



## Misinformation, observational equivalence and the possibility of rationality

Maarten van Doorn

To cite this article: Maarten van Doorn (23 May 2024): Misinformation, observational equivalence and the possibility of rationality, Philosophical Psychology, DOI: [10.1080/09515089.2024.2358089](https://doi.org/10.1080/09515089.2024.2358089)

To link to this article: <https://doi.org/10.1080/09515089.2024.2358089>



Published online: 23 May 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



# Misinformation, observational equivalence and the possibility of rationality

Maarten van Doorn 

Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

## ABSTRACT

In vice epistemology, bad epistemic outcomes, such as maintaining false beliefs, are interpreted as indicators of blame-worthy irrationality. Conversely, a growing trend in philosophical psychology advocates for environmentalist explanations, suggesting these outcomes emerge because rational cognitive processes of faultless individuals falter due to polluted environmental inputs. Building on concrete examples, I first offer a systematic analysis of the relative explanatory merits of that environmentalist project. I then use this analysis to advance the rationality debate, which has recently been identified as stagnating due to an observational equivalence between environmentalist and vice accounts. Although the conceptual imprecision of vice epistemology has frequently been critiqued, the framework developed here reveals that environmentalism is (also) unable to meet various theoretical desiderata. I show why this is so and argue that, to make progress, environmentalism needs a more substantive conception of epistemic rationality. To this end, I propose that a closer engagement with questions of cognitive agency – how rational creatures can “make up” their minds about what to believe – could enable the necessary progress.

## ARTICLE HISTORY

Received 22 November 2023  
Accepted 16 May 2024

## KEYWORDS

Rationality; motivated reasoning; misinformation; bad beliefs; biased assimilation

## 1. Two types of explanations

Especially after 2016, it has been increasingly held that misinformation spread on social media is a grave threat (Aral & Eckles, 2019). A few years ago, for example, the World Health Organisation’s director-general announced during the COVID-19 pandemic that “we’re not just fighting a pandemic; we’re fighting an infodemic” (Zarocostas, 2020).<sup>1</sup> Trying to make sense of why people form and maintain *bad beliefs* – defined as beliefs which conflict with those held by the relevant experts and are maintained despite available counterevidence (Levy, 2021) – some scholars have taken an approach which has come to be known as *vice epistemology* (Nguyen,

2023). According to this view, facts about people's *bad epistemic behavior* account for bad epistemic outcomes

Cassam (2018, p. 5) says: "Epistemic vices are traits, attitudes or thinking styles that systematically obstruct the gaining, keeping or sharing of knowledge". A helpful illustration from Cassam involves Oliver, a conspiracy theorist who insists that 9/11 was a controlled demolition, not a terrorist attack. Since this is a belief debunked by credible research, Oliver, claims Cassam, "should have known better but doesn't" (Cassam, 2016, p. 162). Cassam argues that Oliver's failure to recognize the truth stems from his intellectual shortcomings, describing him as "gullible, cynical, and prejudiced" (Cassam, 2016, p. 163), which collectively account for his unreasonable and debunked beliefs.

McIntyre's (2018) interpretation of our so-called post-truth era can serve as another example of vice epistemology. He diagnoses our era as one characterized by "a growing international trend where some feel emboldened to try to bend reality to fit their opinions, rather than the other way around" (pp. 5–6). He substantially attributes today's alleged truth decay (Kavanagh & Rich, 2018) – a supposed diminishing role of, trust in, and respect for facts – to "logical and other conceptual mistakes that humans routinely make in reasoning" (2018, p. 40). For instance, McIntyre regularly invokes *confirmation bias* – defined as the "tendency to give more weight to information that confirms one of our preexisting beliefs" (p. 173) – to explain why people allegedly get things wrong.

"The vice approach," writes Nguyen (2023, p. 1), "focuses on the thing which believes, ask how it's supposed to behave, and then wonders why it's failed". When confronted with an observation that seems epistemically out of tune – e.g., laypeople maintaining beliefs which conflict with the beliefs held by the relevant epistemic authorities (bad beliefs) or interpreting information so that it confirms their preexisting beliefs (confirmation bias) – it centers the epistemic failings and defects of the agent and, capturing bad takes and belief-consistent information processing as the outcomes of cognitive shortcomings and motivational biases. Consequently, if ordinary people succumb to epistemic vices in these ways, then widespread political participation will not secure social goods. Therefore elites should make collective decisions in a way that bypasses the people's beliefs (Achen & Bartels, 2016; Brennan, 2016).

Recently, several what I shall call *environmentalist* explanations have surfaced as pushback to this approach (Baron & Jost, 2019; Levy, 2021, 2023a; Müller, 2024; Nguyen, 2023; Rini, 2017; Sullivan Bissett, 2022). On these views, ostensibly non-normative epistemic outcomes – such as bad beliefs and belief-consistent information processing – arise from *rational* mechanisms. The locus of the non-normativity some reasoning processes output, is not in motivated irrationality, but in the inputs

with which that reasoning works. Rather than people succumbing to epistemic vices, the cognitive mechanisms they employ in trying to get things right simply misfire when receiving contaminated input. In my reconstruction, a core assumption of environmentalism is that belief-updating mechanisms exclusively accepts information perceived as epistemically relevant (cf. Sommer et al., 2023). In seeking to understand apparent irrationality, it accordingly considers how the mechanisms' perception of epistemic relevance might have been compromised. Much like a computer system susceptible to a Trojan horse, our belief-updating mechanisms can for instance be tricked into updating on misinformation when it imitates the properties they typically filter on (Nguyen, 2023) or they can fail because of unclear communicative practices (Rini, 2017) or dysfunctional but opaque social epistemic norms (Müller, 2024).

One aim of this paper is to bring together and contrast disconnected vice and environmentalist accounts. This enables theoretical progress as it enables generalizations about these two categories of explanations and corresponding hypotheses about their overall explanatory potential. A second goal is to *make* theoretical progress. For example, many commentators on the rationality debate have argued that the controversy has reached an impasse due to the problem of “observational equivalence” (Bullock, 2009; Druckman & McGrath, 2019; B. M. Tappin et al., 2020) – roughly the dual claim that various epistemic behaviors can spring from accuracy and directional processing goals, and that it is difficult to distinguish these alternative processes with current experimental methods (what this means exactly will concern us in due course). The current framework will reveal – perhaps surprisingly – that one overlooked way forward is for environmentalism to develop a more robust positive theory of epistemic rationality in order to make good on its claim that the belief-formation of agents is often largely sensible – in an epistemic sense – and thus not deserving of blame.

I proceed as follows. I start with two examples of epistemic behaviors often seen as evidence for vice epistemology: confirmation-bias like phenomena (Section 2) and people believing false claims (Section 3). For both explananda, I compare the motivated irrationality account with the environmentalist explanation. Then I draw general lessons from these examples to elucidate the characteristics and relative merits of both types of explanation in general (Section 4). This leads to my argument that the extent to which environmentalist accounts succeed in vindicating human rationality is smaller than its proponents assume, because they render it unclear what it *means* to be epistemically rational (Section 5). In the final section I explore the implications of this conclusion for making progress in the rationality debate (Section 6).

## 2. Biased assimilation

The first behavior I wish to discuss concerns *biased assimilation*: the tendency of individuals to interpret new information in a manner that confirms to their preexisting beliefs or attitudes (Hahn & Harris, 2014). It's a tendency very similar to what McIntyre (2018) picks out as confirmation bias (as quoted above). Confirmation bias, however, makes for an unclear umbrella-term which can also refer to e.g., the way individuals *seek out* information sources (Whittlestone, 2017). "Biased assimilation" is the label often used to pick out the more specific phenomenon of individuals being prone to rate information as stronger or more convincing if it confirms vs. contradicts their prior beliefs (e.g., Koehler, 1993; B. Tappin & Gadsby, 2019).<sup>2</sup>

Numerous studies have shown that people rate studies supporting their views as more valid, convincing, and well done than those opposing their views, even when all aspects of the studies are identical except for the direction of the findings (Ditto & Lopez, 1992; Klaczynski & Gordon, 1996; Klaczynski & Narasimham, 1998; Kunda, 1987; Miller et al., 1993; Munro & Ditto, 1997; Taber & Lodge, 2006).<sup>3</sup> The classic example of a study finding biased assimilation in such a *matched-information design* is Lord et al. (1979). In their experiment, participants holding opposing views on capital punishment were presented with two fictional studies—one supporting and the other refuting the effectiveness of capital punishment as a deterrent for crime. Participants rated the study that aligned with their preexisting beliefs as more methodologically sound and convincing, despite the fact that both had similar merits.

Such differential evaluations of the same scientific method depending on whether the result is consistent or inconsistent with one's prior beliefs, are often taken as strong evidence of vice epistemology. Williams (2018, p. 141), for instance, considers it evidence for the pervasive influence of confirmation bias, described as a non-truth-seeking inferential mechanism. On such views, biased assimilation occurs because people are motivated to dismiss counterattitudinal information (Carter & McKenna, 2020). This "defensiveness [...] results in deeper and more favorable elaboration of arguments supporting those attitudes than arguments opposing them" (Lundgren & Prislín 1998, p. 715). Such motivated treatment of evidence, goes the idea, explains biased assimilation and shows people are impervious to information that contradicts their view (e.g., Kahan et al., 2011).

