We have already seen one explanation for the puzzling prevalence of persistence: One path is sufficient to persist, whereas questioning could require no path to be available. For example, Theo can persist in his beliefs about GMOs on purely epistemic grounds. On the other hand, for Theo to question his beliefs about GMOs, he would not only need to become more intellectually humble, but also (i) believe that the healthiness of GMOs is an objective and knowable fact, (ii) estimate that the non-epistemic costs of alternative views don't override other considerations, and (iii) invest cognitive resources to reconsider his views about GMOs. From this standpoint, questioning one's views amid dissent is an edge-case that only happens when people have no paths to persistence; in all other cases, people persist.

This is the simplest explanation of our puzzle, as it assumes that each path operates independently (see Equation 5). However, the possibility that paths trade-off and interact (as described in Equations 6 and 7) suggests that this account is inadequate in some cases. For example, it seems intuitively plausible that Theo might be socially pressured into questioning his views, despite believing that he is smarter and better informed than those who disagree.

Even recognizing these interactions between paths, evidence suggests that the conditions that support persistence often hold across the politicized controversies that motivated our analysis, from abortion to vaccination. Partisans readily assume superior evidence and intelligence over their political opponents (Hartman et al., 2022), resort to subjective framings of key issues when challenged (Friesen et al., 2015), are driven by strong social motives to maintain political views (Golman et al., 2016), and find their attention spread thin across many complex and pressing issues (Williams, 2018). To the extent these paths are mutually reinforcing, rather than independent, the interactions among paths can potentially explain not only why persistence is common in such cases, but

also why it is so entrenched.

Our analysis suggests a third explanation for the prevalence of persistence: Questioning views is hard, persistence is easy, and due to the considerations highlighted above, the juice may rarely *seem to be* worth the squeeze. That is, people’s perceptions of disagreement may be biased, such that people systematically under-estimate the benefits (and over-estimate the costs) of reconsidering their beliefs. Recent work has shown that people do in fact under-estimate the informativeness and value of conversations with strangers (Atir et al., 2022), and it is plausible that people may be especially likely to underestimate how much they could learn from dissent.

Of course, questioning will be optimal in some circumstances. For trivial issues on which people do not have rich prior knowledge or social commitments, for instance, we may expect questioning to be common. Even important controversies may be widely questioned in the right contexts. For instance, in environments where critical thinking is actively rewarded; structured engagement with disagreeing others reveals the limitations of one’s own understanding; and people are jointly engaged in trying to reach an objectively justifiable conclusion, views may be more pliable amid dissent. Higher education can foster such environments: For example, a college moral philosophy course can cause students’ views on important moral controversies, such as the ethics of immigration, slavery reparations, and meat-eating, to flip (Oktar et al., 2023). Such environments cultivate exceptions to the rule of persistence.

The paths thus both explain why persistence is so common for key controversies, and when people may question their views instead. Having addressed our guiding puzzle, we now turn to implications.

### **Theoretical Implications**

The mechanisms underlying disagreement are rich, intertwined, and variable—in a word, complex. Here, we explain why this complexity can cause typical theorizing about disagreement to be misleading, predictively weak, and even harmful. We then describe how



the paths to persistence framework can help scholars accommodate this complexity.

Historically and presently, much psychological research on disagreement revolves around establishing whether particular effects exist—for instance, whether disagreements induce perceptions of bias (Kennedy & Pronin, 2008) and whether subjectivity increases belief revision in the face of dissent (Wijenayake et al., 2020). However, such marginal (i.e., direct) effects of or on disagreement can be unstable and misleading given the large number of factors likely to moderate or otherwise influence effects.

For example, when college students learned about their peers’ opinions, they were more likely to conform to the majority opinion when the issue was perceived to be *subjective* (Blake et al., 1957), but when adults answered questions online, they were more likely to shift their views in line with the majority for issues that were perceived to be *objective* (Wijenayake et al., 2020). As noted previously, such variation in the effects of subjectivity can be explained by taking a more holistic approach to persistence through the PTP framework. In this case, differences in the non-epistemic context are likely responsible for the variation: In Blake et al. (1957), students likely felt social pressure to conform to their peers and subjectivity enabled them to do so without appearing incompetent; in Wijenayake et al. (2020), anonymous strangers provided a context in which one could presumably learn from dissent on objective issues without incurring social costs. However, if we ignored such variation in the non-epistemic context and merely aggregated these two findings in a meta-analysis of meta-epistemic effects, subjectivity would seem unrelated to persistence.

Thus, perfectly valid analyses of the marginal effect of subjectivity can misleadingly support any conclusion—that subjectivity is positively, negatively, or not at all associated with persistence—when the reality is just that *it depends*. In a nutshell, whether an effect relating to disagreement exists is often not a helpful question—instead, we should ask how and why different factors jointly influence responses to disagreement, and the PTP model enables us to do so in a theory-driven manner.

Such misleading conclusions are especially problematic when we consider evidence from correlational, rather than experimental, research. This is for two reasons. The first is that such evidence is often used to draw practically important conclusions about groups of people and their psychological properties.<sup>3</sup> The second is that inverse inferences (from the tendency to persist to potential explanations) are especially likely to result in misattributions. For instance, consider work on the ‘rigidity of the right’ hypothesis (RRH), according to which conservatives are more dogmatic, worse at adapting to novel circumstances, and generally more cognitively rigid than liberals (for a review, see Zmigrod, 2020). Part of the evidence behind the RRH comes from studies investigating how much conservatives and liberals update their views in the face of disconfirmatory empirical evidence (Costello et al., 2023). Does such unresponsiveness to informational interventions mean that conservatives are more rigid or dogmatic in general? Not necessarily.

If conservatives take the meta-epistemic path to persistence, and liberals take the epistemic path, we would expect differences in the efficacy of informational interventions—not because one group is less rigid, but because we happened to target the right path for that group. Establishing rigidity would require comparing responses to a battery of interventions that cover all paths. If we instead wanted to qualify our theoretical conclusions to be about ‘epistemic rigidity,’ we would still run into the challenge of unstable direct effects discussed above. Moreover, because the evidence is correlational, we should expect effects to be even more unstable, as we now have to take into account both interactions from paths and the psychological features that characterize conservatism. Consistent with this concern, a recent adversarial collaboration reveals highly complex

---

<sup>3</sup> Due to the difficulty of experimentally manipulating personal qualities (e.g., race, gender, political affiliation), work relating to these important constructs often investigates group-based differences. Such differences are correlational—even if the statistical analysis is not a correlation, and even if there is an experimental manipulation conducted across groups. This is because observed patterns of differences across groups may be due to any part of the constellation of factors that correlate with groups, for instance, wealth differences (Meehl, 1990).

variation across issues in whether conservatives update more or less than liberals do (Bowes et al., 2023).

