

The skeptical import of motivated reasoning: a closer look at the evidence

Maarten van Doorn

To cite this article: Maarten van Doorn (08 Nov 2023): The skeptical import of motivated reasoning: a closer look at the evidence, Thinking & Reasoning, DOI: [10.1080/13546783.2023.2276975](https://doi.org/10.1080/13546783.2023.2276975)

To link to this article: <https://doi.org/10.1080/13546783.2023.2276975>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Nov 2023.



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



Article views: 288



View related articles [↗](#)




View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

The skeptical import of motivated reasoning: a closer look at the evidence

Maarten van Doorn 

Faculty of Arts, Radboud University, Nijmegen, the Netherlands

ABSTRACT

Central to many discussions of motivated reasoning is the idea that it runs afoul of epistemic normativity. Reasoning differently about information supporting our prior beliefs versus information contradicting those beliefs, is frequently equated with motivated irrationality. By analyzing the normative status of belief polarization, selective scrutiny, biased assimilation and the myside bias, I show this inference is often not adequately supported. Contrary to what's often assumed, these phenomena need not indicate motivated irrationality, even though they are instances of belief-consistent information processing. Second, I engage with arguments purporting to show that belief-consistent information processing does not indicate motivated irrationality because of its mere differential treatment of confirming and non-confirming evidence, but rather because it reveals the undermining presence of an *irrelevant influence*, such as a desire or partisan identity-driven cognition. While linking belief-consistent reasoning to a deeper source of directional motivation to make good on the claim that it indicates motivated irrationality is indeed what's needed, two prominent such arguments fail. The non-normativity of many reasoning processes often taken to indicate motivated irrationality is not in fact well established.

ARTICLE HISTORY Received February 1, 2023; Accepted October 24, 2023

KEYWORDS Motivated reasoning; belief polarization; biased assimilation; rationality; Bayesianism; irrelevant influences

In their landmark study, Lord et al. (1979) asked supporters and opponents of the death penalty to read about two fictional studies. One study supported the idea that the death penalty is an effective crime deterrent. The other study supported the idea that the death penalty is *not* an effective crime deterrent. After the presentation of each study, participants reported how much their belief that the death penalty deters or fails to deter murder and their attitude toward capital punishment (support or

CONTACT Maarten van Doorn  martinus.vandoorn@ru.nl  Faculty of Arts, Radboud University, Nijmegen, the Netherlands.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

opposition) had changed relative to the start of the experiment. It transpired that, after being presented with the *same* mixed evidence on the issue, both groups reported becoming more convinced of their *opposed* initial positions. That is, after being provided with details about both studies, death penalty supporters self-reported that their confidence in the effectiveness of the death penalty as a crime deterrent had increased, and death penalty opponents self-reported they had become more confident of the opposite conclusion.

On the face of it, there's something puzzling about such *belief polarization*. How can two groups see the same information and yet draw opposite conclusions? It seems natural to expect that, after being exposed to evidence of such a mixed character, their disagreement would be reduced. Perhaps it would be unrealistic to expect a perfect convergence of opinion. Still it seems plausible to expect the common evidence to narrow the gap between opposing views. So how can numerous—but by no means all (Velez & Liu, 2023)—subsequent experiments likewise find that exposing groups of subjects who disagree to the same mixed evidence, may cause their initial attitudes to move further apart (Batson, 1975; Munro & Ditto, 1997; Plous, 1991)?

Many scholars have concluded that this polarization shows that people process information in a biased manner, so as to support their preexisting views (Baron, 2008; Munro & Ditto, 1997; Ross & Anderson, 1982). These explanations typically emphasize the role of *motivated reasoning* (Klaczynski, 2000; Kunda, 1990), and suggest the polarization results from people interpreting information in a biased manner to favor conclusions that they would like to be true or are congenial to their political in-group (Dawson et al., 2002; Taber & Lodge, 2006). According to these accounts, belief polarization is strictly irrational behavior. For instance, when discussing the Lord study, Ross and Anderson (1982, p. 145) wrote that polarization is “in contrast to any normative strategy imaginable for incorporating new evidence relevant to one's beliefs.”

However, in a well-known paper Kelly (2008) denies that the polarized beliefs in these particular cases are normatively undermined. Despite appearances to the contrary, and despite being the results of mechanisms that underlie polarization.¹

Following a call for more research on the epistemological import of motivated reasoning (Carter & McKenna, 2020), I will critically assess Kelly's description of the polarizing mechanisms (Section “Kelly on the psychological mechanisms that underwrite polarization”). I argue that Kelly's

¹In making this argument, Kelly assumes that what is reasonable to believe is a function of one's evidence. McWilliams (2021) has argued that, in fact, on plausible version of evidentialism, the polarized beliefs Kelly defends are not justified. How does the argument of the present paper relate to McWilliams' argument? The main differences are that McWilliams focuses on what theories of epistemic justification have to say about the justification of the subjects' belief that are the result of mechanisms that underlie polarization. And then argues that they are in fact not justified on plausible versions of the theory of epistemic justification (evidentialism) that Kelly assumes in arguing for their justification. This paper, by contrast, focuses not primarily on the resulting beliefs but on the underlying mechanism.

normative defense is incomplete, because it conceptualizes selective scrutiny as a *metacognitive* strategy, and so stays objectionably silent about its arguably non-normative effects on *first-order* reasoning. In fact, Avnur and Scott-Karkunes (2015) have argued that beliefs resulting from reasoning subject to exactly such effects are defeated, because they are the result of unreliable desire-based directionally motivated reasoning. In Section “What type of uneven scrutiny, exactly, is rational?”, I consider more deeply whether such directional influences on reasoning convincingly demonstrate unreliability, and conclude that they do not. Along the way, I consider the extent to which selective scrutiny, belief polarization, and so-called wishful thinking indicate motivated irrationality. I argue that this extent is much smaller than commonly assumed. The intermediate conclusion will be that the type of belief-consistent information processing central to selective scrutiny and belief polarization is not in itself (convincingly established as) non-normative. Nor does the evidence allow the conclusion that it reveals the belief-undermining effect of directionally influential desire. I then consider two more patterns of belief-consistent information processing: biased assimilation and the *myside bias*. In Section “The real skeptical import of motivated reasoning”, I argue they are epistemic more above board than often thought. Furthermore, contra Carter and McKenna (2020), I argue that the available evidence does not warrant the conclusion that they reveal that belief-undermining influence of partisan cognition. In Section “Does politically motivated reasoning lead to false beliefs?”, finally, I consider their claim that biased assimilation causally leads to false beliefs. This contention too, I argue, is not adequately supported by the evidence. Section “Conclusion” concludes: researchers are often too quick to assume that reasoning differently about evidence, arguments and statements that are consistent vs. inconsistent with one’s prior beliefs, means this reasoning sprung from motivated irrationality. At least, the conventional wisdom about instances of presumably non-normative reasoning phenomena such as belief polarization, biased assimilation and the *myside bias* needs more precise arguments to make good on the claim that such phenomena indicate irrationality.

Kelly on the psychological mechanisms that underwrite polarization

Kelly (2008) is concerned with the explanandum of *belief polarization in response to mixed evidence*. To take the classic Lord study again, it had two groups of subjects. One of which believed in the deterrent effect of the death penalty and one of which doubted it. Both groups self-reported²

²Self-reported’ meaning here that for example the Lord et al. (1979) did not measure opinions before and after presentation of evidence. Rather, they rely entirely on the subjects’ assessments of whether their views have become more pro- or anti-death-penalty. When pretreatment and posttreatment opinions are measured directly, attitude polarization is less robust (Miller et al., 1993; Anglin, 2019).

becoming more confident in their initial positions after being presented with the same mixed set of studies on the issue. ‘Mixed’ meaning here that some studies seemed to suggest that capital punishment was a deterrent while other studies seemed to suggest it was not. The evidence reflection on which caused subjects with contrasting and relatively firm prior beliefs to polarize thus was properly *ambiguous*: it could legitimately be interpreted as supporting or undermining different viewpoints. It seems that evidence needs to have this property for belief polarization to be a reliable consequence of reflection on it (Anglin, 2019; Benoît & Dubra, 2019; Chaiken & Maheswaran, 1994; Dorst, forthcoming).

Self-reported belief polarization in response to mixed evidence, Kelly (2008, p. 612) tells us, is an “empirically well-confirmed phenomenon.” By and large, as mentioned, belief polarization in response to mixed evidence has been described as an example of irrational behavior. Such irrationalist explanations assume a particular answer to what *mediates* the route from exposure to mixed evidence to belief polarization: people polarize in response to such mixed evidence because their reasoning is insufficiently constrained by uncongenial evidence. Rather than giving the opposing studies their due, people engage in directionally motivated reasoning: their goal is to arrive at a congenial conclusion predetermined by e.g., a desire (Avnur & Scott-Kakures, 2015) or political background beliefs (Carter & McKenna, 2020). And so they will try to incorporate information in ways that are mostly likely to yield that congenial answer.

