

Resistant Beliefs, Responsive Believers*

Carolina Flores

October 23, 2023

Penultimate version. Forthcoming in *The Journal of Philosophy*.

Abstract

Beliefs can be resistant to evidence. Nonetheless, the orthodox view in epistemology analyzes beliefs as evidence-responsive attitudes. I address this tension by deploying analytical tools on capacities and masking to show that the cognitive science of evidence-resistance supports rather than undermines the orthodox view. In doing so, I argue for the claim that belief requires the capacity for evidence-responsiveness. More precisely, if a subject believes that p , then they have the capacity to rationally respond to evidence bearing on p . Because capacities for evidence-responsiveness are fallible and may be masked, beliefs can be held in the face of counter-evidence. Indeed, I will argue that our best science of belief supports the claim that evidence-resistant beliefs result from masks on evidence-responsiveness capacities. This account of belief not only allows for resistance to evidence, but provides us with a framework for describing and explaining actual cases of evidence-resistance.

1 Introduction

Beliefs are often tenaciously held in the face of counter-evidence. For example, we are unlikely to revise political, moral, or religious beliefs (Markus 1986, Leeuwen 2014), beliefs in theories to which we are committed (Chinn and Brewer 1993), and beliefs about ourselves and our talents (Pyszczynski et al. 1985, Gilbert 2006).

The stubborn and commonplace evidence-resistance of belief appears in tension with the idea that belief is constitutively evidence-responsive, the orthodox view in epistemology. This raises a worrisome tension between the roles which belief plays in

***Acknowledgements:** Thanks to Frankie Egan, Danny Forman, Daniel Hoek, Dee Payton, Jake Quilty-Dunn, and Ezra Rubenstein for comments on past drafts. Thanks also to Austin Baker, Eric Bayruns Garcia, D'i Black, Laura Callahan, Andy Egan, Megan Feeney, Adam Gibbons, Samia Hesni, Tyler John, Jackson Kernion, Kelsey Laity D'Agostino, Eric Mandelbaum, Ting-An Lin, Jorge Morales, Albert Newen, and Jill North for helpful discussion. Thanks to audiences at the University of Denver, UC Santa Cruz, NYU, Rutgers, the Oxford Epistemology Group, and York University. Special thanks to Elisabeth Camp, Verónica Gómez Sánchez, Susanna Schellenberg, Ernest Sosa, Christopher Willard-Kyle, and Elise Woodard for detailed feedback on multiple drafts. Finally, thanks to two anonymous referees for comments that much improved this paper.

epistemology and in cognitive science. Perhaps epistemologists and cognitive scientists are talking about different phenomena when they talk about beliefs. Or perhaps, when they talk of beliefs, epistemologists are talking about attitudes that are rarely realized in human psychology—belief-for-angels, not belief-for-humans.

I argue that this tension is only apparent. The evidence-resistance of belief does not force us to abandon the claim that belief is constitutively evidence-responsive. Indeed, surprisingly, findings in the cognitive science of belief *support* that claim—once we properly understand both those findings and what evidence-responsiveness amounts to.

The key contribution of this paper is to deploy analytical tools on capacities and masking to show that the cognitive science of evidence-resistance supports rather than undermines the Evidence-Responsiveness Thesis. In doing so, I will state in a precise way the thesis that cognitive science supports.

To get there, I will proceed as follows. In §2, I briefly motivate the claim that belief is constitutively evidence-responsive and detail two challenges from evidence-resistance to this claim. In §3, I put forward my own account of the evidence-responsiveness of belief. This centers *capacities for evidence-responsiveness*, and incorporates a detailed account of such capacities.

I then argue that this account helps us understand the dynamics of evidence-resistant belief (§4). For this reason, it has promising applications in theorizing about phenomena that have long puzzled philosophers of mind, such as delusions, implicit biases, religious beliefs, and ideologies. At the same time, my way of articulating evidence-responsiveness is robust enough to do justice to the epistemic role of belief, fitting with a naturalistic way of spelling out the claim that belief aims at truth (§5). The result is a conception of belief that unifies epistemology and cognitive science.

2 The Evidence-Responsiveness Thesis and Two Challenges from Cognitive Science

The orthodox view of belief in epistemology is the following:

Evidence-Responsiveness Thesis: Belief is constitutively evidence-responsive.¹

This thesis holds that what makes a mental state a belief is, in part, being responsive to evidence. Nothing can be a belief unless it is evidence-responsive. This fits with traditional functionalist views of belief, on which a folk or broadly *a priori* 'input-side' functional role, and not just the 'output-side' role (that is, downstream effects on behavior, thinking, and feeling), is constitutive of belief.

The wide appeal of this claim is best explained by how it accounts for the epistemic role of belief. For instance, Velleman (2000) moves from the view that belief aims at

1. Proponents of the evidence-responsiveness thesis include Currie and Ravenscroft (2002), Egan (2008), Gendler (2008), Helton (2020), Levy (2015), Mandelbaum (2016), Shah (2003), Shah and Velleman (2005), Smith (2003), Leeuwen (2014), and Velleman (2000), with notable earlier proponents including Davidson (1985), Dennett (1981), and McDowell (1998).

truth to claiming that it is a “conceptual truth” that “the belief that p tends to be... reinforced by additional evidence of it, and to be extinguished by evidence against it” (Shah and Velleman 2005, 500).² Similarly, Tamar Gendler moves from the claim that apportioning beliefs to the evidence is a normative constraint on belief to excluding attitudes from counting as beliefs because they are not *in fact* evidence-responsive (Gendler 2008, 565–566). Finally, Grace Helton has recently appealed to an epistemic ‘ought-implies-can’ principle to argue from the epistemic obligation to rationally revise one’s beliefs in response to relevant counter-evidence to the view that subjects have the ability to revise their beliefs accordingly (Helton 2020).

Simplifying, the Evidence-Responsiveness Thesis is motivated by two desiderata on an account of belief.³ First, it serves as a way of articulating the metaphor that belief aims at truth (or perhaps at other epistemic goods, e.g., knowledge). Second, it factors in explaining for the epistemic assessability of beliefs; more specifically, the fact that beliefs are assessable for how well they respond to evidence. As such, the Evidence-Responsiveness Thesis is supposed to distinguish attitudes that are not assessable in these ways, such as imaginings, from beliefs.

I will assume that these desiderata are important and suffice to motivate the Evidence-Responsiveness Thesis. Unfortunately, the view faces two substantive challenges from the cognitive science of belief.

The first challenge has to do with *extensional adequacy*, i.e., with correctly classifying all and only beliefs as such. The problem is that many of our beliefs are evidence-resistant. A high level of evidence-resistance, the argument goes, is incompatible with evidence-responsiveness. Therefore, highly evidence-resistant beliefs are counterexamples to the evidence-responsiveness thesis.

This quick argument against the evidence-responsiveness thesis has most explicitly been articulated by Lisa Bortolotti (2009). Bortolotti argues that delusions are beliefs (as they are classified in psychiatry), yet are not evidence-responsive. If this is right, then evidence-responsiveness is not necessary for belief. Therefore, belief is not constitutively evidence-responsive.

One might resist this argument on the basis of the fact that whether delusions are beliefs is the object of substantial disagreement (Bortolotti and Miyazono 2015). However, as Bortolotti notes, non-psychiatric evidence-resistant beliefs are ubiquitous. We are unlikely to revise political, moral, or religious beliefs, beliefs in theories to which we are committed, and beliefs about ourselves and our talents, for instance.

2. See also Shah 2003 for further discussion, especially fn. 45.

3. Of course, the motivations for this view are various. For instance, some might want to endorse the Evidence-Responsiveness Thesis on interpretivist grounds (Davidson 2001, Dennett 1981, or because they endorse a strong normativist view of belief, in which epistemic normativity is constitutive of belief (not merely applicable to belief; see McHugh and Whiting (2014) for an overview of such arguments).

The main goal of this paper is to show how the Evidence-Responsiveness Thesis is compatible with results from the cognitive science of belief, so I won’t be able to consider these various arguments for the Evidence-Responsiveness Thesis. My aim here is to show that the Thesis is compatible with cognitive science—on an articulation of the Thesis that is robust enough to meet what I take to be the two least controversial epistemic motivations for it, which I list in the main text.

Similarly, because the focus is on cognitive science, I will not address challenges from thought experiments (cf. Smithies et al. (2022)). Such challenges are not relevant for whether the Evidence-Responsiveness Thesis is empirically adequate, and they rely on questionable intuitions that one may reasonably discount based on commitments from epistemology and cognitive science.

