Bad Beliefs: Automaticity, Arationality, and Intervention

Stephen Gadsby

Centre for Philosophical Psychology, Department of Philosophy, University of Antwerp, Antwerp, Belgium.

gadsby.st@gmail.com

ORCID: 0000-0002-2302-844X

Abstract

Levy (2021) argues that bad beliefs predominately stem from automatic (albeit rational) updating in response to testimonial evidence. To counteract such beliefs, then, we should focus on ridding our epistemic environments of misleading testimony. This paper responds as follows. First, I argue that the suite of automatic processes related to bad beliefs extends well beyond the deference-based processes that Levy identifies. Second, I push back against Levy's claim that bad beliefs stem from wholly rational processes, suggesting that, in many cases, such processes are better characterised as arational. Finally, I note that Levy is too quick to dismiss the role that individuals can play in cleaning up their own epistemic environments, and I suggest one route through which this is possible.

Introduction

Bad beliefs conflict with the beliefs of experts and are held onto despite publicly available evidence to the contrary. Prime examples are the belief that human-driven climate change is not occurring or that vaccines are ineffective or unsafe. Various explanations have been proposed for why people hold such beliefs, many of which refer to motivated cognition (Williams, 2021) or reasoning deficits (Pennycook & Rand, 2019). Levy (2021), however, offers a novel and persuasive explanation for bad beliefs that refers to neither of these factors.

Levy's core claims are as follows. First, knowledge generation is largely outsourced—"For much of what we know about the world, we are deeply dependent on others" (p. 50). We hold so many true beliefs about complex matters because we adopted those beliefs from our peers. Importantly, this is a feature not a bug. Evolution programmed us this way and, all things considered, it's a good way to generate knowledge.

Second, this outsourcing is largely facilitated through *deference*. We defer by updating our beliefs in response to testimonial evidence provided by trusted sources.[1] Such evidence can be first-order (bearing directly on the truth of a proposition) or higher-order (bearing on the reliability of first-order evidence) and it can also be either explicit or implicit. To illustrate, explicit first-order evidence would be the statement "Climate change is not happening", while explicit higher-order

---

[1] At various points, Levy argues that *social referencing*—"looking to others within our social group for cues for what to believe" (p. 48)—also plays an important role. It seems to me, however, that social referencing (defined in this way) merely precedes deference; it allows us to identify the appropriate testimonial evidence to update off. Consequently, I will focus on the more essential stage of (deferential) updating.

evidence would be the statement "Your evidence about climate change should not be trusted". Implicit first-order evidence would involve one going about their life with no apparent concern for climate change. Implicit higher-order evidence would involve becoming increasingly skeptical when climate change evidence is discussed. While deference occurs in response to all four forms of evidence, Levy emphasises the overlooked role that implicit and higher-order evidence play in driving belief.

Deference can be slow and deliberate. For example, we can individually deliberate about complex states of affairs in a way that is heavily weighted towards testimonial evidence. This qualifies as deference, but not of the kind Levy has in mind. Rather, he emphasises deference that is "ingrained and fluent" (p. 83), "smooth and automatic" (p. 133), and which converts evidence to belief "ubiquitously and … routinely" (p. 62). Indeed, it is precisely because of deference's smooth and automatic nature that we often don't notice it. Consequently, "we fail to notice our beliefs are dependent on what others believe, and shift as theirs do" (p. 62).[2] Though deference is automatic, it is nevertheless rational, because it involves responding appropriately to epistemically valuable evidence (see section 2). This leads to Levy's third claim: bad beliefs are, in most cases, rational.

Levy's fourth claim is that because automatic deference to the testimonial evidence present in our epistemic environments is rational, pollution of such environments is to blame for many bad beliefs. If we wish to combat bad beliefs, then, we should focus on cleaning up epistemic pollution. Teaching people how to reason better about complex matters will not succeed because individual reasoning isn't how we form beliefs about such matters anyway (chapter 4). Teaching people to better pick better sources also won't help, as it is notoriously difficult to distinguish genuine experts from those who pose as such (chapter 5). Instead, Levy discusses ways to improve epistemic pollution at the level of policy and collective action (Chapters 5 and 6).

