Belief: Dumb, Cold, & Cynical

Nicolas Porot (Africa Institute for Research in Economics and Social Science, UM6P) & Eric Mandelbaum (Baruch College & The Graduate Center, CUNY)

10/13/23

(Accepted, but awaiting copy editing so please wait to cite, especially if the quote would need a [sic])

Forthcoming in *What is Belief* (eds. Jonathan Jong & Eric Schwitzgebel)

People do not believe just anything. Few of us believe we really can get rich with just two hours of work a week from home by placing tiny classified ads in the newspaper. People can also put their beliefs to rational ends: if you really want more birds to come to your feeder, sooner or later you will realize that you should add some seed to it. Yet often enough, people do believe some very dumb things, such as if the United States builds a wall on their southern border, Mexico will pay for it.

These features of belief are not inconsistent with one another, but their mutual presence calls for explanation. What is it that makes us shrewd in some situations, credulous in others, and seemingly impeccably rational in yet others? What is it about the cognitive processes underlying belief that makes it seem so multifaceted in the first place? One kind of reasonable response here, for example, would be to suggest that human competences are Bayes-approximative, and that appearances to the contrary are performance errors (Lieder & Griffith 2020). Another would be to show that the appearance of multifacetedness is illusory: to write off the evidence for some of these features in favor of others. One might be suspicious of the case for human credulity, and thus favor cognitive shrewdness as a defining feature of human cognition (Mercier 2020); or one might recast evidence for motivated reasoning in a more rationality-friendly light (Kelly 2008).

The aim of this paper is to take a different strategy for explaining how the mind has these seemingly incongruous features: to present distinct mechanistic explanations of each of them. We propose that the cognitive architecture of belief includes processes that are (respectively) discerning, credulous, rational, and irrational. The fact that belief varies with regard to these features turns out to be unsurprising: its constituent mechanisms are each at different ends of those poles. The evidence we provide for these mechanisms will serve a dual function. It will both support our positive view, and build our case against a handful of views that downplay human credulity and irrationality.

We group together these individual mechanisms to argue that belief has three archetypal features: *dumbness, coldness,* and *cynicism.* Belief is dumb because the way we form beliefs is a brute, reflexive-like process; and because we sometimes resist changing beliefs that are at the core of how we see ourselves, even in the face of good evidence that those beliefs are false. It is cold because we seem to be pretty good at updating (at least some of) our beliefs in accordance with the evidence we have, as normatively focused theories of belief prescribe. And it is cynical because despite our remarkable gullibility, we can still reject information that is implausible or of dubious provenance.

In section 1, we describe a view of human cognition that explains how belief seems to be all of these different things. Then, in sections 2 and 3, we focus on alternative views of belief that have aimed to reduce belief to its cynical mechanisms, namely epistemic vigilance (Sperber et al. 2010; Mercier 2020), marketplaces of rationalization (Williams 2022), and beliefs as signals (Funkhouser 2017, 2022). While a constituent mechanism of belief really is cynical, these views neglect other mechanisms of belief that are not. The resulting view incorporates the observation that human cognition is cynical into a larger (less flattering) set of features and raises explanatory challenges for these other views. We end by offering ways forward to unite the differing camps of epistemic theorizing.

1. An Architecture of Belief

We take as a starting point a psychofunctionalist account of belief. Belief is characterized by its actual causal role in cognition: its interactions with perceptual and other cognitive processes, its causal role in behavior, and so forth. This causal role is typically delineated by belief’s place in psychological laws, the laws describing belief’s behavior. Belief’s reality is thus an empirical question. If it should turn out that a posit or posits of cognitive science (typically: mental representations) resemble the folk psychological kind to a sufficient degree, then the folk psychological kind can be identified with the scientific one, cementing belief’s status in our scientific understanding of the world. The match need not be perfect: We might turn out to have been wrong about some features of belief, just as the discovery of gold’s atomic structure supposedly allowed us to tell gold from pyrite. But if it should turn out that no such posit makes the cut, then nothing can be identified with the folk kind, which (despite its then inexplicable usefulness in our day-to-day lives) goes the way of phlogiston, scientifically speaking.

The rest of this section describes evidence for a cognitive architecture of belief that captures much (though not all) of the folk psychological role of belief.

1.1 Dumb Belief Formation

Forty-five years of research on the illusory truth effect (Hasher et al. 1977) have shown that people believe claims simply through repeated exposure to them. Repetition can cause participants to assent to conspiracy theories (Béna et al. 2023), to fake news (Pennycook et al. 2018), to claims they know to be false (Fazio et al. 2015) or unlikely (Fazio et al. 2019); it can even increase truth ratings for sentences as plainly implausible as “The Earth is a perfect square” (Lacassagne et al. 2022). The simplest explanation of this effect is that mere exposure to information causes belief (cf. Bacon 1979, Unkelbach 2006, Mattavelli et al. 2023).

In the same spirit, merely being under cognitive load can make one believe things, meaning that minor shifts in our environment can make us form beliefs that we otherwise wouldn’t. When under cognitive load, people misremember false statements as true far more often than they misremember true ones as false (Gilbert 1990). And this misremembering carries over into processes elsewhere in the mind, for example in domain-general reasoning processes. When participants in Gilbert et al. (1993) were under load, false-labeled exculpatory details in a fictional police report decreased the prison sentences subjects recommended for perpetrators; incriminating ones increased sentences. Information accepted under load isn’t just taken in--it acts as beliefs, serving as premises in inferences.

This kind of evidence has led some (e.g., Mandelbaum 2014, Quilty-Dunn & Mandelbaum 2015, Porot & Mandelbaum 2020/2022, Goffin 2023) to conclude that belief formation is not *Cartesian but* rather is *Spinozan*. On Cartesian models, believing is as effortful and deliberative as rejecting, and one can process information without immediately believing it. On Spinozan models, believing is an automatic and effortless process, and it is prior to rejection, which is effortful. Belief is Spinozan: one can believe any (truth-apt) information one processes, and rejecting that information often takes serious cognitive effort. People will often believe propositions without scrutinizing their prior plausibility, checking the reliability of their sources, or even seeing if the new information coheres with what they antecedently believed.

Taken on its own, Spinozan belief formation appears quite dumb: a kind of brute, mechanistic route to belief, humming along with no rationality checks at all. Under optimal circumstances, it is complemented by a rejection mechanism[[1]](#footnote-1) that scrutinizes incoming information for plausibility, source reliability, consistency with background beliefs, and the like. From an evolutionary perspective, this dual-system is sensible: if initial belief forming processes were wed to perception, taking the deliverances of perceptual processes at face value would have caused little harm, as our perceptual processes are overwhelmingly veridical. As we get to communal belief formation, the situation becomes a bit trickier, but still most testimony seems boringly veridical. It is only much later in the history of the mind that telling truths from falsehoods is likely to have been a problem. To address it, human minds (or some recent ancestors’) developed a secondary effortful rejection mechanism. This minimizes effort in our cognitive economy as its use is optional. However, one problematic quirk of this otherwise sensible system is that activation of rejection processes may be concerningly rare.

1.2 Dumb Belief Updating

Behavioral economists and psychologists of reasoning spend much of their time exploring the pervasiveness of another kind of dumbness, specifically in the way we update beliefs. This includes canonical biases familiar from self-help books and executive trainings. And there may be no more rigorous lack of adherence to the norms of rational belief revision to be found than in belief disconfirmation cases of cognitive dissonance (Festinger 1956; Loughnan et al. 2010). When we encounter information that threatens a belief that is core to our self-image, we tend to reject it in the service of protecting the core belief (Mandelbaum 2019a; Quilty-Dunn forthcoming), since the alternative (accepting the information) would be too psychologically bruising (Mandelbaum & Porot 2022).