Indeed, Stanley (2015) pursues a line of argument similar to Bortolotti's starting from ideological beliefs instead of delusions.

For this reason, it is not clear how the evidence-responsiveness thesis can correctly classify all ordinary beliefs. The first challenge, then, is to specify what evidence-responsiveness amounts to in a way that correctly classifies ordinary evidence-resistant beliefs.

There is a second, less commonly noticed, challenge for the evidence-responsiveness thesis. It corresponds to a second desideratum on a theory of belief: *empirical adequacy*. For a theory of belief to be empirically adequate, it must be compatible with our best scientific findings and generalizations about the cognitive role of belief. This desideratum reflects the fact that beliefs are not only objects of epistemic assessment, but also objects of study in cognitive science, figuring in well-established empirical generalizations (Porot and Mandelbaum 2020).⁴

Troublingly, the evidence-responsiveness thesis is hard to square with a number of key findings in the psychology of belief revision. Beliefs commonly persist in the face of counter-evidence (belief perseverance (Anderson et al. 1980, Anderson et al. 1980, Slusher and Anderson 1989)). Even more worryingly, they sometimes become more entrenched when the subject receives counter-evidence (belief polarization (Festinger et al. 1956, Lord et al. 1979, Liberman and Chaiken 1992, McHoskey 1995, Lodge 2006)). The evidence-responsiveness thesis needs to accommodate these findings, but it is not clear how it can do so.

More generally, proponents should explain how the view is compatible with robust generalizations about belief. Centrally, they should address the two following generalizations: first, "beliefs will generate a negative, motivational, phenomenologically salient discomfort whenever one encounters counterattitudinal evidence," which we will then be moved to assuage "by any easily available route" (Quilty-Dunn and Mandelbaum 2018, 2367); second, if you believe extremely strongly that p and receive information against p , then you will (irrationally) increase your belief that p (Festinger et al. 1956).⁵

Easily available routes to alleviate discomfort are likely to include irrational updating. Therefore, the first generalization suggests that irrational updating is commonplace. The second generalization suggests that irrational updating is nearly inevitable for strongly held beliefs.

Indeed, Quilty-Dunn and Mandelbaum (2018) hold that the two generalizations above are laws that govern belief: beliefs are "governed by, inter alia, the laws of dissonance" (Quilty-Dunn and Mandelbaum 2018, 2367). If they are right, irrational responses to evidence are not a quirk or occasional blip, but a consequence of the (natural) laws that govern belief.⁶ It is hard to see how this is compatible with the

4. Some philosophers reject the view that belief figures in a mature cognitive science (Churchland 1981). Such philosophers will not be worried about empirical adequacy.

5. "Believe strongly," here, refers to how much the belief matters to you, not to how high your credence is. See Leeuwen (2022) for discussion of different senses of "believe strongly."

6. As a referee pointed out, talk of laws might sound strange here, given that these generalizations seem to hold in virtue of contingent features of actual cognitive systems. By 'law' here, these authors have in mind robust regularities of human psychology, not general principles that all beliefs must obey by their nature. The twist is that, according to psychofunctionalism, we can generalize from such regularities to the nature of belief in general (see the next paragraph).

claim that belief is constitutively evidence-responsive.

The problem of empirical adequacy has a special force for psychofunctionalism. According to this view, belief is a functional kind definitionally tied to psychological generalizations about belief (Block and Fodor 1972, Block 1978, Quilty-Dunn and Mandelbaum 2018). Beliefs are whatever satisfies empirical generalizations on belief. If psychofunctionalists are right, then no property can be constitutive of belief unless there is a well-established psychological generalization ascribing that property to beliefs. Unless the empirical laws of belief imply that actual beliefs are evidence-responsive, belief cannot be constitutively evidence-responsive.

To address these two challenges, proponents of the Evidence-Responsiveness Thesis need to show (1) that real-world evidence-resistant beliefs count as evidence-responsive, and (2) that the mechanisms of belief regulation deserve to be labeled as “evidence-responsive.”⁷

To begin, this requires specifying what *evidence-responsiveness* amounts to. Proposals vary here. They include the claim that beliefs “tend to” Shah and Velleman 2005 evidence-responsiveness, are “quickly revisable” (Gendler 2012), or require “the ability” (Helton 2020) to revise.

Second, a fully satisfactory response to the challenge requires spelling out what such a tendency, disposition, or ability to rationally respond to evidence requires. This is what hasn’t yet been done.⁸ Instead, proponents simply note that tendencies, dispositions, and abilities leave wiggle room for failures of evidence-responsiveness. For instance, they might grant that “[given that] belief can be influenced by evidentially irrelevant processes such as wishful thinking, responsiveness to evidence must be weak enough to leave room for such additional influences” (Shah and Velleman 2005, 500).

This is true, but unsatisfying. The problem is that there are cases where this evidence-responsiveness is never, or virtually never, displayed in how the belief is regulated (Quilty-Dunn ms). Why should we think that evidence-responsiveness is still present in such cases? This concern cannot be assuaged simply by insisting that evidence-responsiveness “doesn’t require belief to be governed by truth-seeking mechanisms alone,” (Velleman 2000, 254). The question is why we should continue to think that evidence-responsiveness is involved at all. To address it, one needs to give a positive reason to think that evidence-resistant beliefs are evidence-responsive, in

7. An alternative strategy for handling these cases would focus on showing that subjects rationally respond to evidence in cases of belief perseverance and polarization. Many theorists have recently undertaken to show this (e.g., Dorst (forthcoming), Levy (2021), Poth and Dolega (2022)). If they are right and all cases are covered by such explanations (a tall order), then empirical results are no threat to the Evidence-Responsiveness Thesis.

It is beyond the scope of this paper to consider the rich details of these views. But it is worth noting a serious cost of such a defense of the Evidence-Responsiveness Thesis: it requires us to find ways to re-describe apparent irrationality as actually rational in an extremely wide range of cases. This arguably stretches our concept of rationality beyond recognition, and defangs the power of epistemic standards for criticizing bad beliefs. In contrast, my strategy here grants that there are cases of irrationally evidence-resistant belief.

8. A notable exception here is Helton (2020), who gives a detailed account of the ability to rationally revise. That said, her project is not addressing the two challenges here, leaving an important theoretical gap that my paper addresses. I will return to a comparison with Helton’s view once I have introduced my preferred way of spelling out “evidence-responsiveness.”

a precise sense of evidence-responsiveness. I turn now to this task.

3 Evidence-Responsiveness Capacities and Belief

To address the above challenges for the evidence-responsiveness thesis, then, we need to specify the Evidence-Responsiveness Thesis in a way that makes concrete predictions about what evidence-responsiveness requires. I will now develop a view that does so, in a way that allows us to meet the Extensional and Empirical Adequacy challenges:

Evidence-Responsiveness Capacities View (the Capacities View, for short): Constitutively, if S believes that p , then S has the capacity to respond to evidence bearing on the belief that p by rationally updating their belief that p .⁹

In this view, evidence-responsiveness is a matter of responding to evidence in special conditions—conditions that are suitable for the exercise of evidence-responsiveness capacities. Attitudes that are not updated in the light of the evidence in such conditions do not count as beliefs.

This view covers both full belief and degrees of belief or credence.¹⁰ When one believes that p and receives relevant evidence, one might respond in many ways: by maintaining one's belief, ceasing to believe that p and coming to believe that not- p or coming to suspend on whether p , or increasing or decreasing one's degree of belief in p . Rationally responding to evidence, as I am using the term, is updating in an epistemically permissible way in response to evidence one has.¹¹ I remain neutral on both epistemic permissibility and the nature of evidence; you can fill in the details with your preferred accounts.

9. This view assumes, as standard, that belief is an attitude toward a proposition. But it can be adapted to allow belief to have non-propositional contents (Hunter 2022, Moss 2018, Zimmerman 2018) by extending a notion of relevant evidence to non-propositional contents (with accompanying standards for rationally responding to such evidence).

Further, the view is neutral on whether the correct metaphysics of mind is representationalist (Fodor 1987) or dispositionalist (Schwitzgebel 2002). If you are a dispositionalist, read this view as stating a counterfactual that subjects must satisfy to count as believing that p . If you are a representationalist, read this as constraining the kind of relation to a mental representation that belief involves.