My reply is as follows. First, I argue that deference is just one of a broader array of smooth and automatic belief updating processes. If we're hoping to understand bad beliefs, we ought to focus on this broader category. Second, contrary to Levy's claim that bad beliefs are rational, I suggest that, in many cases, they are better understood as arational. This is because the automatic updating processes that underpin bad beliefs are often arational, and the processes of environmental structuring probably are also. Finally, I show that Levy overlooks a promising route through which people might be taught to improve their own epistemic environments.

1. Automaticity

Consider two kinds of deference. In *wholesale* deference, one completely adopts a new belief (or changes an old belief) based on a single exposure to testimonial evidence. Wholesale deference does occur and may explain some cases of bad belief. For example, I have dinner with a friend who is a geologist, and they tell me that global warming is not caused by human activity. Because I view him as a trusted expert, I change my belief about the cause of global warming. This is a case of wholesale deference, but not the kind Levy has in mind. First, wholesale shifts in belief like this are relatively rare (humans are remarkably stubborn) (Mercier, 2020). Second, wholesale

---

[2] The smooth and automatic nature of deference isn't only endorsed by Levy, it is essential to his argument. Levy claims that deference facilitates the outsourcing of knowledge generation. Outsourcing functions to reduce the complexity of individual cognising. However, carefully reasoning about matters in a way that heavily weights testimonial evidence doesn't facilitate outsourcing because it doesn't reduce the amount of cognitive work involved.

deference is not so smooth and automatic that we lose track of how we formed the belief. If someone asked me why I believe that global warming is part of a natural cycle, I can generally report that my friend the geologist told me.

The kind of deference Levy has in mind is *subtle*, involving small shifts in credence. Unlike wholesale deference, subtle deference is smooth and automatic and therefore can occur without us realising. This subtle nature doesn't undermine its importance for understanding bad beliefs. Because we are consistently bombarded with evidence (first-order and higher-order, implicit and explicit), the effect of subtle deference compounds—slowly but surely nudging our beliefs in the direction of our peers. Through small, indiscernible changes, those of us in polluted epistemic environments form bad beliefs.

In fact, there are several subtle, automatic belief-updating effects. For example, in the illusory truth effect, prior exposure to statements causes people to rate them as more truthful (Hasher, Goldstein, & Toppino, 1977). Though subtle, the effect has been shown to increase the perceived accuracy of false statements (Fazio, Brashier, Payne, & Marsh, 2015), implausible statements (Fazio, Rand, & Pennycook, 2019; Lacassagne, Béna, & Corneille, 2022), and conspiracy theories (Béna, Rihet, Carreras, & Terrier, 2022).

A similar effect was demonstrated by Gilbert and colleagues (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). They presented participants with different statements along with feedback regarding the truth of the statements. Participants who were under cognitive load updated their beliefs in response to statements that they were informed were false.

The above two cases involve automatic updating to testimonial evidence, but they are importantly different to deference. While deference is often automatic, we only to defer to trusted sources (Dutilh Novaes, forthcoming). This is why, for example, those from either side of the U.S political divide don't defer to each other's testimony and why conspiracy theorists don't defer to genuine experts (Levy, 2019). Levy might argue that, in these cases, the experimenters themselves are the trusted sources. But if this were so, then participants should respond to these sources' higher-order evidence about the trustworthiness of the relevant first-order evidence. In Gilbert and colleagues' studies, participants in the cognitive load trial do not take into account evidence that the statements are false. According to Levy, being under cognitive load should cause one to defer to trusted sources more strongly (p. 146), so the absence of updating to discounting evidence suggests that this is not a case of deference.