Accordingly, research on motivated reasoning often claims that "motivated reasoning occurs when we reason differently about evidence that supports our prior beliefs than when it contradicts those beliefs" (Caddick & Feist, 2022, p. 428). However, it's too quick to say that if subjects with different political background beliefs about a fact (say, that the death penalty

is a good idea) provide different assessments of the quality of evidence about that fact (e.g., a study about the efficacy of capital punishment), then they are engaged in motivated reasoning (Van Doorn, 2023b). Individuals with varying political affiliations often hold distinct initial beliefs, which naturally result in divergent subsequent beliefs regarding the credibility of evidence or the accuracy of facts. This divergence is especially evident in contexts such as opinions on the deterrent effects of capital punishment, where differing initial beliefs shape interpretations of new data according to Bayesian updating principles (Gerber & Green, 1999; Hahn & Harris, 2014; Koehler, 1993). Therefore, when a study presents conclusions that contradict a person's existing knowledge, it is reasonable for her to question the reliability of that study.

For example, those in favor of capital punishment (presumably) genuinely think that it serves as a deterrent. Consequently, they naturally reason that studies purporting to show the opposite are more likely to be invalid. It seems more straightforward to assume, then, that the key aspect of the study which leads them to discount it is its presentation of an incorrect conclusion, rather than a conclusion that merely diverges from their personal views or desires (cf. Stanovich, 2021 on knowledge projection). From this perspective, the criterion they consider crucially important for judging the study's validity is that it presents a conclusion opposite to what they believe to be true, represented as not- $p$  when they hold  $p$  (Enoch, 2010). So at least, the fact that people evaluate the quality of otherwise identical studies more favorably when they support their views is consistent with non-motivated cognition (Stanovich, 2021; B. M. Tappin et al., 2020; Van Doorn, 2023b).

Given a set of prior beliefs, characteristics of information ecologies and basic principles of information processing, conclusions appearing profoundly biased from an Archimedian point of view can arise. This can occur despite the fact that the alleged epistemic sinner, from a first-person point of view, was in fact making a good-faith effort to assess new information accurately, and relying on her prior beliefs to guide such judgments. Biased assimilation is a product of this process, occasionally leading us astray by affirming a fallacy. How, according to environmentalism, does our polluted epistemic environment account for its misfirings?

If biased assimilation – i.e., when people condition their evaluation of the reliability of information on their prior beliefs – misfires, say environmentalists, this is not because people have fallen for an irrational confirmation bias, but because the epistemic environment has corrupted a mechanism. How would that work? In the realm of practical reasoning, there's a widely acknowledged need for strategies that help finite minds like ours decide when to halt our deliberations (Bratman, 1987; Nguyen, 2023). This involves making a conscious choice to stop seeking new information. Consider a scenario where I have a free evening with no

obligations the next day. Should I watch several episodes of *Stranger Things*, play video games, or see friends? I could ponder endlessly as new thoughts pop up. However, I'd soon reach a point where further thinking becomes unproductive and drains my cognitive energy. The solution is to set an intention, essentially a decision or plan, that resolves my internal debate (Bratman, 1987). If I choose to watch *Stranger Things* this evening, I effectively seal off my mind from further debate on how to spend my time. This cessation of deliberation is a practical response to the limitations of our cognitive capacities, preventing us from being overwhelmed by excessive information (Nguyen, 2023). Since we are constantly bombarded with more data than we can possibly analyze thoroughly, it's crucial to use heuristics to gauge what merits our attention and what does not.

The extent to which something contradicts one's priors seems a reasonable heuristic, as priors summarize past evidence (Gerber & Green, 1999). When cognitive resource management is guided by prior beliefs, this may be a rational response to the impossibility of complete evidence evaluation. Prior beliefs may this way affect whether or not a piece of information is subjectively informative at all for one's beliefs. If you deem it highly improbable that the information I have provided is accurate, then the expected significance of this evidence is minimal, allowing you to justifiably dismiss it without further scrutiny (Fatollahi, 2022). After all, considering new evidence is not a cost-free exercise, but requires cognitive expenditure (in terms of revising previous beliefs, and so on). So if your credence in a hypothesis is high or your cognitive resources are scarce, it's rational to discount purported evidence contradicting that belief as misleading or unreliable evidence without having invested much time (Fatollahi, 2022). This predicts errors as accuracy-motivated reasoners will still sometimes form beliefs based on the wrong subset of the evidence.

This is how rational management of cognitive resources can explain failings of biased assimilation. Our overwhelming and polluted epistemic environment rationally increases the extent to which we use our prior beliefs as a heuristic for informational quality without investigating further (and opening ourselves up to being misled by misinformation). And so "the idea that there should be a degree of 'stickiness' about the beliefs one already holds", because people condition their evaluation of the reliability of information on their prior beliefs (biased assimilation), "far from showing that individual rationality does not hold, seems to be something we might *predict*, given the assumption that individuals are, broadly speaking, rational" (Borg, 2022, p. 212).

Of course, that such a practice is licensed by practical rationality does not rule out its having adverse consequences on individual credences. But it does mean that belief webs that, from the outside, smell like the product of

vicious epistemology, may have resulted from decisions motivated by accuracy. This is the possibility environmentalism implores us to take seriously.

The fact that environmentalist explanations have the capacity to account for patterns of biased assimilation does not, of course, validate them on its own. It merely suggests that vice epistemology doesn't yet gain (much) plausibility points from experimental data on biased assimilation, absent an argument that its account is superior to environmentalist explanations (this is a symmetrical conclusion). Since both accounts make different predictions about this, one natural follow-up question would be: what was going on in the head of the reasoner? Was she responding to subjective evidence, not particularly motivated to confirm one viewpoint over another, or, alternatively, actively trying to mold the information so as to reach her desired conclusion?

We'll come back to this after discussing a related explanandum of which vice epistemology and environmentalist offer competing accounts.

### 3. Bad beliefs

A similar problem arises with another phenomenon usually taken as clear evidence for vice epistemology: the prevalence of what Levy (2021) calls "bad beliefs" – beliefs at odds with expert consensus and/or widely available scientific evidence, such as climate-change skepticism.

The most prominent vice-epistemology explanation for why people hold such beliefs refers to motivated cognition (Williams, 2021). That is, social concerns motivate people to embrace biased beliefs, powerfully distorting the processes by which they seek out and process related information. Specifically, rather than doing the best we can in making sense of the available evidence, we engage in *identity-protective cognition* instead (Kahan, 2012; Kahan et al., 2017). We engage in motivated reasoning to reject some hypothesis because it is threatening to our group identity, twisting the facts to uphold our tribe's narrative regardless. This explains, in the case of climate-change skepticism, why conservatives reject the science: for them, motivated irrationality kicks in on these topics specifically (Bardon, 2020). So rather than responding to evidence, they stick to the beliefs they share with others similar to them.

This perspective draws from observed links between levels of cognitive skills (such as cognitive reflection, numeracy, and education) and the degree of partisan divergence on controversial topics (Drummond & Fischhoff, 2017; Kahan et al., 2012; Nurse & Grant, 2020). For instance, individuals with higher numerical skills and more education often display more pronounced partisan splits on issues like climate change, instead of converging toward a scientific consensus (Drummond & Fischhoff, 2017; Hamilton, 2011; Kahan et al., 2012; Malka et al., 2009; McPhetres et al., 2019). The



question then arises: why does this occur? This phenomenon can be attributed to the fact that enhanced cognitive capabilities can intensify polarization, as individuals use their mental faculties not just for objective analysis but to bolster their political identities and dismiss conflicting information (Kahan et al., 2012, 2017). Deliberation thus becomes a tool not for uncovering truth but for reinforcing preexisting beliefs and rejecting discordant data, leading to significant disparities in the credibility assigned to evidence and resulting in widespread misconceptions, such as the rejection of climate science by some conservatives, who also generally exhibit lower trust in scientific institutions (Funk & Kennedy, 2019; Gauchat, 2012; Pew Research Center, 2017).

However, again, climate change misbelief might also be explained by people making a good-faith effort to get at the truth in a polluted epistemic environment filled with untransparently misleading epistemic signals. On the environmentalist account put forward by Levy (2021, p. xiii), “those who come to hold bad beliefs do so for roughly the same sort of reasons as those who come to hold good beliefs”. Instead of being the consequence of deficient cognitive habits, these erroneous beliefs are generally rational; they are developed and sustained through the same social learning mechanisms that underpin the majority of our belief-formation activities. (cf. Bago et al., 2023). Rather than employing a distinct kind of identity-protective cognition, climate skeptics might just be responding to everyday subjective evidence.

Specifically, the evidence to which conservatives are responding is the fact that people in their community believe  $p$ , which is higher-order evidence for  $p$ . This practice is not only evolutionarily adaptive but also “directly” rational (Levy, 2021, pp. 131, 142). Since the fact that a proposition is socially approved is higher-order evidence that bears on its truth, it’s not inconsistent with a motivation for accuracy to let oneself be guided by it. But then, the argument goes, those who hold beliefs counter to the scientific consensus on matters like climate change are just deferring to their communities in this appropriate way (Levy, 2021, pp. 82–84). Agents aren’t trying to protect their political identity, but simply following the epistemic norms they’re familiar with, specifying the (ir)relevance of certain kinds of higher-order evidence (Müller, 2024). People might very well deploy social referencing to respond to social cues *as* evidence – not *rather than* evidence. On this view, it’s no surprise that conservatives tend to downplay climate change when those whom they trust tell them that it is a hoax – in fact, by this dynamic, “the acceptance of bad beliefs is to be expected of rational agents under certain conditions” (Levy, 2021, p. 10).

Isomorphic to the previous section, what we have then is a scenario where an epistemic outcome ostensibly lending itself to an explanation in terms of vice epistemology, can be given a reading that does not invoke cognitive

failing as an explanans. But where in fact, the observed outcome can be *expected* assuming rationality.