Kelly (2008, p. 617) disagrees. Why? As it turns out, “individuals who have participated in the relevant experiments typically do not pay less attention to counterevidence than to supporting evidence. Indeed, the opposite seems to be true: far from paying less attention to counterevidence, it seems that we pay *more* attention to it.” This, too, is empirically confirmed (Wyer & Frey, 1983; Velez & Liu, 2023). Rather than immediately dismissing evidence that contradicts our beliefs, we tend to examine it more closely and spend *more* time looking at it, not less.³ As we do, we’ll often find legitimate flaws in the methodology, gaps in the reasoning, or other factors that could explain away the data. So contra the irrationalist account, belief polarization is driven not by out-of-hand dismissal but by *selective scrutiny* (people spend more time looking for flaws with incongruent evidence than congruent evidence). Accordingly, it’s driven not by selective exposure to confirming evidence, but rather by selective exposure to flaws with incongruent evidence (as Dorst (forthcoming) points out). This means there are instances where belief polarization may actually be normative (Stanovich, 2021).

In Lord et al. (1979), for instance, participants scrutinized the study that disagreed with their view. They used their cognitive resources to search for flaws that might discredit the study’s conclusion: problems with its methodology, variables that were not adequately controlled for, and so on. Meanwhile, subjects took the congenial study’s results on board as

³At least, in the relevant reasoning experiments.

further evidence for their view, without scrutinizing it. Consequently, they ended up being aware of plenty possible *alternative explanations* for the uncongenial evidence, but not for the congenial evidence. Subjects take these possible alternative explanations (the study's result was not due to the uncongenial hypothesis being true, but due to its small sample size, flawed analysis, etc.) to largely defeat the evidence the study would otherwise provide for the uncongenial hypothesis. And so their confidence in that hypothesis remains unfaced, while their confidence in the congenial hypothesis increases.

This means the question of the rationality of belief polarization gets pushed back to the question of the normative status of selective scrutiny. This is where I turn now.

Kelly (2008) argues that selective scrutiny *is* in fact rational. As a matter of practical rationality, says Kelly, it is unreasonable to demand equal scrutiny for surprising and unsurprising bits of purported evidence. Any treatment of *genuine* evidence other than adjusting one's beliefs to it is of course irrational. However, not all *purported* evidence is *genuine* evidence. The results of experiments, for instance, are not always treated as genuine evidence right away, by scientists at least. They usually obtain this status if they can be justified as compatible with broader, more accepted data, or if they can be reproduced. In the same way, if you read a newspaper article claiming there have been many instances of misconduct at your department, of which you had not heard anything before, you have *purported* evidence that this in fact happened. But you will probably only accept it as *genuine* evidence after you've sought further evidence about what happened, or about the original piece of evidence.

Unfortunately, we obtain too many pieces of purported evidence to thoroughly investigate them all. We must inevitably prioritize. Kelly (2008) points out that scientists generally prioritize studying anomalous phenomena—those that do not align with currently accepted theories—over those that do align with currently accepted theories. Does that impugn on the rationality of science in any way? There seems to be no reason to believe it's unreasonable for scientists to spend more resources (intellectual or otherwise) attempting to generate novel explanations for anomalous phenomena than they do for phenomena that are already explained by the theory that they currently accept. Indeed, Kelly points out, it seems logical to think that to proceed in any *other* way would be unreasonable.

Another reason why selective scrutiny is rational, is that it maximizes expected accuracy (Dorst, forthcoming). If you don't take your own prior beliefs to be formed irrationally, you should think that counterattitudinal arguments are more likely to contain flaws, and that their flaws will be easier to recognize. As a result, given your prior beliefs and a piece of evidence to scrutinize, there's a positive correlation between how likely you are to find a flaw in the evidence (if there is one) and how accurate you expect scrutinizing the evidence to make you. This means that, as an agent who cares about accuracy (Dorst, forthcoming) and/or understanding

(Levy, 2022), rational management of ambiguous evidence means it makes sense to tend to scrutinize the evidence for which you expect to be able to recognize its flaws, without motivation entering the picture (cf. Fatollahi, 2023; Gerber & Green, 1999).

Selective scrutiny—sometimes also called disconfirmation bias—is often seen as indicating motivated irrationality. But, as argued above, the same empirical pattern can be generated by reasoning motivated by accuracy. Selective scrutiny need not be a mechanism through which motivated irrationality operates. There are other explanations on the table, according to which “disconfirmation bias isn’t so much a bias as a straightforward consequence of thinking that some arguments are stronger than others” (Coppock, 2022, p. 134).

Contrary to what’s often assumed, this means observations of selective scrutiny (or disconfirmation bias) do not in themselves provide evidence for theories of motivated irrationality. At least, a further argument is needed to link mere differential treatment of belief-confirming vs. non-confirming information to motivated irrationality. What precise form such arguments or supporting data should take is somewhat unclear, hampered by what others have described as the “conceptual imprecision of politically motivated reasoning” (Tappin et al., 2020, p. 85). Nevertheless, we will look at two attempts later in this paper.

Belief polarization is likewise often seen as indicating motivated irrationality. There, to summarize, the alternative vindicatory account runs as follows.⁴ By the practically rational process of spending more cognitive resources trying to understand surprising findings, people end up being aware of more alternative explanations for purported evidence that ostensibly conflicts with their prior beliefs vs. alternative explanations for purported evidence that does not so conflict. And, as a matter of normative epistemology, “for a given body of evidence and a given hypothesis that purports to explain that evidence, how confident one should be that the hypothesis is true on the basis of the evidence depends on the space of alternative hypotheses of which one is aware” (Kelly, 2008, p. 620). Put these two together, and a picture emerges on which people who polarize after being exposed to mixed evidence are reasonably devoting greater scrutiny to apparent evidence, and then rationally responding to what this scrutiny reveals (cf. Almagro, 2022, pp. 16–17; Gilovich, 1991, p. 54).

This will be a recurring theme in this paper. Like selective scrutiny and belief polarization, other belief-consistent patterns of reasoning—such as biased assimilation—are often seen as evidence for motivated irrationality as well. Yet, as will become clear throughout this paper, they too can also be explained equally well without positing motivated irrationality. This means they do not in themselves provide evidence for theories of motivated irrationality. Crucially, this non-irrationality account often matches the empirical data just as well as the explanation from motivated

⁴For other such accounts see Koehler (1993) and Stanovich (2021, pp. 62–66).

reasoning. In the case of selective scrutiny, the Bayesian account is not just observationally equivalent to the motivated-irrationality explanation, but has superior observational adequacy. The selective-scrutiny account, but not the irrational-dismissal account, is able to explain the observation that people spend more and not less time looking at evidence inconsistent with their prior beliefs in information-evaluation experiments.

What type of uneven scrutiny, exactly, is rational?

So far, we've established with Kelly that belief polarization is, despite appearances, not irrational because it results from dedicating more investigative resources to scrutinizing purported disconfirming evidence (than to confirming evidence), thereby increasing the chances of finding reasons to dismiss it. This process of selective scrutiny is rational behavior, so, says Kelly, the normative status of the resulting (polarized) beliefs is not undermined. On the one hand, this means that claims that motivated irrationality is driving belief polarization will have to engage with this alternative reading. On the other hand, I will argue in this section that Kelly's defense of the normativity of selective scrutiny (and hence belief polarization) itself might be incomplete.

This is because there are really two forms of uneven scrutiny at work here, and Kelly has defended only one. The first one takes place on the meta-level, as it were. It is not internal to the reasoning process itself—does not concern how one actually investigates confirming and disconfirming evidence—but concerns the higher-order prioritization of one's investigative resources in face of purported evidence against one's beliefs. It is this *quantitative* process of dedicating more cognitive resources to scrutinizing pieces of disconfirming evidence that Kelly (2008) defends. But on top of that, there's another *qualitative* asymmetry in our treatment of confirming vs. disconfirming evidence. This asymmetry *is* internal to one's reasoning process, influencing not just which piece of evidence is scrutinized, but also *how* that piece of evidence is scrutinized (Avnur & Scott-Kakures, 2015, pp. 12–13).

In particular, there's reason to believe that we don't just quantitatively dedicate more resources to scrutinizing disconfirming evidence (versus confirming evidence), but also employ *different qualitative acceptance standards* for congenial and uncongenial purported evidence. A growing body of research suggests that people are uneven skeptics of disconfirming information. They adopt differential judgment criteria when evaluating uncongenial relative to congenial information. And hold arguments they dislike to higher standards that require stronger purported evidence (Ditto et al., 1998; Ditto & Lopez, 1992; Kraft et al., 2015). This application of differential evaluation criteria to belief-consistent and belief-inconsistent information might also explain belief polarization (Sanbonmatsu et al., 1998). Beliefs could polarize because uncongenial information is selectively

held to much higher standards, leaving open the possibility that belief polarization is a result of motivated reasoning rather than rational cognitive resource management in light of ambiguous evidence.

This pattern of asymmetric acceptance thresholds also arises in studies of the *myside bias* in argument generation and evaluation tasks. On the receiving end, people give higher evaluations to argument that support their opinions than those that refute their prior positions (Stanovich & West, 2007, 2008). Which they perhaps do, because, for such arguments, they are rather quickly satisfied, also when producing them. In his original *myside bias* experiments, for example, Perkins (1985, p. 586) concluded that people only generate arguments that make “superficial sense” when asked to justify their point of view, and can often fail to offer genuine evidence (Kuhn, 1991; Sá et al., 2005). However, when people evaluate arguments and evidence with which’ conclusion they disagree, they appear to be a lot more careful and demanding, mostly accepting strong arguments and solid purported evidence. This result has been observed in research on persuasion and attitude change (Petty & Wegener, 1998), as well as in Bayesian studies of argumentation (Hahn & Oaksford, 2007).