By rejecting counterattitudinal information, a *Psychological Immune System* (Gilbert 1998, Mandelbaum 2019a) protects us from the havoc that accepting such information would wreak on our sense of self. Suppose you are an ardent supporter of Jordan Peterson, and sharing his worldview is part of your own sense of who you are. If it is pointed out to you that, based on his own scholarly background, Peterson is not likely to be a competent evolutionary psychologist, biologist, or nutritionist, and that qualified scholars in these areas do not really find his work to be serious, the rational response might be to temper one’s belief in Peterson’s views. But as anyone who has spent time on internet comment threads knows, this is rarely what happens. This is a toy example, but evidence for resistance to evidence of this kind is widespread.[[2]](#footnote-2)

A larger-scale example can be seen by looking at partisanship during election cycles. Information about candidates is most available during elections, and in particular in the short run-up to election day. In an environment with more information in favor of, and against, candidates, partisanship ought to decrease in most cases, as the electorate becomes more informed. Precisely the opposite happens: in a study looking at 86 countries, Michelitch and Utych (2018) find that partisanship is highest just before an election, on average by 6 percentage points compared to down periods. While surely more than one factor contributes to this global pattern, it is a predictable consequence of possessing a psychological immune system that safeguards political views: increased information on both sides activates a system that rejects most counterattitudinal information, but allows pro-attitudinal information. Credence in one’s antecedently held political beliefs consequently increases.

There appears to be at least one mechanism for belief updating that is dumb.[[3]](#footnote-3)

1.3. Belief is Cold

At the same time, for beliefs that do not pose a threat to our sense of self, we have little trouble updating them in the light of new evidence. We easily jettison the belief that the Amazon is the world’s longest river when confronted with evidence that the Nile is longer. Some other mechanism must be responsible for this colder, less self-protection-driven kind of belief updating. The laws of belief updating appear to be sensitive to content; while the Psychological Immune System kicks in for evidence that assails one’s core beliefs, the default updating process—the one that applies to all one’s other beliefs—appears simply to dole out credence as a function of one’s evidence. The co-existence of these very different mechanisms allows us to be swayed by our deep attachments in some situations, and to reason like good Bayesians in others. Belief updating may be dumb, but it is also cold and often accurate.

Cold belief updating covers a great many, perhaps most, of the things we believe. The coldness of belief responds to core features of the folk psychological picture of belief: it allows for rational--perhaps (near) optimal (Lieder & Griffith 2020, cf. Mandelbaum et al. 2020)—adjustment of belief in the light of evidence. By isolating the mechanism responsible for this kind of updating from the psychological immune system, we preserve the assumption that belief can be identified with a particular representational kind, postulating merely that different operations can be performed over that kind, depending on the context and on the content of those states.

 1.4. Belief is Cynical

Despite the automaticity of human believing and despite the sometimes-wishful manner in which we update beliefs, some process allows us to recognize implausible claims, identify cheats and uncover lies—at least intermittently (Cosmides et al. 2010).[[4]](#footnote-4) On the Spinozan view of belief formation, adding a dose of cynicism to belief formation is precisely the role of the belief rejection system. The rejection system is capacity limited--it cannot be engaged when we are under cognitive load, so it is an imperfect remedy to our dumbness. What’s particularly cynical about the belief rejection system is its attention-demanding nature: since rejection is effortful, and since our default state is one of cognitive laziness, we are only apt to be primed for skepticism in places where we expect to receive counterattitudinal information. That is, we are most apt to reject information when we think what we believe may be under attack. This type of system will fall prey to tribalism because, as a default, we apply our powers of rejection only to situations where we expect to be wrong. We are thus inclined towards being argumentative and contrarian, but only in situations where we perceive our beliefs to be under attack. The defensiveness scales with the perception of the attack. If we think our fringe beliefs—ones we don’t much care about (e.g., the authors’ beliefs about the relative populations of Scandinavian countries)—are under attack we may not be reflexively defensive. However, if cherished beliefs may be under attack (say, you’re a theist and find yourself at a dinner party with Dan Dennett and Richard Dawkins), then you will feel like you yourself are under attack, and take a much more cynical view of any incoming information. In that circumstance our rejection system will be primed to screen incoming information for claims that clash with our background beliefs, will assesses the quality of the informational source, and will check whether our interlocutor might have a reason to deceive us.

2. Or Perhaps the Mind is Just Cynical and Cold?

So far, we have characterized belief through a handful of cognitive mechanisms we think characterize the psychofunctional role of a single kind. On this view, the cynicism of belief is just one part of a suite of mechanisms of belief.

2.1 Vigilance, Marketplaces, and Signaling

By contrast with this approach, a cluster of views places cynicism at belief’s core. According to *epistemic vigilance* (Sperber & Mercier 2010, Mascaro & Morin 2014, McKay & Mercier 2023; also called *open vigilance*, Mercier 2020), humans evolved to be shrewd interlocutors, maximizing information extraction from communication, which demands a robust rejection mechanism. Interlocutors can be unknowledgeable, uncooperative, or unreliable; they can cheat and lie. Thus, the vigilance theorists’ reason, our minds must have been fine-tuned to detect incompetence, uncooperativeness, unreliability, and deceit in speakers, and to check incoming information efficiently against our background beliefs. In fact, all this supposed to be required for communication to evolve in the first place (Mercier 2020: 31-2).[[5]](#footnote-5) In support of this evolutionary thesis, they contend, when one looks at the places where one might expect to find human gullibility on full display—propaganda, brainwashing, advertising, fake news, or rumors—evidence for it is weak (Mercier 2020).

Social marketplace theorists place cynicism at the core of belief acquisition (Williams 2022). Like vigilance theorists, they think cynicism emerged through the evolution of communication, and that most of us have beliefs we *want* to hold, if only we could find a way to rationalize them.[[6]](#footnote-6) So if you are a Democrat, you may have wanted to believe that Putin installed Trump. When many people share a desire for rationalizations for the same belief, as with large ingroups such as the Democratic Party, some begin investing energy into producing such rationalizations for wider consumption, giving rise to a kind of marketplace. Demand in this marketplace encourages the production of rationalizations for public consumption, e.g., offering believable theories of Trump-Kremlin connections on social media and cable news networks. This view attempts to explain away evidence of credulity or irrationality as evidencein favor of a different kind of rational calculus–desire satisfaction–by which belief formation occurs at the expense of truth. One trades truth for other social goods, such as status.

Vigilance theorists, marketplace theorists, and others (e.g., Funkhouser 2017, 2022; Mercier 2020: p. 191) argue that belief has a signaling (Lewis 1979, Skyrms 2010) function. Funkhouser (2022) for example maintains that within social groups, some beliefs emerge because they are typical of members of the ingroup and holding them confers the benefits of group membership.[[7]](#footnote-7) He concludes that people are motivated to form beliefs that signal their ingroup status.

Whether or not belief has this signaling function, it cannot be the *only* driver of belief formation. If beliefs have the gatekeeping role that Funkhouser thinks, for example, we must also have beliefs that help us to recognize and acquire the ingroup-typical beliefs we desire (e.g., the expectation that people in one’s network would *approve* of assertions of climate denial). More generally, most of our beliefs–the recognition that these black pants feel too warm for this weather, best guesses about which coffee shops around here are open now, the idea that these chipmunks will like my trail mix–are only distantly related to our ingroup.

All of these views might seem orthogonal to whether belief is Spinozan or Cartesian. After all, they focus on what the authors take to be belief’s social and communicative functions, while we are interested in its underlying architecture. Marketplace and signaling models do not even treat beliefs as mental representations, but as public entities (Williams 2022 p. 104; Funkhouser 2017, p. 16). Vigilance is agnostic about the psychological reality of belief, though as aforementioned vigilance theorists (Mercier 2020: Ch. 3, McKay & Mercier 2023) do claim the mechanisms of belief exist to support reliable signaling.

In fact, marketplace and signaling views, unlike Spinozan models, may not be making psychological generalizations about belief at all. On the Spinozan hypothesis, and for psychofunctionalism more generally, beliefs are mental representations with a distinctive psychofunctional profile, and public signaling and market exchange are not part of it. Beliefs interact with other mental states: they are caused by perception and lead to action; they figure in inferences. It just is not clear how a publicly shared entity could be expected to do these things.