The illusory truth effect also persists when the facts are labelled as contested by fact checkers or are inconsistent with one's political ideology (Pennycook, Cannon, & Rand, 2018). If the effect involved deference, then inconsistency with one's ideology should negate it, as this signals that the source is an (untrustworthy) out-group member. Such effects are consistent with other studies showing that even when experimenters present testimony about the untrustworthiness of the first-order evidence involved in their studies, participants update in response to such evidence anyway (Wegner, Coulton, & Wenzlaff, 1985). In sum, genuine deference should be sensitive to source trustworthiness, but many automatic updating effects are not. So while our beliefs do automatically update in response to testimonial evidence, this often does not involve deference.

Consider another automatic updating effect, related to advocacy goals. Melnikoff & Strohminger (2020) assigned participants to either prosecute or defend a hypothetical client, Harris, charged with embezzling funds. Participants assigned to defend Harris adopted more positive beliefs regarding his innocence and moral character, compared with those assigned to prosecute him.

Participants' beliefs about whether one's behaviour could indicate their "true self" also changed, depending on their advocacy goals: prosecutors who discovered Harris was innocent exhibited lower conviction in a true-self, compared to those who discovered he was guilty.

This research suggests that when advocating for a position, we shift our beliefs to support that goal. We adopt factual beliefs that support our position, and even shift core beliefs, such as those related to the possibility of true-selves. Though subtle, these shifts in belief persist in the face of accuracy incentives and prompts to be wary of biasing oneself. Strohminger & Melnikoff (2022) further showed that advocacy goals shifted beliefs about the truth of far-fetched "crackpot" theories and claims that were inconsistent with strong counter-evidence provided to participants.

The advocacy effect seems even less like deference. In these experiments, participants are not updating in response to testimonial evidence from peers. In fact, fascinatingly, they aren't updating in response to evidence whatsoever. Differences in belief were evident despite participants not actually performing the advocacy, but merely preparing themselves to do so. This rules out the possibility that participants were responding to a kind of (self-generated) evidence, as would occur in actual court rooms when prosecutors and defenders argue their case.[3]

In sum, while some cases of subtle automatic updating may involve deference, not all do. Sometimes automatic updating occurs in response to first-order evidence that is not delivered by trusted in-group members. Other times automatic updating occurs in response to no evidence at all, simply by adopting advocacy goals. To fully understand the role of epistemic environments in driving bad beliefs, then, we should look to the broader category of subtle, automatic updating effects.

2. Arationality

Levy argues that bad beliefs that stem from automatic updating are "the product of genuinely and wholly rational processes" (p. xii). Specifically, he focuses on *direct* rationality, which is met when one updates their beliefs in response to sufficiently strong evidence. Deference could, of course, be directly irrational, if belief updating was too strong, too weak, or occurred in the wrong direction.[4] Levy argues, however, that it is not: the extent and direction of deferential updating is justified by the testimonial evidence that triggers it. There is, however, a third way of characterising automatic belief updating, namely, as (directly) arational. This label is applicable when the relevant processes are, by design, entirely unresponsive to evidence. Consider the advocacy effect. Advocacy goals shift belief without exposing individuals to evidence of any form, and exposure to evidence does not modulate the effect. Because it is irresponsive to evidence, it is neither (directly) rational nor irrational, but arational. If, like the advocacy effect, the kinds of automatic updating processes that underpin bad belief are arational, then bad beliefs may in fact be the product of genuinely and wholly arational processes.

---

[3] Another possibility is that, in preparing to advocate for or against the client, participants generated arguments to prepare, and thinking of such arguments acted as a form of self-persuasion. However, the act of defending or prosecuting Harris did not involve arguments whatsoever. Instead, participants simply prepared to play a game called "attorney at law" which involved quickly pressing the space bar when the word GUILTY (for defenders) or INNOCENT (for prosecutors) appeared on the screen. Given the nature of this game, it is unlikely participants would have generated arguments for their position.

[4] Some suggest that belief updating to testimonial evidence is irrational in this sense, though such results are open to debate (Tappin & Gadsby, 2019).