It seems, then, that people use different acceptance standards for congenial and uncongenial arguments and evidence. It seems plausible that subjects in belief polarization experiments reason this way too. This means that Kelly’s defense of the normativity of belief polarization is incomplete. The uneven *distribution of cognitive resources* might be rational. But the reasoning that gives rise to belief polarization might still be undermined due to the use uneven *acceptance standards* for evidence that confirms vs. disconfirms one’s prior beliefs. The intuition that there’s something unfair—epistemically non-normative—about doing so is hard to shake. It’s worth exploring, then, whether it hints at something. Does this selective use of strict acceptance criteria indicate motivated irrationality?

Avnur and Scott-Kakures (2015) have, in an influential paper, argued that precisely this use of differential decision criteria for confirming versus disconfirming evidence constitutes evidence that one’s reasoning process *was* indeed unreliable. Their argument is situated in a somewhat different epistemological corner—that of irrelevant influences, rather than reasoning experiments. Yet it seems to me that theoretical progress can be made by bringing it to bear on the current issue. Also because, as we saw in the previous section, a further argument is needed to classify differences in the treatment of belief-consistent and belief-inconsistent information as motivated irrationality. The claim that such differences are *grounded in* belief-undermining irrelevant influences could, if successful, be such an argument. This is where I turn now.

Avnur and Scott-Kakures (2015, p. 12, my emphasis) focus on the belief-undermining effects of what they call *directional influences*: an influence that “causes our handling of evidence to favor a particular, predetermined outcome, *where the desires that determine the favored outcome go beyond merely interest in believing truth.*” They claim that when someone

has subjective “evidence that her reasoning was directionally influenced, [this] determines that she has less (or no) justification for currently believing” (p. 22).

We will consider this argument in due course. But first we need to know what subjective evidence they have in mind, the possession of which determines that someone “has less (or no) justification for currently believing.” Here the discussed unequal standards to which congenial and uncongenial evidence are held come back in. For, in general, Avnur and Scott-Kakures (2015, p. 16) say, “what is characteristic of directional influence, is the alternate credulousness and hyper-criticality that is sensitive to one’s interests or error costs.” This just is the just-introduced use of differential decision criteria for likely and unlikely or preferred and non-preferred conclusions, cashed out in terms of “acceptance thresholds” (p. 15). Purported evidence going against our beliefs will have to be highly probable to be accepted, they write, but only moderately improbable to be rejected. Whereas purported evidence compatible with our prior beliefs must be highly improbable to be rejected, but only moderately probable to be accepted (Trope & Liberman, 1996). This way, people often use more stringent criteria when evaluating arguments and evidence they antecedently disagree with than when confronted with congenial information. And doing so, it is claimed, undermines the beliefs formed based on that reasoning.

Why do they think that this is so—that subjective evidence of using different acceptance thresholds is a defeater? To partly repeat a passage quoted earlier, Avnur and Scott-Kakures (2015, p. 12, my emphasis) maintain that these significantly different acceptance thresholds are evidence that reasoning has been such as to favor a particular, predetermined outcome, where *the desires that determine the favored outcome go beyond merely interest in believing the truth*. The idea seems to be that the *explanation* for using asymmetric acceptance thresholds for confirming vs. disconfirming evidence is that one was reasoning with the goal of arriving at a desired conclusion—in the same way as irrationalist interpretations of belief polarization experiments discussed earlier accused the subjects in those studies of doing. This explanatory connection is why asymmetric acceptance thresholds constitute evidence for directionally influential desire(s). Which is an *undermining* influence because it implies that “these desires have shaped one’s management of the evidence in ways that favor a target belief” (p. 21). In this way, Avnur and Scott-Kakures (2015) attempt to make the required further argument linking belief-consistent information processing to motivated irrationality.

Avnur and Scott-Kakures (2015) illustrate their case that differential decision criteria are evidence for a directionally influential desire with an example of someone who desires to believe that some other person is in love with him. This makes him extremely critical of purported evidence that seems to show the contrary, but not of purported evidence that seems consistent with his favored outcome. Citing Mele (1997, p. 94) they

conclude based on this case that “desire or interests directs inquiry, and specifically the way evidence is brought to bear on a hypothesis and toward the acceptance of a doxastic target. Our desiring that p may lead us to interpret data as supporting p that we would easily recognize to count against p in the desire’s absence” (Avnur & Scott-Kakures, 2015, p. 13). The other way around, then, from having used asymmetric acceptance thresholds, “we gain some evidence that our belief forming process is unreliable”—as we can now tell that we’ve incorporated information in ways that were most likely to yield the particular conclusion we desired to arrive at (Avnur & Scott-Kakures, 2015, p. 28).⁵

So in the argument, the irrationality of using different acceptance thresholds for confirming vs. disconfirming evidence hinges on two steps: (a) this way of reasoning is evidence for the directional influence of a desire and (b) such desire-based influence undermines the beliefs that are the result of this reasoning. If this argument goes through, then beliefs in belief polarization experiments might be irrational after all, because they might not result from rational management of ambiguous evidence but instead from desire-based motivated reasoning.

However, both steps are highly dubious. For one, it’s highly uncertain whether the use of differential criteria for confirming vs. disconfirming evidence is in fact subjective evidence for desire-based directionally motivated reasoning in the way Avnur and Scott-Kakures (2015) suggest. This is because such asymmetric acceptance thresholds can also be the result of prior beliefs doing their non-motivational job as Bayesian anchor.

To see why the application of differential evaluation criteria to belief-consistent and belief-inconsistent purported evidence could be Bayesian, consider first that the decision of how to treat purported evidence is, and should be, heavily influenced by the extent to which one takes seriously the possibility that it might be genuine evidence. This level of confidence, in turn, is chiefly determined by one’s higher-order evidence—by how the piece of purported evidence bears on one’s own beliefs (Fatollahi, 2023). There will be many pieces of purported evidence that do not straightforwardly match with one’s beliefs, but are not too distant from them to be outright discounted either. Given one’s lack of confidence in them, these interesting but possible belief change-demanding (i.e., possibly wrong) pieces of purported evidence cannot be accepted as genuine evidence without further scrutiny, during which any rational subject will require them to meet pretty stringent criteria. This makes sense as they contradict what you think you know, and it seems plain silly—and psychologically unrealistic—to ignore what you think you know in assessing new information (Foley & & others, 2001). Without any motivation in the picture, any subject will thus apply more stringent criteria to purported evidence

⁵In such passages, it seems like Avnur and Scott-Kakures (2015) are assuming a subjectivist account of defeat, in which the subject needs to be aware that information d defeats her belief that p for her belief that p to be defeated. As does Kelly (2008, p. 629) at some points. It’s worth nothing, then, that there has been recent pushback against subjectivist accounts of defeat (e.g., Klenk, 2019).

that goes against her past evidence and so against her in-her-mind-not-irrationally-formed beliefs. Indeed, several studies have shown that belief-consistent information processing arises for hypotheses for which people have no stakes in the specific outcome and thus no interest in particular conclusions (e.g., Crocker, 1982; Klayman & Ha, 1987, 1989; Sanbonmatsu et al., 1998; Skov & Sherman, 1986; Snyder & Swann, 1978; Wason, 1960). In other words, the same mechanisms apply, regardless of peoples' interest in the outcome. Being more demanding of evidence that contradicts your priors vs. evidence that does not, in short, is not behavior that seems particularly diagnostic of motivated reasoning. Reasoners solely interested in accuracy will also exhibit it (cf. Dorst, forthcoming).

Recall Avnur and Scott-Kakures (2015) claim that subjective evidence that one was using such differential acceptance criteria was *ipso facto* subjective evidence that one's reasoning was a puppet of a directionally influential desire. Since such belief-consistent information processing also takes place when people are not motivated to confirm their belief, we might ask: noticing that she is applying such asymmetric acceptance thresholds, how is any subject to decide whether she was doing the best she rationally could or, alternatively, that her information processing was tailored to achieve some non-truth goal? "Belief-consistent information processing seems to be a fundamental principle in human information processing [...] a *conditio humana*" (Oeberst & Imhoff, 2023, p. 4). So without a psychologically realistic criterion here, it would seem that the position that all beliefs that are the result of evidential management that favored some (prior) belief, are defeated, leads to global skepticism (cf. Enoch, 2010). Authors who claim there's a distinct skeptical import of belief-consistent information processing will want to avoid this result. Avnur and Scott-Kakures (2015) gambit now is to claim that such information processing is evidence of a directionally influential desire making one's reasoning unreliable. But the use of differential decision criteria is an underspecified criterion for detecting directionally motivated reasoning.