Importantly, our goal here is not to evaluate more than 100 years of research on rigidity (Schultz & Searleman, 2002). Instead, we intend our analysis of RRH to serve as an example of the general principle that establishing the marginal effects of disagreement—whether situational, informational, or personal—is unlikely to be helpful, and could even be harmful; leading to groups being labelled dogmatic or inflexible on the basis of potentially insufficient evidence.

Critically, this complexity is not insurmountable, and the PTP model helps point us to fruitful ways forward. Our conceptual analysis in Part 2 outlines the ingredients of persistence, and our computational framework in Part 3 provides us with tools for combining these ingredients into testable hypotheses about disagreement. The flexibility of our analysis allows us to consider much richer hypotheses than alternative frameworks can accommodate. For instance, explaining persistence through confirmation bias would obscure the richness of epistemic considerations (Kappes et al., 2020); solely Bayesian explanations would miss out on the non-epistemic (Gershman, 2019); and accounts that integrate practical and epistemic value, such as value-based belief (Sharot & Sunstein, 2020), would miss out on the consequences of bounded cognitive resources. The paths to persistence framework outlines how these critical components come together to form the mechanisms of persistence.

## **Empirical Implications**

Our theoretical analysis highlights new questions and opportunities for future empirical research. In particular, the integrated PTP model generates predictions about the functional profile of persistence, and has direct implications for the design of interventions aimed towards changing beliefs about important issues.

**Shallow Justifications for Persistence.** What predictions does our model generate about the observable characteristics and behavioral consequences of persistence?

While the specific drivers of persistence will vary across cases (as we discuss next), our model suggests that in virtually all cases, persistence itself will be relatively *shallow*. By shallow, we mean that people will generally lack access to sophisticated, deep explanations for why they persist—and to the extent that they can articulate the reasons why, these explanations will only be at the level of granularity of the latent features used in meta-reasoning.

For instance, consider a column from 1956, “Can Fifty Million Americans be Wrong?” which describes the reflections of a musician deeply troubled by Elvis’s success—80% of television owners had tuned into his latest live performance, despite the musician believing that Elvis is not a talented artist (Brown, 1956). In explaining his reasons for persisting in his view that Elvis is a sham, despite a large mass of Americans disagreeing, the writer first observes: “The answer, it seems to us, is that it’s not really a matter of right and wrong. Except in clear-cut matters of fact and morality, it’s presumptuous for any man to declare another right or wrong”, apparently pursuing the meta-epistemic path (by identifying the issue as subjective). And a couple paragraphs later, he appeals to the epistemic path: “It’s a bandwagon-conscious public, and most persons, perhaps for reasons of personal insecurity, feel a compulsion to get aboard every time.” He finally claims that disagreeing others, in addition to being biased, lack the intellectual expertise necessary for evaluating music: “Intellectual reasoning rarely succeeds in opening an emotional or anti-intellectual vise (...)” This is exactly the kind of explanation we would expect if people persisted through a coarse evaluation of the latent features described by our paths. Contrast this with what we may expect if persistence resulted from a sophisticated joint-inference process: At a minimum, the incoherent appeal to both subjectivity and ignorance would be flagged by any sophisticated evaluation, and we might expect much more nuanced consideration of how the factors jointly informed his belief update.

**Interpreting and Designing Interventions.** As Ross and Anderson (1982) remarked, “beliefs are remarkably resilient in the face of empirical challenges that seem

logically devastating.” Accumulating evidence since then has shown that interventions on controversial beliefs and attitudes tend to have small effects across domains, intervention types, and measures (around two to three tenths of a standard deviation; see Albarracin & Shavitt, 2018). These facts seem to establish a pessimistic baseline for the potential efficacy of persistence-reducing interventions. Indeed, prominent scholars in the behavioral sciences have begun arguing for a shift away from individual-level interventions due to their inefficacy in important domains (Chater & Loewenstein, 2023).

But why do typical belief-change interventions fail? A common explanation is that people’s views on important issues are simply too robust: Haidt (2001), for instance, writes that moral judgments will change “primarily in cases in which the initial intuition is weak.” PTP suggests another explanation: Belief change interventions have the best chances of succeeding when they address the *sources* of persistence in a *targeted* manner. Much as precision medicine aims to further the efficacy and efficiency of healthcare interventions by tailoring the selection of drugs, dosage, timing, and additional treatments (Kosorok & Laber, 2019), such *precision interventions* would optimally tailor belief-change interventions to the specific set of mechanisms driving individual beliefs.

Yet typical belief-change interventions are more akin to fast-fashion than tailored interventions. The vast majority are epistemic interventions that aim to be cheap and scalable: For example, providing factual corrections (e.g., Brashier et al., 2021), sharing relevant arguments (e.g., Jolley & Douglas, 2017), or communicating expert consensus (e.g., van Stekelenburg et al., 2022). Other paths have received much less attention—we are not aware of any work on meta-epistemic interventions for changing beliefs, for instance, except for some related work in educational psychology (Klopp & Stark, 2022). Relatedly, the vast majority of interventions examine the effects of just one intervention: as Ecker et al. (2022) point out in a review of misinformation research, “most research to date has considered each approach separately and more research is required to test synergies between these strategies.” The PTP explains why: Persistence is driven by multiple

interacting mechanisms, with potential heterogeneity in mechanisms across issues and across the population. We should therefore avoid drawing premature conclusions (e.g., that beliefs are too robust to change) without examining a broader set of intervention strategies. Much as conservatives' views may be less rigid than previous research suggests, tailored interventions may generally be more effective at fostering scientifically-informed beliefs than past interventions indicate. Moreover, there is much room for rich empirical exploration in the design of tailored interventions: Beyond developing effective ways to target the right set of paths, social scientists could draw inspiration from precision medicine in exploring optimal intervention sequences, timing, and delivery. Recent intervention tournaments can be seen as unstructured first steps towards exploring this rich space (Gelfand et al., 2022), and work on 'integrative experiments' outlines how they can be scaled up in coordinated fashion (Almaatouq et al., 2022).

In sum, when designing interventions to address belief persistence, we need to pay attention to which paths underlie particular cases, and aim to address all of them.

### **Normative Implications**

In the preceding sections, we have presented an extensive analysis of the descriptive questions relating to disagreement. Here, we outline some of the key normative questions that emerge from this discussion. These questions are important for psychologists to be aware of, because how individuals *ought to* react to disagreement influences (i) how researchers *should* try to influence individual's beliefs through interventions, and (ii) which psychological questions *should* be pursued with greater urgency and attention. The answers to these normative questions also have implications for accounts of rationality and for epistemology more generally.