10. As a referee pointed out, belief and credence are governed by different evidence-responsiveness norms. Specifically, coarse-grained attitudes such as beliefs can rationally persist in light of some amount of counter-evidence, unlike credences (cf. Buchak 2014, Buchak 2021). For this reason, the capacities involved in credence and belief will have different profiles. A promising avenue for distinguishing belief and credence might be in terms of the profile of the evidence-responsiveness capacities they involve. I leave this for future work.

11. Note the restriction to responding to evidence *one has*. If no evidence bears on p —as, on some views, is the case for necessary propositions—then the conditional trivially holds. Further, an agent can satisfy the condition while systematically failing to rationally respond to easily available evidence due to failing to gather such evidence. In my view, this observation motivates thinking that there are epistemic norms on evidence-gathering. See [redacted] for a defense.

3.1 Evidence-responsiveness capacities

“Capacity” is a term of art. My usage adapts the points that Schellenberg (2018) makes about mental capacities.¹² At a general level, mental capacities are mental tools that perform some function. Evidence-responsiveness capacities, in particular, are mental tools that serve to rationally respond to evidence: when successfully deployed, the result is a rational response. Capacities can be low-level and deployed automatically.

Having the capacity to Φ does not imply that one Φ s whenever one engages in the relevant activity, or whenever one tries to Φ . For example, having the capacity to run 10k in under 40 minutes does not imply that one always runs at that pace, and having the capacity to score a goal in soccer does not mean that one’s every shot results in a goal. Indeed, having a capacity does not even require one to succeed reliably, i.e. most of the time that one exercises that capacity. The runner may mostly run at a slower pace, and the soccer player may only occasionally score. Having a capacity is compatible with a track-record of little success.¹³

Having the capacity to Φ involves Φ -ing in specific conditions that suit that capacity. For example, what matters to whether one has the capacity to run a 40-minute 10k is whether one would do so when exerting serious effort, well-rested, highly motivated, and so on—even if one would fail when these conditions are not met.¹⁴ Having the capacity to Φ is a matter of satisfying counterfactuals of the form “If special conditions C were in place, then the subject would successfully Φ .”

Applying this to the capacity to rationally respond to evidence e , having such a capacity does not require always responding to e when one has e . Instead, it involves satisfying the following counterfactual: if one were to receive evidence e in some set of special conditions, one would rationally respond to e . More precisely:

Evidence-Responsiveness Capacities. A subject S has the capacity to rationally respond to evidence e from overall doxastic state D just in case S would rationally respond to e were the following conditions to hold:

- (a) S is in D , and has evidence e ,
- (b) S is minimally cognitively capable,
- (c) S ’s brain state shares all relevant aspects with S ’s brain state in the actual world, i.e. all aspects which that capacity supervenes, or is grounded, on,
- (d) no finking, mimicking, or masking occurs.

Condition (a) specifies that, to determine whether a subject has the capacity to respond to evidence e from the overall doxastic state D , we need only consider instances

-
- 12. Schellenberg (2018) offers a more detailed account. I extract only the points needed for my purposes, and remain neutral on additional features of capacities. Interested readers should refer to chapter 3 of Schellenberg 2018.
 - 13. Others (e.g. Sosa (2015)) understand reliability to allow for a low proportion of success. Nothing hangs on this terminological difference.
 - 14. At the same time, having the potential to run a 40-minute 10k if one were to train for many months does not suffice for having the capacity. If one would not do it (under any conditions) as one is right now, one might have the capacity to acquire the capacity to run a 40-minute 10k, but not the capacity to run a 40-minute 10k. See §3.2 and Schellenberg (2018) for more discussion that elucidates this distinction.

in which the subject has the relevant evidence and is in the overall doxastic state which the capacity operates on. A subject's overall doxastic state includes beliefs, suspensions, and credences.

We can think of the subject's overall doxastic state in a fragmentationist or unificationist framework. On a unificationist reading, we should count *all* the subject's doxastic attitudes as part of the overall doxastic state that these capacities operate on. In contrast, fragmentationists (Lewis 1982, Borgoni et al. 2021) about belief highlight that not all of a subject's beliefs are always "on". We do not bring all the information we have to each action. Instead, different sets of beliefs (fragments) are activated in different circumstances. On a fragmentationist view, capacities to respond to evidence are capacities to rationally integrate evidence with one's active beliefs. Though I favor fragmentationism, the Capacities View can be spelled out either way.

Condition (b) says that not responding when asleep, unconscious, or seriously cognitively impaired (e.g. highly sleep-deprived or having a panic attack) is compatible with having the capacity to rationally respond.

Condition (c) tells us that we should not consider counterfactual scenarios in which there has been interference with the capacity's neural basis in determining whether the subject has the capacity here and now. The background assumption here is that evidence-responsiveness capacities, though multiply realizable, are implemented physically. As we come to understand the human brain better, we should come to know what the implementation of these capacities is in us, and therefore be able to identify when there have been the kinds of neural changes that condition (c) excludes.

Condition (d) tells us that how subjects respond where finking, mimicking, or masking are involved is irrelevant to whether they have the capacity. Finking occurs when the conditions for *S* acquiring or losing a capacity are the very same conditions as that capacity's manifestation conditions. The classic example is Martin (1994)'s dead wire connected to a device that senses when the wire is about to be touched by a conductor and makes the wire live in every such circumstance. Mimicking occurs when there is an interfering factor—something other than the exercise of *S*'s capacity to Φ —in virtue of which *S* Φ s. The classic case is Lewis (1997)'s example of the Styrofoam glass that the Hater of Styrofoam breaks whenever it is struck because they detest the sound it makes when struck (see also Smith 1977, Prior et al. 1982). Masking occurs when *S* exercises the capacity to Φ but does not succeed at Φ -ing, due to the interference of a masking factor. For example, bubble-wrap masks glass's disposition to break when struck (Johnston 1992, Bird 1998).

Applying these points to evidence-responsiveness, condition (d) tells us that, to determine whether a subject has a specific evidence-responsiveness capacity, we can exclude the following: cases in which when they acquire or lose the capacity to rationally respond to evidence at the same time as they receive evidence; cases in which they respond to evidence in virtue of something other than the exercise of their capacities (e.g. a neurological fluke); and cases in which a masking factor interferes with the capacity's exercise.

Excluding such cases implies that this is not a *reductive* counterfactual analysis of capacities. I doubt that we can fully specify the conditions under which a subject would successfully exercise a capacity: the literature on dispositions and abili-

ties (Choi and Fara 2018) suggests that counterexamples to such analyses are forever forthcoming. Nevertheless, a counterfactual account remains epistemically useful in determining whether a subject has the target capacity. Specifically, if a subject rationally responds to evidence in many instances where conditions (a)-(c) are met, we have good reason to ascribe the capacity. And, if (a)-(c) are met but the subject does not rationally respond to evidence, then, unless we can detect a specific mask, fink, or mimicking factor, we have grounds to think that the subject lacks the corresponding capacity.

Before turning to the resources that this view of evidence-responsiveness capacities brings to the Extensional and Empirical Adequacy Challenges, I want to briefly pause and contrast my view with others in the literature.

Other ways of spelling out the Evidence-Responsiveness Thesis appeal to dispositions or tendencies (Velleman 2000, Gendler 2008) or abilities (Helton 2020). Now, dispositions and abilities both standardly receive a (non-reductive) counterfactual analysis. Could these theorists borrow the resources I develop here and appeal to the solution I offer to the Extensional and Empirical Adequacy Challenges? In particular, could they say that the account I offer above of capacities is an account of dispositions or abilities?¹⁵

In my view, proponents of the abilities view can do so naturally, but it is harder for those who appeal to dispositions or tendencies to do so. Let me take the latter case first.

Intuitively, dispositions and capacities correspond to different counterfactuals. On a natural reading, a disposition to revise requires a good track record of revision (Millikan 2000). Much as it would be unnatural to say that someone is disposed to respond stoically to criticism if they start crying almost every time they are criticized, it can sound strange to say that a belief tends to rational revision if it stays put nearly every time counter-evidence is offered.

