

# Evidence and Bias\*

Nick Hughes  
University of Oxford  
[nickhowellhughes@gmail.com](mailto:nickhowellhughes@gmail.com)

## §I. Evidentialism

The core claim of evidentialism is that one ought to believe what one's evidence supports:

EVIDENTIALISM: One ought to believe what one's evidence supports

Many epistemologists think this is so obviously true as to be almost platitudinous.<sup>1</sup> However, as it stands it doesn't tell us much. What is

---

\* This is a draft of a chapter for the forthcoming Routledge Handbook of the Philosophy of Evidence. Feel free to cite it but bear in mind that it might change. If you're going to quote it directly, it's probably a good idea to email me first.

<sup>1</sup> The canonical defence of evidentialism is Conee & Feldman (2004), who remark that they initially thought it was "...sufficiently obvious to be in little need of defense" (2004: 1). The observation that it has the air of a platitude is often made. See, for instance, Dougherty (2011: 6), and Piller (2016). The term 'evidentialism' is sometimes used specifically to describe Conee and Feldman's view. Here we use it more broadly.

evidence? When is evidence *your* evidence? How strongly must it support a proposition for belief to be the correct attitude to adopt? What is the evidential support relation? How should we interpret the ‘ought’ in question? What is its logical form? To whom does evidentialism apply? What is its modal profile? Does it only apply to belief or does it also apply to other doxastic attitudes? Which ones? A fully-fledged evidentialist theory must answer these questions and many more. Thus, EVIDENTIALISM is schematic. It is compatible with a wide variety of answers, each of which will give us a different evidentialist theory. This may explain its appeal. Perhaps part of the reason evidentialism is so popular is that, with so many flavours on offer, most epistemologists have been able to find at least one to their taste.

To see just how flexible EVIDENTIALISM is, consider two contrasting forms it could take. According to an externalist view we’ll call EXACTING EVIDENTIALISM, one’s evidence consists of all and only those propositions one knows to be true, and one may believe that *p* if and only if one’s evidence entails that *p*:

EXACTING EVIDENTIALISM: One’s evidence consists of all and only those propositions one knows to be true. One may believe that *p* if and only if one’s evidence entails that *p*

Since no falsehood is known and no set of truths entail a falsehood, one consequence of EXACTING EVIDENTIALISM is that one must never believe a falsehood. Many evidentialists will be unhappy with this consequence.<sup>2</sup> Evidentialists usually think that you ought to be rational, and that

---

<sup>2</sup> One exception is Williamson (2013, forthcoming) who embraces it. For criticism of Williamson, see Cohen & Comesana (2013, forthcoming)

rationality is a matter of believing in accordance with your evidence.<sup>3</sup> Since it is sometimes rational to believe falsehoods, they will say that this form of evidentialism is too demanding.

Consider instead, then, an evidentialism at the other end of the spectrum. According to an internalist view we'll call EASY-GOING EVIDENTIALISM, one's evidence consists of a subset of one's non-factive mental states (one's beliefs, phenomenal states, 'intellectual seemings', etc.), and one may believe that  $p$  if  $p$  is more probable on one's evidence than not- $p$ :

EASY-GOING EVIDENTIALISM: One's evidence consists of a subset of one's non-factive mental states. One may believe that  $p$  if  $p$  is more probable on one's evidence than not- $p$ .

EASY-GOING EVIDENTIALISM does not entail that one must never believe a falsehood. In fact, it is very undemanding indeed. Suppose you believe that the coin you are about to flip is biased 51% towards Heads, and none of your other non-factive mental states conflicts with this belief. EASY-GOING EVIDENTIALISM says that you may believe that the coin will land heads even before flipping it and seeing the result. Many evidentialists will be unhappy with this consequence. Intuitively, rationality requires you to suspend judgement on how the coin will land until you see the result. For this reason, they will say, EASY-GOING EVIDENTIALISM is not demanding enough.

There is a vast gulf between these two forms evidentialism could take.<sup>4</sup> Most epistemologists, whether they are internalists or externalists, would

---

<sup>3</sup> We will take on-board the assumption that you ought to be rational here. Williamson (2017) distinguishes between two notions of rationality and argues that on only one of them is rationality a matter of believing in accordance with you evidence.