As will become clearer in the next section, the main difference between environmentalism and vice epistemology lies in the *epistemically relevant feature of the situation* they take reasoners to be responding to; what they hypothesize is the feature that in subject's minds justifies e.g., demoting a pro-climate-change study. Motivated irrationality says conservatives believe what they do because they're responding to *identity-protective* properties of (potential) beliefs. Environmentalism claims these people reject the science of climate change because the social mechanisms of belief updating provide them with *epistemic* reasons to do so. So rather than being blinded by social concerns, climate deniers may in fact be making a good-faith effort to form accurate beliefs. It just happens that their epistemic environment is polluted such that this rational practice of deference leads them to false beliefs. The blame for climate denial would then rest on polluters of the epistemic environment – producers and distributors of climate misinformation – rather than on motivated irrationality.

#### 4. Observational equivalence and other comparisons

In this section, I'll generalize over the two case studies (Sections 2 and 3) to infer general properties of vice- and environmentalist explanations. Characterizing those contrasting categories will help us to get a clear sense of how to make progress in the rationality debate.

Perhaps the most obvious difference between these two<sup>4</sup> types of explanation is the amount of blame they assign to individual reasoners. For example, vice explanations understand biased assimilation as an irrational confirmation bias run amok. Environmentalists see it as the result of agents with prior beliefs they *actually believe* who increasingly have to rely on them to gauge new information in an environment with too much (mis)information to critically evaluate in a comprehensive and balanced manner. In explaining bad beliefs, one side may thus invoke our nature as “vigilant, devious, self-deceiving, coalitional apes” (Williams, 2023a), and the other proposes not to understand bad beliefs as the products of motivated irrationality, but as the products of gamed accuracy-oriented processes.

By contrast, both sides might agree that, what we *are* responsible for, is responding appropriately to signs of error. However, this process of error correction can only take place if we have access to the evidence of our errors, which can be concealed. If we are brought to trust and distrust wrongly, then “if we have been convinced to settle our minds in certain directions [we] can miss, or dismiss, the evidence of our error” (Nguyen, 2023, p. 21). Being constantly on the alert for errors is also an unrealistic expectation. We cannot investigate every possibility or scrutinize every detail in search of

our missteps, as our capacities are limited. Our cognitive resources are constrained, and many things can remain hidden outside the narrow beam of our attentive gaze (cf. Smith & Archer, 2020). An individual will consider (first-order) evidence in forming beliefs only if it captures her attention sufficiently; evidence that is not noticed is effectively non-existent from the perspective of the agent (Munton, 2023). Furthermore, inferring the evidential value of new information requires the activation of relevant prior beliefs. Because people cannot notice the relevance of all the information they encounter for all of the beliefs that information is related to, much of what we experience will not be updated on. This way, the manipulation of an epistemic environment's salience structure – understood as an ordering of information by accessibility in a given epistemic environment (Munton, 2023) – can lead to individuals forming false beliefs, even when the information available to them is accurate. For example, media's disproportionate reporting on rare occurrences, such as crimes (truly) committed by immigrants, can (nonetheless) create false perceptions about a particular group. In such a context, environmentalists will argue, the individual reasoner is not primarily at fault for these misconceptions. Rather, the blame lies with those who misleadingly organize and emphasize information to influence belief-forming processes: “By making these events disproportionately salient to the public, false beliefs about this group are induced” (Dutilh Novaes, 2023, p. 8).

Vice epistemologists, however, might protest that talking about beliefs in this way – as things that are *induced* by an environment – misses a vital facet of how beliefs are shaped. We do not only form beliefs by passively absorbing information. We consciously decide what or whom to trust, guided by factors like our personality, core beliefs, identity, and epistemic strategies. Merely highlighting which views are conventional and supported by science will, for instance, primarily benefit those who have already opted to align their trust with mainstream scientific perspectives. In this view, social deference can be (ir)rational not in virtue of being a passive mechanism absorbing the prevalent or opinion within one's environment, but in virtue of a powerful active component: one's capacity to (in)adequately pick who to trust (Murphy-Hollies & Caporuscio, 2023; Worsnip, 2019). Similarly, insofar as attention also involves some degree of voluntary control, bad beliefs and unwarranted mind-settling may be mitigated not only by epistemic engineering of the environment but also by agents carefully curating their attentional regime (Gardiner, 2022). So “bad believers” may well be at least partially responsible for how they (mis)manage their attention and trust.

This difference is echoed in how environmentalist and vice epistemology differ with regards to their preferred locus of intervention. Should the environment be acknowledged as the principal determinant in shaping

beliefs, we can artificially generate better beliefs by cleaning up the epistemic environment (Levy, 2021). By emphasizing the perspectives that are grounded in science as mainstream, and nudging individuals toward these well-founded positions, we could not only mitigate the spread of conspiracy theories but also contribute to the cultivation of more proficient and effective epistemic individuals. By contrast, if individuals bear a greater degree of responsibility, a simple epistemic depolluting won't be enough. Changes must be made to accomplish other things, such as preventing people from developing anti-scientific identities or increasing their critical thinking skills and open-mindedness (Hare, 2003).<sup>5</sup>

A more subtle difference concerns the explanatory vantage point, as it were. Environmentalist accounts – the label notwithstanding – emphasize that people respond to *subjective* evidence. And so bad beliefs can arise because “evidence that one rationally interprets as leading to a certain conclusion may be false, because what seems true may not be” (Levy, 2023a, p. 3). After all, if we aim for rationality, we make ourselves vulnerable for deception (Vavova, 2023). Evidence can often be subtly deceptive yet highly persuasive. Depending on our epistemic environment, attempting to be rational can thus mislead us, and this is simply a consequence of being fallible creatures. Though we may wish to be open-minded toward truth and not falsehood, open-mindedness does not favor one over the other. The very characteristic that elevates it as a virtue is what can also prompt us to believe in untruthful evidence. If individuals possessed infinite time and cognitive resources, greater open-mindedness could invariably be advantageous. However, this scenario is neither practical nor particularly useful. Consequently, open-mindedness should be viewed not as an ultimate goal, but as part of a trade-off involving the stability of holding fixed beliefs versus the flexibility of adjusting one's perspectives. What's often called “confirmation bias” is therefore not a quirky anomaly of human cognition but simply a skewed but not necessarily unreasonable configuration of this trade-off. It sits on a continuum with rational Bayesian reasoning. Accordingly, say environmentalists, there's no quick inference to the best explanation from observing epistemic outcomes that appear non-normative from an external point of view – the vantage point preferred by vice epistemology – to the conclusion that motivated irrationality was involved. Looking at biased assimilation or bad beliefs from an Archimedian point of view, vice epistemology must therefore resist the temptation to “presume directional motivations exist without direct evidence that such a motive played a role among a sample of participants” (Bayes et al., 2020, p. 1036).

The topic of motivation brings me to the next comparative factor. Environmentalism usually purports to show that behavior (e.g., biased assimilation) and outcomes (e.g., arriving at bad beliefs) that look irrational from the outside are actually generated by good-faith responses to subjective

evidence rather than by motivated irrationality. So even though “almost everyone continues to form and revise their beliefs in the light of genuine evidence” (Levy, 2023b, p. 2), they arrive at bad beliefs because the epistemic environment abounds with misleading evidence. As Taber and Lodge (2006, p. 767) write, people are “often unable to escape the pull of their prior beliefs, which guide the processing of new information in predictable ways” (Taber & Lodge, 2006, p. 767). Even though such observations are often interpreted as the result of motivated biased reasoning – such as in Taber and Lodge (2006) – environmentalism emphasizes that they might just as well be the product of an attempt to form accurate beliefs in which people judge new evidence in light of coherence with beliefs they’ve formed so far based on past information – a process potentially consistent with Bayesian rationality. In these ways, environmentalist explanations assume people are often motivated – if they are motivated at all – to have beliefs that are true (enough). Rather than to, for example, protect their identity or get their opinions confirmed.

By contrast, vice epistemology tends to assume people *are* usually motivated by something like that, rather than by accuracy or not motivated at all. From its Archimedean viewpoint, it sees behavior Y, and uses that as evidence to infer some Y-motive that closely resembles it. For example, one observes that people are prone to rate information as stronger or more convincing if it confirms vs. contradicts their prior beliefs (biased assimilation), and uses this to infer that they’re unreasonably motivated to seek confirmation of their beliefs – that the presence of biased assimilation must be *caused* by an unreasonable motivation to have one’s beliefs confirmed. As we’ve seen, such motives are a sufficient cause of biased assimilation – just like identity-protective cognition would be a sufficient cause for bad beliefs – but neither is a *necessary* cause. The same, of course, goes for the unobserved motivations stipulated by environmentalist accounts.

So: according to the latter, in experiments showing biased assimilation like Lord et al. (1979), what’s going in the subject’s reasoning when they rate the belief-confirming study higher than the uncongenial one, is that it suggests a *false* conclusion, not that it suggests a conclusion that deviates from their worldview, desires or tribal commitments. By contrast, vice epistemology takes the epistemically relevant feature of the study that in the subject’s mind justifies demoting it to be exactly something like that: people downgrade the uncongenial study precisely because they don’t like it (it disconfirms their ideas or tribal core beliefs). On this view, the feature of the situation subjects take to be of normative epistemic significance – what their reason is for making up their mind about the study’s reliability – is not simply accuracy – the study suggests *not-p*, whereas *p* – but crucially contains an ingredient of motivated irrationality such as confirmation bias or identity-protective cognition.