Consequently, it might seem that vigilance theorists should be very open to Spinozan belief formation. They are simply agreeing that there exists a robust rejection mechanism somewhere in the mind, separate from an acquisition mechanism. As Mercier writes, the “mechanisms [of belief] are at least as much about being open to communicated information as being vigilant toward it” (2020, Ch. 3; he also attributes this position to Sperber et al. 2010 in a footnote).

Yet Mercier and others take Spinozan models to be rivals to their own (Sperber et al. 2010: p. 362-4; Mercier 2020: 36-37, 43-44). For example, after lumping Spinozan models in with dual process theories, Mercier (2020: 43-44) claims, “far from being gullible and biased to believe (Gilbert 1990), System 1 is, if anything, biased to reject any message incompatible with our background belief, but also ambiguous messages or messages coming from untrustworthy sources” and “[t]here is no experimental evidence suggesting a systematic association between being less analytically inclined–less likely to use one’s System 2–and being more likely to accept empirically dubious beliefs” (p. 44).

We doubt the empirical claims made by Sperber, Mercier, and others arguing against Spinozan accounts (see Mandelbaum 2014 for some counterarguments) and stress there is strong and replicable experimental evidence in favor of the traditional link between analytic thinking and System 2 (see, e.g., Knowles and Condon 1999; Pennycook et al. 2012; Pennycook et al. 2013; Pennycook et al. 2015; Pennycook & Rand 2020; Pennycook et al. 2022; Binnendyk & Pennycook, 2022; Hattersley et al., 2022; 2023). Nonetheless, there is a more recent argument against dumb mechanics of belief formation vigilance proponents could rely on here, namely that the evidence for them is due to biases in guessing strategies that favor “true”-responding (Street & Kingston 2017). These arguments are usually based on apples-to-oranges comparisons of the Gilbert and Gilbert et al. results with other studies that do not involve cognitive load manipulations.[[8]](#footnote-8) This is important, since it is the cognitive load manipulations that are the critical causal manipulations in the Spinozan architecture: load is the force that makes rejection, but not acceptance, more difficult. In addition, skin conductance responses provide compelling empirical evidence that they do not offer an alternative explanation of existing results, as misremembering false sentences as true appears to have a different physiological signature from misremembering true sentences as false, and from correctly remembering the truth value of a sentence (Asp under review).[[9]](#footnote-9)

In short, marketplaces, signaling, and vigilance all share a core commitment to cynicism in belief and a failure to commit to a specific cognitive architecture to explain that ability. The first two views are not easily compatible with any psychological treatment of belief, as they treat beliefs as public entities. Proponents of vigilance mistakenly treat Spinozan belief as a rival view to their own, rather than a complement to it, and they offer only thin arguments against it.

2.2 Cognition: Both Dumb and Cynical

A general concern with epistemic vigilance, marketplaces of rationalization, and signaling views is that they do not speak to questions about the cognitive sources of cynicism in individuals. Proponents of these views might favor a different cognitive architecture of belief from the one that we have sketched here, and this is indeed the case for at least some proponents of epistemic vigilance (Mercier & Sperber 2018), who favor a massively modular architecture (Tooby & Cosmides 1992, Sperber 1995, Carruthers 2006).[[10]](#footnote-10) However, one could endorse the mechanisms that we have described here as the underlying structure of individual minds, while also accepting cynicism’s fitness-enhancing role. The project that proponents of these views are engaged in is important, but also quite different from the project of describing the structure of human minds.

This avoidance of mechanistic psychological explanation converges with an assumption of these views mentioned earlier, which is that believing bears some deep connection to communicating.[[11]](#footnote-11) For example, Mercier (2020, Ch. 3) claims that belief emerged to ensure reliable communication. Non-human animals have simpler, more constrained communication than we do, he notes, just as the eucalyptus-eating koala has a simpler, more constrained diet than our omnivorous one. This explains the richness of our vigilance system: vigilance evolved in lockstep with communication to help us navigate the risks of communication, much like our preparedness to learn nausea-gustatory stimuli associations evolved with our omnivorousness to help us avoid risky foodstuffs (Mandelbaum 2015/2020). For Mercier, there was no “arms-race” (Mercier 2021, p. 260) through which increasingly sophisticated rejection abilities were added on to a more ancient dumb belief formation system.

Many animal species have beliefs. Beyond merely forming perceptual beliefs or seeking out food and mates, many anticipate cache-theft (Cheke & Clayton 2010), learn from conspecifics (Bridges et al. 2023), navigate complex hierarchical social environments (Cheney & Seyfarth 2008; Camp 2009), and engage in dissonance-based reasoning (Zentall 2016; Harmon-Jones et al. 2017 discuss effort justification; Egan et al. 2008 provide evidence for the spreading-of-alternatives). Correspondingly, recent decades have witnessed an explosion of mentalistic language in comparative psychology journal-article titles (Whissell et al. 2013). While we suspect non-human communication may be more complex than the dietary preferences of the koala, if animals accomplish these complex cognitive tasks without human-like communication, then surely our pre-linguistic ancestors could have done the same.

It is likely that the ability to scrutinize incoming information became important as we began to communicate in more complex ways, including through natural language. But the core of belief’s psychofunctional role predates that development. While there is much we still do not know about the cognitive complexity of non-human minds, we share a lot of our psychological faculties with many species: we have shared faculties for forming perceptual beliefs, performing practical inferences, and representing many features of our social and physical environments without the aid of rich communication. These are not merely peripheral aspects of believing.

This presents a dilemma for vigilance theorists. They seem to think the ability to form beliefs evolved with the function to be vigilant against misinformation, i.e., to reject information. But they also think that vigilance emerged as communication became more complex. This seems to entail that belief should only arise with humans, or at least in species with rich communicative abilities. Excluding beliefs from animals is, in our opinion, a rather extreme position to hold.

Advocates of vigilance might respond that their explanation of human vigilance applies to much animal and preverbal infant behavior, too (Mercier 2020, Ch. 3). The idea is that in order for signaling to aid fitness, the signals should be generally reliable. For example, white-tailed deer show ambush predators their tails because this reliably signals they’ve been detected (as well as readies them to run just in case the signal is ignored). But this mostly restates the challenge to Mercier and others: do animals signal because they have beliefs (about, e.g., nearby cheetahs), or for some other reason (innate behavioral responses, operant conditioning)? If it is because of some set of beliefs, then the core of belief really is older than vigilance; if not, then the explanation of animal behavior here is not really the same as the explanation of vigilance, since one is mediated by belief and the other is not.

An alternative view is that animals—even from widely separated branches of the phylogenetic tree— have the capacity to believe, predicated on a shared language of thought (Quilty-Dunn et al. 2022). However, the emergence of rich human communication created selective pressures on the mostly dumb belief mechanisms of our ancestors. New belief mechanisms emerged that allow us to assess the quality of testimony. In other words, communication created the need for a rejection mechanism separate from the existing dumb acceptance mechanism, as Spinozan views predict.[[12]](#footnote-12)

To be clear, the argument here is not that animals have beliefs but lack natural language, and thus that vigilance cannot have arisen at the same time as belief. Many animals communicate without natural language, as Mercier’s white-tailed deer example demonstrates. Instead, it is that belief’s psychofunctional role is almost completely independent of communication, however one understands it. This is most visible in animals, who possess full-fledged beliefs despite heavy constraints on their communicative signals. But the argument works just as well in the human case. Should we really suppose, for example, that our ability to form the belief that these pants are too warm for this weather has anything to do with our ability to communicate?