Perhaps the foundational question here is whether it is good for people to persist. Our rational analysis provides first steps towards an answer to this question: The learning mechanisms outlined in Part 3 lead to individually optimal policies that balance various costs and benefits. But the question of whether it is good for people to persist is much

broader than the individual—is it *societally* optimal for individuals to persist? Scholars across disciplines have long noted that individually optimal behavior does not necessarily result in societally optimal outcomes and vice versa (Axelrod, 1980; Kitcher, 1990; Ostrom, 1999). In the case of disagreement, persistence plays a particularly important role in enabling transient diversity of opinion across key issues, which can facilitate effective problem-solving (though entrenched disagreements can be harmful; Smaldino et al., 2023; Zollman, 2010). On the other hand, whether such diversity is productive depends on more complex considerations. For example, ethically, we may not want diversity on core human values, such as freedom from slavery, and epistemologically, we may want diversity on scientific questions to be proportional to the state of current evidence (Kitcher, 1990).

A related question is whether there are better or worse ways to persist. For example, different paths may lead to differentially truth-promoting or prosocial behavior. Some paths seem inferior on both counts: Attributions of bias or dishonesty may be particularly disabling of truth-promoting deliberation and community building interaction. Supporting this claim, Kennedy and Pronin (2008) show that disagreement-induced perceptions of bias lower the perceived effectiveness of communication and lead to more aggressive interactions. On the other hand, some paths, such as attributions of evidential inferiority, may be truth-promoting both individually and societally. If someone believes that a large group disagrees with them because they do not have access to some key evidence, they may be inclined to try to share this evidence with members of that group. Perhaps for these reasons, philosophers have long emphasized the importance of prioritizing evidence and arguments over personal attributions in critical dialogue (Tindale, 2007).

We note that both of these normative questions—whether we should, and if so, how we should persist—involve a mixture of epistemic and instrumental considerations. This mixture is also reflected in our meta-reasoning analysis, which incorporates both epistemic and non-epistemic elements in deriving optimal policies. However, these are distinct standards for evaluating rationality (Kelly, 2003), which raises interesting questions about

how we should evaluate the rationality of meta-reasoning more broadly.

### Conclusion

In this paper, we developed a model of how individuals persist in their beliefs amid societal controversy. Starting with a definition of disagreement, we situated persistence among other possible responses, and introduced a Bayesian analysis of belief revision in the face of disagreement. The limitations of this analysis in explaining persistence motivated a broader conceptual analysis, which resulted in the broader paths to persistence framework. We explained how each of four distinct paths (epistemic, meta-epistemic, non-epistemic, and bounded) can individually drive persistence, and then introduced a rational analysis that integrated these considerations into a flexible model of persistence grounded in meta-reasoning. We finally considered theoretical, empirical, and normative implications; from why typical theories of disagreement are likely to result in misleading conclusions, to how empirical research needs to evolve to precisely address the sources of persistence, and whether there are better or worse ways to persist. We hope that the paths to persistence model will guide much-needed empirical inquiry into the psychology of persistence—from its prevalence to its variation across domains—and thus set the stage for the development of effective interventions that bridge the widening rifts in our societies.



## References

- Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? Children consider informants' sources when deciding whom to believe. *Journal of Experimental Psychology: General*. <https://doi.org/https://doi.org/10.1037/xge0001198>
- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*(2), 358–372. <https://doi.org/10.1037/0033-295X.96.2.358>
- Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology*, *69*, 299–327. <https://doi.org/https://doi.org/10.1146/annurev-psych-122216-011911>
- Alister, M., Ransom, K. J., & Perfors, A. (2023). Inferring the truth from deception: What can people learn from helpful and unhelpful information providers? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). <https://escholarship.org/uc/item/8vt661bv>
- Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2022). Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences. *Behavioral and Brain Sciences*, 1–55. <https://doi.org/https://doi.org/10.1017/S0140525X22002874>
- Altay, S., Majima, Y., & Mercier, H. (2023). Happy thoughts: The role of communion in accepting and sharing (mis) beliefs. *British Journal of Social Psychology*, *62*(4), 1672–1692. <https://doi.org/https://doi.org/10.1111/bjso.12650>
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Angouri, J., & Locher, M. A. (2012). Theorising disagreement. *Journal of Pragmatics*, *44*(12), 1549–1553. <https://doi.org/https://doi.org/10.1016/j.pragma.2012.06.011>
- Atir, S., Wald, K. A., & Epley, N. (2022). Talking with strangers is surprisingly informative. *Proceedings of the National Academy of Sciences*, *119*(34), e2206992119. <https://doi.org/https://doi.org/10.1073/pnas.2206992119>

- Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4(6), 1236–1239.
- Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, 24(1), 3–25.  
<https://doi.org/https://doi.org/10.1177/00220027800240010>
- Ayars, A., & Nichols, S. (2020). Rational learners and metaethics: Universalism, relativism, and evidence from consensus. *Mind & Language*, 35(1), 67–89.  
<https://doi.org/https://doi.org/10.1111/mila.12232>
- Bandura, A. (2010). Self-Efficacy. In *The Corsini Encyclopedia of Psychology*.  
<https://doi.org/10.1002/9780470479216.corpsy0836>
- Bénabou, R., & Tirole, J. (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, 30(3), 141–164.  
<https://doi.org/10.1257/jep.30.3.141>
- Bendana, J., & Mandelbaum, E. (2021). The Fragmentation of Belief. In C. Borgoni, D. Kindermann, & A. Onofri (Eds.), *The Fragmented Mind*.
- Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, 41, 15–21.  
<https://doi.org/10.1016/j.cobeha.2021.02.015>
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3), 151–170. <https://doi.org/https://doi.org/10.1257/jep.12.3.151>
- Blake, R. R., Helson, H., & Mouton, J. S. (1957). The generality of conformity behavior as a function of factual anchorage, difficulty of task, and amount of social pressure. *Journal of Personality*. <https://doi.org/10.1111/j.1467-6494.1957.tb01528.x>
- Boothby, E. J., Cooney, G., & Schweitzer, M. E. (2023). Embracing complexity: A review of negotiation research. *Annual Review of Psychology*, 74, 299–332.  
<https://doi.org/https://doi.org/10.1146/annurev-psych-033020-014116>
- Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology*. Oxford University Press.