This might be the sharpest point of disagreement. It's unfortunate then, that it's not clear how to adjudicate it. This is because it's hard to know what cognitive mechanism people employ just by looking at their behavior. Environmentalist accounts have rightly pointed out that phenomena like biased assimilation and bad beliefs *need* not indicate that people take belief-confirmation or identity-protection as the epistemically relevant feature, as such patterns could also arise from subjects making a good-faith effort at forming accurate beliefs. What reasoning *does* underlie such outcomes remains unclear, as the conclusions that participants are assumed to be motivated to reach in such experiments correlate with what they in fact believe (B. M. Tappin et al., 2020). This is known as the observational equivalence paradox of motivated reasoning (Bullock, 2009; Druckman & McGrath, 2019).

For example, studies have often pointed to a correlation between conservative political beliefs and the rejection of climate science. Their designs, however, predominantly provide insight into group-level phenomena, suggesting that conservatives as a cohort engage in motivated reasoning by dismissing scientific evidence that contradicts their cultural identity. While group-level data might indicate that conservatives reject climate science, this aggregate observation may stem from a variety of individual-level psychological mechanisms. Perhaps motivated irrationality distorts the processes by which they seek out and process information. But it seems equally compatible with the data that such patterns arise not from motivated irrationality, but from people with different beliefs about – among other things – what constitutes accurate information, seeking to accurately interpret novel information in light of these divergent beliefs. Many commentators have consequently argued that the evidence cannot distinguish between the hypothesis of politically motivated reasoning and the alternative hypothesis of a non-motivated effect of prior beliefs (e.g., B. M. Tappin et al., 2021, p. 1112).

However, scientific theories should not be assessed solely on their ability to match observational data. Instead, a variety of factors should be taken into account, including the simplicity of the theory, its alignment with well-established scientific findings, its capacity to connect various findings and reveal underlying patterns, and its potential to lead to unexpected predictions (Lipton, 2017; Newton-Smith, 2002). For example, environmentalist theories might try to invoke Ockham's razor here. They explain the data relying only on the assumption that humans are straightforward Bayesian information processors using their model of the world to make sense of incoming information (cf. Hohwy, 2017). We then don't need to *further* stipulate that their information processing is causally influenced by motivated irrationality, since phenomena justifying this ontological addition can in fact also be explained by Bayesian mechanisms (Sommer et al., 2023,

p. 30). And if e.g., identity-protective motivations are explanatory superfluous why should we add to our ontology these motivations as an extra force causally affecting human information processing?

We can, for instance, more parsimoniously point out that testimonial networks are by default taken to add epistemic value to the information that propagates through them, so any conservative need not be busy protecting her identity when she defers to networks of laypeople she has good reason to think are appropriately plugged into experts (Levy, 2023b). When a belief that reverberates around this network is near-universally held with high confidence, her deference is perhaps not a case that requires a special motivation for its explanation. She simply conforms to the beliefs of those around her, whose behavior and assertions are best explained by attributing to them a matching belief, secured by testimony from news sources and politicians, which, she has good reason to believe, receive testimony from genuine experts. This behavior doesn't seem to require an extra (motivated-irrationality) element in its explanation on top of (accuracy-oriented) bread-and-butter updating in response to higher-order evidence.<sup>6</sup>

This argument, however, rests heavily on what should be our default assumption. Vice epistemology might object that ontological sparseness is a criterion for theory evaluation that is relatively less important. Especially since it might not be immediately clear how it applies to mental states such as motivations. And so we should compare plausibility points for other criteria first, before going there. It's worth noting, of course, that a theory that says people's opinion is sensitive to changes in social cues because they see other people's opinion as higher-order evidence would seem to make the same predictions as a theory saying that beliefs are similarly sensitive to social cues but then because people believe what they believe to fit in rather than to get things right and so they must respond to social cues to achieve this goal. Because reasoning experiments focus on aggregate data, they give limited insights into individual-level mechanisms. Just by looking at differences between conditions, it is difficult to know how individuals were thinking (cf. Molenaar, 2004).<sup>7</sup>

Accounts of identity-protective cognition infer motivated reasoning when e.g., co-partisan sources are deemed more credible than uncongenial ones. This way of thinking is also consistent with an accuracy motivation, however, as co-partisan sources *are*, from the subject's point of view, often more likely to get things right and act reliably (Lepoutre, 2020; Levy, 2021; Rini, 2017; Van Doorn, *in press*). To address the observational equivalence paradox, experimental evidence must therefore specifically isolate social motivations from other cognitive mechanisms. Here it becomes crucial that, to my knowledge, identity-protective interpretations of the data simply *assume* on theoretical grounds that social motivations strongly influence belief formation (e.g., Kahan, 2015). Accordingly, a relatively simple – and

to my knowledge, untried – experimental innovation could involve directly querying participants about their concerns regarding social standing or friendships in relation to changing their beliefs on contentious issues.<sup>8</sup> If individual variation there predicts variation in reasoning outcomes, that would be evidence for identity-protective cognition. By contrast, absence of such a relation would *prima facie* constitute weak Bayesian evidence for environmentalism.<sup>9</sup>

Harnessing such potential experimental results for theoretical progress, however, is somewhat hampered by what others have called the “conceptual imprecision of politically motivated reasoning” (B. M. Tappin et al., 2020, p. 85). We see this, for instance, in McIntyre’s quoted definition of confirmation bias. Confirmation bias is a prominent element in explanations offered by motivated reasoning theory but, as argued, it seems consistent with Bayesian reasoning. So it’s not clear what precise phenomena its alleged existence allows vice epistemology to give a superior explanation of.

It has less often been noted, on the other hand, that a sophisticated theory is similarly missing for environmentalism. In the next section, I use the above analysis to show that, if its identified characteristics are on the right track, environmentalism lacks the resources to make sense of rationality (in any robust way). This undermines its aspirations, and renders its status as an alternative to vice epistemology less than clear.

## 5. Rationality worth wanting?

On environmentalist accounts, to recap, ostensible epistemic failings arise not because humans reason poorly or have lost concern for the truth. Rather, they reflect rational cognitive processes given bad inputs. The local conclusion then is supposed to be that phenomena such as biased assimilation and bad beliefs do not convincingly demonstrate human irrationality. The more general aim is to vindicate human reasoning, or at least fend off pessimistic pictures of it (such as vice epistemology and post-truth narratives). However, a rarely noted paradox is that the explanatory centrality of epistemic environments frustrates that larger aim. In this section, I present several arguments for this novel conclusion drawing on the explanatory features identified above.

As an example, let’s take a closer look at Levy’s (2021) claim that instead of stemming from deficient individual cognitive habits, erroneous beliefs are largely rational. They arise and persist through the same social learning mechanisms that shape most of our belief-forming processes, primarily through the updating of our beliefs based on testimonial evidence from sources we trust. Levy emphasizes this deference “ingrained and fluent” (p. 83), “smooth and automatic” (p. 133). It converts evidence to belief “ubiquitously and . . . routinely” (p. 62). But



then, as Gadsby (2023, p. 782) points out, one might wonder whether bad beliefs are more accurately characterized as “the product of genuinely and wholly arational processes”. Specifically, they now appear as an automatic belief-updating effect, instances of which have been documented in the psychological literature. For example, participants tend to revise their beliefs in response to evidence even when they are informed of its unreliability (Wegner et al., 1985). A widely accepted view of this behavior is that individuals tend to adjust their beliefs automatically upon hearing new statements, regardless of their validity (Mandelbaum, 2014). This suggests that the updating of beliefs might happen irrespective of the actual evidential value of the statements presented. The belief-updating mechanism underpinning bad beliefs might then not be a genuine case of evidence-responsiveness, nor evidence for accuracy motivation, but simply arational updating.<sup>10</sup>

I’m personally not extremely confident in the empirical details of automatic updating.<sup>11</sup> But I bring it up to show that the arationality-conclusion arises from the general structure of environmentalist explanations (and so does not depend on these details). Here’s another example. While Levy argues that the extent and direction of deferential updating is often justified by the *testimonial* evidence that triggers it, Coppock (2022) documents many cases where the direction of non-deferential updating is justified by the *first-order* evidence that triggers it. His basic claim is that the effects of persuasive information on attitudes are “small, positive, and durable for everyone” (Coppock, 2022, p. 51). “Positive” here means that people move in the direction of the information provided, regardless of their prior belief. That is, if supporters of the death penalty and opponents see new information indicating that the death penalty reduces crime, they both will become more favorable toward the death penalty. Coppock’s work seems to indicate that people are more responsive to reasons and evidence than often presumed, and, conversely, cases of belief polarization and backfire effects are more exceptional than often assumed (Anglin, 2019; Velez & Liu, 2023). If true, that would seem to count against vice epistemology and in favor of environmentalism. But in fact, the deeper picture emerging from such accounts again seems a peg *too* environmentalist for its own purpose of vindicating human rationality. In essence, the image is one on which our beliefs passively emerge as the sum of the persuasive information we’ve encountered. Unbeknownst to us, we continuously update in the direction of the information, by a small amount. This “happens many millions of times a day as people scroll through social media feeds and talk with friends and coworkers about the news” (Coppock, 2022, p. 5). Given the constant influx of both explicit and implicit evidence, from first-order to higher-order, our beliefs are gradually shaped by the surrounding information landscape.

This ongoing exposure leads to subtle, often unnoticed shifts that, particularly in a polluted epistemic environment, can culminate in the formation of erroneous beliefs. If such accounts succeed in reclaiming explanatory territory from vice epistemology, they do so, I want to suggest, at the cost of our rational agency. I argue for this conclusion in the remainder of this section.