When one focuses on individual psychology, Spinozan theories are needed in order to explain the relevant data at hand. Never mind the memory asymmetries (see Mandelbaum 2014; Porot and Mandelbaum 2020/2022), basic facts of misinformation need a Spinozan system. First, if people are forever epistemically vigilant then why would receiving a single exposure to fake news raise people’s credence in news they’ve already seen flagged as false (approximately doubling “true”-responses, Pennycook et al. 2018)? And it’s not just gullible acceptance that epistemic vigilance fails to explain, but also specific problems with misinformation retraction. For instance, why are fact checks (e.g., true and false tags) more effective at countering misinformation when they come *after* the misinformation than when they are concurrent with it (Brashier et al. 2021)? The vigilance camp should see this as a surprise--if we are selected to be vigilant, then why should timing matter? If anything, errors should be not just rare, but randomly distributed regardless of the timing of the warning (within bounds; e.g., the warning can’t come a decade after the message). The Spinozan account can explain this data. It’s because the rejection process is fragile (i.e., easily short circuited), that it needs attention. Information reception (for startling information) can itself be load imparting, so synchronous fact-checking will lead to more errors. And warnings given before misinformation (‘prebriefing’, Wegner et al. 1985) create more load because one must keep the warning in mind when receiving the new information. However, receiving the warning after the headline allows one to focus on the misinformation without the need to hold the false tag in memory, and without any extra stream of incoming information to interfere with the false tag application.

Evidence for a *bias blind spot* (BBS) might seem like just the evidence epistemic vigilance would want (Mata et al. 2013). In BBS, people attribute bias more readily to others than to themselves (Pronin et al. 2002) and assess others’ reasoning more effectively than their own, suggesting differences in the way that testimony and thought are scrutinized, as epistemic vigilance would predict (ibid).

Crucially, however, when reasoning about issues people care deeply about, this asymmetry disappears. Cusimano & Lombrozo (2023) asked participants to assess real sociology papers that either confirmed or denied racial differences in tipping behavior in Texas. When asked whether they thought their reasoning about the quality of the studies or the reliability of their conclusions was biased by their moral outlook, participants were generally pretty accurate. Not only did motivated reasoners noticed they were biased, when asked to evaluate their reasoning, they endorsed their bias, judging it was *good* to reason in the way they had reasoned. Clearly the same mechanisms of rejection are available for one’s own thought as for testimony, as participants recognized when they were biased. But some other factor keeps them from rejecting their biased reasoning, even as they appear to be conscious of it. This suggests the bias blind spot is not evidence for some special suite of testimony-assessing mechanisms. Instead, it offers yet more evidence that incoming information is assessed in a normatively non-ideal way when it undercuts beliefs we care about—just as predicted on a view that pairs Spinozan belief with a Psychological Immune System.

Summing up, the main problems facing epistemic vigilance are: how could communication be the origin of belief? And how do proponents explain away the behavioral and neural evidence for a Spinozan system, one where we effortlessly accept information, but effortfully reject information? The real dialectical problems for them run even deeper, as the explanatory goods here are asymmetric. The psychofunctionalist position we are advocating for[[13]](#footnote-13) can capture the cynicism that vigilance theory captures, while also explaining patterns of acceptance and rejection and the ubiquity of belief in the animal world. Similarly, the psychofunctional theory is consistent with the claims from the social marketplace collection of views, the inverse isn’t true: the epistemic vigilance and social marketplace camps can’t explain the asymmetries in belief that the Spinozan theory captures.

2.3 Uncles on Facebook: Rampant Misinformation

An added difficulty of placing cynicism at the core of the mechanics of belief is that it becomes difficult to explain the truly dumb mechanics of belief updating. Supposing that human belief is, in general, cynical, what *explains* performance of participants in illusory truth studies, or in the cognitive load paradigms run by Gilbert and colleagues? How do we explain what our uncles post on Facebook? The systematicity of human idiocy suggests an important structural feature of the human belief competence.

One increasingly popular strategy for theorists to undertake is to downplay the existence of dumb belief. Altay et al. (2023), for example, take explicit aim at the idea of an “infodemic,” arguing that both the amplitude and effectiveness of misinformation online are largely overblown (and misinformation is often shared because it is news that would be interesting-if-true, Altay 2022). In support of this idea, a recent series of papers from a collaboration between Meta and non-industry researchers suggest that suppressing resharing tools (Guess et al. 2023a) and eliminating feed-order algorithms (Guess et al. 2023b)[[14]](#footnote-14) had little effect on misinformation exposure, compared to controls.

This strategy has largely relied on observations of Western politics, and in particular on the run-up to the 2016 and 2020 U.S. elections. Mercier’s (2020, Ch. 13) case against the widespreadedness of fake news relies exclusively on analysis of pro-Trump articles from 2015-16. This is not due to cherry-picking. Most of the research to date has been conducted on these samples, with quantitative research outside these areas especially limited. Nonetheless the recent politics of Nigeria and the Sahel states offers a different perspective, providing evidence against both of Altay et al.’s generalizations. First, though fake news’ exact reach in the region is hard to quantify, it appears to be widespread. Nigeria’s 2019 election faced many barriers, including rescheduled national elections and suspension of the nation’s Chief Justice. But not least of these was a tsunami of pro-government fake news just preceding the elections. Fact checkers at CrossCheck Nigeria described their own efforts as “a drop in the ocean” (CNN 2019). By all measures the informational ecosystem is getting worse, with fake news deemed an “existential” threat to the nation leading up to the recent elections (DW 2023), and even mainstream news outlets peddling falsehoods from WhatsApp (Reuters 2023). Across the Sahel, and especially in Burkina Faso, Mali, and Niger, stories misrepresenting the French military’s regional anti-terrorism “Barkhane” operation are rampant on WhatsApp (Kirwin et al. 2022). The September 2022 coup d’état in Burkina Faso witnessed exaggerations on social media of previous president Paul-Henri Damiba’s coziness with France. Rumors that Damiba had taken refuge in the French embassy were used to justify an attempted attack by protesters there (France 24, 2022). And this isn’t to mention the political violence perpetuated by fake news on Facebook and Whatsapp in India, Brazil, Myanmar, the Philippines, and even Germany; or the current war in Israel and the deluge of fake news and­ deepfakes on X.[[15]](#footnote-15)

As for whether fake news really changes people’s minds, rigorous research in countries outside the Global North is just beginning. But it is plausible, for example, that in the Global South, where smartphones and social media dominate internet usage, and trust of traditional journalism is already low in many countries, fake news may be even easier to disseminate than in the North. The more one is skeptical of incoming information, the more one is forced to rely on their prior beliefs. Kirwin et al. (2022) find, for example, that across four Sahel countries, WhatsApp users showed significantly higher anti-French sentiment than non-WhatsApp users. This suggests increased access to anti-French-military fake news shifted attitudes toward France.[[16]](#footnote-16) Crucially, frequent users also showed more of this sentiment than infrequent users, partly deflating demographic confounds due to rural-urban divides in access to 3/4G towers and smartphones. Clearly much work remains to be done to better understand the reach and impact of misinformation globally. Expanding social research on fake news, especially to the Global South (where it is arguably a deeper-rooted tradition, Wasserman & Madrid-Morales 2022) is a critical step in understanding the reach and efficacy of fake news.

The strategy of downplaying fake news has been of limited value in experimental settings. Even though fake news stories on Twitter spread faster, more deeply, and to more users than true ones (Vosoughi et al. 2018), Pennycook and Rand’s work on fake news shows that people will share information that, if asked to focus on accuracy, they may disavow (Pennycook & Rand 2021). Theorists of the ‘social marketplace’ ilk (Funkhouser 2022, Williams 2022), take evidence like this to show that people don’t really believe a lot of claims on social media (Altay et al. 2022). However, this work of Pennycook and Rand only shows that when people attend to a claim and are asked to assess accuracy, they can do so. From this we can infer that there isn’t some set of monolithic motivational biases that make people unable to assess evidence, and that we have the capacity to correctly assess information, which is consistent with the psychofunctional project. But what this work doesn’t show is how the original exposure affects subjects’ beliefs and loads of work in this area—including Pennycook and Rand’s—show that a single exposure raises credences (e.g., doubling credence in fake news headlines (Pennycook et al. 2018). Again, the question is how the vigilance and social signaling camps can explain why mere exposure should have any effect on belief if Spinozan views are wrong.[[17]](#footnote-17)

An alternative to downplaying dumb belief is to explain it away by attributing it to a cognitive attitude that is merely belief-*like*. Mercier (2020), for example, acknowledges the existence of genuinely dumb belief in his book length defense of vigilance. Following Sperber (1997), he divides belief into *intuitive* and *reflective* varieties (Baumard & Boyer 2013, Mercier & Sperber 2009; Sperber & Mercier 2018; Mercier 2020, p. 152). The difference between these attitudes, the idea goes, is obscured by our overly-coarse-grained folk psychology.