Proponents of the view that beliefs are disposed or tend to rationally respond to evidence seem to embrace this consequence. For example, Gendler (2008) uses the claim that certain attitudes do not change in response to evidence in a wide enough range of conditions in the actual world to exclude those attitudes from the belief category. Similarly, a number of theorists have excluded delusions from counting as beliefs because they do not easily enough change in the face of counter-evidence in ordinary circumstances (cf. Bortolotti 2009)

It is fair to say, then, that one could have the capacity to rationally respond to evidence without having the disposition or tendency to do so. The version of the Evidence-Responsiveness Thesis I propose here allows for more evidence-resistance than such views.¹⁶

15. Thanks to an anonymous referee for raising these questions.

16. Proponents of the disposition view could push back in two ways. First, they could revise their view of what counterfactuals are involved in dispositions to match my account of capacities. Arguably, there would be no remaining substantive disagreement if they made this change. I still find the term ‘capacities’ to be more perspicuous because it does not suggest the requirement of a good track-record. Second, they could claim that capacities as I describe them just are special kinds of dispositions. Even if that is right, *which* dispositions one appeals to matters. The disposition to rationally respond to evidence is not the same as the disposition to rationally respond to evidence *when special conditions are met*. And there is a clear divergence between the conditions of success I specify and those that proponents of the

Grace Helton, in contrast, articulates the Evidence-Responsiveness Thesis in terms of abilities. Like “capacity”, “ability” is a term of art: Helton and I might be referring to the same underlying phenomenon (indeed, they are often used interchangeably; cf. Millikan (2000)).¹⁷ Helton specifies that a subject who has some ability “exemplifies [that ability] on some range of counterfactual circumstances in which her overall psychology is similar to its current state” (Helton 2020, 513). She clarifies that the relevant counterfactual circumstances are ones where the subject’s current cognitive mechanisms and skills are held fixed, but where the subject’s mental states might be different from their current ones. Translating this to my terms, only mental states and not cognitive mechanisms and skills can count as masks on abilities.

I am inclined to agree with this point about masks. However, this is not enough to tell whether abilities (in Helton’s sense) involve the same counterfactuals as capacities. For all that Helton says, a subject could have an ability to revise that fails when the subject receives evidence in good conditions and in the absence of masks.

My account does not allow for this: subjects must succeed in all such circumstances. For this reason, it provides a precise test for determining when such a capacity is present. This difference in specificity is a result of the difference in our underlying projects. Helton does not need to undertake such commitments on abilities to account for epistemic oughts. In contrast, we need this level of specificity to address the Extensional and Empirical Adequacy Challenges. It is open to Helton to borrow the resources I develop here to explain how her account can handle those challenges. Indeed, I take it to be a strength of my account that it can provide theoretical resources of use to other projects.

3.2 Masks on capacities

I turn now to masking cases, which are crucial to my response to the extensional and empirical adequacy challenges. In masking cases, the subject has the capacity to respond to evidence, but some factor masks its successful exercise (Johnston 1992, Bird 1998). If this factor were removed, the capacity would be successfully exercised, resulting in a rational change of mind. More precisely, we can define an instance of masking as follows. An agent’s capacity to rationally respond to evidence e from their overall doxastic state D is masked just in case (i) the agent has the capacity, (ii) conditions (a)-(c) above obtain (i.e., the agent has evidence e and is in overall doxastic state D and not cognitively incapacitated, and there has been no neural tampering) yet (iii) the agent does not rationally respond to e .¹⁸

Legitimate appeals to masks are not *ad hoc*: they are justified by inference to the best explanation. We should appeal to masks when (and only when) our best science of belief supports thinking that the overall behavior we witness is best explained by a masked capacity, and not by the subject lacking the capacity.

Consider the standard example of a mask: bubble wrap around a fragile object.

dispositionalist view assume, as made salient by the way in which they use their account to exclude certain attitudes from the belief category.

17. Why not opt for the term “ability,” then? The reason is that I draw on Schellenberg (2018)’s analysis of perceptual *capacities*, and using the term “capacities” highlights that continuity.

18. This is adapted from Fara (2008)’s account of masked abilities.

Why should we think that the bubble wrap masks the object's disposition to break when struck, instead of removing it? The answer is that we bubble-wrap glass precisely because it has that disposition (Bird 2007). Taking the glass to remain fragile when it is wrapped explains why we wrap it to *protect* it by *preventing* it from breaking—not to make the glass stronger, so that it ceases to be disposed to break.

Appeals to masks as the best explanation for some behavioral pattern are commonplace. Consider appeals to cognitive load to explain failures at tasks that require working memory (System 2 tasks (Kahneman 2011)), such as probabilistic reasoning. Given subjects' success when they are not under cognitive load, these cases are well-described as ones where subjects have the capacity to engage in probabilistic reasoning, but this capacity is masked by cognitive load. Similarly, in epistemology, appeals to competences are committed to masks. For example, in Sosa (2015)'s account, factors that put the agent in bad shape function as masks on their competences, which are retained even when the agent is in bad shape.

As these cases illustrate, despite the traditional denial that there can be intrinsic masks (Choi 2005, Cohen and Handfield 2007, Handfield and Bird 2008), factors that are internal to the agent can function as masks. Appeals to intrinsic masks are justified in the same way as appeals to extrinsic masks: because the best explanation for the presence of some intrinsic feature of an entity is that it prevents the manifestation of one of the entity's dispositions, capacities, or abilities (Ashwell 2010). Where we find a removable feature of a subject that functions in such a way, we are licensed in describing it as a mask.¹⁹

4 The Extensional and Empirical Adequacy Challenges Addressed

Having the capacity to rationally respond to evidence only requires responding when the conditions discussed in §3.1 are met. But these conditions are not always met, making room for failures to rationally respond to evidence. There might be easily available evidence in their environment that agents do not have, in which case they are not in the conditions in which the capacity to respond to that evidence would be applied. Or, assuming a fragmentationist model, they may fail to revise despite having counter-evidence because the belief is not in the active fragment when they receive that evidence. Or they might be incapacitated when they receive evidence.

Indeed, capacities to rationally respond to evidence need not be reliable in the subject's environment. They may fail most of the time they are exercised. This will be the case if one's environment (including the subject's internal states) is one where it is statistically normal for one or more of the conditions (a)-(d) discussed above to fail to hold. Specifically, such capacities may be quite systematically masked. I will now argue that this—not the absence of such capacities—best explains real-world evidence-resistance.

19. A feature that is always present in the actual world might be removable. It is beyond the scope of this paper to detail the modal force of "removable" here; all the masks to which I will appeal are clear cases of masks that can be removed.

4.1 The extensional adequacy challenge addressed

I will now argue that belief perseverance and polarization are typically explained by motivational factors masking evidence-responsiveness capacities.²⁰ This shows that evidence-resistant beliefs—i.e. beliefs that systematically display perseverance and polarization—involve evidence-responsiveness capacities. Therefore, they count as beliefs according to the Capacities View.

The central insight in the literature on cognitive dissonance is that receiving counter-evidence to one's beliefs hurts (i.e., it generates an unpleasant feeling of cognitive dissonance) and it hurts more the more central the threatened beliefs are to the subject (Festinger et al. 1956, Devine 1994, Cooper 2007, Harmon-Jones and Harmon-Jones 2007). Paradigmatic central beliefs for most subjects include those that constitute a positive and stable self-image (e.g., beliefs that one is moral, smart, or attractive) (Gilbert 2006), and beliefs that give meaning to one's life, such as those associated with group affiliations or meaningful activities (Pyszczynski et al. 2015). They may also include beliefs that are especially relevant to actions the subject needs to perform (Harmon-Jones and Harmon-Jones 2007).

The role of the unpleasant feeling of cognitive dissonance is motivational: it motivates the subject to find ways to accommodate counter-evidence while maintaining cherished beliefs. The need to reduce negative affect motivates the subject to engage in the psychological work of incorporating evidence so as to maintain those beliefs, resulting in belief perseverance and polarization. Subjects do this psychological work in two main ways: through biased assimilation and the belief polarization effect.

In biased assimilation, people receive evidence both for and against their beliefs but *strengthen* their beliefs. In the most-cited study of this phenomenon (Lord et al. 1979), subjects with strong views on the death penalty were given mixed evidence on its efficacy. Instead of becoming more uncertain of their views, subjects became more strongly convinced of whichever view they initially held.

Counter-evidence to the subject's view on the death penalty threatens their beliefs that they are moral and good at reasoning. Therefore, it threatens their positive self-image. Because subjects are motivated to avoid blows to their self-esteem, they experience high levels of cognitive dissonance in these cases. To alleviate dissonance without damaging their self-esteem, they are willing to put in significant effort. This effort takes the form of scrutinizing studies that go against their views, without doing the same with studies that support their views. Because subjects focus on refuting counter-attitudinal studies, they find many arguments against those studies and very few against studies supporting their view. This leads them to take their view to be on even more solid ground than they thought before receiving the evidence.