<sup>4</sup> Equally clearly, neither is a fully-fledged theory, but this is unimportant for our purposes.

reject both and go for something in-between. But how should we narrow down the options? In this chapter, we'll look at how empirical work on bias in cognitive and social psychology can help us to do so, and, ultimately, whether this work motivates rejecting evidentialism altogether. As we will see, the existence of biases has far-reaching implications for epistemology, though not for the reasons one might have expected.

## §II. Bias

One often-touted virtue of evidentialism is that it provides a principle for belief-formation that we are, despite all our flaws, capable of following, whilst at the same time being suitably connected to the epistemic *telos* of truth.<sup>5</sup> Is that right? Empirical work on bias gives us a reason to be sceptical about the first claim. Over the last fifty years, cognitive and social psychologists have repeatedly put to the test Aristotle's definition of man as a rational animal. Many have viewed the results as a blow to our self-satisfied self-image. According to the dominant 'irrationalist' narrative, empirical work shows that we are prone to a range of irrational biases in reasoning and doxastic attitude formation.<sup>6</sup> Since a comprehensive overview of this large body of research is impossible, we will focus here on a handful of biases that should be of special interest to epistemologists, and to evidentialists in particular.

---

<sup>5</sup> Conee & Feldman (2004), *et al.* It is sometimes argued that we should reject more demanding forms of evidentialism like EXACTING EVIDENTIALISM because we are not capable of following them. See Srinivasan (2015), Hughes (2019a) and Lasonen-Aarnio (2020) critical discussion of this idea.

<sup>6</sup> For a good overview, see Gilovich *et al.* (2002)

## §2.1. Hindsight Bias

First up: hindsight bias. This is a phenomenon whereby people who know the outcome of an event judge it to be more probable on the evidence available before the outcome than do people who are ignorant of the outcome. Neal Roese and Kathleen Vohs (2012) offer an example: “A voter might believe that after accepting the Democratic nomination for president in August 2008, Barack Obama’s chances of winning the U.S. presidency was about 60%. After Obama’s victory in November 2008, this same voter might look back, see the victory as more predictable than it was before the outcome was known, and conclude that Obama’s chances were at least 80% at the time of the convention”. Hundreds of studies have confirmed that people exhibit hindsight bias.<sup>7</sup> It affects judgements about topics as diverse as terrorist attacks, medical diagnoses, and accounting and auditing decisions.<sup>8</sup> It has been documented in people from a variety of cultures, and across age-groups.<sup>9</sup> It is widely regarded by psychologists as irrational.

## §2.2. Biased Assimilation

Next: biased assimilation. As Charles Lord *et al.* (1979) describe it, this is a tendency people have to “accept confirming evidence at face value whilst subjecting disconfirming evidence to critical evaluation, with the result that they draw undue support for their initial positions”. An example: Scott Plous (1991) selected two groups of experimental participants. The first group strongly supported the use of nuclear energy, the second group strongly opposed it. Both groups were given the same literature on the

---

<sup>7</sup> Roese & Vohs (2012)

<sup>8</sup> For terrorist attacks, see Fischhoff *et al.* (2005). For medical diagnoses, see Arkes *et al.* (1988). For auditing decisions, see Peecher & Piercey (2010)

<sup>9</sup> RF. Pohl *et al.* (2002), Bernstein *et al.* (2011).

risks and rewards of nuclear energy. Each group drew support for their pre-existing views from the literature – the pro-nuclear-energy group judged that on balance the arguments within the literature supported its use, the anti-nuclear-energy group drew the opposite conclusion. Like hindsight bias, biased assimilation is robustly documented and widely regarded by psychologists as irrational. Biased assimilation is often thought of as one part of a broader phenomenon known as ‘confirmation bias’.

### §2.3. Selective Exposure

Raymond Nickerson describes confirmation bias as “...the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (1998: 175). The ‘interpreting’ part is biased assimilation. The remainder – the seeking – is known as ‘selective exposure’. The observation that we prefer to gather evidence that confirms our pre-existing beliefs and avoid evidence that disconfirms them is not new. In 1620 Francis Bacon wrote that “The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it”. (2008). A retro example: Sweeney and Gruber (1984) found that relative to Undecideds and supporters of George McGovern, Richard Nixon’s supporters reported less interest in, and paid less attention to, information related to the Watergate scandal. More recently, selective exposure has been invoked to explain the existence of online echo chambers.<sup>10</sup> As with the other biases we have looked at it is empirically well-confirmed and widely regarded as irrational.