In Mercier’s and Sperber’s idiolect, intuitive beliefs are the kinds of beliefs that guide our everyday behavior. For example, one might intuitively believe that on Fridays, there is usually couscous served in the faculty restaurant starting at 1 PM. Seeing that it is 12:45, and being pretty hungry, one then might head down to the restaurant. Reflective beliefs, on the other hand, have at best a tenuous connection to our day-to-day behavior. We can usually endorse them publicly, but they do not seem to direct behavior, and their inferential promiscuity is limited. On the initial characterization of the view (Sperber 1997), reflective beliefs are attitudes we bear to contents (understood as that-clauses in a LoT) that we do not “disquote”: we can think with them without believing them, just as we supposedly think about attitude reports without believing their contents. This allows us to express them without acting on them. A key example is belief in the Trinity, which while widely proclaimed by Christians, may not be fully grasped by anyone, and so cannot be the kind of thing that drives day-to-day behavior, or lead to non-trivial inferences.[[18]](#footnote-18)

According to Mercier, epistemic vigilance applies to intuitive belief, but not to reflective belief. To believe something, the idea goes, one must act on one’s beliefs, not merely proclaim them. Mercier illustrates this using Pizzagate (2020, p154-5). While followers of Q across social media claimed Comet Ping Pong Pizza was the site of a D.C.-elite pedophilia ring, only one person stormed the establishment to rescue the supposed victims. Shouldn’t all followers of Q have done the same, Mercier wonders? The vigilante in Pizzagate intuitively believed Q’s conspiracy theory, while others only reflectively believed it. (Unfortunately, Mercier’s book was published less than a year before the events of January 6, 2021—though other anecdotes could be used here for the same point).

A key aim of the cognitive science of belief is to isolate a natural kind that best matches the one used in folk psychology and capture the generalizations that hold over it (Porot & Mandelbaum 2020/2022; Mandelbaum & Porot 2022). There will inevitably be ways in which the natural kind diverges from the folk one,[[19]](#footnote-19) and so it is an open question whether one ought to split belief in the way epistemic vigilance requires. But the barrier to entry on splitting cognitive attitudes should be high.[[20]](#footnote-20) There is no kind more central to folk psychology than belief. All else equal, competing explanations of cases such as Pizzagate that allow for the unity of belief should be preferable to explanations that suggest our basic folk psychology is deeply misguided.

But there are more than methodological reasons for thinking we should keep a single kind. For one thing, there is no evidence that illusory truth effects show any difference for reflective or intuitive beliefs. The fact that the same laws hold over both putatively separate states implies that there is only a single state here. Similarly, the immune system generalizations hold over contents one self-identifies with regardless of whether they are “disquoted” or not (e.g., belief that Jesus is God or the belief that caffeine isn’t bad for you will both work just fine, Mandelbaum 2019a), thus crosscutting the intuitive/reflective divide.

Furthermore, there are alternative explanations for the cases Mercier focuses on. The evidence for reflective belief boils down to context dependence. People act as if they believe *P* in some contexts, and as if they believe *~P* in others. Some people speak as if they believe the Clintons are chaining up kids at a pizzeria, but they do not act as if they believe that when they are walking past that pizzeria. We think there is nothing mysterious about this, certainly nothing more mysterious than people who claim they cannot live without their partners in one context, and assert that they can’t stand them in another one. Context sensitivity is widespread in human belief.

This is not because we are just sometimes acting *as if* we believe, but because we actually believe p in some contexts and not-p in others. Thefragmentationof belief (Bendaña & Mandelbaum 2021) offers a general explanation of this context-sensitivity. Fragmentation maintains that beliefs are stored in causally isolated databases in long-term memory (i.e., fragments) each of which is activated by a specific context or contexts. Because of the causal isolation of fragments, the full set of one’s beliefs is almost certain to contain both redundancies and contradictions. By hypothesis, fragments are internally consistent. Because fragments are activated by specific contexts, fragmentation is tailor-made to explain inconsistency across contexts.

We will not delve into the full literature supporting the fragmentation of belief, but it should suffice to point out that evidence of incompatible beliefs is, in general, hard to explain without fragmentation. What fragmentation allows for is a doxastic network where we can both harbor inconsistent beliefs (in different fragments) and still abhor inconsistency (as each fragment is consistent and the default situation is to only be thinking with one fragment at a time). As it happens, there is ample evidence from across cognitive science for incompatible beliefs within a single mind (see Sommer et al. 2022, Bendaña & Mandelbaum 2021 for reviews), casting doubt on the web of belief view, and favoring a fragmented storage of belief.

Let’s step back and look at where this leaves proponents of epistemic vigilance who want to explain dumbness using the intuitive-reflective distinction. If belief is fragmented, it must be context-sensitive, too. Little motivation remains for partitioning our most fundamental folk psychological kind into intuitive and reflective attitudes, as the partitioning of memory already does that work. (Dividing up belief storage into fragments is also a natural move, as memory is already divided functionally–perceptual, short term, long term, working, visual-working, procedural, prospective, semantic, episodic, and so forth. For a quite different take on memory see Aronowitz 2018). Dumbness is a feature of belief simpliciter and need not require appeal to novel belief-like cognitive attitudes.

3. Conclusion

We have argued for a general view of belief. Beliefs are acquired in an automatic, uncomplicated way, and one where a person is primed to reject information only given a certain cynical sort of mindset. Changing beliefs can be done in (at least) two ways: one cold and rational, the other hot and ego-protective. Our dumb, cold, cynical view of belief is grounded in a specific architecture of the mind. This architecture is tendentious (not unlike…the vast majority of cognitive science posits). But in doing so we put our necks out for very specific views of how the mind works. This allows us to make falsifiable claims. Pinpointing exactly where we go wrong moves the dialectic forward. We are trying to ensure that our goal doesn’t fall short of our reach.

We do not believe that negative articles—articles that merely criticize others’ views—are particularly fruitful, especially without putting forward a positive proposal. Yet here we have also criticized competitor views. We do not intend this criticism as a way of convincing these other theorists to stop forwarding their programs. Their theses are interesting, and—as we’ve read—competition in the marketplace of ideas often helps sharpen ideas.[[21]](#footnote-21) Instead of gladiatorial debate, we welcome competitor theories that offer up specifics for how the cognitive architecture of belief works. We don’t want, e.g., Williams to stop pushing the social marketplace line; rather, it’s better for him to flesh out how he thinks it works, what the specific evidence for rationalization is, and how status-seeking is supposed to interplay with other examples of the functional role of belief offered here.

For example, it is unclear to us whether Mercier would agree that beliefs are endemic to humans. Mercier—like Funkhauser and Williams—talk of *signaling,* e.g., predator-deterrent signals. Gazelles will signal to predators that they have spotted the predator by stotting: jumping very high, which is an unfakeable signal of physical fitness. Stotting deters wild dogs from attacking. But what are the mechanisms of this intricate communication? We just wrote “gazelles will signal to predators that they have spotted the predator”—don’t the gazelles *believe* they have spotted the predator, and doesn’t the predator believe that the gazelles believe they have been spotted? If not, how does this system work?

We see three possible classes of mechanism for this communication. Perhaps there is an innate connection, knowledge shared through species-specific inheritance, that when a gazelle stots, wild dogs automatically stay still (or inhibit their penchant to chase) so that they conserve energy. Or perhaps this is a learned connection such that over the course of the animal’s lifespan they have been co-conditioned: every time a gazelle jumped high and dogs chased, all the creatures ended up enervated and the dogs dinner-less.