In the belief disconfirmation effect, subjects receive only counter-evidence to their beliefs and respond to it by becoming more convinced of those beliefs. In the classic discussion of this effect, Festinger et al. (1956) tracked members of a cult that had as its central tenet that the world would end on December 21, 1954. Members appeared

20. Other masks, such as abnormal perceptual experiences and cognitive biases, also occur. However, following the vast literature on cognitive dissonance, I take motivational factors to be the central culprit for ordinary cases of evidence-resistance, and will therefore focus on establishing that they function as masks.

to genuinely believe this: they had quit their jobs, donated their savings to the cult, and started preparing for the end of the world. When the date came and went, many became even more strongly attached to the cult and the views it prescribed. They reasoned that aliens had given planet Earth a second chance, and turned to environmentalism to prevent damage to the planet.

As we have seen, subjects are motivated to avoid accepting the conclusion that a view that matters deeply to their identity is wrong. To protect against this, they explain away discomfiting evidence—in the case of the belief disconfirmation effect, by changing other views to maintain cherished fixed points. For example, in the case above, members devised an explanation for why the cult’s prediction failed—namely, that aliens had changed their mind—which allowed them to continue adhering to the cult.²¹

I will now argue that the motivational factors that ground the belief perseverance and polarization effects (e.g., desire to defend one’s self-image) function as masks on evidence-responsiveness capacities.

First, if we remove or alleviate the weight of these motivational factors, subjects become more likely to rationally revise in light of the evidence they receive.²² This is just what we would expect if these motivational factors function as masks. More specifically, it is what we would expect if they function as masks *on evidence-responsiveness capacities*, because the responses we get when motivational factors are alleviated approach rational responses.

We can see this by considering the effects of self-affirmation and cognitive load.

Self-affirmations are reminders of the subject’s important values, skills, or past achievements, i.e. evidence for their positive self-image. Once self-affirmed, one can admit to having been wrong on cherished views while still retaining the beliefs that one is good and competent, because the balance of evidence one has still supports those beliefs. Self-affirmation hence reduces the subject’s motivation to hold on to their views, making it easier to abandon those views in response to counter-evidence (Steele 1988).

Cognitive load results from having to maintain multiple items of information in working memory (e.g. when one attempts two different problems at once). Cognitive load impedes effortful cognitive processes that require access to working memory. Interpreting evidence in biased ways, or coming up with alternative explanations for it, is effortful, and therefore hindered by cognitive load (Ditto et al. 1998, Valdesolo

21. There is another mechanism that may underwrite some instances of the belief disconfirmation effect: evidential pre-emption (Begby 2021), where the subject’s belief system leads to the prediction that they will get counter-evidence, at least from certain sources. In such cases, arguably, given background beliefs, properly exercising one’s evidence-responsiveness capacities leads to strengthening beliefs when one receives counter-evidence from those sources.

22. In biased assimilation and the belief polarization effect, subjects may end up rationally responding to *different* evidence. Kelly (2008) argues that, given the differential scrutiny in cases of biased assimilation, subjects arrive at a body of evidence that supports increased confidence in their beliefs. Once they have such evidence, they exercise their capacity to rationally respond to it. Similarly, it is widely recognized that how strong counter-evidence is depends on the range of alternative explanations for it: if there are many plausible such explanations, the evidence is weak. Perhaps, in the belief disconfirmation effect, once subjects have generated alternative explanations for the evidence, they update rationally. In other words, though motivational factors may mask some evidence-responsiveness capacities, it may be that the subject ends up exercising different ones.

and DeSteno 2008).

As predicted by the hypothesis that motivational factors function as masks, we find that biased assimilation and belief disconfirmation (and the resulting belief perseverance or polarization) are attenuated or eliminated by self-affirmation (Cohen et al. 2000, Reed and Aspinwall 1998, Sherman et al. 2000, Sherman and Cohen 2002) and cognitive load (Ditto et al. 1998, Moreno and Bodenhausen 1999). In other words, these factors are removable, and, if they are removed, subjects rationally respond to evidence (or get much closer to doing so).²³

At this point, an objector might point out that the facts I adduced are compatible with the alternative hypothesis that the subject lacks the capacity to respond to evidence, with self-affirmation and increased cognitive load causing the subject to acquire evidence-responsiveness capacities. To put it differently: perhaps what explains the data on evidence-resistance are just motivational factors, with no underlying capacities needed to explain it.

In response, I will follow the strategy for justifying masks outlined in §3.2, and argue that the best explanation for the presence and operation of these motivational factors appeals to evidence-responsiveness capacities.

The key idea is this: Why would our beliefs be carefully wrapped up in defensive motivational systems if they were not fragile to the evidence? If beliefs were entirely evidence-insensitive, then counter-evidence would be no threat, for it would not lead to one abandoning cherished beliefs. Cognitive dissonance would play no useful function in the cognitive system. The pattern of behavior we would expect to see when subjects get counter-evidence would be something more like indifference to that evidence; what we see in the case of imaginings, as I will discuss in §5.

In contrast, suppose beliefs are evidence-responsive in the sense I have detailed. Then, in the absence of motivational factors (and other masks), they will be revised in accordance with the evidence. But such revisions would, in some cases, be unsettling or painful. In those cases, it would be a good idea—from the point of view of self-esteem and effective action—for motivational factors to kick in and mask those capacities for revision. That is precisely the role of feelings of dissonance: they serve to motivate the subject to seek out alternative ways of accommodating evidence to escape discomfort.

Similarly, the best explanation for confirmation bias in evidence-gathering—for why subjects fail to gather available evidence bearing on their beliefs, especially where that evidence might support a discomfiting belief—appeals to evidence-responsiveness capacities. Here is a good reason to actively avoid gathering evidence: wanting to keep central beliefs that one would revise in the light of that evidence. In contrast, if one lacked the capacity to respond to the evidence, gathering it would make no difference to what one believes. Evidence-avoidance would be puzzling for fully evidence-insensitive attitudes.²⁴ This supports the claim that the beliefs at stake

23. Removable motivational masks are compatible with the existence of a permanent *psychological immune system* (Gilbert 2006, Mandelbaum 2019, Porot and Mandelbaum 2020, Quilty-Dunn ms), which functions to defend us against unhappiness and maintain stable motivation in the face of a hostile world. Such a system is flexible and provides motivation to hold on to different beliefs depending on the context. Specific motivational masks will therefore be removable.

24. There is an alternative explanation that is compatible with full evidence-insensitivity: perhaps sub-

are evidence-responsive.

The presence of evidence-responsiveness capacities not only explains the existence, but also the *modulation* of cognitive dissonance. It explains the fact that the stronger the counter-evidence to a cherished belief, the more subjects experience cognitive dissonance. Where counter-evidence is weak, it is easy enough for the subject to find ways of integrating it while maintaining their cherished beliefs, and therefore not much motivation is needed to do so. Weak feelings of dissonance suffice. In contrast, when the counter-evidence is strong, it takes significant cognitive maneuvering to avoid revision. Strong feelings of dissonance are needed to provide sufficient motivation. Without such strong feelings motivating difficult psychological work, the subject's evidence-responsiveness capacities would force the abandonment of cherished beliefs.

Finally, why should we think that the relevant capacities are capacities *to rationally respond to evidence*?²⁵

In response, note that, when masking factors are removed, we get rational responses to evidence (or something close to that). This is best explained if the underlying capacities are capacities to rationally respond to evidence. Further, the beliefs that form the vast background of our cognition are easily rationally revised, showing that they involve the capacity to rationally respond to evidence. If you believe that your cat is in the living room and then hear a meow from the kitchen, you rationally revise; if you believe the train leaves at 5:30 and Google Maps says it leaves at 5:40, you rationally revise; and so on. If we think these cases are of a kind, this supports thinking evidence-responsiveness capacities are at play across beliefs in general.²⁶

Let's recap. In the absence of capacities to rationally respond to evidence, what we would expect is indifference to such evidence—not effortful processes to resist such evidence, which is what we see in cases that threaten the Evidence-Responsiveness Thesis. In contrast, a combination of capacities and masks predicts effortful evidence-resistance. (Compare: if someone had a plastic glass, we would not expect them to bother to wrap it when they want to prevent it from breaking, as we would if the glass were made of glass.) Appealing to capacities and masks jointly offers a better explanation of the behavior we see in cases of evidence-resistance than appealing to only masking factors.