Consider again the studies by Gilbert and colleagues. There is a way to interpret such effects in terms of direct rationality. By updating their beliefs in response to the statements, participants are deferring to a kind of testimonial evidence, and they even update in response to the higher-order testimony about the accuracy of the statements, except for when this response is disrupted by cognitive load. But—and here is the important point—to truly assess whether a process is rational, we need to consider how it operates in various contexts. For example, a dominant interpretation of Gilbert and colleagues' results is that participants automatically update their beliefs in response to simply entertaining the statements (Mandelbaum, 2014). If this is accurate, then the effect would occur regardless of whether such statements constituted evidence or not. If a random word generator produced the relevant statements, for example, the effect would still hold. Of course, a process that is responsive to statements regardless of whether they are evidence or simply random words is not responding to evidence, but language. Consequently, such a process is neither (directly) rational nor irrational, but arational (see also: Levy & Mandelbaum, 2014).

This point generalises for Levy's arguments for the rationality of deference-based updating in the case of nudges (Chapter 6). He argues that responses to nudges are rational because nudges supply implicit testimonial evidence. One example is setting the default option for retirement contribution on employee contracts. Levy argues that listing an option as the default provides a kind of testimonial evidence, namely, an implicit recommendation in favour of that option (p. 143). Levy justifies this by asking the reader to imagine a case where the default was 98% of the salary: "Wouldn't you think whoever drew up the contract was incompetent?" (p. 143). One might, however, think that the role of contract defaults is to save peoples' time by selecting the option that they would most likely pick. This would entail that anyone who selected 98% as a default was incompetent, but not because they gave a bad recommendation.

A better counterfactual scenario—and one that could be empirically tested—is one where people are informed that the default was picked by random. If the nudge loses its power, then this suggests they were responding to the default option as a kind of evidence, if not, they weren't responding to evidence after all (see also: Sullivan-Bissett, 2022). In sum, to show that cases of automatic updating are rational, rather than arational, Levy must do more than show that people are responding to evidence: he must show that they are responding to it as evidence. While I have discussed just one of Levy's examples, the broader point is that it is an open empirical question whether all cases of apparent deference to testimony are genuine cases of evidence-responsiveness, or simply arational updating.

Putting aside automatic updating, Levy's emphasis on epistemic environments brings up an interesting point. If we automatically update in response to the evidence in our environments, then perhaps whether our beliefs are "the product of genuinely and wholly rational processes" should depend on how we structure those environments. Of course, narrow views like that of direct rationality don't apply to how one structures their epistemic environment. But there is reason to think that epistemic norms shouldn't be restricted to belief updating in response to evidence (Flores & Woodard, 2022) and Levy's illustration of the importance of epistemic environments chimes with this. If beliefs could be rational or irrational in virtue of the environmental structuring that spawned them, then even beliefs stemming from arational updating could be rational or irrational.

What would it mean to structure one's environment irrationality? Political history offers a few examples. Lyndon Johnson was well known for firing anyone who disagreed with him. As a result, "fewer and fewer people who had Johnson's ear told him the truth as they saw it", and Johnson dug his heels in on various bad (political) beliefs (Caro, 2012, p. 83). In a similar vein,

George Bush refused to meet with anyone expressing dissenting views about the Iraq wars' potential for success (including experts with firsthand experience of the war). In doing so, he insulated himself from the truth and developed an unjustifiably rosy picture of the war's progress (McClellan, 2008, p. 253). Both Johnson and Bush structured their environments in ways that led to bad beliefs and that kind of behaviour seems irrational. The label seems apt, however, because they ought to have been doing otherwise. That is, given the kinds of projects they were involved in (deciding which policies to implement), they ought to have been constructing their environments in ways that were maximally conducive to truth tracking. Consequently, the charge of irrationality applies to the way in which they structured their environments. At a minimum, this suggests that bad beliefs might be irrational even if the belief updating that underpins them is not.