This resembles the motivated reasoning observational equivalence problem (Druckman & McGrath, 2019). Some theorists have long maintained that motivational constructs must be invoked to explain certain patterns of reasoning. Others claimed that information-processing variables could adequately these phenomena. As Tetlock and Levi (1982, p. 68) wrote more than thirty years ago on the inconclusiveness of the cognition-motivation debate: "Cognitive and motivational theories are currently empirically indistinguishable. In particular, it's possible to construct information-processing explanations for virtually all evidence for motivated bias." Thirty years later, Little (2022, p. 15) still draws the same conclusion—all that seems to have changed is the Bayesian twist—writing that "directional motives and different priors are observationally equivalent." A person might be driven to critique an argument more rigorously because they are motivated to disbelieve its conclusion. Alternatively, a person might do this because being motivated by accuracy they objectively assess

it as being weaker or less credible. The problem is that both motivations can lead to the same observable behavior—more intense scrutiny of certain arguments over others.

It's worth noting here that although motivated reasoning is often defined as biased information processing driven by the desire to reach a particular conclusion (e.g., Hart & Nisbet, 2012; Kahan, 2013), it was originally conceptualized as serving both accuracy and directional goals, emphasizing the importance that both types of motivation play in information processing (Kunda, 1990). People are motivated to maintain their beliefs, but they are also motivated to be accurate (Hart et al., 2009; Klaczynski, 2000). They can only arrive at desired conclusions if they are justifiable (Epley & Gilovich, 2016; Hahn & Harris, 2014, p. 82; Kunda, 1990). In her classic paper, Kunda concluded that even directionally motivated reasoning does not constitute a *carte blanche* to believe whatever one desires; the desired conclusion is only drawn if it can be supported by evidence—indeed, if that evidence could “persuade a dispassionate observer” (Kunda, 1990, pp. 482–483). This evidential constraint on motivated reasoning explains why a significant body of empirical evidence demonstrates that motivated reasoners revise their beliefs in response to clear contrary evidence, even when such evidence reflects unfavorably on their desired outcome (Bisgaard, 2019; Nyhan, 2021; Tappin et al., 2023).

This brings us to step (b) in the argument by Avnur and Scott-Kakures (2015): the claim that desire-based directional influence on reasoning undermines the beliefs that are the result of this reasoning. Here, even if we grant that a desire was being directionally influential, and that a subject can get clear evidence of this and recognize it *as* evidence of desire-based motivated reasoning, it's *still* doubtful whether this would defeat the beliefs in question in the way Avnur and Scott-Kakures (2015) suggest. They claim that evidence that one's reasoning was directionally influenced is evidence that it was less reliable, hence undermined: “directionally influenced reasoning is, all else equal, less reliable than non-directionally influenced reasoning, so that evidence that one's belief-forming process was so influenced is evidence that one is less reliable than one otherwise would be” (p. 22). They continue: “Why, then, does the directional influence of the desire that *p* render the process less reliable? All else equal (that is, absent any additional information about the correlation between one's wanting that *p* and *p*'s being true), believing according to one's desires is about as reliable as believing randomly, or by chance” (pp. 22–23). They allege that “forming a belief under directionally influential desire is a way of thinking wishfully,” as it is “believing according to one's desires” (p. 22).

In such passages—given the justificatory constraint on desire-based belief choice—it seems to me they move too quickly from talk of a desire being a directional influence on reasoning (making some outcome *more likely*), and a desire *fixing the outcome* of a reasoning process. The cases in which a directionally influential desire fixes the outcome, such that we

end up believing according to them every time they make some reasoning outcome *more likely*, are much more exceptional than the authors assume. Reasoners can only bring themselves to believe things (they want to believe) for which they can find genuinely epistemic reasons (i.e., reasons that justify the truth of the relevant beliefs; Kunda, 1990). Consequentially, desire-based directional influence is not as epistemically pernicious as supposed.

In fact, robust evidence for such a biasing effect of desires on judgments has been hard to come by. It has even been dubbed “the elusive wishful thinking effect” (Bar-Hillel & Budescu, 1995). Studies on wishful thinking have generally failed to find evidence for it under well-controlled laboratory conditions. There have been some observations of wishful thinking outside the lab (Babad & Katz, 1991; Simmons & Massey, 2012). These, however, seem well explained as “an unbiased evaluation of a biased body of evidence” (Bar-Hillel & Budescu, 1995, p. 100). For instance, Bar-Hillel et al. (2008) found potential evidence of wishful thinking in the prediction of results in the 2002 and 2006 football World Cups, but further investigation revealed that these results were more likely caused by a salience effect than by a “magical wishful thinking effect” (Bar-Hillel et al., 2008, p. 282). In particular, they seemed to arise from a specific division of cognitive resources that influences information accumulation and not from any direct biasing effects of desirability. In general, there is little evidence for a general “I wish for, therefore I believe...” relationship (Bar-Hillel et al., 2008, p. 283).

Taken together, this sets up a dilemma for Avnur and Scott-Kakures (2015). Either their argument that desire-based directional influence forms a defeater is well taken, but only applies to a tiny minority of beliefs. Or, if the intended scope of their argument about the debunking force of directional influence is wider, it seems that the arguments they offer about asymmetric acceptance thresholds (which are not inconsistent with Bayesian reasoning) and desire-induced unreliability (which seems to be a confined effect) are not able to carry it.

The real skeptical import of motivated reasoning

It’s time to zoom out. In the literature on motivated reasoning, one is likely to find claims along the lines of “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). In the paper so far, we’ve been discussing examples of why this inference is too quick. Giving more critical attention to purported evidence contradicting our beliefs and holding it to higher acceptance standards—quantitative and qualitative selective scrutiny—are examples of reasoning differently about evidence that supports our prior beliefs than when it contradicts those beliefs but need *not* indicate motivated irrationality. They are

reasoning patterns which can just as well emerge from caring about accuracy. Even if they lead to belief polarization.

More generally, contra what Caddick and Feist (2022) seem to imply in the quote above, such belief-consistent information processing can perfectly well occur without motivation or irrationality (cf. Oeberst & Imhoff, 2023). At least, then, *something more* is needed to link reasoning differently about evidence that supports our prior beliefs than when it contradicts those beliefs, to motivated irrationality. Because by itself, such reasoning is undiagnostic evidence of motivated irrationality. So understood, we can see the argument by Avnur and Scott-Kakures (2015) as an attempt to supply this needed extra element. According to them, one particular instance of reasoning differently about evidence that supports our prior beliefs than when it contradicts those beliefs—qualitative selective scrutiny—reveals the belief-undermining directional influence of a *desire*. However, their argument is ultimately unconvincing because (a) qualitative selective scrutiny might very well not indicate this and (b) the non-normative influence of desire on reasoning is significantly constrained.

The upshot: while Kelly's (2008) defense of the normativity of belief polarization is incomplete because it only covered quantitative selective scrutiny, it still stands. Belief polarization and both forms of selective scrutiny are not convincing evidence for motivated irrationality. And common claims like the one by Caddick and Feist (2022, p. 428) are too trigger-happy in diagnosing motivated reasoning. More precision is needed.

In this section, let us consider one more attempt to link reasoning differently about evidence that supports one's prior beliefs than that contradicts those beliefs to a deeper source of motivated irrationality. Carter and McKenna (2020), namely, argue for what they call "the skeptical import of motivated reasoning" not because it implicates desire-based unreliability, but because it indicates a troubling influence of our *political identities* on reasoning. They specify they're interested in "the impact that our political beliefs and convictions have on our assessment of arguments that pertain to those beliefs and convictions." And propose to "call this politically motivated reasoning" (Carter & McKenna, 2020, p. 703). Like its directional superset, politically motivated reasoning is typically contrasted with a motivation for accuracy when reasoning. The authors clarify that "the operative notion of "political belief" is very broad indeed," such that the world 'political' hardly seems an important constraint and their argument ties into the discussion of motivated reasoning generally (p. 704).

According to Carter and McKenna (2020, p. 706), "if a subject engages in politically motivated reasoning when assessing some evidence or argument, their assessment of that evidence or argument is nontrivially influenced by their background political beliefs" (in the very broad sense of 'political belief' they have in mind). "If the evidence or argument causes trouble for those beliefs, [people] try to reject it, explain it away, or minimize its importance; if the evidence or argument supports those beliefs, they enthusiastically endorse it, and exaggerate its importance." This by

now familiar pattern of selective scrutiny and asymmetric acceptance thresholds, they tell us, “leads many of us to form beliefs about scientific topics that conflict with the scientific consensus.”⁶ Because of this, Carter and McKenna (2020, p. 714) conclude—directly positioning themselves against Kelly—that “the phenomenon of motivated reasoning raises insidious skeptical problems—and accordingly that the epistemological ramifications of motivated reasoning are much more serious than one may initially think (see, e.g., Kelly, 2008).”

While that quote might suggest differently since Kelly (2008) is mentioned as main opponent, their argument does not revolve around belief polarization and selective scrutiny. Rather, Kelly (2008) is treated in a more general way as a proponent of the position that skeptical epistemological ramifications of motivated reasoning remain doubtful. The specific psychological experiments Carter and McKenna (2020) rely on to challenge that view are not about belief polarization, but about *biased assimilation*.⁷ Terms like motivated reasoning, disconfirmation belief polarization, and biased assimilation are frequently mixed up or treated as synonyms (as also noted e.g., by Van der Linden, 2023, p. 44; Stanovich, 2021, p. 15), but they are not, so we should be precise about their meanings. The normative analysis of these phenomena, I believe, is hampered by their inconsistent use in the literature. The difference between the latter two is the dependent variable they refer to. In experiments on belief polarization, researchers typically measure as outcome variable the extent to which presented information *changes* people’s (self-reported confidence in) relevant *beliefs*. Whereas in studies on biased assimilation, the dependent variables typically include people’s subjective *evaluations* of the new *information*. Biased assimilation refers to individuals’ predisposition to evaluate information that contradicts their priors more negatively than information that confirms their priors (Hahn & Harris, 2014).