Environmentalism redescribes unfortunate epistemic outcomes as being mainly due to a polluted environment “hacking” cognitive mechanisms. It seems a logical consequence that such theories focus on the environment in terms of intervention: by cleaning up epistemic pollution, those mechanisms will be less likely to yield bad beliefs. While that may be true, it should be disconcerting to opponents of vice epistemology that their alternative is to paint a picture on which the degree to which our beliefs have positive epistemic status depends on factors entirely outside of our own cognitive agency. The quality of beliefs now just seems to reduce to the quality of the “environmental structuring that spawned them” (Gadsby, 2023, p. 783). If this is how beliefs are formed, it seems to imply we should treat our own beliefs as the sober and logical results of millions of tiny shifts in our credences produced by our circumstances, rather than as entities connected to our agency for which we are (therefore) responsible.

Indeed, Levy (2023b, p. 855) admits that, on his environmentalist view, “the difference between the agent whose inherited epistemic resources are reliable and the agent who lacks such resources” comes down to luck. He bites this bullet, but I’m not sure he fully appreciates the costs of doing so.

This firstly wed environmentalism to a conception of our cognitive agency which seems to imply we should treat ourselves as mere “truthometers” (Enoch, 2010) – merely devices with a certain probability of tracking the truth. On this picture, we are each just a set of cognitive mechanisms with a certain probability of issuing a true “reading” of whatever claim, depending on the quality of our inputs. But if this is how we should see ourselves, we should treat our reflective response to evidence as a mere *seeming*: we can merely observe the outcome of our belief mechanisms – read our own truthometer – but not exercise agency in its coming to be. Just like when we each taste a sample of food; I find one sample spicier than the other, whereas you perceive them as equally mild. Upon learning about our differing perceptions, you might reconsider and adjust your initial judgment to align more closely with the shared data. Similarly, if we view ourselves as merely systems calibrated to various accuracies in sensing truth, the way to revise your own views given the evidence of other truthometers, is similarly spineless.

In matters of surface-level appearances, the first-person perspective can be easily dismissed. However, in disputes centered around deeply rational beliefs – those informed by thorough reflection on the evidence, where the

individual's full cognitive engagement is evident – the first-person perspective is not unproblematically eliminable. As Enoch (2010, p. 963) puts it: “Once you reflect on a question, asking yourself, as it were, what is the truth of the matter, and so what is to be believed once the believing self is fully engaged – you can no longer eliminate yourself and your reflection in the way apparently called for by the truthometer view.” This is a requirement that cannot be universally complied with, but it seems the logical upshot of arguing against vice epistemology by making epistemic environment the ultimate explanans.

Especially when considering the first-person question: “What am I to believe in the face of this conflict?”, this seems to me an unsatisfactory account. Here, the issue is not merely about neutral arbitration between conflicting truthometers, but involves a personal, internal deliberation wherein one employs their own faculties, procedures, and opinions – even those that might be under scrutiny – to arrive at a belief. This self-reflective process is not just an optional layer of cognitive activity; it is essential for the exercise of rationality. In such cases, it seems misleading to think about the disagreement in terms of a model of neutral arbitration between conflicting automata (Foley, 2001, p. 79). Yet it seems to me that is the exact reading proposed by environmentalism when it conceives of beliefs as the result of automatic mechanisms and a particular epistemic environment.

A related problem I see is that, insofar as environmentalists emphasize the automatic nature of belief-updating to move the blame for phenomena like bad beliefs from the individual to the environment, the normative focal point of the rationality debate seems to disappear from view. This is because talk of justification or rationality has a target only where we are willing to impute responsibility (Owens, 2002, p. 5). Even if belief is mostly automatic, for genuine normative epistemic assessment, there must be cognitive acts such as deliberation and judgment within the purview of human agency. That would be where epistemic norms get a grip on us. It's not clear what it means to say that people are “rational” if, at the same time, they hardly have cognitive agency: saying that the quality of one's cognitive life is a matter of luck is saying that it's not a matter of responsibility. So rather than pushing back against vice epistemology by giving a positive account of first-person reasoning that meets certain epistemic norms and so is not adequately characterized as vicious, environmentalism's gambit is to remove the space for any individual epistemic contributions and corresponding rationality assessments. This is an important difference. On the assumption that genuine normative demands concern things in some way connected to our agency, and insofar as environmentalism seeks to uphold human rationality, I take this to be a major weakness of the approach.

Accusing someone of irrationality involves critiquing their beliefs, inherently assuming that the person holds responsibility for holding those beliefs

(Schmidt, 2020). In my view, environmentalism loses much of its appeal if this agential aspect of rationality needs to be sacrificed to counter these charges. Yet strictly speaking it would be open to defenders of the account to insist they're not seeking to *vindicate* our rationality, but merely to *deny* charges of irrationality by pointing out that evaluations of rational and irrational require that the subject who is so evaluated bears responsibility (Kiesewetter, 2017), and then questioning this assumption. This however, it seems to me, sits ill with the spirit of the view.

A further problem has to do with extensional adequacy. On the environmentalist view, how an agent makes up her mind seems to be a function of (1) built-in mechanisms for belief updating that automatically adjust beliefs when presented with specific types of information and (2) the environment supplying these inputs. Probably we can account for quite some variation in epistemic outcomes in this way, but there are also cases in which explaining away first-person reasoning leaves the view with a deficit. Cases like Mary Wollstonecraft (1792), for example, who disputed the almost universal consensus in her eighteenth-century society concerning the supposed cognitive inferiority of women – a consensus affirmed by almost all those socially designated with epistemic authority – constitute instances of epistemic rationality that seem hard to plausibly explain within the environmentalist framework (Williams, 2023a). Her cognitive mechanisms must have received different inputs; different propositions had the higher-order evidence of being believed by similar people counting in favor for her, or something like that. But that just seems the wrong diagnosis.<sup>12</sup>

A related problem, I think, is that it might not be possible to give a satisfactory account of what rationality-conferring mechanisms exists, that succeeds in avoiding a proliferation of cognitive-processes-working-as-they're-supposed-to that, given bad inputs, can nonetheless yield epistemic behavior that looks irrational from an Archimedian point of view. In order to make plausible the existence of a plurality of rationality-conferring cognitive mechanisms, one needs to develop a systematic theory that explains how these mechanisms work and how they are to be individuated. If all reasoning outcomes could be accounted for this way, the theory comes close to being unfalsifiable and we lose out on a useful picture of both irrationality and rationality (cf. Murphy-Hollies & Caporuscio, 2023).

To alleviate this concern, we'd need to know more about the postulated suite of rational cognitive mechanisms, what input they accept as informative and so how the mechanism can be gamed while still updating on perceived epistemic relevance (and not, e.g., desirability). That would allow us to make specific predictions about in which circumstances bad beliefs can arise in the absence of motivated irrationality. Without specifics on how this works, the environmentalist explanation risks being a just-so story – an ad-hoc explanation tailored to fit the outcome, rather than

a predictive or explanatory theory. For example, Levy's (2021) claim that Republican climate denial can be explained without positing motivated irrationality because the fact that a proposition is socially approved is higher-order evidence that bears on its truth, seems fine as far as it goes. But it seems unsatisfactory as a place to turn our spades. For obviously not all Republicans use this mechanism to determine their views on climate change while (*ex hypothesi*) their social reference group is similar (cf. Bago et al., 2023). Does that make them irrational? Or are they relying on *another* rational mechanism? If we accept that a myriad of combinations of rational mechanisms and manipulated inputs can lead to similar belief outcomes, we're left with a model that lacks predictive specificity. We need to know which cognitive phenomena *cannot* be adequately explained by a polluted epistemic environment mimicking the properties rational belief-updating mechanisms are designed to filter on. Otherwise it seems to me environmentalism is unable to accommodate the possibility of culpable irrationality. We could construct environmentalist explanations for all these violations and, environmentalism would lack the resources to accommodate the existence of a normatively relevant difference between them.

## 6. Conclusion

A brief recap seems in order before I respond to some objections and sketch a way forward. The surveyed environmentalist accounts have in common that they argue we do not lose ourselves in motivated irrationality. But they do so by demonstrating that we, as individual reasoners, barely contribute to belief formation. And so, if I'm right, the contrast between vice epistemology and environmentalist accounts is not what it seems. Either the agent contributes nothing of significance (environmentalism) or she contributes by being irrational (vice epistemology).

For environmentalism, the problem is that recasting non-normative epistemic outcomes as nothing more than the output of our cognitive system given certain inputs seems to imply a bleak picture of cognitive agency, at odds with its vindicatory spirit. It's winning the battle but losing the rationality wars: A theory on which the reasoning of individual agents has no place – is irrelevant – does not obviously paint a more positive picture of human rationality than vice epistemology. This, as mentioned, is not in itself an argument against environmentalist explanations. The point is rather that (as things stand), we should not be too quick to see vice epistemology's main competitor as a view that does make place for a substantive positive conception of human rationality – even though it defends the claim that most (bad) beliefs arise through rational mechanisms.

If it's not such a direct counterargument, doesn't that make my point moot? In response, consider that, historically, a range of philosophers have

defended Kantian views along the lines that “judging, making up our minds what to think, is something for which we are, in principle, responsible – something we freely do, as opposed to something that merely happens in our lives [. . .] But even when a belief is not freely adopted, it is an actualisation of capacities of a kind, the conceptual, whose paradigmatic mode of actualisation is in the exercise of freedom that judging is. This freedom, exemplified in responsible acts of judging, is essentially a matter of being answerable to criticism in the light of rationally relevant considerations” (McDowell, 1998, p. 434). It’s very much worth noting, I believe, that an important theory in the philosophy of social sciences essentially argues for human rationality by minimizing the extent to which belief is governed by reason in this way. Vindicating human thought by removing it from the picture seems to be a pyrrhic victory at most. Theory choice in the grand rationality debate would then seem to come down to: people are reasoning poorly or they’re not reasoning at all. This is what I believe we should aim to avoid.