Shifting away from politicians, do people, by and large, structure their epistemic environment in rational or irrational ways? I suspect the answer is neither. People don't structure their epistemic environments in ways conducive to truth, but it is not clear that they ought to. Why do I hang out with one group of people at the pub on Friday evenings, rather than a different group? Because of the jokes they tell, a shared past that we can reflect on, the trust that everyone will buy a round when it is their turn, and so on. Structuring my epistemic environments according to these principles isn't conducive to truth-tracking—as Nguyen puts it, "Friends make for good parties, but poor information networks" (2020, p. 144)—nevertheless, it doesn't seem as if it ought be. That isn't the goal I ought to be pursuing on a Friday night. Consequently, how I structure my epistemic environment on Friday nights is neither rational nor irrational, but arational.

If belief updating is as automatic as Levy, and myself, have characterised it then who we spend time with at the pub will affect our beliefs, but this doesn't make those beliefs any more or less rational. Of course, I haven't offered an account of rationality as it extends to cases of structuring one's environments and, consequently, I can't offer much more than intuition for why my decision to spend time with my friends at the pub is arational, while Lyndon Johnson's decision to surround himself with yes-men is irrational. Nevertheless, it is plausible that notions of (epistemic) rationality should only apply in contexts where epistemic goals are relevant; deciding political policies is one such context, drinking beer is not. This returns us to the claim that many beliefs that are automatically formed off testimonial evidence are simply arational.

3. Intervention

Levy points to epistemic pollution as a primary cause of bad beliefs, and, as such, the best focus-point for intervention. While we must clean-up epistemic pollution, this is a collective action problem—"individuals can't make a significant difference to it by themselves" (p. 130). Consequently, combating epistemic pollution is a task for governments and institutions.

While structural reforms to reduce epistemic pollution are important, Levy overlooks the role that individuals can play in engineering their own epistemic environments. When newspapers and television were the main sources of information about the world, there were only so many sources to choose from and people were restricted in how they could tailor their epistemic environments. Today, epistemic environments are predominately digital, and we therefore have remarkable capacities for customisation. Of course, much of the information we are exposed to is determined by algorithms, whose inner workings we do not have access to (Pariser, 2011). Nevertheless, those algorithms are fed by our behaviour: we tell them what we want to see, through our clicks, likes, and shares. So while customising our epistemic environments may not be straightforward, it is nevertheless possible.

Of course, it doesn't particularly matter how much control we have over our epistemic environments. To combat pollution, we must recognise it as such. Levy argues that recognising that one occupies a polluted epistemic environment is too difficult, as it requires distinguishing between clever frauds and genuine experts. In many cases, however, bad beliefs stem from epistemic environments in which it is easy to recognise their compromised nature. Specifically, they emerge from communities where the communicative norms are such that they encourage epistemic pollution. While it may be difficult to distinguish experts from frauds, it is often straightforward to identify the norms of your community and assess whether those norms promote or reduce epistemic pollution.

4chan is an excellent example of an epistemically polluted environment. It was, at one point, one of the most popular sites on the internet and originated some paradigmatic examples of bad beliefs, such as the pizzagate conspiracy theory along with various ideological beliefs associated with white nationalism. When you look to the history of 4chan and other epistemically polluted online environments, they often didn't start that way (Koebler, 2016). Over time, the norms of the community shifted, in recognisable ways, which in turn led to epistemic pollution. 4chan, for example, started as a place to talk about Japanese anime and was a good place to gain knowledge about such matters. Over time, however, the norms of the community changed. This was encouraged by the board's design. The enforced anonymity of users, automatic deletion of posts, and the way in which attention-grabbing content was visually prioritised all contributed to norms of communication that undermined truth (Beran, 2019; Nagle, 2017; Wendling, 2018):

> Users weren't debating topics to reach some sort of shared understanding or consensus. They were elaborating on jokes, sharing files, or generating … inverted discussions where the point was not communication, but the performance of sliced-up gibberish … (Wendling, 2018, pp. 51-52)

The community began to prize, above all else, attention-drawing content. The content that attracts the most attention is that which involves shocking and outlandish claims or breaks social taboos, such as racist content. In this way, competing with one another in terms of how offensive and outlandish one could be became a core feature of the communities' identity, leading to the proliferation of hate speech and conspiracy theories—content which then filtered out to other internet communities (Beran, 2019). The proliferation of such content constitutes a kind of epistemic pollution, though (mostly) involving first-order pollution (i.e. misinformation about facts), rather than the kind of pollution that Levy focuses on (evidence that undermines trust).