---

<sup>10</sup> Quattrociocchi *et al.* (2016)

## §2.4. Implicit Bias

Finally: implicit bias. This is a multifaceted concept. One aspect of it is a phenomenon whereby people's judgements are unconsciously influenced by prejudices and stereotypes.<sup>11</sup> For example, Jennifer Saul notes that "...the same CV is considered much better when it has a typically white rather than typically black name, a typically Swedish rather than typically Arab name, a typically male rather than typically female name, and so on". (2013: 244) Unsurprisingly given its obvious ethical and political implications, implicit bias is one of the most studied and discussed human biases. Once again, beliefs formed as a result of the influence of the bias are widely taken to be irrational.

## §III. Introspection

These are just a handful of a large number of biases in reasoning and doxastic attitude formation that empirical work in cognitive and social psychology appears to have uncovered. Taken at face value, this work seems to show that we aren't nearly as good at believing what our evidence supports as many philosophers have assumed. Bad news for the Aristotelian image of man. Worse still, empirical work also suggests that we are usually quite clueless about our biases. Not because we're lazy, but, because most biases operate at a sub-personal level we have little to no introspective access to their influence on us.<sup>12</sup> As Timothy Wilson and Nancy Brekke put it, unlike bad food "Human judgements – even very bad ones – do not smell" (1994: 121).

---

<sup>11</sup> For a useful introduction to the philosophical dimensions of implicit bias, see Brownstein (2019)

<sup>12</sup> Wilson (2002)

You might think that you're an exception: others might not be able to recognise their biases, but you have a good nose, and your beliefs aren't off. This is predictable. Studies by Emily Pronin and colleagues suggest that whilst people are reasonably good at detecting bias in others, they are, by-and-large, blind to the influence of their own biases.<sup>13</sup> They hypothesize that this asymmetry can be explained by what they call 'introspection illusion' – the illusion that introspection is more reliable a method of gaining knowledge of one's thoughts, feelings, motives, intentions, and other aspects of one's mind than it actually is. Kristoffer Ahlstrom-Vij summarises the point nicely: "Since the processes that give rise to bias tend to operate on a sub-personal level, outside the scope of our introspective gaze, our search [for them] tends to come up empty. From the fact that our search comes up empty, we then infer an absence of bias – despite the fact that such a search is more or less *guaranteed* to come up empty, given the inaccessibility of the relevant mechanisms" (2013: 280).<sup>14</sup>

The fact that the influence of many of our biases is introspectively undetectable makes debiasing difficult. Ahlstrom-Vij identifies two hurdles. First, since people don't see themselves as biased, they are not motivated to engage in debiasing in the first place – if it ain't broke, why fix it? Second, even when people can be persuaded to engage in debiasing efforts there is a risk of both undercorrection and overcorrection. Studies on debiasing have shown that we are prone to both errors.<sup>15</sup> In general, the results of work on debiasing are disappointing. It is possible to reduce (if not eradicate) the influence of certain biases, but only with time and effort, and the results are usually only temporary. There is no silver bullet.

---

<sup>13</sup> Pronin (2009)

<sup>14</sup> Kornblith (2012) also makes this point

<sup>15</sup> Wilson (2002)

## §IV. Optimism?

According to the irrationalist narrative we are riddled with all manner of biases and there is little we can do about it. It is not difficult to find implications for evidentialism within this story. For one thing, it is hard to see how it can be squared with internalism. Ultimately, I think this is correct. Indeed, I think that unconscious biases have profound implications for the formulation of evidentialism and may even require us to give it up altogether. But before we see just why this is, it will be useful to first look at how some psychologists and philosophers have pushed back against the irrationalist narrative. Not everyone agrees that the situation is as bleak as irrationalists portray it to be.

One source of resistance points to the recent replication crisis in psychology. Many of the seminal studies which kick-started the irrationalist narrative have not been successfully replicated.<sup>16</sup> This casts some doubt on the narrative. Another does not question the studies themselves, but rather their interpretation. According to one school of thought, most closely associated with Gerd Gigerenzer and the ABC research group, many of our so-called biases should not be thought of as irrational, but rather ‘ecologically rational’ – rational for creatures like ourselves with limited time and computational powers, who cannot conform to the canons of ideal rationality.<sup>17</sup> In effect, the proposal is that the cognitive rules we employ that give rise to biases are the best rules for us to employ, given our limitations. We would do worse at achieving our epistemic goals (truth, knowledge, etc.) were we to attempt to use more

---

<sup>16</sup> Camerer *et al.* (2018).