A related objection might call into question the distinction between social and accuracy motivations.<sup>13</sup> The motivations that lead people to adopt their priors might be (partly) social. These beliefs then function as the priors from which Bayesian updating starts. This would blur the distinction central to the way I’ve presented the theoretical landscape. But even if priors are adopted for partly social reasons, accounts according to which people are (mainly) trying to form their beliefs for gains in social standing and accounts according to which they’re (mainly) trying to get things right still make different experimental predictions. For example, the former category would seem to predict causal effect of persuasive messaging is undermined by countervailing party leader cues from party leader, as exposure to such cues activates people’s party identification and loyalty, producing an emotional reaction and directional motivation to adopt the party position. Whereas the latter would, in accordance with experimental findings (B. M. Tappin et al., 2023), predict the impact of partisan motivated conformity to be more constrained. So while the accounts might be reconcilable in some cases, there still seems to be a valid contrast here.

This paper presented a unified framework for diverging of explanations of different reasoning phenomena such as biased assimilation and acceptance of bad beliefs. Bringing those together shows, clearer than the literature would suggest, that we don’t *need* to postulate truth-independent cognitive mechanisms to account for them. This does not of course falsify accounts which rely on such mechanisms to account for these phenomena. But it suggests their explanatory advantage is smaller than it may have seemed. On the other hand, while environmentalist explanations make a solid case that evidence indicating motivated irrationality to be a good model of information processing might not indicate that at all, they also

minimize cognitive agency and make it unclear what it could still *mean* to be rational. Making progress in the rationality debate, then, in my view, requires deeper engagement with the question of how people actually make up their mind.

Our knowledge of this process is actually relatively minor, as most reasoning experiments disallow investigation of person-level phenomena (McManus et al., 2023).<sup>14</sup> To make further progress, it seems we need experimental innovations untangling the effects of directional goals from non-motivated (or accuracy-motivated) reasoning in order to yield less ambiguous insight in what people take to be the feature of the situation that is of normative epistemic significance in updating their beliefs could help to tease apart environmentalist and irrationalist explanations. I've proposed one in this paper. But, as the previous section made clear, such empirical data only help if we have better theory. The onus is on environmentalism to delineate its rational cognitive mechanisms, the criteria by which these mechanisms discern epistemically relevant inputs from those that are not, and the specific conditions under which they can lead to flawed belief systems absent motivated irrationality. Experimental innovations are only useful if we know what experimental outcomes environmentalism predicts and what these would mean for its assessments of blame.

This is not to say that all the room for progress lies at the empirical border. As an example of philosophical unclarity surrounding the rationality debate, consider that environmentalism and vice epistemology might simply be talking past each other in the sense that they might be picking out different relations when talking about the epistemic reason someone *has*. Being an epistemic reason *for* someone to believe a proposition and being an epistemic reason one *has* could be two different relations (Schroeder, 2008). Vice epistemology might say that e.g., the scientific consensus on climate change counts in favor of believing in it, so it's an epistemic reason *for everyone*. But in a more subjective sense of having a reason, the stereotypical Republican climate denier does not *have* this reason, environmentalists might retort. That is: given her background beliefs about who to trust, and so on, we would not expect her to think that updating on that consensus effectively promotes the goal of her now having accurate beliefs. Conversely, her climate beliefs are immune to intellectual self-criticism and so epistemically rational in one important (coherentist) sense (Foley, 2001).<sup>15</sup> On the other hand, she maybe *has* an epistemic reason (by virtue of the rational social learning mechanism) to update on the higher-order evidence provided by the social approval around her, without that being an epistemic reason *for* her in a more objective sense (it doesn't lead to true beliefs in this case). If vice epistemology is to be understood as pointing out that the stereotypical climate denier is not responding to considerations that should be

epistemic reasons *for* her and environmentalism as retorting that she could nonetheless be responding to epistemic reasons she *has*, the debate would benefit from more clarity about which of those two relations is the touchstone of rationality.<sup>16</sup>

Earlier, I argued that assessments of rationality only have a point when the object of assessment is in some way connected to our agency. There's much (philosophical) research on whether and in what way individuals have agency over what feature of a situation they take to be of normative epistemic significance and so how they could be responsible for their beliefs. According to a long-standing philosophical tradition, humans' ability to reflect on our thoughts and perceptions sets us apart from other conscious beings. This capacity for a special sort of cognitive self-determination is said to make our own relation to our own mental lives fundamentally different compared to nonrational animals (e.g., Korsgaard, 1996, p. 93). Because of our rational intellect we can "step back" from the mental happenings that *would* directly determine the beliefs and actions of a nonrational creature. One outstanding task for environmentalism is to find a way to preserve this idea, and so give substance to its claims of rationality. The theory now comes dangerously close to conceptualizing beliefs as the mere effects of brute causal incursions from the environment. But then they don't have the right kind of etiology to be (ir)rational. It seems to me an open question how to square environmentalism with an account of our responsibility for our beliefs that is able to do the required work.

Solving this requires a closer engagement with the literature on cognitive agency, which contains accounts of responsibility for believing around that could potentially be harnessed for this (Hieronymi, 2008; Owens, 2002). More generally, the lack of engagement with questions of cognitive agency has been detrimental to the rationality debate. For example, De Cruz (2020, p. 441) claims that the explanation of climate skepticism in terms of higher-evidence responsiveness fails to account for the mechanism of social belonging, which according to her "can explain why people sometimes accept testimony even if they deem the source inaccurate". But this is something we cannot do. We simply don't have the power to form beliefs based on a source we know to be unreliable. Forming such a belief necessitates that we either manage to forget that our belief originated from a decision based on non-epistemic reasons, or we adjust our perspective on the evidence to view our belief as justified, regardless of its dubious origins. However, we lack the direct capability to accomplish either of these adjustments. Environmentalism will likely acknowledge this – as it seems to conceptualize beliefs as happening to people, not products of actual thought. But then in what way can they be normatively assessed as more or less rational? Environmentalism owes us an answer, but it's not clear how it could give one.



## Notes

1. Questions have been raised about the extent of the so-called infodemic (e.g., Altay et al., 2023). For instance, one study observed that only 1% of users were responsible for 80% of the engagements with recognized fake news sources, and a mere 0.1% of users accounted for nearly 80% of the shares from these sources on Twitter (Grinberg et al., 2019). Nonetheless, these findings often depend on a limited definition of misinformation, focusing solely on established fake news website (Guess et al., 2020). But this definition overlooks the truly problematic category of misinformation: when it is strictly true but misleading (rather than false or fake) and comes from the top (rather than from problematic web domains). When this is taken into account, conclusions of low misinformation levels seem premature (Van Doorn, 2023a). My thanks to an anonymous reviewer for raising this.
2. It's not an ideal term either because it misleadingly implies that disliked evidence doesn't get assimilated in people's worldviews (Coppock, 2022, p. 28). See Van Doorn (2023b) on the need for improved conceptual precision in the reasoning literature.
3. An important qualification is that such findings do not generally replicate when subjects are presented with unequivocal rather than ambiguous evidence. Far from simply clinging to partisan beliefs – or even doubling down (Nyhan & Reifler, 2010) – when such beliefs are challenged, people tend to recognize clear contrary evidence and persuasive arguments, even when such evidence and arguments reflect unfavorably on their side, and change their mind accordingly (Bisgaard, 2019; Nyhan, 2021; B. M. Tappin et al., 2023).
4. This comparison is not (meant to be) exhaustive. For example, a third alternative posits that the goal of belief-adoption is not social in an identity-protective sense, but seeks to satisfy social interests on the group rather than the individual level, such that group membership unconsciously biases cognition by generating motivations to advocate for group interests or solve social challenges such as coordination (Blancke, 2023; Funkhouser, 2017; Williams, 2023b). For reasons of space, I could not expand the comparison to adequately include this family of views, and so I'd like to emphasize that the current comparison is not exhaustive. My thanks to an anonymous reviewer for pointing this out.
5. What about hybrid models which argue that the spread of bad beliefs is due both to the behavior of agents and systemic factors and so remedial strategies should operate at both levels (e.g., Bortolotti, 2022)? Vice epistemology, it seems to me, can accommodate a role for environmental factors. But, as I argue in the next section, it's not immediately clear how environmentalism can make sense of agential contributions to epistemic outcomes. This might be another way in which the current comparison is not exhaustive.
6. This paragraph is meant to serve as an example of other desiderata that are relevant in theory evaluation; partly as a prelude to the next section, where I argue that there are some such desiderata environmentalism has a hard time meeting. It's not meant to be an exhaustive treatment of the relevant considerations surrounding the observational equivalence paradox. Thanks to an anonymous reviewer for pressing me to be clear on this.
7. More generally, McManus et al. (2023) claim that many psychology experiments disallow investigation of person-level phenomena.
8. Thanks to Hugo Snijders for verifying that this indeed could be a plausible empirical test. Thanks to an anonymous reviewer for pressing me to be clearer on the needed experimental evidence.