Of course, not all 4chan users believed the extreme content that was posted on the board. However, as noted, engaging and identifying with a community where such content is common can lead to bad beliefs through various routes. One such route is both deferential and rational, wherein one takes other 4channers to be their trustworthy peers, takes their testimonial evidence regarding race and conspiracies theories as genuine, and updates appropriately. One might suggest that such a route is implausible, as 4channers were aware that the claims made on the board weren't made in earnest. However, as Wendling (2018, p. 11) notes, "it's often not possible to tell whether [those] slinging racial and gay slurs on 4chan actually hold extreme beliefs, or think that shouting "nigger" and "faggot" is funny, or are just trying to shock or offend for effect".

In fact, one would be right to assume that many of the board's extreme posts were genuinely endorsed by those who posted them, as many 4chan regulars were in fact white supremacists and

radical conspiracy theorists (Wendling, 2018). As an illustration of the more earnest racist content on the board, consider the following post:

> Lets try to rationalize why certain discrepancies exist /pol/ starting with blacks. Why do blacks have such low iq's compared to whites? I think it's because life isn't able to flourish in Africa, therefor the generations that stay will evolve into low intelligence people, why? because realizing your environment is shit for development is a sign of intelligence. (Wendling, 2018, p. 56)

This post offers testimonial evidence related to the topic of the IQ differences amongst race, some of which is first-order evidence, bearing directly on the truth of the relevant claim (IQ differences). Such a post may also be associated with various forms of higher-order evidence, for example, that the post was socially approved by fellow board members, by being engaged with or persisting on the list of posts (Levy, 2021, p. 84). If one was predisposed to defer to evidence from fellow 4channers, then one would shift their beliefs about the topic of differences in intelligence between races.

Another route through which such content could alter belief involves automatic, non-deferential updating. Just as people increase their confidence in implausible statements and conspiracy theories when exposed to them in experiments, the same effect likely occurs when exposure is through a message board.

Finally, one may be pushed towards bad beliefs by identifying with and advocating for a community who espouses such beliefs. The advocacy effect shows that people shift their factual beliefs in ways that aid their advocacy goals. Believing in the legitimacy of claims about racially based differences in intelligence facilitates advocating for communities that trades in such claims. By advocating for a community that trades in extreme claims (even for the sake of humour), one may sway themselves towards believing those claims.

It is important to note that for 4chan, and similar highly polluted epistemic environments, the pernicious norms of the community are often recognisable. The norms of 4chan, for example, were codified into the list of the boards' conventions, titled "the rules of the internet". The list included, "There are no rules about posting …", "anything you can say can be ignored", "All your carefully picked arguments can easily be ignored". In fact, the board even featured a warning at the top of its page, reading "The stories and information posted here are artistic works of fiction and falsehood … Only a fool would take anything posted here as fact". While discounting cues such as these may not be effective in stopping users from automatically updating in response to evidence from the board, they do serve as indicators that the community promotes rather than restricts epistemic pollution.

Just as we have increased awareness about the personal harms associated with environmental pollution, we can increase awareness about epistemic harms associated with epistemic pollution. If people were aware of such harms, then they may look to the social norms of their community to assess whether it is conducive to epistemic pollution. 4chan is an extreme example, but this point generalises. It may be difficult (or pointless) to teach people to reason better or to distinguish experts from frauds, but the preceding discussion suggests an easier and more effective route. The first step is teaching people about the pernicious effects of epistemic pollution. This does not seem difficult. As noted, people learnt about the harms of environmental pollution and are generally responsive to learning facts about their own psychology. The second step is teaching people to recognise the communicative norms of their communities and whether those encourage or discourage epistemic pollution. If people were

aware of the pernicious epistemic consequences of engaging in polluted environments and how to recognise that the norms of their community are of a kind that give rise to epistemic pollution, then they may take steps to restructure their environments and therefore counter their own bad beliefs.