- Bowes, S. M., Clark, C. J., Conway, I., Lucian G, Costello, T. H., Osborne, D., Tetlock, P., & van Prooijen, J.-W. (2023, June). An adversarial collaboration on the rigidity-of-the-right, rigidity-of-extremes, or symmetry: The answer depends on the question. <https://doi.org/10.31234/osf.io/4wmx2>
- Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5), e2020043118. <https://doi.org/10.1073/pnas.2020043118>
- Bromberg-Martin, E. S., & Sharot, T. (2020). The value of beliefs. *Neuron*, 106(4), 561–565. <https://doi.org/https://doi.org/10.1016/j.neuron.2020.05.001>
- Broomell, S. B. (2020). Global–local incompatibility: The misperception of reliability in judgment regarding global variables. *Cognitive Science*, 44(4), e12831. <https://doi.org/10.1111/cogs.12831>
- Brown, G. D., Lewandowsky, S., & Huang, Z. (2022). Social sampling and expressed attitudes: Authenticity preference and social extremeness aversion lead to social norm effects and polarization. *Psychological Review*, 129(1), 18. <https://doi.org/https://doi.org/10.1037/rev0000342>
- Brown, L. (1956). Can fifty million americans be wrong [Septmber 19]. *Down Beat*, 41. <https://www.elvis-atouchofgold.com/fifty-million/>
- Carothers, T., & O'Donohue, A. (Eds.). (2019). *Democracies divided: The global challenge of political polarization*. The Brookings Institution.
- Chater, N. (2018). *The mind is flat: The illusion of mental depth and the improvised mind*. Penguin UK.
- Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 46, e147. <https://doi.org/10.1017/S0140525X22002023>

- Cheek, N. N., Blackman, S. F., & Pronin, E. (2021). Seeing the subjective as objective: People perceive the taste of those they disagree with as biased and wrong. *Journal of Behavioral Decision Making*, 34(2), 167–182. <https://doi.org/10.1002/bdm.2201>
- Chlup, D. T., & Collins, T. E. (2010). Breaking the ice: Using ice-breakers and re-energizers with adult learners. *Adult Learning*, 21(3-4), 34–39. <https://doi.org/10.1177/104515951002100305>
- Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2), 187–217. <https://www.jstor.org/stable/20446955>
- Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81(1), 185–215. <https://doi.org/https://doi.org/10.1111/j.1933-1592.2010.00366.x>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrocioni, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology*, 85(5), 808–822. <https://doi.org/10.1037/0022-3514.85.5.808>
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin*, 26(9), 1151–1164. <https://doi.org/10.1177/01461672002611011>
- Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. *Critical Review*, 18(1-3), 1–74. <https://doi.org/10.1080/08913810608443650>
- Costello, T. H., Bowes, S. M., Baldwin, M. W., Malka, A., & Tasimi, A. (2023). Revisiting the rigidity-of-the-right hypothesis: A meta-analytic review. *Journal of Personality and Social Psychology*, 124(5), 1025. <https://doi.org/https://doi.org/10.1037/pspp0000446>

- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.  
[https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Cunningham, V. (2021). The Rioters in the Senate Chamber. *The New Yorker*. Retrieved May 2, 2022, from <https://www.newyorker.com/culture/annals-of-appearances/the-rioters-in-the-senate-chamber>
- Cusimano, C., & Lombrozo, T. (2023). People recognize and condone their own morally motivated reasoning. *Cognition*, 234, 105379.  
<https://doi.org/10.1016/j.cognition.2023.105379>
- Davoodi, T., & Lombrozo, T. (2022a). Explaining the existential: Scientific and religious explanations play different functional roles. *Journal of Experimental Psychology: General*, 151(5), 1199. <https://doi.org/https://doi.org/10.1037/xge0001129>
- Davoodi, T., & Lombrozo, T. (2022b). Varieties of Ignorance: Mystery and the Unknown in Science and Religion. *Cognitive Science*, 46(4), e13129.  
<https://doi.org/10.1111/cogs.13129>
- Dawkins, R. (1989). In Short: Nonfiction. *The New York Times*. Retrieved April 16, 2021, from <https://www.nytimes.com/1989/04/09/books/in-short-nonfiction.html>
- Desai, S. C., Xie, B., & Hayes, B. K. (2022). Getting to the source of the illusion of consensus. *Cognition*, 223, 105023. <https://doi.org/10.1016/j.cognition.2022.105023>
- Dinas, E. (2014). Why Does the Apple Fall Far from the Tree? How Early Political Socialization Prompts Parent-Child Dissimilarity. *British Journal of Political Science*, 44(4), 827–852. <https://doi.org/10.1017/S0007123413000033>
- Dion, K. L. (2003). Prejudice, racism, and discrimination. In *Handbook of psychology: Personality and social psychology*, Vol. 5. (pp. 507–536). John Wiley & Sons, Inc.  
<https://doi.org/10.1002/0471264385.wei0521>
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312–325.  
<https://doi.org/10.1016/j.neuron.2013.09.007>

- Dorst, K. (2023). Rational polarization (1st edition). *Philosophical Review*, 132(3), 355–458. <https://doi.org/10.1215/00318108-10469499>
- Duck, J. M., & Mullin, B.-A. (1995). The perceived impact of the mass media: Reconsidering the third person effect. *European Journal of Social Psychology*, 25(1), 77–93. <https://doi.org/https://doi.org/10.1002/ejsp.2420250107>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Egan, A. (2010). Disputing about taste. In T. Warfield & R. Feldman (Eds.), *Disagreement* (pp. 247–286). Oxford University Press.
- Enke, B., & Zimmermann, F. (2019). Correlation Neglect in Belief Formation. *The Review of Economic Studies*, 86(1), 313–332. <https://doi.org/10.1093/restud/rdx081>
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Erickson, R. J. (1995). The importance of authenticity for self and society. *Symbolic interaction*, 18(2), 121–144. <https://doi.org/https://doi.org/10.1525/si.1995.18.2.121>
- Erikson, R. S., & Tedin, K. L. (2019). *American public opinion: Its origins, content, and impact*. Routledge.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Feldman, R. (2007). Epistemological Puzzles about Disagreement. In S. Heatherington (Ed.), *Epistemological Futures* (pp. 216–36). Oxford University Press.