Opposing both the nativist and behaviorist hypotheses is the belief-based reasoning hypothesis: gazelles jump high when they see dogs because they believe this will show the dogs that they a) know the dogs are there and b) are too fit to be trifled with. Dogs know stotting gazelles know they have been spotted. Practical rationality wins out and the dogs leave the gazelle alone. We are willing to bet that belief is already in the picture in this communicative dance, as it seems like all of these animals are reasoning. In this way of seeing the world, signals aren’t beliefs, yet beliefs still sometimes have a kind of signaling function as they set off chains of reasoning. Does Mercier think that all there is to belief is the signal? Or that there is just information but not belief here (perhaps because of the lack of vigilance?) Or perhaps one is inclined to become a knowledge-firster (Williamson 2002) or otherwise a knowledge-before-belief theorist who think there’s often knowledge in the picture but no belief (Drayton & Santos 2016; Phillips et al. 2021)?

We are no foes of nativism, yet of course recognize that hard-coding information like this is costly. The behaviorist idea predicts many fruitless interactions to build the foundations of the learned connection between stotting and not-chasing, a learning curve for which we know of no evidence. Still either hypothesis can be true, and the belief-reasoning one false. We are not experts in gazelle behavior---perhaps there will be evidence brought forward showing the trial-and-error learning between gazelles and dogs. We can make progress by having competitor views specify how they think these links work, and where beliefs enter the picture.

The larger moral is that specific hypotheses about cognitive architecture would deepen the dialectic. Mercier relies on a massive modular model of the mind but the specific details--how massive modularity links to vigilance--are hard to discern. Perhaps he is skeptical of the psychological reality of belief, for example, or perhaps the ‘information’ stored in our minds’ many modules will turn out to be just beliefs encoded in a LoT after all. Social marketplace theorists offer almost no details about any cognitive processing. We see our theories as possibly being compatible—there is no particular reason insights from these more social-facing theories cannot be merged with our architecture. For example, it is intuitive that beliefs can act as costly social signals for forging in-group bonds. If our interlocutors object to our framework of belief acquisition, storage, and change we welcome them to offer competitor architectures that we can assess and test. Doing so would help flesh out their theories, pulling out their deeper commitments. Seeing where any of these views go wrong—including our own—would spur the cognitive science of belief forward.

Acknowledgements: Thanks to Imane Badari for discussion of the bias blind spot, and Josh Shepherd for general discussion.

References

Altay, S., de Araujo, E., & Mercier, H. (2022). “If This account is True, It is Most Enormously Wonderful”: Interestingness-If-True and the Sharing of True and False News. *Digital Journalism*, *10*(3), 373–394.<https://doi.org/10.1080/21670811.2021.1941163>

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media + Society*, *9*(1), 20563051221150412.

<https://doi.org/10.1177/20563051221150412>

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 241–252. [https://doi.org/10.1037/0278-7393.5.3.241](https://psycnet.apa.org/doi/10.1037/0278-7393.5.3.241)

Béna, J., Rihet, M., Carreras, O., & Terrier, P. (2022). *Repetition could increase the perceived truth of conspiracy theories*. PsyArXiv.<https://doi.org/10.31234/osf.io/3gc6k>

Bendaña, J., & Mandelbaum, E. (2021). The fragmentation of belief. In C. Borgoni, D. Kinderman, A. Onofri (Eds.) *The fragmented mind*. Oxford: Oxford University Press

Binnendyk, J., & Pennycook, G. (2022). Intuition, reason, and conspiracy beliefs. *Current Opinion in Psychology*, 101387.

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5), e2020043118.<https://doi.org/10.1073/pnas.2020043118>

Bridges, A. D., MaBouDi, H., Procenko, O., Lockwood, C., Mohammed, Y., Kowalewska, A., González, J. E. R., Woodgate, J. L., & Chittka, L. (2023). Bumblebees acquire alternative puzzle-box solutions via social learning. *PLOS Biology*, *21*(3), e3002019.<https://doi.org/10.1371/journal.pbio.3002019>

Camp, E. (2009). A language of baboon thought? In R. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 108-127). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511819001.007

Cheke, L. G., & Clayton, N. S. (2010). Mental time travel in animals. *WIREs Cognitive Science*, *1*(6), 915–930.<https://doi.org/10.1002/wcs.59>

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon Metaphysics: The Evolution of a Social Mind*. University of Chicago Press.<https://press.uchicago.edu/ucp/books/book/chicago/B/bo5387727.html>

CNN. (2019). Nigeria Election 2019: How fake news was weaponized. https://edition.cnn.com/2019/02/15/Baafrica/fake-news-nigeria-elections-intl/index.html

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adapted mind: Evolutionary psychology and the generation of culture*, *163*, 163-228.

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences*, *107*(supplement\_2), 9007-9014.

Cusimano, C., & Lombrozo, T. (2023). People recognize and condone their own morally motivated reasoning. *Cognition*, *234*, 105379. <https://doi.org/10.1016/j.cognition.2023.105379>

DW. (2023). Nigeria 2023 elections: Countering fake news. https://www.dw.com/en/nigeria-2023-election-countering-fake-news/a-64651081

Davidson, D. (1982). Rational animals. *dialectica*, *36*(4), 317-327.

Egan, L. C., Santos, L. R., & Bloom, P. (2007). The Origins of Cognitive Dissonance: Evidence From Children and Monkeys. *Psychological Science*, *18*(11), 978–983.<https://doi.org/10.1111/j.1467-9280.2007.02012.x>

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002.

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review*, *26*(5), 1705–1710. <https://doi.org/10.3758/s13423-019-01651-4>

Funkhouser, E. (2017). Beliefs as signals: A new function for belief. *Philosophical Psychology*, *30*(6), 809–831.<https://doi.org/10.1080/09515089.2017.1291929>

Funkhouser, E. (2022). A tribal mind: Beliefs that signal group identity or commitment. *Mind & Language*, *37*(3), 444–464.<https://doi.org/10.1111/mila.12326>

Jern, A., Chang, K. M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological review*, *121*(2), 206.

Festinger, L., Riecken, H. W. & Schachter, S. (1956). When prophecy fails. Minneapolis: University of Minnesota Press. https://doi.org/10.1037/10030-000

France 24. Coup d’État au Burkina Faso : les raisons de la chute du putschiste Paul-Henri Damiba. https://www.france24.com/fr/afrique/20221003-coup-d-%C3%A9tat-au-burkina-faso-les-raisons-de-la-chute-du-colonel-damiba

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*(4), 601–613. [https://doi.org/10.1037/0022-3514.59.4.601](https://psycnet.apa.org/doi/10.1037/0022-3514.59.4.601)

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can’t not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221–233.<https://doi.org/10.1037/0022-3514.65.2.221>

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*(3), 617–638. [https://doi.org/10.1037/0022-3514.75.3.617](https://psycnet.apa.org/doi/10.1037/0022-3514.75.3.617)

Goffin, K. (2022). Feeling is Believing: Recalcitrant Emotion & Spinozan Belief Formation. *Synthese*, *200*(6), 1–14.<https://doi.org/10.1007/s11229-022-03980-9>

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023a). *Reshares on social media amplify political news but do not detectably affect beliefs or opinions*.

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023b). *How do social media feed algorithms affect attitudes and behavior in an election campaign?*

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review. 108, 814-834.

Harmon-Jones, C., Haslam, N., & Bastian, B. (2017). [Dissonance reduction in nonhuman animals: Implications for cognitive dissonance theory](https://www.wellbeingintlstudiesrepository.org/cgi/viewcontent.cgi?article=1191&context=animsent). *Animal Sentience,* 12(4)

Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. *The Journal of philosophy*, *105*(10), 611-633.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. [https://doi.org/10.1016/S0022-5371(77)80012-1](https://doi.org/10.1016/S0022-5371%2877%2980012-1)

Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. The Journal of Philosophy, 105(10), 611–633. https://doi.org/10.5840/jphil20081051024

Kirwin, M., Ouedraogo, L., & Warner, J. (2022). Fake News in the Sahel: “Afrancaux News,” French Counterterrorism, and the Logics of User-Generated Media. *African Studies Review*, *65*(4), 911–938.<https://doi.org/10.1017/asr.2022.63>

Knowles, E. S., & Condon, C. A. (1999). Why people say" yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379.