<sup>17</sup> Gigerenzer *et al.* (2001). Bishop and Trout have put this idea to work in epistemology (Bishop 2000, Bishop & Trout 2004).

reliable rules.<sup>18</sup> This line of thought should be of interest to those engaged in non-ideal epistemology.

Finally, some epistemologists have questioned the philosophical assumptions made by irrationalists. In a recent article Brian Hedden (2019) has argued that, contrary to received wisdom, hindsight bias is often a perfectly rational response to one's evidence, and not (just) because it is ecologically rational – Hedden maintains that hindsight bias need not even constitute a violation of ideal rationality. Simplifying greatly, he argues that knowledge of an outcome often provides one with a certain kind of higher-order evidence: in particular, it provides one with evidence *about* the evidence that was available before the outcome. Hedden argues that correctly responding to this higher-order evidence will often lead one to display hindsight bias, and that, as a result, this 'bias' need not be in breach of evidentialism's imperative. In a similar vein, Thomas Kelly (2008) and Kevin Dorst (ms.) have used sophisticated theories of higher-order evidence to push back against the idea that confirmation bias is irrational.

It is worth also noting that some evidentialists will be unbothered by selective exposure in the first place. Earl Conee and Richard Feldman (2004, 2011) maintain that even if a person's evidence only favours believing that *p* because they avoided evidence against *p*, the resulting belief that *p* is still epistemically rational. This is because they regard norms of evidence-gathering as practical, rather than epistemic norms (Kelly agrees). Hence, they will deny that a propensity for selective exposure casts doubt on our ability to conform with evidentialism. Evidentialism, they will say, is a thesis about what you should believe given the evidence you actually have; not about what evidence you should have. Finally, Bernhard Salow (2017) has argued that it is *impossible* to

---

<sup>18</sup> Of course, even rules that take into account our limitations will sometimes be misapplied.

intentionally, knowingly, seek out confirming evidence and avoid disconfirming evidence. However, this is compatible with the possibility of unintentionally, unwittingly, doing so, as Salow himself acknowledges.

## §V. Possibility

None of Hedden, Kelly, or Dorst would claim that they have debunked the irrationalist narrative. They do not claim to have shown that every purported cognitive bias is, or could be, epistemically rational. Nor do they claim that the biases they discuss are necessarily rational; only that they can be under certain circumstances. Even if we accept Hedden's arguments, it may be that our actual hindsight judgements do not normally fit with the evidence because we systematically overestimate the import of the higher-order evidence. Similarly, even if we accept Kelly and Dorst's arguments it may be that our actual tendency to seek confirming evidence over disconfirming evidence is too pronounced, or that we give too little weight to disconfirming evidence and too much weight to confirming evidence, with the result that we violate the evidentialist imperative.

Nevertheless, this work does at least call the irrationalist narrative into question. Arguably, it shows that psychologists have been philosophically naïve. When we also take into account the replication crisis and alternative 'ecological' interpretations of the data, perhaps the most reasonable stance to take towards the irrationalist narrative is to suspend judgement. Maybe it is correct, but the case has not yet been decisively made.

Does this mean that reflection on cognitive biases cannot tell us anything about the formulation and prospects of evidentialism? Not necessarily. To

see why we need to consider its modal profile. Proponents of the view are not always clear on this, but there are good reasons to think that if evidentialism is true, it is necessarily true. Otherwise a lot of the literature would be hard to make sense of. It is difficult to see how modally remote worlds populated by envatted agents, evil epistemic demons, Boltzmann brains, philosophical zombies, infallible oracles, and the like could tell us anything about evidentialism were it merely contingently true: true for us in the actual world but not necessarily true for other agents in other worlds. These worlds, being modally remote, would simply be irrelevant. Yet, evidentialists do not treat them as irrelevant; they go to great lengths to accommodate them.<sup>19</sup>

This raises the question of what kind of necessity evidentialism is supposed to have. The obvious answer is: metaphysical. But whatever it is, the fact that it has at least enough modal strength to encompass these exotic worlds means that we need not wait for psychology to vindicate or debunk the irrationalist narrative to make use of it in shaping and assessing evidentialism. All we need is for the narrative to describe a *possibility*. And that it surely does. Whether or not we are actually prone to hindsight, confirmation, and implicit bias, we *could* have been, and these biases *could* have operated in a way that leads to violations of evidentialism. Moreover, their influence on us *could* have been introspectively inaccessible.