- Feldman, R. (2009). Evidentialism, Higher-Order Evidence, and Disagreement. *Episteme*, 6(3), 294–312. <https://doi.org/10.3366/E1742360009000720>
- Fields, J. M., & Schuman, H. (1976). Public Beliefs About the Beliefs of the Public. *Public Opinion Quarterly*, 40(4), 427–448. <https://doi.org/10.1086/268330>
- Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of Social Influence: Towards the Next Frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2. <https://doi.org/10.18564/jasss.3521>
- Frances, B. (2014). *Disagreement*. Polity.
- Frances, B., & Matheson, J. (2019). Disagreement. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. Retrieved March 6, 2021, from <https://plato.stanford.edu/archives/win2019/entries/disagreement/>
- Frantz, C. M. (2006). I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides. *Basic and Applied Social Psychology*, 28(2), 157–167. [https://doi.org/10.1207/s15324834basp2802\\_5](https://doi.org/10.1207/s15324834basp2802_5)
- Friesen, J. P., Campbell, T. H., & Kay, A. C. (2015). The psychological advantage of unfalsifiability: The appeal of untestable religious and political ideologies. *Journal of Personality and Social Psychology*, 108, 515–529. <https://doi.org/10.1037/pspp0000018>
- Gelfand, M., Li, R., Stamkou, E., Pieper, D., Denison, E., Fernandez, J., Choi, V., Chatman, J., Jackson, J., & Dimant, E. (2022). Persuading republicans and democrats to comply with mask wearing: An intervention tournament. *Journal of Experimental Social Psychology*, 101, 104299. <https://doi.org/10.1016/j.jesp.2022.104299>
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409. <https://doi.org/10.1016/j.tics.2004.07.001>

- Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, 26, 13–28. <https://doi.org/https://doi.org/10.3758/s13423-018-1488-8>
- Gollwitzer, A., & Oettingen, G. (2019). Paradoxical knowing: A shortcut to knowledge and its antisocial correlates. *Social Psychology*, 50(3), 145–161. <https://doi.org/10.1027/1864-9335/a000368>
- Golman, R. (2023). Acceptable discourse: Social norms of beliefs and opinions. *European Economic Review*, 160, 104588. <https://doi.org/https://doi.org/10.1016/j.euroecorev.2023.104588>
- Golman, R., Loewenstein, G., Moene, K. O., & Zarri, L. (2016). The Preference for Belief Consonance. *Journal of Economic Perspectives*, 30(3), 165–188. <https://doi.org/10.1257/jep.30.3.165>
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48(1), 250–256. <https://doi.org/10.1016/j.jesp.2011.08.006>
- Gottlieb, S., & Lombrozo, T. (2018). Can Science Explain the Human Mind? Intuitive Judgments About the Limits of Science. *Psychological Science*, 29(1), 121–130. <https://doi.org/10.1177/0956797617722609>
- Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, 24(11), 873–883. <https://doi.org/https://doi.org/10.1016/j.tics.2020.09.001>
- Hafer, C., & Sutton, R. (2016). Belief in a Just World. [https://doi.org/10.1007/978-1-4939-3216-0\\_8](https://doi.org/10.1007/978-1-4939-3216-0_8)
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814. <https://doi.org/10.1037/0033-295x.108.4.814>
- Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.),



- Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd, pp. 3–24). American Psychological Association. <https://doi.org/10.1037/0000135-001>
- Harris, P. L. (2012). *Trusting What You're Told: How Children Learn from Others*. Harvard University Press.
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive Foundations of Learning from Testimony. *Annual Review of Psychology*, 69(1), 251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>
- Hartman, R., Hester, N., & Gray, K. (2022). People See Political Opponents as More Stupid Than Evil. *Personality and Social Psychology Bulletin*, 01461672221089451. <https://doi.org/10.1177/01461672221089451>
- Hartmann, S., Martini, C., & Sprenger, J. (2009). Consensual decision-making among epistemic peers. *Episteme*, 6(2), 110–129.
- Hedden, B., & Dorst, K. (2022). (almost) all evidence is higher-order evidence. *Analysis*, 82(3), 417–425. <https://doi.org/10.1093/analys/anab081>
- Heiphetz, L., Landers, C. L., & Van Leeuwen, N. (2021). Does think mean the same thing as believe? Linguistic insights into religious cognition. *Psychology of Religion and Spirituality*, 13(3), 287–297. <https://doi.org/10.1037/rel0000238>
- Heiphetz, L., & Young, L. L. (2017). Can only one person be right? the development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition*, 167, 78–90. <https://doi.org/10.1016/j.cognition.2016.05.014>
- Hemp, P. (2009). Death by information overload. *Harvard business review*, 87(9), 82–121.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hetherington, M. J., & Rudolph, T. J. (2015). Why washington won't work: Polarization, political trust, and the governing crisis. In *Why washington won't work*. University of Chicago Press. <https://doi.org/10.7208/9780226299358>

- Hippel, W. v., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*(1), 1–16.  
<https://doi.org/10.1017/S0140525X10001354>
- Hsiaw, A., & Ing-Haw, C. (2022). Distrust in experts and the origins of disagreement. *Journal of Economic Theory*, *200*, 105–1010.  
<https://doi.org/https://doi.org/10.1016/j.jet.2021.105401>
- Hyman, H. H., & Sheatsley, P. B. (1950). The current status of american public opinion. In J. C. Payne (Ed.), *The teaching of contemporary affairs: 21st yearbook of the national council of social studies* (pp. 11–34). National Council of Social Studies.
- Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141.  
<https://doi.org/https://doi.org/10.1037/0022-3514.50.6.1141>
- Iyengar, S., & Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, *59*(3), 690–707. <https://doi.org/https://doi.org/10.1111/ajps.12152>
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naive utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604. <https://doi.org/10.1016/j.tics.2016.05.011>
- Jern, A., Chang, K.-m. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, *121*(2), 206–224. <https://doi.org/10.1037/a0035941>
- Johnson, S. G., Kim, K., & Keil, F. (2016). The determinants of knowability. In: *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.
- Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology*, *47*(8), 459–469. <https://doi.org/10.1111/jasp.12453>

- Jones, J. (2021). Democratic, republican confidence in science diverges.  
<https://news.gallup.com/poll/352397/democratic-republican-confidence-science-diverges.aspx>
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100(1), 109.  
<https://doi.org/https://doi.org/10.1037/0033-295X.100.1.109>
- Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, 22(1), 51–66. <https://doi.org/10.1037/h0032356>
- Kahan, D. (2010). Fixing the communications failure. *Nature*, 463(7279), 296–297.  
<https://doi.org/10.1038/463296a>
- Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2), 410–425.  
<https://doi.org/10.1017/S0003055419000923>
- Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020). Confirmation bias in the utilization of others' opinion strength. *Nature Neuroscience*, 23(1), 130–137.  
<https://doi.org/https://doi.org/10.1038/s41593-019-0549-2>
- Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, 24(2), 163–204. <https://doi.org/https://doi.org/10.1086/266945>
- Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, 66(3), 612–640.  
<https://doi.org/https://doi.org/10.1111/j.1933-1592.2003.tb00281.x>
- Kelly, T. (2005). The Epistemic Significance of Disagreement. In J. Hawthorne & T. Gendler (Eds.), *Oxford Studies in Epistemology, Volume 1* (pp. 167–196). Oxford University Press.