Lewis, D. (1969). *Convention: A philosophical study*. New York, NY: Cambridge University Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*, e1.

Loughnan, S., Haslam, N., & Bastian, B. (2010). The role of meat consumption in the denial of moral status and mind to meat animals. *Appetite*, *55*(1), 156–159.<https://doi.org/10.1016/j.appet.2010.05.043>

Mandelbaum, E. (2013). Against Alief. Philosophical Studies. 165: 197-211.

Mandelbaum, E. (2014). *Thinking is Believing*. *Inquiry*. 57(1). 55-96.

Mandelbaum, E. (2015). The automatic and the ballistic: Modularity beyond perceptual processes. *Philosophical Psychology*, *28*(8), 1147-1156.

Mandelbaum, E. (2015/2020) Associationist theories of thought. In Zalta, Edward N. (Ed.). The Stanford Encyclopedia of Philosophy https://plato.stanford.edu/entries/associationist-thought/

Mandelbaum, E. (2017). Seeing and Conceptualizing: Modularity and the Shallow Contents of Perception. *Philosophy and Phenomenological Research.* 97(2): 267-283.

Mandelbaum, E., Won, I., Gross, S., & Firestone, C. (2020). Can resources save rationality?‘ Anti-Bayesian’ updating in cognition and perception.

Mandelbaum, E. (2019a). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language*, *34*(2), 141–157.<https://doi.org/10.1111/mila.12205>

Mandlebaum E. (2019b). Modularist explanations of experience and other illusions. *Consciousness & Cognition* 76: 102828

Mandelbaum, E. & Porot, N. (2022). How the cognitive science of belief can transform the study of mental health. *JAMA: Psychiatry*.

Mascaro, O., & Morin, O. (2014). Gullible’s travel: How honest and trustful children become vigilant communicators. In *Trust and Skepticism*. Psychology Press.

Mattavelli, S., Corneille, O., & Unkelbach, C. (2022). Truth by repetition … Without repetition: Testing the effect of instructed repetition on truth judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. [https://doi.org/10.1037/xlm0001170](https://psycnet.apa.org/doi/10.1037/xlm0001170)

Mayo, R. (2019). Knowledge and distrust may go a long way in the battle with disinformation: Mental processes of spontaneous disbelief. *Current Directions in Psychological Science*, *28*(4), 409-414.

McKay, R., & Mercier, H. (2023). Delusions as Epistemic Hypervigilance. *Current Directions in Psychological Science*, *32*(2), 125–130.<https://doi.org/10.1177/09637214221128320>

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 149–170). Oxford University Press. [https://doi.org/10.1093/acprof:oso/9780199230167.003.0007](https://psycnet.apa.org/doi/10.1093/acprof%3Aoso/9780199230167.003.0007)

Mercier, H. & Sperber, D. (2018). *The Enigma of Reason*. Harvard University Press.

Mercier, H. (2020). *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton University Press.

Mercier, H. (2021). How Good Are We At Evaluating Communicated Information? *Royal Institute of Philosophy Supplements*, *89*, 257–272.<https://doi.org/10.1017/S1358246121000096>

Michelitch, K., & Utych, S. (2018). Electoral Cycle Fluctuations in Partisanship: Global Evidence from Eighty-Six Countries. *The Journal of Politics*, *80*(2), 412–427.<https://doi.org/10.1086/694783>

Müller, Karsten, and Carlo Schwarz. "Fanning the flames of hate: Social media and hate crime." *Journal of the European Economic Association* 19.4 (2021): 2131-2167.

Quilty-Dunn, J., Porot, N. & Mandelbaum, E. (2022). The best game in town: The re-emergence of the

Language of Thought Hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335-346.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2013). Belief bias during reasoning among religious believers and skeptics. *Psychonomic Bulletin & Review*, *20*, 806-811.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current directions in psychological science*, *24*(6), 425-432.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, *147*(12), 1865.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality*, *88*(2), 185-200.

Pennycook, G., Binnendyk, J., & Rand, D. (2022). Overconfidently conspiratorial: Conspiracy believers are dispositionally overconfident and massively overestimate how much others agree with them.

Pennycook, G. (2022). A framework for understanding reasoning errors: From fake news to climate change and beyond. *Advances in Experimental Social Psychology vol 67.*

Porot, N. & Mandelbaum, E. (2020/2022). The science of belief: A progress report.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. Personality and Social Psychology Bulletin, 28(3), 369-381

Reuters. (2023). Nigeria's social media fact-checkers fight fake news as vote nears. https://www.reuters.com/article/nigeria-socialmedia-election-idUSL8N33Y0BH

Sommer, J., Musolino, J., & Hemmer, P. (2022). A hobgoblin of large minds: Troubles with consistency in belief. *WIREs Cognitive Science*.<https://doi.org/10.1002/wcs.1639>

Sperber, D. (1997). Intuitive and Reflective Beliefs. *Mind & Language*, *12*(1), 67–83.<https://doi.org/10.1111/j.1468-0017.1997.tb00062.x>

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393.<https://doi.org/10.1111/j.1468-0017.2010.01394.x>

Sun, L. (2023). *The Liars of Nature and the Nature of Liars: Cheating and Deception in the Living World*.

Princeton University Press.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. Psychological Science, 17(4),

339–345 <https://doi.org/10.1111/j.1467-9280.2006.01708.x>

Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition, 133*(3), 698-715.

<https://doi.org/10.1016/j.cognition.2014.08.015>

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.<https://doi.org/10.1126/science.aap9559>

Wasserman, H., Madrid-Morales, D. (Eds.). (2022). Disinformation in the Global South. Wiley-Blackwell.

Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology*, *49*(2), 338.

Whissell, C., Abramson, C. I., & Barber, K. R. (2013). The search for cognitive terminology: an analysis of comparative psychology journal titles. *Behavioral sciences (Basel, Switzerland)*, *3*(1), 133–142. <https://doi.org/10.3390/bs3010133>

Williams, D. (2022). The marketplace of rationalizations. *Economics & Philosophy*.

https://doi.org/10.1017/S0266267121000389

Zentall, T.R. (2016) [Cognitive dissonance or contrast?](https://www.wellbeingintlstudiesrepository.org/animsent/vol1/iss12/1). *Animal Sentience* 12(1)