With this in mind, consider a possible biased agent, Blixia. Let us stipulate that Blixia is unconsciously influenced by hindsight, confirmation, and implicit bias, with the result that he often unwittingly forms irrational beliefs that are not supported by his evidence – he goes beyond anything that ecological rationality or Hedden, Kelly, and Dorst's, arguments

---

<sup>19</sup> Smithies (2019) defence of internalist evidentialism, for instance, discusses all of these worlds.

would license. Blixia might be an actual person. But even if he is not, he is a *possible* person, and that suffices for our purposes. What does the possibility of someone like Blixia tell us about evidentialism?

## §VI. Internalism

Firstly, it tells us that access internalist versions of evidentialism are untenable. According to access internalists, one is always in a position to know which propositions it is rational for one to believe at any given time.<sup>20</sup> Many of Blixia's beliefs are irrational in virtue of their biased etiologies. But since the influence of these biases is unconscious, Blixia is not in a position to know that they are irrational. It follows that access internalism is false.<sup>21</sup> Call this 'the problem of inaccessibility'.

The possibility of Blixia also causes trouble for other internalist evidentialisms. According to internalist 'mentalism', the rationality of a belief supervenes on one's non-factive mental states.<sup>22</sup> Plugging this idea into the evidentialist schema, we get the result that one's evidence is comprised entirely of one's non-factive mental states and that what one

---

<sup>20</sup> See Smithies (2019) for an extended argument for access internalism. Many access internalists, including Smithies, use the language of 'justification' rather than 'rationality'. However, they almost always take the two to be synonymous.

<sup>21</sup> I am not the first to point out that the possibility of bad unconscious etiologies causes problems for internalism. Puddifoot (2016) argues that implicit bias creates difficulties for the view. Srinivasan (forthcoming) argues that it cannot handle beliefs caused by internalised classist, racist, and sexist ideologies. Siegel (2017) argues that beliefs formed on the basis of perceptual experiences with 'hijacked' etiologies are irrational. Since hijacked agents are not usually in a position to know that their experiences have been hijacked, this is a problem for access internalism.

<sup>22</sup> This is Conee & Feldman's (2004) view.

ought to believe is determined entirely by these states. It is not clear how this proposal can be reconciled with the idea that whether a belief is the result of bias can affect its rational standing. Reconciliation requires proponents of mentalism to show that hindsight, confirmation, and implicit bias can be adequately described solely in terms of non-factive mental states. It is far from obvious that this can be done. Psychologists working on bias do not take evidence to consist of non-factive mental states. Rather, they almost always (implicitly) take it to consist of worldly facts. So, any internalist re-description of the relevant biases will need to 'internalise' the evidence. The worry is that this process of internalisation will result in non-factive mental states that are *caused by* bias counting as evidence. For example, internalists often take an agent's evidence to consist in, *inter alia*, how things 'seem' or 'appear' to the agent. There is a risk of cases in which it seems or appears to the agent that p only because of the influence of bias. A study by B. Keith Payne (2001) found that participants primed with pictures of black faces were more likely to incorrectly identify pictures of tools as guns than were participants primed with pictures of white faces. Payne hypothesises that this can be explained by the influence of implicit bias. Did it seem or appear to the participants primed with black faces that the picture is of a gun? The natural answer is 'yes'. If so, then this study illustrates the problem. Call it the 'problem of laundered biases'.<sup>23</sup>

Internalist evidentialism faces another, related, problem (one that, as we will see, it shares with certain kinds of externalist evidentialism). Suppose that, unconsciously wanting to believe that people like him, Blixia engages in selective exposure: he gathers up all the evidence he can to support this belief and avoids acquiring counterevidence.<sup>24</sup> Is it then rational for Blixia

---

<sup>23</sup> C.f. Siegel (2017).

<sup>24</sup> In order to side-step Salow's (2017) arguments, we can stipulate that Blixia does not do this intentionally.

to believe that people like him? Arguably not. The resulting belief is tainted by the means by which the evidence for it was acquired. To avoid having to say that the belief is rational, evidentialists will need to complicate their view. In addition to the condition that one must believe what one's evidence supports they will need to add a further condition: one must not selectively expose oneself to evidence.<sup>25</sup> Insofar as the motivation for introducing this condition is *ad hoc*, this is a theoretical cost; the condition is, it seems, not introduced as a natural extension of the core evidentialist idea, but rather simply to patch up the theory against criticism. Call this the 'problem of selective exposure'. As mentioned earlier, some evidentialists will insist that Blix's belief is rational, reasoning that the rationality of evidence-gathering is a practical rather than an epistemic matter. But there are reasons to be sceptical of this response. Later on, we will look at a theory of epistemic rationality ('dispositionalism') which suggests that it presents us with a false dichotomy.