Further, the best interpretation of the overall system of cognitive dissonance, and its strength-of-evidence-sensitive modulation, is that it serves to mask capacities to rationally respond to evidence. It is logically possible that such a system functions to remove capacities to rationally respond to counter-evidence. Perhaps motivation to hold on to a belief removes evidence-responsiveness capacities by increasing effortful processing of counter-evidence in a way tailored to the strength of that evidence. The

jects avoid gathering evidence because they do not want to receive evidence of their own evidence-insensitivity. This explanation imputes beliefs about the degree of evidence-sensitivity of beliefs, and it does not match the phenomenology of confirmation bias, which is one of wanting to avoid being forced to accept a view.

25. Thanks to an anonymous referee for pressing me on this.

26. Additionally, if you think that rationally responding to evidence is equal to Bayesian updating (and thus, only directly applies to degrees of belief), then the success of the Bayesian program in cognitive science (Tenenbaum et al. 2011) at explaining the vast majority of cases of belief updating supports thinking that the underlying capacities are capacities to rationally respond to evidence (i.e., to do Bayesian updating).

capacity re-appears when motivation to hold on to a belief is reduced. This explanation postulates rapid appearance and disappearance of capacities in a way that is itself sensitive to the strength of the evidence. This is functionally far-fetched, and in tension with the idea that cognitive capacities are robust features of cognitive systems. It is analogous to explaining sealing fragile glasses in a safe container by claiming that this makes the glasses non-breakable, where they become breakable as soon as removed from the container.

In sum, the surface behavior and mechanisms at play in real-world evidence-resistance are best explained if evidence-responsiveness capacities are involved in evidence-resistant beliefs. This supports the hypothesis that evidence-resistance is the result of masked evidence-responsiveness capacities.

Real-world evidence-resistance, then, is compatible with the view that belief is constitutively evidence-responsive, understood along the lines of the Capacities View.²⁷

This provides new tools for investigating puzzling evidence-resistant attitudes, such as delusions, implicit biases, religious faith, conspiracy theories, and ideological commitments. The question to ask is whether it is the case that the subject changes their attitude in the specific circumstances articulated in §3. In providing a concrete test for whether these attitudes are sufficiently evidence-responsive to count as beliefs, my development of the Evidence-Responsiveness Thesis makes substantive progress.

I conjecture that it is likely that the evidence-resistance of many of these problem cases can be explained in terms of masks—motivational and otherwise—on evidence-responsiveness capacities. Indeed, I have argued elsewhere that this is true for the most intractably evidence-resistant of these cases—delusions.

4.2 The empirical adequacy challenge addressed

In §2, I highlighted two generalizations that seem especially problematic for the Evidence-Responsiveness Thesis. The first one is: “beliefs will generate a negative, motivational, phenomenologically salient discomfort whenever one encounters counterattitudinal evidence” which we will then be moved to assuage “by any easily available route” (Quilty-Dunn and Mandelbaum 2018, 2367). The second one is: if you believe extremely strongly that p and receive information against p , then you will increase your belief that p (Festinger et al. 1956).

If the discussion in §4.1 is along the right lines, we can model the generalizations that hold of human belief updating as the result of a dual-layered system of belief revision, with evidence-responsiveness capacities as one layer and motivational masks as another. When you believe extremely strongly that p , such masks are very likely to be present. When they are present, you will either maintain or increase your belief that p , depending on how strongly motivated you are to believe that p . For this reason,

27. Cognitive dissonance has also been used to defend the claim that our beliefs are “minimally rational” not in the sense of evidence-responsive but in the sense that “they respond to perceived irrationality by re-establishing coherence” (Ganapini 2020, 10). The two views are different in that re-establishing coherence does not entail rationally responding to evidence; and that Ganapini’s view requires *actually re-establishing* coherence, not just having the capacity to do so.

the second generalization above holds in this model, and it is therefore compatible with the presence of evidence-responsiveness capacities.

In agreement with the first generalization, the motivational layer involved in belief revision is implemented through the generation of discomfort in the face of counter-evidence. And we are indeed moved to assuage this discomfort “by any easily available route.” In an optimistic note, my view indicates that the most easily available routes will often be rationally responding to evidence, given the psychological work it takes to respond in other ways.

This dual-layered model of belief revision has two noteworthy benefits. First, it unifies beliefs that seem to be revised in very different ways: primarily in response to evidence and in ways that seem motivationally and affectively driven. From the point of view of studying the mechanics of belief revision, we do not need to split up the category of belief into different sub-kinds, as [Leeuwen and Lombrozo \(2023\)](#) suggest. Second, equipped with this dual-layer model of belief revision, we can in fact turn concerns about empirical adequacy on their head. Generalizations about belief revision—in particular, about the role and modulation of cognitive dissonance—are not just compatible with the presence of evidence-responsiveness capacities. They are, if the argument in §4.1 succeeds, *best explained* by the claim that evidence-responsiveness capacities are present.

Specifically, the generalizations about belief revision that contemporary psychofunctionalists endorse ([Quilty-Dunn and Mandelbaum 2018](#), [Porot and Mandelbaum 2020](#)) commit them to evidence-responsiveness capacities underlying belief. Given that psychofunctionalists claim that belief is a functional kind definitionally tied to our best psychological generalizations about belief ([Block and Fodor 1972](#), [Block 1978](#), [Quilty-Dunn and Mandelbaum 2018](#)), this implies that they ought to embrace the Evidence-Responsiveness Thesis.²⁸ Psychofunctionalists who think that the laws of dissonance are constitutive of belief should also think that evidence-responsiveness capacities are constitutive of belief. On such a view, belief is constitutively both evidence-responsive and evidence-resistant.²⁹

This is significant. Traditionally, as described in §2, the motivation for claiming that belief is constitutively evidence-responsive has come from epistemology. Cognitive science is taken to present a serious problem for that view. Against this, I have here shown that taking cognitive science as our primary guide to the nature of belief supports thinking that belief is constitutively evidence-responsive.

28. This is conditional on their acceptance of the generalizations I mentioned. This means that I am not here doing justice to different versions of psychofunctionalism. Nonetheless, the point is significant, given that this is currently the most developed psychofunctionalist proposal.

29. A referee noted that the empirical data might be compatible with belief having an epistemic function other than rationally responding to evidence (perhaps maintaining coherence). Indeed, the empirical data on belief arguably allows for running this style of argument for other epistemic roles for belief. If theorists identify some epistemic role of belief *X* such that (a) most beliefs straightforwardly fulfill it, and (b) when we unmask in the cases I have described, unmasked beliefs fulfill that function, then the empirical data will also support the view that belief is constitutively *X*. This is not a problem for my view: belief could have multiple constitutive epistemic roles, or, alternatively, we might need to appeal to non-empirical considerations (e.g., to considerations from epistemology, as discussed in §2) to decide which are central to belief.

5 Capturing the Epistemic Role of Belief

I will now argue that the Capacities View does justice to the epistemic role of belief. Specifically, I will show that it satisfies the two desiderata mentioned in §2: it distinguishes beliefs from attitudes that are not epistemically assessable for their responses to evidence, such as imaginings, and it captures the claim that belief aims at truth. This shows that the sense of “evidence-responsiveness” detailed above is strong enough to account for the epistemic role of belief—what initially motivated theorists to embrace the Evidence-Responsiveness Thesis.

My view distinguishes belief from on-doxastic cognitive attitudes—attitudes that present their content as true but which are not beliefs, and which are not straightforwardly assessable on purely epistemic grounds. Such attitudes are not evidence-responsive in the capacities sense. To show this, I will focus on imaginings.³⁰

Consider imagining that you are on a luxury tropical vacation while receiving decisive counter-evidence: your senses present you with evidence that you are grading in your apartment while it rains outside. You believe that you are grading in your apartment and not on a luxury tropical vacation. This reflects the fact that you successfully exercise your capacities to rationally respond to evidence. But the successful exercise of these capacities, made manifest in your beliefs, leaves the imagining intact.

This illustrates that it is possible (in fact, ordinary) to fail to revise one’s imaginings in response to counter-evidence in good conditions for the exercise of one’s evidence-responsiveness capacities—indeed, in conditions where you successfully exercise those capacities by revising your beliefs. Imaginings, then, are not constitutively evidence-responsive.

Indeed, *not* being evidence-responsive may be constitutive of imagining. In diametric opposition to beliefs, imagination plays the cognitive role of helping us get away from reality and explore alternative ways things could be. Imagining is standardly taken to be subject to a decoupling mechanism that insulates them from counter-evidence (Leslie 1987, Perner 1991). If that is right, then there are irremovable barriers to evidence-responsiveness capacities operating on imagining, “irremovable” in the sense that, if removed, the attitude would cease to be an imagining.