9. At any rate, it seems to me this information is crucial in validating the abductive inference to motivated reasoning that is often made when “researchers leap from an observation that respondents deviate from some standard of source credibility to an inference about the respondents’ motive” (Druckman & McGrath, 2019, p. 116).
10. One reviewer points out that accounts of basic trust (e.g., Nguyen, 2022) argue that a source of information is reliable and so automatically updating one’s beliefs is rational makes us entitled to justification. I thank the reviewer for mentioning this important nuance, but, as far as I can see, the arguments in the text of why the emerging conception of rationality is too thin to serve environmentalism’s purposes are not negatively affected by it.
11. For my reasons, see Mercier (2020) and Vorms et al. (2022).
12. As Williams (2023a, p. 11) writes, what distinguished those who historically have successfully challenged the myths that prevalent in their epistemic environments, was, rather, “how they approached the claims and beliefs of others: with skepticism and vigilance, attentive to the hidden interests that often drive and corrupt majority and elite opinion, and sensitive to bad arguments and unfounded appeals to epistemic authority.”
13. Thanks to an anonymous reviewer for this objection.
14. Also see Rozin (2001) on the “yellow-ball problem”.
15. Compare also Meylan and Schmidt (2023).
16. Of course, any forthcoming agreement does not automatically resolve the underlying disagreement about what people usually take to be the epistemically relevant feature of the situation.

## Acknowledgements

Thanks to Philipp Schmid and members of the Radboud Theoretical Psychology reading group for helpful discussion. Thanks to two anonymous reviewers for helpful comments.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Maarten van Doorn  <http://orcid.org/0000-0002-6943-462X>

## References

- Achen, C., & Bartels, L. (Eds.). (2016). Partisan hearts and spleens: Social identities and political change. In *Democracy for realists* (pp. 232–266). Princeton University Press.
- Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media + Society*, 9(1), 205630512211504. <https://doi.org/10.31234/osf.io/edqc8>
- Anglin, S. M. (2019). Do beliefs yield to evidence? Examining belief perseverance vs. change in response to congruent empirical findings. *Journal of Experimental Social Psychology*, 82, 176–199. <https://doi.org/10.1016/j.jesp.2019.02.004>

- Aral, S., & Eckles, D. (2019). Protecting elections from social media manipulation. *Science*, 365(6456), 858–861. <https://doi.org/10.1126/science.aaw8243>
- Bago, B., Rand, D., & Pennycook, G. (2023). Reasoning about climate change. *Proceedings of the National Academy of Sciences Nexus*, 2(5), 100. <https://doi.org/10.1093/pnasnexus/pgad100>
- Bardon, A. (2020). *The truth about denial: Bias and self-deception in science, politics, and religion*. Oxford University Press.
- Baron, J., & Jost, J. T. (2019). False equivalence: Are liberals and conservatives in the United States equally biased? *Perspectives on Psychological Science*, 14(2), 292–303. <https://doi.org/10.1177/1745691618788876>
- Bayes, R., Druckman, J. N., Goods, A., & Molden, D. C. (2020). When and how different motives can drive motivated political reasoning. *Political Psychology*, 41(5), 1031–1052. <https://doi.org/10.1111/pops.12663>
- Bisgaard, M. (2019). How getting the facts right can fuel partisan-motivated reasoning. *American Journal of Political Science*, 63(4), 824–839. <https://doi.org/10.1111/ajps.12432>
- Blancke, S. (2023). Nothing about collective irrationalities makes sense except in the light of cooperation. *Philosophical Psychology*, 36(5), 990–1010. <https://doi.org/10.1080/09515089.2023.2169602>
- Borg, E. (2022). IX—In defence of individual rationality. *Proceedings of the Aristotelian Society*, 122(3), 195–217. <https://doi.org/10.1093/arisoc/aoac009>
- Bortolotti, L. (2022). Sharing responsibility for conspiracy beliefs: The agency-in-context model. *Resistances Journal of the Philosophy of History*, 3(6), e210103. <https://doi.org/10.46652/resistances.v3i6.103>
- Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.
- Brennan, J. (2016). *Against democracy*. Princeton University Press.
- Bullock, J. G. (2009). Partisan bias and the Bayesian ideal in the study of public opinion. *The Journal of Politics*, 71(3), 1109–1124. <https://doi.org/10.1017/S0022381609090914>
- Caddick, Z. A., & Feist, G. J. (2022). When beliefs and evidence collide: Psychological and ideological predictors of motivated reasoning about climate change. *Thinking & Reasoning*, 28(3), 428–464. <https://doi.org/10.1080/13546783.2021.1994009>
- Carter, J. A., & McKenna, R. (2020). Skepticism motivated: On the skeptical import of motivated reasoning. *Canadian Journal of Philosophy*, 50(6), 702–718. <https://doi.org/10.1017/can.2020.16>
- Cassam, Q. (2016). Vice epistemology. *The Monist*, 99(2), 159–180. <https://doi.org/10.1093/monist/onv034>
- Cassam, Q. (2018). Epistemic insouciance. *Journal of Philosophical Research*, 43, 1–20. <https://doi.org/10.5840/jpr2018828131>
- Coppock, A. (2022). *Persuasion in parallel. How information changes minds about politics*. The University of Chicago Press.
- De Cruz, H. (2020). Believing to belong: Addressing the novice-expert problem in polarized scientific communication. *Social Epistemology*, 34(5), 440–452. <https://doi.org/10.1080/02691728.2020.1739778>
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568–585. <https://doi.org/10.1037/0022-3514.63.4.568>
- Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2), 111–119. <https://doi.org/10.1038/s41558-018-0360-1>
- Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the*

- National Academy of Sciences*, 114(36), 9587–9592. <https://doi.org/10.1073/pnas.1704882114>
- Dutilh Novaes, C. (2023). The (higher-order) evidential significance of attention and trust—comments on Levy’s bad beliefs. *Philosophical Psychology*, 36(4), 792–807. <https://doi.org/10.1080/09515089.2023.2174845>
- Enoch, D. (2010). Not just a truthometer: Taking oneself seriously (but not Too Seriously) in cases of peer disagreement. *Mind*, 119(476), 953–997. <https://doi.org/10.1093/mind/fzq070>
- Fatollahi, A. (2022). Conservative Treatment of Evidence. *Episteme*, 1–16. <https://doi.org/10.1017/epi.2022.29>
- Foley, R. (2001). *Intellectual trust in oneself and others*. Cambridge University Press.
- Funkhouser, E. (2017). Beliefs as signals: A new function for belief. *Philosophical Psychology*, 30(6), 809–831. <https://doi.org/10.1080/09515089.2017.1291929>
- Funk, C., & Kennedy, B. (2019). *How Americans see climate change in 5 charts*. Pew Research Center. Retrieved December 8, 2023, from <https://www.pewresearch.org/fact-tank/2019/04/19/how-americans-see-climate-change-in-5-charts/>
- Gadsby, S. (2023). Bad beliefs: Automaticity, arationality, and intervention. *Philosophical Psychology*, 36(4), 778–791. <https://doi.org/10.1080/09515089.2023.2173060>
- Gardiner, G. (2022). Attunement: On the Cognitive Virtues of Attention. In M. Alfano, C. Klein, & J. de Ridder (Eds.), *Social virtue epistemology* (pp. 48–72). Routledge.
- Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review*, 77(2), 167–187. <https://doi.org/10.1177/0003122412438225>
- Gerber, A., & Green, D. (1999). Misperceptions about perceptual bias. *Annual Review of Political Science*, 2(1), 189–210. <https://doi.org/10.1146/annurev.polisci.2.1.189>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378. <https://doi.org/10.1126/science.aau2706>
- Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, 4(5), 472–480. <https://doi.org/10.1038/s41562-020-0833-x>
- Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased. *Psychology of Learning and Motivation*, 61, 41–102. <https://doi.org/10.1016/B978-0-12-800283-4.00002-2>
- Hamilton, L. C. (2011). Education, politics and opinions about climate change evidence for interaction effects. *Climatic Change*, 104(2), 231–242. <https://doi.org/10.1007/s10584-010-9957-8>
- Hare, W. (2003). The ideal of open-mindedness and its place in education. *Journal of Thought*, 38(2), 3–10.
- Hieronymi, P. (2008). Responsibility for believing. *Synthese*, 161(3), 357–373. <https://doi.org/10.1007/s11229-006-9089-x>
- Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition*, 47, 75–85. <https://doi.org/10.1016/j.concog.2016.09.004>
- Kahan, D. M. (2012). Cultural cognition as a conception of the cultural theory of risk. In S. Roeser, R. Hillerbrand, P. Sandin, & Peterson (Eds.), *Handbook of risk theory* (pp. 725–729). Springer.
- Kahan, D. M. (2015). What is the “science of science communication”? *Journal of Science Communication*, 14(3), 1–10. <https://doi.org/10.22323/2.14030404>
- Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2973067>