Typically, an experiment on biased assimilation involves randomly assigning people to one of two treatments. In each condition, people receive some information. Across treatments, almost all characteristics of the information are held constant, except the upshot of the information—which is manipulated to be consistent with either one type of outcome or another (e.g., that death penalty laws reduce crime or do not reduce crime in Lord et al. (1979)). Researchers measure peoples’ evaluations of the reliability of the information on self-report scales, and, typically, also assess covariates (e.g., political identity or prior beliefs). The critical inferential test is then conducted on the interaction between treatment (i.e., information) and covariate (e.g., political identity, prior beliefs). If peoples’ evaluations of information reliability in these *matched-information designs* are observed to be conditional on their preference or identities, biased assimilation is said to have taken place. And this, in turn, regularly serves

⁶I consider this causal claim in the next section.

⁷Like belief polarization, one more term made prominent by Lord et al. (1979).

as ground for inferring politically motivated reasoning, as for Carter and McKenna (2020).

In their support, numerous studies have shown that people are indeed prone to 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; Miller et al., 1993; Munro & Ditto, 1997; Taber & Lodge, 2006). Let's look at some examples. As evidence for the "biasing effects" (Avnur & Scott-Kakures, 2015, p. 13) of motivated reasoning, both Avnur and Scott-Kakures (2015) and Carter and McKenna (2020) build their case on a widely cited study by Kunda (1987). In this study, Kunda gave subjects a study showing that women who are heavy drinkers of coffee are at high risk of developing fibrocystic disease. She asked the subjects to indicate how convincing the study was. In one treatment, fibrocystic disease was characterized as a serious health risk, and women who are heavy coffee drinkers rated the article as less convincing than women who are light coffee drinkers (and than men). In another condition, the disease was described as common and innocuous and both groups of women rated the article as equally convincing. This resembles the phenomenon of biased assimilation, but in a non-political context, and possibly implies that heavy coffee-drinking women were resistant to the new, undesirable information. Indeed, Kunda's interpretation of her findings is that subjects engage in motivated reasoning and discount the article when it clashes with what they want to believe.

In addition to measuring subjects posterior confidence in their views on the death penalty, Lord et al. (1979) also inquired into biased assimilation. They found that—indeed—participants self-reported evaluations of the mixed evidence they received seemed to depend heavily on their (prior) position on the death penalty. Subjects who favored capital punishment were more likely to endorse a particular methodology if the study that used it found evidence for the deterrent effect of the death penalty; the same methodology was regarded as inferior when it generated the opposite conclusion.

A third example comes from studies on myside bias in evaluation tasks (Stanovich & West, 2007, 2008). "Natural myside bias" is typically defined as "the tendency to evaluate propositions from within one's own perspective" (Stanovich & West, 2007, p. 225). Or as when "we evaluate evidence, generate evidence, and test hypotheses in a manner biased toward our own prior beliefs" (Stanovich, 2021, p. 7). Such a tendency is presumably problematic because the ability to deal with evidence in an unbiased manner and the ability to take multiple perspectives when thinking about a problem are important yet conflict with my-side thinking (Stanovich et al., 2013). For example, Stanovich and West (2008, p. 138) cite as evidence of the myside bias their observation that, in their sample, females were significantly more favorable towards the proposition "There is bias in favor

of males in admissions to medical school, law school, and graduate school” than men. They and others find more such cases where “people with a particular stance or group status evaluate propositions [involving that group] differently from those having the opposite group status” (Stanovich & West, 2008, p. 140). By contrast, “the critical thinking literature [...] strongly emphasizes our ability to decouple prior beliefs and opinions from the evaluation of new evidence and arguments” (Stanovich, 2021, p. 46). So from that perspective, “it seems natural to see myside bias as a dysfunctional thinking style” (Stanovich, 2021, p. 46).

One might feel this normative diagnosis needs more precision. After all, in any decision-making process where trust in others or any kind of knowledge-based deliberation is involved, your starting point is and cannot but be your own beliefs, conditional probabilities, epistemic procedures and so on. If this is a cause for concern, it is a cause for much more general concern—if this fact undermines justification, the most radical of skepticisms seems to follow (Van Cleve, 2003). But if at least sometimes justification can be had despite the fact that your starting point is *your* starting point, if starting there does not amount to begging the (or a) question in any objectionable way, then it’s hard to see why precisely the particular tendency picked out by the myside bias should be non-normative reasoning.

After all, social group membership might very well be epistemically relevant to forming judgments about political issues like whether “there is bias in favour of males in admissions to medical school, law school, and graduate school” (Stanovich & West, 2008, p. 138). This is because part of what it means to belong to a social group defined along an axis such as gender is that one occupies a distinctive position in the social structure (Young, 2000). This, in turn, means that one is subjected to a distinctive set of social constraints and enablements by the laws, norms, and physical infrastructure that constitute the social context. The social group ‘women’, on this view, is partly defined by exposure to a distinctive set of shared constraints, such as some cultural norms that discourage potentially high-status females from pursuing prestigious careers, culminating in biased admissions to medical, law, and graduate schools. Now, because they experience group-specific constraints and enablements, members of a social group have distinctive knowledge that members of other groups may lack (Lepoutre, 2020). Members of contrasting social groups often have different experiences, inhabit different social networks, and more generally encounter different information, including different forms of misinformation (Pennycook et al., 2022). They will therefore have different prior beliefs, and evaluate factual statements differently even when they’re Bayesian reasoners not particularly motivated to embrace biased myside beliefs.

Indeed, persuasion research (Hoeken et al., 2020), research on motivated reasoning (Tappin et al., 2020) and psychological studies generally (Hahn & Harris, 2014) have been criticized for failing to clearly articulate a

standard of rationality, comparison with which would deem some observed behavior irrational. As Elqayam and Evans (2011) point out, it is becoming increasingly rare to find “single norm paradigms” in reasoning and decision making research—tasks where a single normative model is undisputed. Evans (1993) refers to this as the “normative system problem,” and Stanovich (2011) similarly talks of the “inappropriate norm argument.”

Typically, for instance, studies of belief polarization have not explicitly included normative models of how people should interpret information and update their beliefs, simply relying on the supposedly common-sense assumption that belief polarization is irrational. The same can be said of studies on biased assimilation, where the assumption is that participants’ evaluations of the mixed evidence should not diverge based on those participants’ different (political) priors. Rather, they should rate the overall diagnostic value of the evidence the same, as the positive and the negative evidence balances each other out. But what epistemological principle might vindicate this intuition?

One candidate is what Baron (2008, pp. 208–211) calls the “neutral evidence principle”: “Neutral evidence should not strengthen belief.” Mixed evidence, such as in Lord et al. (1979), should not change our beliefs, because positive and negative evidence should balance each other out. Regardless of our prior beliefs, the diagnostic impact of (mixed) evidence should be the same.⁸ This neutral evidence principle is presumably violated when ambiguous evidence is interpreted as supporting a favored belief.

Another often-used standard for assessing the normative status of some mode of reasoning is the Bayesian framework. Bayesian updating is a theoretical model of the process for incorporating new information into prior beliefs to arrive at an updated belief (Bullock, 2009). Bayes rule is frequently described as the model according to which “rational” people ought to process new information. According to this model, when individuals encounter new information, they incorporate the new information with their prior beliefs to form an updated posterior belief. In the Bayesian framework, we can calculate how individuals “should” update their beliefs as a function of the specific prior belief of the individual plus the diagnosticity of any new information provided. This means we can ask to what extent subjects, given their prior beliefs and the information they’re given, update their beliefs in accordance with this benchmark.

Let us look at an example to see how that works. In studies like Hill (2017), psychologists try to answer this question by investigating people’s reactions to noisy but informative information about factual political questions. They first gather data on people’s prior beliefs and their perceptions

⁸It is worth noting here that Jern et al. (2014) have shown that belief polarization can be consistent with a normative account of belief revision. In some cases, rational agents with opposing beliefs should both strengthen their positions as a result of reading the same information. When information is experimentally crafted to be ambiguous, polarization may arise, but this too seems compatible with a model on which reasoners are motivated to get to the truth of the matter rather than to arrive at a particular conclusion (cf. Benoit & Dubra, 2019).

of the informativeness of the information. Using this data, they calculate the posterior beliefs that would be expected according to Bayes' rule. They then compare the observed posterior beliefs of individuals to this Bayesian benchmark, examining the extent to which subjects' posterior beliefs diverge from the benchmark, perhaps as a function of the political favorability of the new information. For example, if individuals' posterior beliefs are "too far" in the direction of politically favorable information and/or "not far enough" in the direction of politically unfavorable information, this could be taken as evidence of a skeptical import of politically motivated reasoning. Hill (2017) observed that both US Republicans and Democrats updated their beliefs after receiving evidence about the truth (or falsity) of various partisan political facts—even if the evidence was politically uncongenial.