1. If there is more than one type of updating mechanism (as we’ll argue), might there also be more than one rejection mechanism? We find the question intriguing but have no firm stance here. Perhaps there isn’t a single rejection mechanism. Maybe Bayesian-based rejection and Psychological Immune System rejection are different ways of rejecting, both being somewhat fragile (we say ‘somewhat’ because when people are in skeptical mindsets they seem to be able to flexibly reject information quickly). [↑](#footnote-ref-1)
2. Importantly, this type of perverse updating happens in two separate ways. One way is by selectively attending to pro-attitudinal information. This behavior leads to a type of polarization that can be modeled as rational (Jern et al 2014; Kelly 2008, Dorst 2022). However, the other route to polarization--the “belief disconfirmation route”--occurs only after one has accepted the validity of the disconfirming information (see Mandelbaum 2019 for discussion). This situation—where one believes that p, receives evidence that not-p and then raises their beliefs in p—has been mostly ignored by the ‘polarization is rational’ contingent. Until this type of disconfirmation is modellable, it’s misleading to claim that polarization is rational. [↑](#footnote-ref-2)
3. Sometimes we speak as if there is a belief updating mechanism, in the same sense as there is a belief rejection mechanism. However, this is a bit misleading. As we see it there is a separate faculty for rejecting information but that isn’t quite meant to imply that there is a faculty for belief acceptance. Instead, we think that the default architecture of the mind is that every (e.g.,) sentence that is tokened in the language of thought is automatically accepted as acceptance is just the default state (in which case there is technically no specific mechanism for acceptance at all, which is why its effortless: it’s just the default state of the mind). [↑](#footnote-ref-3)
4. What we are calling ‘cynicism’ and ‘coldness’ are deeply linked, in that a rational agent should be able to update beliefs in the right sort of way and also do a good job at assessing the quality of information sources, and these abilities may not be very useful in isolation. However, ‘cynicism’ here is meant to pick out a kind of social disposition in reasoning that is stressed by several authors in the literature (see Section 2 below). [↑](#footnote-ref-4)
5. Mercier adds: “If, from a very early stage in their (pre)history, humans stood to gain enormous benefits by communicating with each other, they have also been at risk from the abuse of communication. More than any other primate species, we are in danger of being misled and manipulated by communication. The existence of an evolutionarily relevant problem creates selection pressures that favor the development of cognitive mechanisms dedicated to solving this problem. The same is true of communication, with its promises and its perils. Indeed, the stakes are so high that it would be puzzling if we hadn’t evolved specialized cognitive mechanisms to deal not only with the potential but also with the danger of communication” (Mercier, 2020, p. 30). [↑](#footnote-ref-5)
6. Williams borrows from economic theories of belief (or ‘belief’, depending on your druthers) and assumes that “that belief ‘choice’ is subject to a rationalization constraint, such that individuals can only bring themselves to believe things for which they can find appropriate rationalizations” (Williams 2022 p2). This is an extremely provocative claim that is taken on as a premise and offered without argument (which is particularly disappointing as the claim is intriguing and, we suspect, dubious. For example, one’s evidence for one’s political or religious beliefs often strikes us as: *shut up*). [↑](#footnote-ref-6)
7. The precise function of belief for different authors is different. Mercier, for example, takes the role to be one of reducing unreliable signaling. [↑](#footnote-ref-7)
8. To be clear, we are not doubting that when one is primed to be in a skeptical mindset, e.g., when one thinks they are in an informationally hostile environment, they will be focused and primed to reject incoming information (see, e.g., Mayo 2019). [↑](#footnote-ref-8)
9. Even if the guessing explanation were true, it would not greatly help the case of epistemic vigilance. If their point is that people are critically assessing all incoming testimony, a near-blind guessing strategy with a bias toward assenting to what others say is hardly preferable to actual gullibility. [↑](#footnote-ref-9)
10. Mercier and Sperber (2009, 2018) claim their architecture quantifies over “information,” rather than beliefs structured in a LoT (cf. Sperber 1997). But their proposal quantifies over mental representations that have most of the hallmarks of beliefs and, breaking with the spirit of massive modularity, at least one of its “modules” is completely domain general. Sperber and Mercier posit what they call the *argumentative* (2009) or (ironically-named) *reason* (2018) module, which operates over representations of arguments: it builds post hoc rationalizations for one’s prior beliefs. It also breaks down representations of arguments into premises and conclusions, such that they can be individually processed. This mechanism can operate over arguments from any content domain and is thus domain general.

Moreover, whenever we are talking about modularity, the normal questions arise (Mandelbaum 2015; 2017; 2019b): what are the inputs to the module, what are the outputs, and, most importantly, what specific computations does the module run? Mercier and Sperber court an even more specific question: why isn’t their system a LoT system? The putative reason module seems to operate over pieces of information that feature in inferences and have a logical or at least probabilistic structure. It is sensitive to logical-inferential relations between pieces of information (such as what would make a good set of premises for a given conclusion). It operates over all manner of content domains. In short, it seems exactly the kind of system that would rely on a LoT (see Quilty-Dunn et al. 2023 for discussion).

To our knowledge this massively modular position is not explicitly endorsed by proponents of signaling or marketplaces, but one can easily imagine pairing such a view with them. Signaling itself could have emerged through modular additions to an already massively modular mind. Moreover marketplaces—whether economic or biological—are ripe places for cheating (Sun 2023). For example, brood parasitic birds (like the jerk Cowbird) will leave their eggs in unsuspecting birds’ nests to have other birds do the hard work of child rearing. The flagship case for massive modularity is the success of cheater detection modules, modules that owe their existence to the evolutionary pressures of dealing with deception like brood parasitism, so one could see how the tools of evolutionary psychology could help marketplace theorists solve the looming problems of cheaters and free riders. (Note another way these views may interact: there may be free riding from people signaling that they believe p—e.g., your religious leader who is actually an atheist—just to get status. But of course this situation—where one signals p but believes not-p—isn’t even possible on Williams style views where the role of belief is just tied to signaling.) [↑](#footnote-ref-10)
11. The two views needn’t be held together—communication mostly happens in the mind, after all, and a detailed account of such processes should be part of a psychofunctional treatment of the mind—but they are natural bedfellows if one focuses on the public signals of communication. [↑](#footnote-ref-11)
12. The view under discussion thus suggests that rejection is endemic to humans, perhaps a defining feature of what it is to be human. But on the vigilance view it seems that *belief* in general is endemic to humans, a rather extreme thesis though not an unprecedented one, e.g., (Davidson 1982). [↑](#footnote-ref-12)
13. Sometimes referred to as ‘the Manhattan Project’ as it is a series of theses loosely affiliated with people who are or were working in the NYC area. [↑](#footnote-ref-13)
14. I.e., swapping the algorithm-based feed order for a chronologically ordered one. [↑](#footnote-ref-14)
15. For Germany see, Muller and Schwarz (2021). For Brazil:

https://www.nytimes.com/2018/10/19/technology/whatsapp-brazil-presidential-election.html

For Myanmar: <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>

For India: <https://www.nytimes.com/interactive/2018/07/18/technology/whatsapp-india-killings.html>

For Israel: https://www.nytimes.com/2023/10/11/world/middleeast/israel-attacks-video-disinformation.html?smid=nytcore-ios-share&referringSource=articleShare [↑](#footnote-ref-15)
16. While access to smartphones and cell phone towers correlate with income and proximity to cities, respectively, Kiriwin et al. also note that anti-French sentiment in the region cuts across social class and region. [↑](#footnote-ref-16)
17. This is a particularly pressing problem for views like Williams’. Williams view relies heavily on analogies from economics. Those theories often have questionable applicability to actual individual human behavior. But more to the point, in Williams' hands the view is interesting, but impressionistic with little attempt made to tie it to any of the results in the cognitive science of belief. [↑](#footnote-ref-17)
18. See also van Leeuwen (2014), who takes this failure to drive action to be paradigmatic of religious belief, which he also considers to be a distinct attitude. As we see it, the main problem for Van Leeuwen is that putative factual beliefs also are recalcitrant (see, e.g., Mandelbaum 2019; Porot & Mandelbaum (2020) fn21). [↑](#footnote-ref-18)
19. How much separation is possible between the folk and scientific kinds before it no longer counts as an extension of belief? In other words, how much similarity must the two states share to make the theoretical identification? We think such a question is a verbal debate. Instead, what matters is that the lack of overlap matches the actual variance of behavior explained by the requisite states. So, for instance, folk belief is really good at predicting behavior, but still gets it wrong in many cases, and in others doesn’t offer any judgment at all (e.g., you just broke up with your partner--will you be more likely to be angry with them, or pity them?). The cognitive scientific notion of belief should, in the fullness of time, explain a greater variance than the folk kind--that’s part of what science is for. The overarching goal is to have enough overlap between the folk and scientific kind to explain how people get it right, but also enough slippage to explain why folk psychology often gets it wrong or fails to render a judgment. [↑](#footnote-ref-19)
20. This problem isn’t specific to Mercier, but is a totally general point. The same morals apply to alief (see Mandelbaum 2013) and other belief-like ‘cognitive attitudes’ (van Leeuwen 2014; see fn 20). [↑](#footnote-ref-20)
21. To be sure, the market is no panacea. Often less-than-ideal products win out either because they are promoted by celebrities (this handsome twit loves these peanuts, so you will too!), or have the shiniest packaging (5e--now with an extra e!), or are the thirstiest on social media. For any number of reasons, bad products and ideas appeal to less-than-ideally-rational people. [↑](#footnote-ref-21)