Though my view correctly excludes imaginings and other non-doxastic attitudes, there may be constitutively evidence-responsive attitudes other than belief, such as suspension. This is an advantage of the view. To the extent that we think suspension, like belief, is constitutively subject to epistemic standards (Miracchi 2019), it is natural to think that it is also constitutively evidence-responsive. This raises the intriguing possibility of delimiting the class of doxastic attitudes, including but not limited to

30. The points generalize straightforwardly to non-doxastic acceptances.

What about non-cognitive attitudes, i.e. attitudes which do not present their content as true (such as desires)? Desires are clearly not evidence-responsive in the same sense as beliefs, i.e., in the sense of constitutively requiring the capacity to revise in light of evidence that what one desires does not hold. Indeed, we typically desire things that do not hold, in full knowledge that they do not hold. A more fitting notion of evidence-responsiveness for desires would appeal to sensitivity to evidence on how likely those desires are to be achieved (as opposed to evidence that their content is true already). I am neutral on whether desires are constitutively evidence-responsive in this sense. See Smith (2003) for a view on which they are.

belief, in terms of the involvement of evidence-responsiveness capacities. And it supports the claim that the necessary condition I place on believing tracks epistemically assessability.

Another reason to think that the Capacities View captures the epistemic role of belief is that it illuminates the claim that belief aims at truth.³¹ A popular way of cashing out the “belief aims at truth” metaphor appeals to the claim that beliefs are regulated by evidence-responsive systems. In David Velleman’s words, belief aims at truth in that it is necessarily in the province of systems which “regulate cognitions in ways designed to ensure that they are true, by forming, revising, and extinguishing them in response to evidence and argument” (Velleman 2000, 253).

My discussion vindicates this view. I have offered an account of how, even in cases where how we respond to evidence is decisively influenced by motivational factors, evidence-responsiveness capacities remain involved. My account makes it clear why we should think that beliefs are regulated as Velleman claims even when these capacities are never, or virtually never, successfully employed.

More strongly, my discussion leaves room for arguing that belief itself aims exclusively at truth. Regardless of how human belief is regulated, it is plausible that there could be epistemic agents whose beliefs are fully insulated from motivational factors, and who always respond to evidence in epistemically permissible ways. Theists presumably hold that God is such a believer.

If that is right, *pace* psychofunctionalism, the connections we find in humans between motivation and belief change are not constitutive features of belief, but contingent features of human cognitive systems. On such a view, there is an asymmetry between the belief–epistemic aims connection on the one hand, and the belief–motivation link on the other. Belief itself aims exclusively at truth but is, in our case, embedded in cognitive systems that have other aims, such as maintaining motivation.

This is the view I favor. But the Capacities View does not necessitate it. Given the discussion of psychofunctionalism in §4.2, the View is compatible with taking either cognitive science or epistemology as our central guides to the nature of belief. Epistemology and cognitive science can operate with a unified notion of belief.

6 Conclusion

This paper addressed the question: In the light of evidence-resistant beliefs, what is the connection between belief and evidence-responsiveness? I have argued that, counter-intuitively, the cognitive science of belief supports the claim that capacities to respond to evidence are involved in belief. This supports an account of belief—the Evidence-Responsiveness Capacities View—according to which belief is constitutively underwritten by capacities to respond to evidence. This account provides a clear sense in which belief, unlike imaginings and acceptances, aims at truth, and provides resources for analyzing real-world evidence-resistant beliefs. More importantly, it allows us to acknowledge both the epistemic role of belief and the ways in which beliefs often fail to live up to our epistemic aspirations.

31. Or other epistemic goods; my discussion applies to any view that (a) takes belief to have an epistemic aim and (b) spells that out in terms of regulation by evidence-responsive systems.

Works Cited

- Craig A Anderson, Mark R Lepper, and Lee Ross. Perseverance of social theories: the role of explanation in the persistence of discredited information. *Journal of personality and social psychology*, 39(6):1037, 1980.
- Lauren Ashwell. Superficial dispositionalism. *Australasian Journal of Philosophy*, 88(4):635–653, 2010. doi: [10.1080/00048400903352892](https://doi.org/10.1080/00048400903352892).
- Endre Begby. Evidential preemption. *Philosophy and Phenomenological Research*, 2021. doi: [10.1111/phpr.12654](https://doi.org/10.1111/phpr.12654).
- Alexander Bird. Dispositions and antidotes. *The Philosophical Quarterly*, 48(191):227–234, 1998.
- Alexander Bird. *Nature's metaphysics: Laws and properties*. Oxford University Press on Demand, 2007.
- Ned Block. Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9:261–325, 1978.
- Ned Block and Jerry A. Fodor. What psychological states are not. *Philosophical Review*, 81(April):159–81, 1972. doi: [10.2307/2183991](https://doi.org/10.2307/2183991).
- Cristina Borgoni, Dirk Kindermann, and Andrea Onofri. *The Fragmented Mind*. Oxford University Press, 2021.
- Lisa Bortolotti. *Delusions and Other Irrational Beliefs*. Oxford University Press, 2009.
- Lisa Bortolotti and Kengo Miyazono. Recent work on the nature and development of delusions. *Philosophy Compass*, 10(9):636–645, 2015. doi: [10.1111/phc3.12249](https://doi.org/10.1111/phc3.12249).
- Lara Buchak. Belief, credence, and norms. *Philosophical Studies*, 169(2):1–27, 2014. doi: [10.1007/s11098-013-0182-y](https://doi.org/10.1007/s11098-013-0182-y).
- Lara Buchak. A faithful response to disagreement. *The Philosophical Review*, 130(2):191–226, 2021. doi: [10.1215/00318108-8809893](https://doi.org/10.1215/00318108-8809893).
- Clark A Chinn and William F Brewer. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research*, 63(1):1–49, 1993.
- Sungcho Choi. Do categorical ascriptions entail counterfactual conditionals? *The Philosophical Quarterly*, 55(220):495–503, 2005.
- Sungcho Choi and Michael Fara. Dispositions. *The Stanford Encyclopedia of Philosophy*, 2018.
- Paul M Churchland. Eliminative materialism and propositional attitudes. *the Journal of Philosophy*, 78(2):67–90, 1981.

- Daniel Cohen and Toby Handfield. Finking Frankfurt. *Philosophical Studies*, 135(3): 363–374, 2007.
- Geoffrey L Cohen, Joshua Aronson, and Claude M Steele. When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and social psychology bulletin*, 26(9):1151–1164, 2000.
- Joel Cooper. *Cognitive dissonance: 50 years of a classic theory*. Sage, 2007.
- Gregory Currie and Ian Ravenscroft. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press, 2002.
- Donald Davidson. Incoherence and irrationality. *Dialectica*, 39(4):345–54, 1985. doi: [10.1111/j.1746-8361.1985.tb01603.x](https://doi.org/10.1111/j.1746-8361.1985.tb01603.x).
- Donald Davidson. *Essays on Actions and Events: Philosophical Essays Volume 1*. Clarendon Press, 2001.
- Daniel C. Dennett. True believers: The intentional strategy and why it works. In A. F. Heath, editor, *Scientific Explanation: Papers Based on Herbert Spencer Lectures Given in the University of Oxford*, pages 150–167. Clarendon Press, 1981.
- A.J. Elliot & P.G. Devine. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67(3):384–394, 1994. doi: [10.1037/0022-3514.67.3.382](https://doi.org/10.1037/0022-3514.67.3.382).
- Peter H Ditto, James A Scepansky, Geoffrey D Munro, Anne Marie Apanovitch, and Lisa K Lockhart. Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 75(1):53, 1998.
- Kevin Dorst. Rational polarization. *The Philosophical Review*, forthcoming.
- Andy Egan. Imagination, delusion, and self-deception. In Tim Bayne and Jordi Fernandez, editors, *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation (Macquarie Monographs in Cognitive Science)*. Psychology Press, 2008.
- Michael Fara. Masked abilities and compatibilism. *Mind*, 117(468):843–865, 2008.
- L. Festinger, H. W. Riecken, and S. Schachter. *When prophecy fails*. University of Minnesota Press, 1956. doi: [10.1037/10030-000](https://doi.org/10.1037/10030-000).
- Jerry A Fodor. *Psychosemantics: The problem of meaning in the philosophy of mind*, volume 2. MIT press, 1987.
- Marianna Bergamaschi Ganapini. Belief’s minimal rationality. *Philosophical Studies*, pages 1–20, 2020. doi: [10.1007/s11098-019-01369-y](https://doi.org/10.1007/s11098-019-01369-y).
- Tamar Szabó Gendler. Alief in action (and reaction). *Mind & Language*, 23(5):552–585, 2008.