Given the popularity of Bayesian models of cognition, it's worth noting that the neutral evidence principle is in fact not an adequate normative principle from a Bayesian point of view. The strength of a piece of evidence doesn't remain static across a spectrum of priors, and Bayesian inference has its relationship with accuracy not just despite the fact that judgment is influenced by priors, but also because of it (Hahn & Harris, 2014, p. 89). Where information is received sequentially, as it is in reality, priors summarize past evidence, and it seems plain silly to ignore what we know when assessing new information.

Some authors have interpreted findings on biased assimilation as showing that people reject evidence with which they disagree and are therefore impervious to information that contradicts their views (e.g., Carter & McKenna, 2020; Kahan et al., 2011). But to me these findings show that people *actually do* have prior beliefs that they *actually believe*. Capital punishment proponents *actually think* that the death penalty has a deterrent effect, so it seems reasonable for them to reason that studies purporting to show the opposite are more likely to be incorrect. So the epistemically relevant feature of the study that in their mind justifies demoting it, is that it suggests a *false* conclusion, not that it suggests a conclusion that is merely different from their worldview or desires (cf. Stanovich, 2021, 2023 on knowledge projection). The feature of the situation they take to be of normative epistemic significance—what their reason is for making up their mind about the study's reliability—is that it suggests *not-p*, whereas *p* (Enoch, 2010).

This seems likely, as such reasoning seems to fall out of what it means to have prior beliefs in the first place. Individuals with different prior beliefs about the deterrent effect of capital punishment *should* have different views about the quality of new evidence if they update belief by Bayes' rule. If a study provides a conclusion that is inconsistent with what a subject already knows, she is right to be skeptical of its quality. So why, exactly, would biased assimilation—the finding that subjects on either side of an issue both report that evidence that matches their view is more credible than contrary evidence—have the skeptical import that's often

attributed to it? As Lord et al. (1979) themselves remark, this asymmetry is not in and of itself problematic, as it may be rational for a person to have greater confidence in a finding that confirms something she believes than a finding that disconfirms her belief.

This way, the standard model of rational learning—using Bayes rule, without any directional motives—can make predictions which are qualitatively consistent with many findings on biased assimilation, meaning those findings in general do not seem to provide convincing evidence for the presence of politically motivated reasoning (Bullock, 2009). A complexity which too often goes unnoticed in the literature on motivated reasoning, where one is likely to find claims along the lines of the previously cited “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). It’s too quick to say that if subjects with different political background beliefs about a fact (say, that global warming is real) provide different assessments of the quality of evidence about that fact (e.g., a study about the severity of global warming), this indicates that they are engaged in motivated reasoning. Those with different partisan leanings may simply have different prior beliefs, which will generally lead to different posterior beliefs about the quality of the evidence or soundness of factual statements.

Conclusion: we can add biased assimilation and the *myside bias* to our list of reasoning phenomena which are, like selective scrutiny and belief polarization, too thoughtlessly categorized as non-normative. It is not actually clear why differential ratings of information reliability or factual statements constitute evidence of motivated identity-based irrationality rather than a Bayesian influence of prior beliefs. All in all, then, it seems premature to conclude that biased assimilation—subjective evaluations of argument and information quality varying with (political background) beliefs—has the dire skeptical implications suggested by Carter and McKenna (2020). The same observational equivalence problem that we encountered on a first-person level in section 2, makes it tricky to infer something about people’s reasoning process from their information evaluations. Disagreeing about the quality of a study as a result of motivated reasoning is observationally equivalent to disagreement driven by different prior beliefs about the fact in question (Little, 2022). As Druckman and McGrath (2019, p. 111) conclude: “There is scant evidence for directional motivated reasoning when it comes to climate change: the evidence put forth cannot be distinguished from a model in which people aim for accurate beliefs, but vary in how they assess the credibility of different pieces of information.”

Does politically motivated reasoning lead to false beliefs?

However, biased assimilation is not the only way that, according to Carter and McKenna (2020, p. 702), motivated reasoning might have “negative

import for the epistemic statuses of beliefs formed in part through such reasoning.” Motivated reasoning might still prove to be an unreliable way of forming beliefs, not primarily by influencing people’s *assessment of information*, but because the stricter criteria applied to belief-inconsistent information might lead individuals to form or retain *false beliefs*. The Bayesian idea is that, if our personal probabilities are erroneous, their impact will get washed out in due time, as the number of observations increases. But, this suggestion runs, this won’t happen if uncongenial information is constantly being dismissed as unreliable after we scrutinize it carefully and employ stringent acceptance thresholds. For example, a popular explanation for the public divide in beliefs about climate change attributes it to people engaging in directional motivated reasoning (Hart & Nisbet, 2012). According to this account, individuals skeptical about climate change reject ostensibly credible information because it counters their standing beliefs. Along the same lines and using the same example (climate change), Carter and McKenna (2020, p. 703) claim “there is empirical evidence which suggests that [politically motivated reasoning] leads many of us to form beliefs about scientific topics that conflict with the scientific consensus.”

Notice the causal element in this claim. Politically motivated reasoning—defined as the impact that our political beliefs (in a very broad sense of ‘political’) have on our assessment of evidence and arguments—is said to *lead to* false beliefs. Because our information evaluations go awry, our posterior beliefs end up inadequate as well.

We have seen, in the last section, that it’s not clear that there’s anything non-normative and skepticism-warranting about biased assimilation on its own. In this section, I’ll argue that the claim that it leads to false beliefs is equally unwarranted.

In arguing for the claim that politically motivated reasoning tends to distort our reasoning, Carter and McKenna (2020) rely heavily on work by Dan Kahan on so-called cultural cognition. Roughly, cultural cognition seeks to explain why groups with different values tend to disagree about important societal issues. In particular, the cultural cognition thesis argues that public disagreement over key societal risks (e.g., climate change, nuclear power) arises not because people fail to understand the science or lack relevant information, but rather as a result of the fact that “people endorse whichever position reinforces their connection to others with whom they share important ties” (Kahan, 2010, p. 296). This latter notion is central to much of the cultural cognition thesis and is generally referred to as a specific form of motivated reasoning (but see Van der Linden (2016)). As Carter and McKenna (2020, p. 704) write: “The thought is that the motivation or goal that is served by politically motivated reasoning is, broadly speaking, the goal of identity protection—that is, the goal of forming beliefs that protect and maintain our status within a group that defines our identity and whose members are united by a shared set of values.” A key prediction that flows from this theory is that when people

are exposed to (new) information, “culturally” biased cognition will merely reinforce existing predispositions and cause groups with opposing values to become even more polarized on the respective issue—a prediction ostensibly confirmed by studies like Lord et al. (1979).

Research on cultural cognition typically uses the same design as the discussed studies on biased assimilation. In general, subjects are randomly assigned to receive one of two pieces of information; where the substantive detail of the information is held constant across conditions, but its implication for subjects’ political identities or preferences is varied between conditions. Concretely, Democrats and Republican participants are exposed to a piece of information on, usually, climate change. This piece of information is identical for both groups, except for the truth implied by its conclusion: identical methods or sources are described as reaching politically congenial versus uncongenial conclusions. The key result, typically, is that subjects’ evaluation of the information differs by condition, and, in particular, that this difference is correlated with their political identities or preferences. Specifically, people evaluate the information less favorably when it is discordant with their political identities or preferences than when it is concordant with their political identities or preferences. For example, a key result in the cultural cognition paradigm is that Democrats (or what Kahan calls “cultural egalitarians”) see environmentalist climate scientists as more trustworthy on the topic of climate change, while Republicans (Kahan speaks of “hierarchical individualists”) rate them as less reliable than scientists who are more skeptical of global warming (Kahan et al., 2011). Like many scholars, Carter and McKenna (2020) conclude from such results that politically motivated reasoning was involved: the subjects were motivated to reach one political conclusion over another. That explains the patterning of the information ratings.

However, this inference is not warranted by data gathered using this type of study design. Causal inferences of politically motivated reasoning assume that the information treatment affects subject’s reasoning only insofar as it activates politically motivated reasoning (or identity-protective cognition) or not. However, people’s political group identity is typically correlated with their prior beliefs about the specific issue under study. This means that prior beliefs undermine inferences of politically motivated reasoning (Tappin et al., 2020). The results from these designs are susceptible to (confounding) explanations based on prior beliefs. The random assignment of information not only varies the consistency of said information with peoples’ desires, political identities, and so on, but also with their prior beliefs. Empirical evidence supports the idea that the correlation between group identities and reasoning is due to prior factual beliefs (Tappin et al., 2021). The tendency for reasoning to be affected by the coherence between new information and prior factual beliefs is a feature of human psychology that is independent of political group motivation (Markovits & Nantel, 1989; Trippas et al., 2018). It’s hard to distinguish

cultural cognition from straightforward Bayesian updating on general beliefs (Greco, 2021).

In this light, then, the observed patterns of information evaluation may reduce to “people are more receptive to evidence that confirms their prior beliefs” (Williams, 2018, p. 142). And as pointed out in the previous section, such biased assimilation in the interpretation of new information does not provide particularly convincing evidence of a violation of Bayesian inference. Which means that cultural cognition studies don’t provide evidence that politically motivated reasoning *leads to* false beliefs, contra to what Carter and McKenna (2020) suppose (cf. Tappin et al., 2021). As this is their main empirical evidence, this makes their case for a skeptical import of motivated reasoning less than clear.