Conclusion

Levy outlines an important route through which epistemic pollution can spawn bad belief. I responded to his argument with three points. First, the suite of automatic updating processes that contribute to bad beliefs is much broader than those involving deference-like updating. Second, while Levy argues for the rationality of bad beliefs, I suspect that, in many cases, they are better understood as arational. Finally, Levy claims that his book "challenges the view that we ought to step back and carefully deliberate about important issues" (Levy, 2022, p. 2). In contrast, I suggest we ought to do so even more carefully. Perhaps we needn't deliberate about first-order evidence directly, or about who the real experts are, but about the epistemic environments that we construct for ourselves. A worthy philosophical contribution to this task involves developing epistemic virtues for the digital age (Frost-Arnold, 2023; Heersmink, 2018), focusing on, for example, identifying the communicative norms of online communities.

Acknowledgments

References

Béna, J., Rihet, M., Carreras, O., & Terrier, P. (2022). Repetition could increase the perceived truth of conspiracy theories.

Beran, D. (2019). *It came from something awful: how a toxic troll army accidentally memed Donald Trump into office*: All Points Books.

Caro, R. A. (2012). *The Passage of Power: The Years of Lyndon Johnson IV* (Vol. 4): Vintage.

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.

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.

Flores, C., & Woodard, E. (2022). Epistemic Norms on Evidence-Gathering. *Philosophical Studies*.

Frost-Arnold, K. (2023). *Who Should We Be Online?: A Social Epistemology for the Internet*: Oxford University Press.

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.

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*(2), 221.

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.

Heersmink, R. (2018). A virtue epistemology of the Internet: Search engines, intellectual virtues and education. *Social Epistemology, 32*(1), 1-12.

Koebler, J. (2016). How r/the_donald became a melting pot of frustration and hate. *Vice, July*.

Lacassagne, D., Béna, J., & Corneille, O. (2022). Is Earth a perfect square? Repetition increases the perceived truth of highly implausible statements. *Cognition, 223*, 105052.

Levy, N. (2019). Is conspiracy theorising irrational. *Social Epistemology Review and Reply Collective, 8*(10), 65-76.

Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford: Oxford University Press.

Levy, N. (2022). Bad beliefs–a precis. *Philosophical Psychology*, 1-6.

Levy, N., & Mandelbaum, E. (2014). The powers that bind: Doxastic voluntarism and epistemic obligation. In J. Matheson & R. Vitz (Eds.), *The Ethics of belief: Individual and social* (pp. 15-32).

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

McClellan, S. (2008). *What happened: Inside the Bush White House and Washington's culture of deception*: PublicAffairs.

Melnikoff, D. E., & Strohminger, N. (2020). The automatic influence of advocacy on lawyers and novices. *Nature Human Behaviour, 4*(12), 1258-1264.

Mercier, H. (2020). Not born yesterday. In *Not Born Yesterday*: Princeton University Press.

Nagle, A. (2017). *Kill all normies: The online culture wars from Tumblr and 4chan to the alt-right and Trump*: Zero Books.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme, 17*(2), 141-161.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*: penguin UK.

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. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39-50.

Strohminger, N., & Melnikoff, D. (2022). Breaking reality's constraints on motivated cognition.

Sullivan-Bissett, E. (2022). Believing badly ain't so bad. *Philosophical Psychology*, 1-9.

Tappin, B. M., & Gadsby, S. (2019). Biased belief in the Bayesian brain: A deeper look at the evidence. *Consciousness and cognition, 68*, 107-114.

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.

Wendling, M. (2018). *Alt-right: From 4chan to the White House*: Fernwood Publishing.

Williams, D. (2021). Socially adaptive belief. *Mind & Language, 36*(3), 333-354.