- Tamar Szabó Gendler. Intuition, imagination, and philosophical methodology: A summary. *Analysis*, 72:759–764, 2012. doi: <https://doi.org/10.1093/analys/anso87>.
- Daniel Gilbert. *Stumbling on happiness*. Knopf, 2006.
- Toby Handfield and Alexander Bird. Dispositions, rules, and finks. *Philosophical Studies*, 140(2):285–298, 2008.
- Eddie Harmon-Jones and Cindy Harmon-Jones. Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1):7–16, 2007.
- Grace Helton. If you can't change what you believe, you don't believe it. *Noûs*, 2020. doi: [10.1111/nous.12265](https://doi.org/10.1111/nous.12265).
- David Hunter. *On Believing: being right in a world of possibilities*. Oxford University Press, 2022.
- Mark Johnston. How to speak of the colors. *Philosophical studies*, 68(3):221–263, 1992.
- Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- Thomas Kelly. Disagreement, dogmatism, and belief polarization. *Journal of Philosophy*, 105(10):611–633, 2008. doi: [10.5840/jphil20081051024](https://doi.org/10.5840/jphil20081051024).
- Neil Van Leeuwen. Religious credence is not factual belief. *Cognition*, 133(3):698–715, 2014. doi: [10.1016/j.cognition.2014.08.015](https://doi.org/10.1016/j.cognition.2014.08.015).
- Neil Van Leeuwen. Two concepts of belief strength: Epistemic confidence and identity centrality. *Frontiers in Psychology*, 13:1–4, 2022.
- Neil Van Leeuwen and Tania Lombrozo. The puzzle of belief. *Cognitive Science*, 47(2): e13245, 2023. doi: [10.1111/cogs.13245](https://doi.org/10.1111/cogs.13245).
- Alan M Leslie. Pretense and representation: The origins of “theory of mind.”. *Psychological review*, 94(4):412, 1987.
- Neil Levy. Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Noûs*, 49(4):800–823, 2015.
- Neil Levy. *Bad Beliefs: Why They Happen to Good People*. Oxford University Press, 2021.
- David Lewis. Finkish dispositions. *The Philosophical Quarterly*, 47(187):143–158, 1997.
- David K. Lewis. Logic for equivocators. *Noûs*, 16(3):431–441, 1982. doi: [10.2307/2216219](https://doi.org/10.2307/2216219).
- Akiva Liberman and Shelly Chaiken. Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin*, 18(6):669–679, 1992.
- C.S. Taber & M. Lodge. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50:755–769, 2006. doi: [10.1111/j.1540-5907.2006.00214.x](https://doi.org/10.1111/j.1540-5907.2006.00214.x).

- Charles Lord, Lee Ross, and Mark Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37:2098–2109, 11 1979. doi: [10.1037/0022-3514.37.11.2098](https://doi.org/10.1037/0022-3514.37.11.2098).
- Eric Mandelbaum. Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3):629–658, 2016.
- Eric Mandelbaum. Troubles with Bayesianism: An introduction to the psychological immune system. *Mind and Language*, 34(2):141–157, 2019. doi: [10.1111/mila.12205](https://doi.org/10.1111/mila.12205).
- Gregory B Markus. Stability and change in political attitudes: Observed, recalled, and “explained”. *Political behavior*, 8(1):21–44, 1986.
- Charles B Martin. Dispositions and conditionals. *The Philosophical Quarterly (1950-)*, 44(174):1–8, 1994.
- John McDowell. Having the world in view: Sellars, Kant, and intentionality. *Journal of Philosophy*, 95(9):431–492, 1998.
- John W. McHoskey. Case closed? on the John F. Kennedy assassination: Biased assimilation of evidence and attitude polarization. *Basic & Applied Social Psychology*, 17(3):395–409, 1995.
- Conor McHugh and Daniel Whiting. The normativity of belief. *Analysis*, 74(4):698–713, 2014.
- Ruth Garrett Millikan. *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press, 2000.
- Lisa Miracchi. When evidence isn’t enough: Suspension, evidentialism, and knowledge-first virtue epistemology. *Episteme*, 16(4):413–437, 2019.
- Kristen N. Moreno and Galen V. Bodenhausen. Resisting stereotype change: The role of motivation and attentional capacity in defending social beliefs. *Group Processes & Intergroup Relations*, 2(1):5–16, 1999. doi: [10.1177/1368430299021001](https://doi.org/10.1177/1368430299021001).
- Sarah Moss. *Probabilistic knowledge*. Oxford University Press, 2018.
- Josef Perner. *Understanding the representational mind*. The MIT Press, 1991.
- Nicolas Porot and Eric Mandelbaum. The science of belief: A progress report. *Wiley Interdisciplinary Reviews: Cognitive Science*, page e1539, 2020.
- Nina Poth and Krzysztof Dolega. Bayesian belief protection: A study of belief in conspiracy theories. *Philosophical Psychology*, pages 1–22, 2022.
- Elizabeth W Prior, Robert Pargetter, and Frank Jackson. Three theses about dispositions. *American Philosophical Quarterly*, 19(3):251–257, 1982.

- Tom Pyszczynski, Jeff Greenberg, and Kathleen Holt. Maintaining consistency between self-serving beliefs and available data: A bias in information evaluation. *Personality and Social Psychology Bulletin*, 11(2):179–190, 1985.
- Tom Pyszczynski, Sheldon Solomon, and Jeff Greenberg. Thirty years of terror management theory: From genesis to revelation. In *Advances in experimental social psychology*, volume 52, pages 1–70. Elsevier, 2015.
- Jake Quilty-Dunn. Unconscious rationalization, or: How not to think about awfulness and death. manuscript, ms.
- Jake Quilty-Dunn and Eric Mandelbaum. Against dispositionalism: belief in cognitive science. *Philosophical Studies*, 175(9):2353–2372, 2018.
- Mark B Reed and Lisa G Aspinwall. Self-affirmation reduces biased processing of health-risk information. *Motivation and emotion*, 22(2):99–132, 1998.
- Susanna Schellenberg. *The unity of perception: Content, consciousness, evidence*. Oxford University Press, 2018.
- Eric Schwitzgebel. A phenomenal, dispositional account of belief. *Noûs*, 36(2):249–275, 2002. doi: [10.1111/1468-0068.00370](https://doi.org/10.1111/1468-0068.00370).
- Nishi Shah. How truth governs belief. *The Philosophical Review*, 112(4):447–482, 2003.
- Nishi Shah and J David Velleman. Doxastic deliberation. *The Philosophical Review*, 114(4):497–534, 2005.
- David AK Sherman, Leif D Nelson, and Claude M Steele. Do messages about health risks threaten the self? *Personality and Social Psychology Bulletin*, 26(9):1046–1058, 2000.
- David K Sherman and Geoffrey L Cohen. Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current directions in psychological science*, 11(4):119–123, 2002.
- Morgan P Slusher and Craig A Anderson. Belief perseverance and self-defeating behavior. In *Self-defeating Behaviors*, pages 11–40. Springer, 1989.
- A David Smith. Dispositional properties. *Mind*, 86(343):439–445, 1977.
- Michael Smith. Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In Sarah Stroud and Christine Tappolet, editors, *Weakness of will and practical irrationality*, pages 17–38. Oxford: Clarendon, 2003.
- Declan Smithies, Preston Lennon, and Richard Samuels. Delusions and madmen: Against rationality constraints on belief. *Synthese*, 200(3):1–30, 2022. doi: [10.1007/s11229-022-03632-y](https://doi.org/10.1007/s11229-022-03632-y).
- Ernest Sosa. *Judgment & Agency*. Oxford University Press UK, 2015.

- Jason Stanley. *How propaganda works*. Princeton University Press, 2015.
- Claude M Steele. The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology*, 21(2):261–302, 1988.
- Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.
- Piercarlo Valdesolo and David DeSteno. The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5):1334–1338, 2008.
- David Velleman. On the aim of belief. In David Velleman, editor, *The Possibility of Practical Reason*, pages 244–81. Oxford University Press, 2000.
- Aaron Z. Zimmerman. *Belief: A Pragmatic Picture*. Oxford University Press, 2018.