Cultural cognition can be seen as motivated reasoning made social (Levy, 2021, p. 30). It says we are motivated to reject some hypothesis because it is threatening to our group identity. Such identity-protective cognition explains, on this view, why some social groups reject the science of climate change. But according to an alternative account (Levy, 2021), these people reject the science of climate change because the social mechanisms of belief updating provide them with epistemic reasons to do so. They deploy social referencing, asking themselves what people like them believe. Multiple cues tell them that people like them reject the science (think of how merchants of doubt play on cues to identity). Rather than thinking of social referencing as identity-protective, people might very well deploy it to respond to social cues *as* evidence. Not *rather than* evidence: “The fact that a proposition is socially approved is higher-order evidence that bears on its truth, and there’s nothing irrational in being guided by it. The primary purpose for which we deploy these mechanisms is to get things *right*, not (just) to fit in” (Levy, 2021, p. 81).

This seems to match the reasoning of theorists working on cultural evolution. Joseph Henrich, for instance, argues that “[l]ike natural selection, our cultural learning abilities give rise to “dumb” processes that can, operating over generations, produce practices that are smarter than any individual or even group” (Henrich, 2016, p. 12). The idea that cumulative cultural evolution and cultural transmission is crucial to our intellectual and cognitive abilities, and that this requires highly developed social learning skills, is generally accepted (Boyd & Richerson, 2005; Herrmann et al., 2007; Heyes, 2018; Sterelny, 2012). So at least we should not be too quick to jump to the conclusion that mechanisms that *seem* partisan stem from identity protection rather than social learning.⁹

Nevertheless, several researchers have theorized that biased evaluation processes contribute to belief polarization in response to mixed evidence (e.g., Lord et al., 1979; Taber & Lodge, 2006). Fatollahi (2023, p. 2) even asserts there is a “well-documented causal linkage” between biased assimilation (evidence ratings) and belief polarization (opinion changes). But that

⁹And see Rini (2017) on the reasonableness of relying on co-partisanship in determining whom to trust.

statement seems unwarranted. The two outcome variables of (i) information evaluations versus (ii) posterior beliefs following exposure to the information can yield divergent results (Anglin, 2019), and sometimes dramatically so (Kim, 2020). Gerber and Green (1999, p. 206) even concluded that “making sense of the literature on biased learning requires a sharp distinction between studies that examine the credence subjects place in an argument and studies that examine how new evidence changes existing beliefs.” Those are two normative targets, and they’re often not clearly enough distinguished in arguments on the epistemological import on motivated reasoning.

To illustrate: Kunda’s (1987) results, as discussed, showed that the heavy coffee-drinking women made substantial belief updates towards the information, indicating that they incorporated the information into their prior beliefs despite their negative evaluations. That is, although heavy coffee drinkers in the serious health risk treatment describe the article as less convincing than in the innocuous risk treatment, they seem to be equally convinced in the two treatments. This pattern was not evident from the information evaluations alone and sheds doubt on the idea that biased assimilation accounts for belief polarization through a causal link. It makes the case for irrationality motivated reasoning based on biased assimilation again less than clear, even for contentious topics where motivated reasoning would seem to be an intuitive explanation. The fact that biased assimilation doesn’t (reliably) bias posterior beliefs, weakens the case for inferring motivated irrationality on its basis. And it suggests that ‘biased assimilation’ is a misleading label. While term suggests that disliked evidence doesn’t get ‘assimilated’ into subject’s posterior beliefs, as we’ve seen, even disliked evidence can exert persuasive influence on attitudes. Accordingly, contra Carter and McKenna (2020) second argument, the empirical evidence does not support the claim that motivated reasoning leads many of us to form false beliefs.

Conclusion

When people reason differently about information that confirms vs. disconfirms their prior beliefs, it is often inferred that they’re engaging in motivated reasoning. Most research on motivated reasoning so defined, in turn, proceeds on the assumption that it runs afoul of one or another epistemic norm—that it is an “important source of epistemic irrationality in human cognition” (Williams, 2020, p. 17), to be regarded as a mark against the quality of human judgment (Ellis, 2022). Hence, instances of belief-consistent information processing such as selective scrutiny, belief polarization, the myside bias and biased assimilation are often explained by, and seen as evidence for, motivated irrationality. The non-normativity of many such cases of belief-consistent information processing is often assumed without argument. But, in general—and that’s the central claim I’ve made—common inferences of motivated irrationality based on belief-consistent information

processing are not adequately supported. Once the details are fleshed out, it turns out to not actually be straightforward to make good on the claim that many reasoning patterns often seen as non-normative arise from, and are thus evidence for, motivated irrationality.

In arguing this, the paper covered a lot of ground. Starting with Kelly's (2008) claim that, in experiments where subjects polarize after exposure to ambiguous evidence, the justification of their beliefs is not undermined by their being the result of mechanisms that underlie polarization because they result from rational management of cognitive resources. Kelly's defense of selective scrutiny, however, was incomplete. This is because subjects also use differential acceptance criteria for congenial and uncongenial purported evidence, whose presence, Avnur and Scott-Kakures (2015) argue, constitute evidence for desire-based directionally motivated reasoning, making it evidence that defeats the justification of a belief. In fact, such asymmetric acceptance thresholds turned out to be undiagnostic with regard to the presence of a belief-undermining desire. Moreover, a deeper examination of the empirical evidence underlying these phenomena showed that it's far from clear that a presumed wishful thinking effect makes our belief-forming process unreliable to the extent Avnur and Scott-Kakures (2015) suggest. In fact, then, selective scrutiny and belief-polarization in response to mixed, ambiguous evidence—instances of belief-consistent information processing—have not convincingly been established as non-normative reasoning. Nor does the empirical evidence support the conclusion that they are grounded in a deeper source of motivated irrationality such as a belief-undermining desire.

In their related case for the skeptical import of motivated reasoning, Carter and McKenna (2020) claim that patterns of biased assimilation indicate motivated irrationality, because they are evidence one's reasoning was shaped by political motivations. In response, I highlighted that most evidence for this claim derives from study designs that do not permit causal inferences on the role of motivation. Rather, these study designs reveal that people condition their evaluation of new information on their prior beliefs. These results seem consistent with (rational) non-motivational Bayesian inference. Similarly, when members of contrasting social groups evaluate a factual proposition involving these groups differently, it is typically inferred they are biased towards their own side. Here too, however, it appeared doubtful whether the inference that a non-normative bias rather than a Bayesian influence of prior beliefs was at work, was in fact warranted by the evidence. Finally, the causal inference that motivated reasoning leads us to form false beliefs was not found to be warranted by the cited empirical evidence either.

Belief-consistent information processing is frequently associated with motivated reasoning. Such claims are common in the literature, but are not in fact well-established. The arguments put forward in this paper call for a more nuanced engagement with questions about the epistemically normative status of belief-consistent information processing.

Acknowledgements

Thanks to the audience at the OZSW conference 2023, Jos Hornikx and Simon Rippon for helpful discussion. Thanks to Henri Markovits and an anonymous reviewer from this journal for helpful comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

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

References

- Almagro, M. (2022). Political polarization: Radicalism and immune beliefs. *Philosophy & Social Criticism*, 49(3), 309–331. <https://doi.org/10.1177/01914537211066859>
- 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>
- Avnur, Y., & Scott-Kakures, D. (2015). How irrelevant influences bias belief. *Philosophical Perspectives*, 29(1), 7–39. <https://doi.org/10.1111/phpe.12060>
- Babad, E., & Katz, Y. (1991). Wishful thinking—Against all odds. *Journal of Applied Social Psychology*, 21(23), 1921–1938. <https://doi.org/10.1111/j.1559-1816.1991.tb00514.x>
- Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning*, 1(1), 71–103. <https://doi.org/10.1080/13546789508256906>
- Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008). Predicting World Cup results: Do goals seem more likely when they pay off? *Psychonomic Bulletin & Review*, 15(2), 278–283. <https://doi.org/10.3758/pbr.15.2.278>
- Baron, J. (2008). *Thinking and deciding*. Cambridge: Cambridge University Press.
- Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology*, 32(1), 176–184. <https://doi.org/10.1037/h0076771>
- Benoît, J., & Dubra, J. (2019). Apparent bias: What does attitude polarization show? *International Economic Review*, 60(4), 1675–1703. <https://doi.org/10.1111/iere.12400>
- 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>
- Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures*. Oxford 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>
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460–473. <https://doi.org/10.1037/0022-3514.66.3.460>
- Coppock, A. (2022). *Persuasion in parallel. How information changes minds about politics*. The University of Chicago Press.
- Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, 8(2), 214–220. <https://doi.org/10.1177/0146167282082005>
- Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the was on selection task. *Personality and Social Psychology Bulletin*, 28(10), 1379–1387. <https://doi.org/10.1177/014616702236869>
- 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), 568–584. <https://doi.org/10.1037/0022-3514.63.4.568>
- Ditto, P. H., Munro, G. D., Lockhart, L. K., Scepansky, J. A., & Apanovitch, A. M. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 75(1), 53–69. <https://doi.org/10.1037/0022-3514.75.1.53>
- Dorst, K. (Forthcoming). Rational polarization. *Philosophy and Phenomenological Research*.
- Druckman, J., & McGrath, M. (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>
- Ellis, J. (2022). Motivated reasoning and the ethics of belief. *Philosophy Compass*, 17(6), e12828. <https://doi.org/10.1111/phc3.12828>
- Elqayam, S., & Evans, J. (2011). Subtracting “ought” from “is”: Descriptivism versus normativism in the study of human thinking. *The Behavioral and Brain Sciences*, 34(5), 233–248. <https://doi.org/10.1017/S0140525X1100001X>
- 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>
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Evans, J. S. B. T. (1993). Bias and rationality. In D. E. Over (Ed.), *Rationality: Psychological and philosophical perspectives* (pp. 6–30). Taylor & Francis/Routledge.
- Fatollahi, A. (2023). Conservative treatment of evidence. *Episteme*, 20(3), 568–583. <https://doi.org/10.1017/epi.2022.29>
- Foley, R. (2001). *Intellectual trust in oneself and others*. Cambridge University Press.
- 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>
- Gilovich, T. (1991). *How we know what isn't so*. Free Press.
- Greco, D. (2021). Climate change and cultural cognition. In M. Budolfson, T. McPherson, & D. Plunkett (Eds.), *Philosophy and climate change* (pp. 178–197). Oxford University Press. <https://doi.org/10.1093/oso/9780198796282.003.0009>
- 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>

- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732. <https://doi.org/10.1037/0033-295X.114.3.704>
- Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, 39(6), 701–723. <https://doi.org/10.1177/0093650211416646>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Henrich, J. (2016). *The secret of our success. How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science (New York, N.Y.)*, 317(5843), 1360–1366. <https://doi.org/10.1126/science.1146282>
- Heyes, C. (2018). *Cognitive gadgets. The cultural evolution of thinking*. Harvard University Press.
- Hill, S. J. (2017). Learning together slowly: Bayesian learning about political facts. *The Journal of Politics*, 79(4), 1403–1418. <https://doi.org/10.1086/692739>
- Hoeken, H., Hornikx, J., & Linders, Y. (2020). The importance and use of normative criteria to manipulate argument quality. *Journal of Advertising*, 49(2), 195–201. <https://doi.org/10.1080/00913367.2019.1663317>
- Jern, A., Chang, K. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, 121(2), 206–224. <https://doi.org/10.1037/a0035941>
- Kahan, D. (2010). Fixing the communications failure. *Nature*, 463(7279), 296–297. <https://doi.org/10.1038/463296a>
- Kahan, D. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424. <https://doi.org/10.1017/S1930297500005271>
- Kahan, D., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174. <https://doi.org/10.1080/13669877.2010.511246>
- Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. *Journal of Philosophy*, 105(10), 611–633. <https://doi.org/10.5840/jphil20081051024>
- Kim, J. W. (2020). *Evidence can change partisan minds rethinking the bounds of partisan-motivated reasoning*. <https://jinwookimqssdotcom.files.wordpress.com/2020/08/aca-paper.pdf>
- Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, 71(5), 1347–1366. <https://doi.org/10.1111/1467-8624.00232>
- 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>
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211–228. <https://doi.org/10.1037/0033-295X.94.2.211>

- Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 596–604. <https://doi.org/10.1037/0278-7393.15.4.596>
- Klenk, M. (2019). Objectivist conditions for defeat and evolutionary debunking arguments. *Ratio*, 32(4), 246–259. <https://doi.org/10.1111/rati.12230>
- 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>
- Kraft, P. W., Lodge, M., & Taber, C. S. (2015). Why people “don’t trust the evidence”: Motivated reasoning and scientific beliefs. *The ANNALS of the American Academy of Political and Social Science*, 658(1), 121–133. <https://doi.org/10.1177/0002716214554758>
- Kuhn, D. (1991). *The skills of arguments*. 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>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- 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. (2022). Do your own research!. *Synthese*, 200(5), 356. <https://doi.org/10.1007/s11229-022-03793-w>
- Little, A. (2022). Detecting motivated reasoning. *OSF Preprints* <https://doi.org/10.31219/osf>
- Lord, C., Ross, L., & Lepper, M. (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>
- Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, 17(1), 11–17. <https://doi.org/10.3758/BF03199552>
- McWilliams, E. (2021). Evidentialism and belief polarization. *Synthese*, 198(8), 7165–7196. <https://doi.org/10.1007/s11229-019-02515-z>
- Mele, A. R. (1997). Real self-deception. *The Behavioral and Brain Sciences*, 20(1), 91–102. <https://doi.org/10.1017/s0140525x97000034>
- 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>
- 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>
- 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>
- Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of bias-

- es. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*. <https://doi.org/10.1177/17456916221148147>
- Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2022). Beliefs about COVID-19 in Canada, the United Kingdom, and the United States: A novel test of political polarization and motivated reasoning. *Personality & Social Psychology Bulletin*, 48(5), 750–765. <https://doi.org/10.1177/01461672211023652>
- Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, 77(5), 562–571. <https://doi.org/10.1037/0022-0663.77.5.562>
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 323–390). McGraw-Hill.
- Plous, S. (1991). Biases in the assimilation of technological breakdowns: Do accidents make us safer? *Journal of Applied Social Psychology*, 21(13), 1058–1082. <https://doi.org/10.1111/j.1559-1816.1991.tb00459.x>
- Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, 27(25), E-43–E-64. <https://doi.org/10.1353/ken.2017.0025>
- Ross, L., & Anderson, C. A. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 129–152). Cambridge University Press.
- Sá, W. C., Kelley, C. N., Ho, C., & Stanovich, K. E. (2005). *Actively open-minded thinking scale* [Database record]. APA PsycTests. <https://doi.org/10.1037/t12030-000>
- Sanbonmatsu, D. M., Posavac, S. S., Kardes, F. R., & Mantel, S. P. (1998). Selective hypothesis testing. *Psychonomic Bulletin & Review*, 5(2), 197–220. <https://doi.org/10.3758/BF03212944>
- Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology. General*, 141(4), 630–634. <https://doi.org/10.1037/a0027405>
- Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, 22(2), 93–121. [https://doi.org/10.1016/0022-1031\(86\)90031-4](https://doi.org/10.1016/0022-1031(86)90031-4)
- Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, 36(11), 1202–1212. <https://doi.org/10.1037/0022-3514.36.11.1202>
- Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.
- Stanovich, K. (2021). *The bias that divides us: The science and politics of myside thinking*. MIT Press.
- Stanovich, K. E. (2023). Myside bias in individuals and institutions. In H. Samaržija & Q. Cassam (Eds.), *The epistemology of democracy* (pp. 170–194). Routledge.
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13(3), 225–247. <https://doi.org/10.1080/13546780600780796>
- Stanovich, K. E., & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, 14(2), 129–167. <https://doi.org/10.1080/13546780701679764>
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, 22(4), 259–264. <https://doi.org/10.1177/0963721413480174>
- Sterelny, K. (2012). *The evolved apprentice. How evolution made humans unique*. MIT Press.

- 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., 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/10.31234/osf.io/yuzfj>
- Tappin, B. M., Berinsky, A., & Rand, D. (2023). Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues. *Nature Human Behaviour*, 7(4), 568–582. <https://doi.org/10.1038/s41562-023-01551-7>
- Tetlock, P., & Levi, A. (1982). Attribution bias: On the inconclusiveness of the cognition-motivation debate. *Journal of Experimental Social Psychology*, 18(1), 68–88. [https://doi.org/10.1016/0022-1031\(82\)90082-8](https://doi.org/10.1016/0022-1031(82)90082-8)
- Trippas, D., Kellen, D., Singmann, H., Pennycook, G., Koehler, D. J., Fugelsang, J. A., & Dubé, C. (2018). Characterizing belief bias in syllogistic reasoning: A hierarchical Bayesian meta-analysis of ROC data. *Psychonomic Bulletin & Review*, 25(6), 2141–2174. <https://doi.org/10.3758/s13423-018-1460-7>
- Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). Guilford.
- Van Cleve, J. (2003). Is knowledge easy-or impossible? Externalism as the only alternative to skepticism. In S. Luper (Ed.), *The sceptics: Contemporary essays* (pp. 45–59). Ashgate.
- Van der Linden, S. (2016). A conceptual critique of the cultural cognition thesis. *Science Communication*, 38(1), 128–138. <https://doi.org/10.1177/1075547015614970>
- Van der Linden, S. (2023). *Foolproof. Why misinformation infects our minds and how to build immunity*. W. W. Norton & Company.
- 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>
- 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. (2020). Epistemic irrationality in the Bayesian brain. *The British Journal for the Philosophy of Science*, 72(4), 913–938. <https://doi.org/10.1093/bjps/axz044>
- Wyer, R. S., Jr., & Frey, D. (1983). The effects of feedback about self and others on the recall and judgments of feedback-relevant information. *Journal of Experimental Social Psychology*, 19(6), 540–559. [https://doi.org/10.1016/0022-1031\(83\)90015-X](https://doi.org/10.1016/0022-1031(83)90015-X)