- Kennedy, K. A., & Pronin, E. (2008). When Disagreement Gets Ugly: Perceptions of Bias and the Escalation of Conflict. *Personality and Social Psychology Bulletin*, 34(6), 833–848. <https://doi.org/10.1177/0146167208315158>
- Kitcher, P. (1990). The division of cognitive labor. *The journal of philosophy*, 87(1), 5–22. <https://doi.org/https://doi.org/10.2307/2026796>
- Kivy, P. (2015). *De gustibus: Arguing about taste and why we do it*. Oxford University Press.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., et al. (2014). Investigating variation in replicability. *Social Psychology*. <https://doi.org/https://doi.org/10.1027/1864-9335/a000178>
- Klofstad, C. A., Sokhey, A. E., & McClurg, S. D. (2013). Disagreeing about Disagreement: How Conflict in Social Networks Affects Political Behavior. *American Journal of Political Science*, 57(1), 120–134. <https://doi.org/10.1111/j.1540-5907.2012.00620.x>
- Klopp, E., & Stark, R. (2022). How to change epistemological beliefs? effects of scientific controversies, epistemological sensitization, and critical thinking instructions on epistemological change. *Education Sciences*, 12(7), 499. <https://doi.org/https://doi.org/10.3390/educsci12070499>
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development*, 76(6), 1261–1277. <https://doi.org/10.1111/j.1467-8624.2005.00849.x>
- Kominsky, J. F., Langthorne, P., & Keil, F. C. (2016). The better part of not knowing: Virtuous ignorance. *Developmental Psychology*, 52(1), 31–45. <https://doi.org/10.1037/dev0000065>
- Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7), 319–326. <https://doi.org/10.1016/j.tics.2006.05.003>

- Kosorok, M. R., & Laber, E. B. (2019). Precision medicine. *Annual review of Statistics and Its Application*, 6, 263–286.  
<https://doi.org/https://doi.org/10.1146/annurev-statistics-030718-105251>
- Kruglanski, A. W. (2004). *The Psychology of Closed Mindedness*. Taylor & Francis Group.
- Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., Mannetti, L., et al. (2005). Says who? epistemic authority effects in social judgment. *Advances in Experimental Social Psychology*, 37(37), 345–392.  
[https://doi.org/https://doi.org/10.1016/S0065-2601\(05\)37006-7](https://doi.org/https://doi.org/10.1016/S0065-2601(05)37006-7)
- Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, 15(3), 309–328.  
[https://doi.org/10.1016/S0885-2014\(00\)00030-7](https://doi.org/10.1016/S0885-2014(00)00030-7)
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lenz, G. S. (2013). *Follow the Leader?: How Voters Respond to Politicians' Policies and Performance*. University of Chicago Press.
- Levitsky, S., & Ziblatt, D. (2018). *How Democracies Die*. Crown.
- Lewandowsky, S., Armaos, K., Bruns, H., Schmid, P., Holford, D. L., Hahn, U., Al-Rawi, A., Sah, S., & Cook, J. (2022). When science becomes embroiled in conflict: Recognizing the public's need for debate while combating conspiracies and misinformation. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 26–40. <https://doi.org/https://doi.org/10.1177/00027162221084663>
- Lewandowsky, S., Cook, J., & Lloyd, E. (2018). The 'Alice in Wonderland' mechanics of the rejection of (climate) science: Simulating coherence by conspiracism. *Synthese*, 195, 175–196. <https://doi.org/https://doi.org/10.1007/s11229-016-1198-6>
- Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, 124(6), 762–794. <https://doi.org/10.1037/rev0000075>

- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, e1. <https://doi.org/10.1017/S0140525X1900061X>
- Liquin, E. G., Metz, S. E., & Lombrozo, T. (2020). Science demands explanation, religion tolerates mystery. *Cognition*, 204, 104398. <https://doi.org/10.1016/j.cognition.2020.104398>
- Martini, C., & Sprenger, J. (2017). Opinion aggregation and individual expertise. In *Scientific collaboration and collective knowledge: New essays* (pp. 180–201). Oxford University Press USA.
- Matheson, J. (2015). *The epistemic significance of disagreement*. Springer.
- McCombs, M., & Valenzuela, S. (2020). *Setting the Agenda: Mass Media and Public Opinion*. John Wiley & Sons.
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195–244. <https://doi.org/https://doi.org/10.2466/PR0.66.1.195-244>
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
- Miller, D. T., & Prentice, D. A. (1994). Collective errors and errors about the collective. *Personality and Social Psychology Bulletin*, 20(5), 541–550. <https://doi.org/10.1177/0146167294205011>
- Minson, J. A., Bendersky, C., de Dreu, C., Halperin, E., & Schroeder, J. (2023). Experimental studies of conflict: Challenges, solutions, and advice to junior scholars. *Organizational Behavior and Human Decision Processes*, 177, 104257. <https://doi.org/https://doi.org/10.1016/j.obhdp.2023.104257>
- Minson, J. A., & Dorison, C. A. (2022). Why is exposure to opposing views aversive? Reconciling three theoretical perspectives. *Current Opinion in Psychology*, 47, 101435. <https://doi.org/10.1016/j.copsyc.2022.101435>

- Molinaro, G., & Collins, A. G. (2023). A goal-centric outlook on learning. *Trends in Cognitive Sciences*, 27(12), 1150–1164. <https://doi.org/10.1016/j.tics.2023.08.011>
- Molnar, A., & Loewenstein, G. (2020). The false and the furious: People are more disturbed by others' false beliefs than by differences in beliefs. *Available at SSRN 3524651*. <https://doi.org/http://dx.doi.org/10.2139/ssrn.3524651>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, 21(3), 262–283. [https://doi.org/10.1016/0022-1031\(85\)90020-4](https://doi.org/10.1016/0022-1031(85)90020-4)
- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4), 602. <https://doi.org/https://doi.org/10.1037/0033-2909.83.4.602>
- Newport, F. (2023). Update: Partisan gaps expand most on government power, climate. <https://news.gallup.com/poll/509129/update-partisan-gaps-expand-government-power-climate.aspx>
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Noelle-Neumann, E. (1977). Turbulences in the Climate of Opinion: Methodological Applications of the Spiral of Silence Theory. *Public Opinion Quarterly*, 41(2), 143–158. <https://doi.org/10.1086/268371>
- Norenzayan, A. (2013). *Big gods: How religion transformed cooperation and conflict*. Princeton University Press.
- Oaksford, M., & Chater, N. (2007). Rationality and rational analysis. In M. Oaksford & N. Chater (Eds.), *Bayesian rationality: The probabilistic approach to human*