- Kahan, D. M., Hank, C. J.-S., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174. <https://doi.org/10.1080/13669877.2010.511246>
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86. <https://doi.org/10.1017/bpp.2016.2>
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10), 732–735. <https://doi.org/10.1038/nclimate1547>
- Kavanagh, J., & Rich, M. (2018). *Truth Decay: An initial exploration of the diminishing role of facts and analysis in American public life*. RAND Corporation. <https://doi.org/10.7249/RR2314>
- Kiesewetter, B. (2017). *The normativity of rationality*. Oxford University Press.
- Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology*, 62(3), 317–339. <https://doi.org/10.1006/jecp.1996.0033>
- Klaczynski, P. A., & Narasimham, G. (1998). Development of scientific reasoning biases: Cognitive versus ego-protective explanations. *Developmental Psychology*, 34(1), 175–187. <https://doi.org/10.1037/0012-1649.34.1.175>
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, 56(1), 28–55. <https://doi.org/10.1006/obhd.1993.1044>
- Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press.
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53(4), 636–647. <https://doi.org/10.1037/0022-3514.53.4.636>
- Lepoutre, M. (2020). Democratic Group Cognition. *Philosophy & Public Affairs*, 48(1), 40–78. <https://doi.org/10.1111/papa.12157>
- Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.
- Levy, N. (2023a). It's our epistemic environment, not our attitude toward truth, that matters. *Critical Review*, 35(1–2), 94–111. <https://doi.org/10.1080/08913811.2022.2149108>
- Levy, N. (2023b). Response to commentators. *Philosophical Psychology*, 0(0), 1–14. <https://doi.org/10.1080/09515089.2023.2188051>
- Lipton, P. (2017). Inference to the best explanation. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (1st ed. pp. 184–193). Wiley. <https://doi.org/10.1002/9781405164481.ch29>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. <https://doi.org/10.1037/0022-3514.37.11.2098>
- Lundgren, S. R., & Prislun, R. (1998). Motivated cognitive processing and attitude change. *Personality & Social Psychology Bulletin*, 24(7), 715–726.
- Malka, A., Krosnick, J. A., & Langer, G. (2009). The association of knowledge with concern about global warming: Trusted information sources shape public thinking. *Risk Analysis: An International Journal*, 29(5), 633–647. <https://doi.org/10.1111/j.1539-6924.2009.01220.x>
- Mandelbaum, E. (2014). Thinking is believing. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 57(1), 55–96. <https://doi.org/10.1080/0020174X.2014.858417>

- McDowell, J. (1998). Having the world in view: Lecture one. *Journal of Philosophy*, 95(9), 431–450. <https://doi.org/10.5840/jphil199895937>
- McIntyre, L. (2018). *Post-truth*. MIT Press.
- McManus, R. M., Young, L., & Sweetman, J. (2023). Psychology is a property of persons, not averages or distributions: Confronting the group-to-person generalizability problem in experimental psychology. *Advances in Methods and Practices in Psychological Science*, 6(3). <https://doi.org/10.1177/25152459231186615>
- McPhetres, J., Bago, B., & Pennycook, G. (2019). *Science beliefs, political ideology, and cognitive sophistication* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/ad9v7>
- Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.
- Meylan, A., & Schmidt, S. (2023). Refusing the COVID-19 vaccine: What's wrong with that? *Philosophical Psychology*, 36(6), 1102–1124. <https://doi.org/10.1080/09515089.2023.2181151>
- Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology*, 64(4), 561–574. <https://doi.org/10.1037/0022-3514.64.4.561>
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2(4), 201–218. [https://doi.org/10.1207/s15366359mea0204\\_1](https://doi.org/10.1207/s15366359mea0204_1)
- Müller, B. (2024). *Bad social norms rather than bad believers—Examining the role of social norms in bad beliefs*. <https://philsci-archive.pitt.edu/22959/>
- Munro, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Personality and Social Psychology Bulletin*, 23(6), 636–653. <https://doi.org/10.1177/0146167297236007>
- Munton, J. (2023). Prejudice as the misattribution of salience☆. *Analytic Philosophy*, 64(1), 1–19. <https://doi.org/10.1111/phib.12250>
- Murphy-Hollies, K., & Caporuscio, C. (2023). What is left of irrationality? *Philosophical Psychology*, 36(4), 808–818. <https://doi.org/10.1080/09515089.2023.2186220>
- Newton-Smith, W. H. (2002). *The rationality of science*. Routledge.
- Nguyen, C. T. (2022). Trust as an unquestioning attitude. *Oxford Studies in Epistemology*, 7, 214–244. <https://doi.org/10.1093/oso/9780192868978.003.0007>
- Nguyen, C. T. (2023). Hostile epistemology. *Social Philosophy Today*, 39, 9–32. <https://doi.org/10.5840/socphiltoday2023391>
- Nurse, M. S., & Grant, W. J. (2020). I'll see it when I believe it: Motivated numeracy in perceptions of climate change risk. *Environmental Communication*, 14(2), 184–201. <https://doi.org/10.1080/17524032.2019.1618364>
- Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15), e1912440117. <https://doi.org/10.1073/pnas.1912440117>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Owens, D. (2002). *Reason without freedom* (0 ed.). Routledge. <https://doi.org/10.4324/9780203464601>
- Pew Research Center. (2017). *Sharp partisan divisions in views of national institutions*. Pew Research Center for the People and the Press. Retrieved November 30, 2023, from <https://>

- [www.people-press.org/2017/07/10/sharp-partisandivisions-in-views-of-national-institutions/](http://www.people-press.org/2017/07/10/sharp-partisandivisions-in-views-of-national-institutions/)
- Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, 27 (2S), E-43–E-64. <https://doi.org/10.1353/ken.2017.0025>
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14. [https://doi.org/10.1207/S15327957PSPR0501\\_1](https://doi.org/10.1207/S15327957PSPR0501_1)
- Schmidt, S. (2020). Rationality and responsibility. *Australasian Philosophical Review*, 4(4), 379–385. <https://doi.org/10.1080/24740500.2021.1964245>
- Schroeder, M. (2008). Having reasons. *Philosophical Studies*, 139(1), 57–71. <https://doi.org/10.1007/s11098-007-9102-3>
- Smith, L., & Archer, A. (2020). Epistemic injustice and the attention economy. *Ethical Theory and Moral Practice*, 23(5), 777–795. <https://doi.org/10.1007/s10677-020-10123-x>
- Sommer, J., Musolino, J., & Hemmer, P. (2023). Updating, evidence evaluation, and operator availability: A theoretical framework for understanding belief. *Psychological Review*, 0(0), OnlineFirst. <https://doi.org/10.1037/rev0000444>
- Stanovich, K. E. (2021). *The bias that divides us: The science and politics of myside thinking*. MIT Press. [https://books.google.com/books?hl=en&lr=&id=e\\_gLEAAQBAJ&oi=fnd&pg=PR9&dq=stanovich+bias+that+divides+us&ots=cnNko5Hb7i&sig=zOz\\_cWfYsa-6o6A26pH4mNrpokw](https://books.google.com/books?hl=en&lr=&id=e_gLEAAQBAJ&oi=fnd&pg=PR9&dq=stanovich+bias+that+divides+us&ots=cnNko5Hb7i&sig=zOz_cWfYsa-6o6A26pH4mNrpokw)
- Sullivan Bissett, E. (2022). Believing badly ain't so bad. *Philosophical Psychology*, 36(6), 1–9. <https://doi.org/10.1080/09515089.2022.2077717>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Tappin, B. M., Berinsky, A., & Rand, D. (2023). Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues. *Nature Human Behavior*, 7(4), 568–582. <https://doi.org/10.1038/s41562-023-01551-7>
- Tappin, B., & Gadsby, S. (2019). Biased belief in the Bayesian brain: A deeper look at the evidence. *Consciousness and Cognition*, 68, 107–114. <https://doi.org/10.1016/j.concog.2019.01.006>
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences*, 34, 81–87. <https://doi.org/10.1016/j.cobeha.2020.01.003>
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2021). Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General*, 150(6), 1095–1114. <https://doi.org/10.1037/xge0000974>
- Van Doorn, M. (2023a). Advancing the debate on the consequences of misinformation: Clarifying why it's not (just) about false beliefs. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 1–27. <https://doi.org/10.1080/0020174X.2023.2289137>
- Van Doorn, M. (2023b). The skeptical import of motivated reasoning: A closer look at the evidence. *Thinking & Reasoning*, 1–31. <https://doi.org/10.1080/13546783.2023.2276975>
- Van Doorn, M. (in press). How partisanship can moderate the influence of communicated information on the beliefs of agents aiming to form true beliefs. *Social Epistemology*.
- Vavova, K. (2023). Open-mindedness, rational confidence, and belief change. *Social Epistemology Review & Reply Collective*, 12(2), 33–44.
- Velez, Y., & Liu, P. (2023). *Confronting core issues: A critical test of attitude polarization*. APSA Preprints. <https://doi.org/10.33774/apsa-2023-gxh3l-v2>

- Vorms, M., Harris, A. J. L., Topf, S., & Hahn, U. (2022). Plausibility matters: A challenge to Gilbert's "Spinozan" account of belief formation. *Cognition*, 220, 104990. <https://doi.org/10.1016/j.cognition.2021.104990>
- Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology*, 49(2), 338–346. <https://doi.org/10.1037/0022-3514.49.2.338>
- Whittlestone, J. (2017). *The importance of making assumptions: Why confirmation is not necessarily a bias*. University of Warwick.
- Williams, D. (2018). Hierarchical Bayesian models of delusion. *Consciousness and Cognition*, 61, 129–147. <https://doi.org/10.1016/j.concog.2018.03.003>
- Williams, D. (2021). Motivated ignorance, rationality, and democratic politics. *Synthese*, 198(8), 7807–7827. <https://doi.org/10.1007/s11229-020-02549-8>
- Williams, D. (2023a). Bad beliefs: Why they happen to highly intelligent, vigilant, devious, self-deceiving, coalitional apes. *Philosophical Psychology*, 36(4), 819–833. <https://doi.org/10.1080/09515089.2023.2186844>
- Williams, D. (2023b). The case for partisan motivated reasoning. *Synthese*, 202(3). Article number 89. <https://doi.org/10.1007/s11229-023-04223-1>
- Wollstonecraft, M. (1792). *A vindication of the rights of woman: With strictures on political and moral subjects*. Thomas and Andrews.
- Worsnip, A. (2019). The obligation to diversify one's sources: Against epistemic partisanship in the consumption of news media. In C. Fox & J. Saunders (Eds.), *Media ethics: Free speech and the requirements of democracy* (pp. 240–264). Routledge.
- Zarocostas, J. (2020). How to fight an Infodemic. *The Lancet*, 395(10225), 676.