- reasoning* (pp. 10–40). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198524496.003.0002>
- Oktar, K., Lerner, A., Malaviya, M., & Lombrozo, T. (2023). Philosophy instruction changes views on moral controversies by decreasing reliance on intuition. *Cognition*, 236, 105434. <https://doi.org/10.1016/j.cognition.2023.105434>
- Oktar, K., & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition*, 223, 105021.  
<https://doi.org/10.1016/j.cognition.2022.105021>
- Ostrom, E. (1999). Coping with tragedies of the commons. *Annual Review of Political Science*, 2(1), 493–535.  
<https://doi.org/http://dx.doi.org/10.1146/annurev.polisci.2.1.493>
- Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, 58(1), 97–118.  
<https://doi.org/10.1146/annurev.psych.58.110405.085555>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion*, 29(1), 1–26.  
<https://doi.org/10.1080/02699931.2014.967183>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.  
[https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Pew Research Center. (2016). U.S. Public Divides Over Food Science. Retrieved September 23, 2021, from <https://www.pewresearch.org/science/2016/12/01/the-new-food-fights/>



- Pew Research Center. (2019). How partisans view each other. Retrieved August 8, 2022, from  
<https://www.pewresearch.org/politics/2019/10/10/how-partisans-view-each-other/>
- Plunkett, D., Buchak, L., & Lombrozo, T. (2020). When and why people think beliefs are “debunked” by scientific explanations of their origins. *Mind & Language*, 35(1), 3–28. <https://doi.org/10.1111/mila.12238>
- Poliakov, L. (2003). *The History of Anti-Semitism, Volume 3: From Voltaire to Wagner*. University of Pennsylvania Press.
- Pool, G. J., Wood, W., & Leck, K. (1998). The self-esteem motive in social influence: Agreement with valued majorities and disagreement with derogated minorities. *Journal of Personality and Social Psychology*, 75(4), 967–975.  
<https://doi.org/10.1037/0022-3514.75.4.967>
- Powell, D., Weisman, K., & Markman, E. M. (2023). Modeling and leveraging intuitive theories to improve vaccine attitudes. *Journal of Experimental Psychology: General*, 152(5), 1379–1395. <https://doi.org/https://doi.org/10.1037/xge0001324>
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review*, 111(3), 781–799. <https://doi.org/10.1037/0033-295X.111.3.781>
- Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of personality and social psychology*, 38(1), 141–152.  
<https://doi.org/https://doi.org/10.1037/0022-3514.38.1.141>
- Rabb, N., Fernbach, P. M., & Sloman, S. A. (2019). Individual Representation in a Community of Knowledge. *Trends in Cognitive Sciences*, 23(10), 891–902.  
<https://doi.org/10.1016/j.tics.2019.07.011>
- Reeder, G. D., Pryor, J. B., Wohl, M. J. A., & Griswell, M. L. (2005). On Attributing Negative Motives to Others Who Disagree With Our Opinions. *Personality and*

- Social Psychology Bulletin*, 31(11), 1498–1510.  
<https://doi.org/10.1177/0146167205277093>
- Reynolds, J. W. (2020). Talking about abortion (listening optional). *Texas A&M Law Review*, 8, 141–162. <https://doi.org/10.37419/LR.V8.I1.4>
- Rieskamp, J., & Otto, P. E. (2006). Ssl: A theory of how people learn to select strategies. *Journal of experimental psychology: General*, 135(2), 207–236.  
<https://doi.org/https://doi.org/10.1037/0096-3445.135.2.207>
- Roberts, S. O., Ho, A. K., & Gelman, S. A. (2021). Should Individuals Think Like Their Group? A Descriptive-to-Prescriptive Tendency Toward Group-Based Beliefs. *Child Development*, 92(2), 201–220. <https://doi.org/10.1111/cdev.13448>
- Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, 68(3), 404–417.  
<https://doi.org/10.1037/0022-3514.68.3.404>
- Ross, L., & Ward, A. (1996). Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. In *Values and Knowledge*. Psychology Press.
- Ross, L., & Anderson, C. A. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 129–152). Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511809477.010>
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X)
- Rothschild, J. E., Howat, A. J., Shafranek, R. M., & Busby, E. C. (2019). Pigeonholing partisans: Stereotypes of party supporters and partisan polarization. *Political Behavior*, 41, 423–443. <https://doi.org/https://doi.org/10.1007/s11109-018-9457-5>

- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562.  
[https://doi.org/10.1207/s15516709cog2605\\_1](https://doi.org/10.1207/s15516709cog2605_1)
- Rubin, M., & Badea, C. (2012). They're All the Same!. . . but for Several Different Reasons: A Review of the Multicausal Nature of Perceived Group Variability. *Current Directions in Psychological Science*, 21(6), 367–372.  
<https://doi.org/10.1177/0963721412457363>
- Russell, S., & Wefald, E. (1991). Principles of metareasoning. *Artificial Intelligence*, 49(1), 361–395. [https://doi.org/10.1016/0004-3702\(91\)90015-C](https://doi.org/10.1016/0004-3702(91)90015-C)
- Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk Moral Relativism. *Mind and Language*, 26(4), 482–505.  
<https://doi.org/10.1111/j.1468-0017.2011.01428.x>
- Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs*, 128(2), 165–207.
- Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, 3(10), 1055–1061. <https://doi.org/10.1038/s41562-019-0666-7>
- Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning From Others: The Consequences of Psychological Reasoning for Human Learning. *Perspectives on Psychological Science*, 7(4), 341–351. <https://doi.org/10.1177/1745691612448481>
- Shamir, J., & Shamir, M. (1997). Pluralistic ignorance across issues and over time: Information cues and biases. *Public Opinion Quarterly*, 227–260.  
<https://www.jstor.org/stable/2749551>
- Shanteau, J. (2015). Why task domains (still) matter for understanding expertise. *Journal of Applied Research in Memory and Cognition*, 4(3), 169–175.  
<https://doi.org/10.1016/j.jarmac.2015.07.003>

- Sharot, T., Rollwage, M., Sunstein, C. R., & Fleming, S. M. (2023). Why and when beliefs change. *Perspectives on Psychological Science*, 18(1), 142–151.  
<https://doi.org/10.1177/17456916221082967>
- Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, 4(1), 14–19. <https://doi.org/10.1038/s41562-019-0793-1>
- Shavitt, S. (1989). Operationalizing functional theories of attitude. In *Attitude structure and function* (pp. 311–337). Psychology Press.
- Sherif, M. (1956). Experiments in group conflict. *Scientific American*, 195(5), 54–59.  
<https://www.jstor.org/stable/24941808>
- Shields, M. (2021). On the pragmatics of deep disagreement. *Topoi*, 40(5), 999–1015.  
<https://doi.org/10.1007/s11245-018-9602-0>
- Shrager, J., & Siegler, R. S. (1998). SCADS: a model of children’s strategy choices and strategy discoveries. *Psychological Science*, 9(5), 405–410.  
<https://doi.org/https://doi.org/10.1111/1467-9280.00076>
- Simon, H. A. (1990). Bounded rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and probability* (pp. 15–18). Palgrave Macmillan UK.  
[https://doi.org/10.1007/978-1-349-20568-4\\_5](https://doi.org/10.1007/978-1-349-20568-4_5)
- Singer, M. T., & Lalich, J. (1995). *Cults in our midst*. San Francisco: Jossey-Bass Publishers.
- Smaldino, P. E., Moser, C., Pérez Velilla, A., & Werling, M. (2023). Maintaining transient diversity is a general principle for improving collective problem solving. *Perspectives on Psychological Science*, 18, 17456916231180100.  
<https://doi.org/10.1177/17456916231180100>
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others’ opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>

- Sommer, J., Musolino, J., & Hemmer, P. (2023). A hobgoblin of large minds: Troubles with consistency in belief. *Wiley Interdisciplinary Reviews: Cognitive Science*, 14(4), e1639. <https://doi.org/https://doi.org/10.1002/wcs.1639>
- Son, J.-Y., Bhandari, A., & FeldmanHall, O. (2021). Cognitive maps of social features enable flexible inference in social networks. *Proceedings of the National Academy of Sciences*, 118(39), e2021699118. <https://doi.org/10.1073/pnas.2021699118>
- Spear, A. D. (2019). Epistemic dimensions of gaslighting: Peer-disagreement, self-trust, and epistemic injustice. *Inquiry*, 0(0), 1–24.  
<https://doi.org/10.1080/0020174X.2019.1610051>
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, 25(4), 359–393.  
<https://doi.org/10.1111/j.1468-0017.2010.01394.x>
- Sun, Y., Pan, Z., & Shen, L. (2008). Understanding the third-person perception: Evidence from a meta-analysis. *Journal of Communication*, 58(2), 280–300.  
<https://doi.org/https://doi.org/10.1111/j.1460-2466.2008.00385.x>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Svensson, I. (2013). One God, Many Wars: Religious Dimensions of Armed Conflict in the Middle East and North Africa. *Civil Wars*, 15(4), 411–430.  
<https://doi.org/10.1080/13698249.2013.853409>
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109(3), 451–471.  
<https://doi.org/https://doi.org/10.1037/0033-295X.109.3.451>
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324.  
[https://doi.org/10.1016/S1364-6613\(03\)00135-9](https://doi.org/10.1016/S1364-6613(03)00135-9)

- The Cato Institute. (2020). Poll: 62% of Americans Say They Have Political Views They're Afraid to Share. Retrieved April 7, 2021, from <https://www.cato.org/survey-reports/poll-62-americans-say-they-have-political-views-theyre-afraid-share>
- Thoits, P. A. (2011). Mechanisms Linking Social Ties and Support to Physical and Mental Health. *Journal of Health and Social Behavior*, 52(2), 145–161.  
<https://doi.org/10.1177/0022146510395592>
- Tindale, C. W. (2007). *Fallacies and Argument Appraisal*. Cambridge University Press.
- Toseland, N. (2019). *Truth, “conspiracy theorists”, and theories: An ethnographic study of “truth-seeking” in contemporary Britain* [Doctoral dissertation, Durham University].  
<http://etheses.dur.ac.uk/13147/>
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3), 299–314.  
<https://doi.org/10.1037/0033-2909.103.3.299>
- van Lieshout, L. L., de Lange, F. P., & Cools, R. (2020). Why so curious? quantifying mechanisms of information seeking. *Current Opinion in Behavioral Sciences*, 35, 112–117. <https://doi.org/https://doi.org/10.1016/j.cobeha.2020.08.005>
- van Prooijen, A.-M., & Sparks, P. (2014). Attenuating initial beliefs: Increasing the acceptance of anthropogenic climate change information by reflecting on values. *Risk Analysis*, 34(5), 929–936. <https://doi.org/https://doi.org/10.1111/risa.12152>
- van Stekelenburg, A., Schaap, G., Veling, H., van't Riet, J., & Buijzen, M. (2022). Scientific-consensus communication about contested science: A preregistered meta-analysis. *Psychological Science*, 33(12), 1989–2008.  
<https://doi.org/10.1177/09567976221083219>
- Wagner-Pacifi, R., & Hall, M. (2012). Resolution of social conflict. *Annual Review of Sociology*, 38, 181–199.  
<https://doi.org/https://doi.org/10.1146/annurev-soc-081309-150110>

- Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). Children's Thinking About Diversity of Belief in the Early School Years: Judgments of Relativism, Tolerance, and Disagreeing Persons. *Child Development*, 75(3), 687–703. <https://doi.org/10.1111/j.1467-8624.2004.00701.x>
- Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology*, 37(3), 373–386. <https://doi.org/10.1037/0012-1649.37.3.373>
- Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 10(2), 145–158. <https://doi.org/10.1177/1745691615569849>
- Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to Shared Information in Social Learning. *Cognitive Science*, 42(1), 168–187. <https://doi.org/10.1111/cogs.12485>
- Whiting, D. (2020). Higher-Order Evidence. *Analysis*, 80(4), 789–807. <https://doi.org/10.1093/analys/anaa056>
- Wijenayake, S., Van Berkel, N., Kostakos, V., & Goncalves, J. (2020). Quantifying the effect of social presence on online social conformity. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1–22. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2021.102743>
- Williams, J. (2018). *Stand out of our light: Freedom and resistance in the attention economy*. Cambridge University Press.
- Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological and Personality Science*, 3(6), 767–773. <https://doi.org/https://doi.org/10.1177/1948550611434786>

- Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The Illusion of Consensus: A Failure to Distinguish Between True and False Consensus. *Psychological Science*, 30(8), 1195–1204. <https://doi.org/10.1177/0956797619856844>
- Yudkin, D. A., Hawkins, S., & Dixon, T. (2019, September). The perception gap: How false impressions are pulling americans apart. <https://doi.org/10.31234/osf.io/r3h5q>
- Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.
- Zmigrod, L. (2020). The role of cognitive rigidity in political ideologies: Theory, evidence, and future directions. *Current Opinion in Behavioral Sciences*, 34, 34–39. <https://doi.org/https://doi.org/10.1016/j.cobeha.2019.10.016>
- Zollman, K. J. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, 72(1), 17–35.