## §VII. Rationality

Internalists might bite the bullet, of course. They might argue that the influence of unconscious biases has no effect on the rational standing of a belief: biased beliefs can nevertheless be rational beliefs. The claim would not be that such beliefs can be rational because they are ecologically rational (Gigerenzer's line), nor because the alleged 'biases' are no such thing (Hedden, Kelly, and Dorst's line). Rather, internalists might acknowledge that biased beliefs are epistemically suboptimal but maintain that since the influence of the relevant biases is unconscious it

---

<sup>25</sup> C.f. Greco (2005), Baehr (2011), Miracchi (2019), Flores & Woodard (ms.)

cannot make a difference to rationality, nor, *ipso facto*, to what one ought to believe. They might, for instance, insist that seemings and appearances with biased etiologies *are* evidence, and that beliefs based on them are rational in virtue of being supported by the evidence.

As uncomfortable a position as this is to adopt, a certain kind of internalist may be drawn to it. Many philosophers see a tight connection between obligation, blame, and control. According to one broadly Kantian line of thought, at least as popular in ethics as it is in epistemology, if one is blameless for  $\varphi$ -ing, it cannot be that one ought to have done anything other than  $\varphi$ , and an outcome that is beyond one's control, no matter how unfortunate, is not an outcome for which one can be blamed. It follows that it is always in one's control to do as one ought to. Internalists who endorse this line of thought and maintain that one ought to be rational will infer that one cannot be blamelessly irrational, and that the rationality of a belief cannot depend on factors beyond one's control. Since unconscious influences on belief *are*, in some fairly strong sense, outside of one's control, these internalists will say that they cannot make a difference to rationality. And since they cannot make a difference to rationality, they cannot make a difference to what one ought to believe.

The reasoning here is suspect in a number of ways. Here we will only look at its implausible consequences. Consider these passages from a short story by Vladimir Nabokov:<sup>26</sup>

“What he really wanted to do was to tear a hole in his world and escape...”Referential mania,” Herman Brink had called it. In these very rare cases the patient imagines that everything happening around him is a veiled reference to his personality and existence...Phenomenal nature shadows him wherever he goes.

---

<sup>26</sup> ‘Signs and Symbols’ (1948)

Clouds in the staring sky transmit to one another, by means of slow signs, incredibly detailed information regarding him. His inmost thoughts are discussed at nightfall, in manual alphabet, by darkly gesticulating trees. Pebbles or stains or sun flecks form patterns representing in some awful way messages which he must intercept. Everything is a cipher and of everything he is the theme. Some of the spies are detached observers, such are glass surfaces and still pools; others, such as coats in store windows, are prejudiced witnesses, lynchers at heart; others again (running water, storms) are hysterical to the point of insanity, have a distorted opinion of him and grotesquely misinterpret his actions. He must be always on his guard and devote every minute and module of life to the decoding of the undulation of things."<sup>27</sup>

The correct response here is not blame, but pity. This is, in part, because the man has no control over his delusions – he is more a victim of them than he is their architect. Yet it is perfectly clear that he is profoundly irrational. There is more to rationality than mere blamelessness. So there is no motivation for rejecting the idea that unconscious biases can affect the rationality of belief on the grounds that such biases are beyond our control.

---

<sup>27</sup> 'Referential Mania' is a fictional disorder, but we could easily replace it with documented real-world monothematic delusions. People with somatoparaphrenia believe that one or more of their limbs belongs to someone else. People with Cotard delusion believe that they are dead or do not exist. People with Capgras delusion believe that a loved-one has been replaced by an identical-looking imposter. In each case, sufferers produce elaborate confabulations to explain away evidence to the contrary.

## §VIII. Externalism

Internalist versions of evidentialism have a hard time accounting for the irrationality of biased beliefs. Do externalist versions fare better? At first glance at least some of them seem to. Consider, for instance, EXACTING EVIDENTIALISM. A reminder, it says that:

EXACTING EVIDENTIALISM: One's evidence consists of all and only those propositions one knows to be true. One may believe that  $p$  iff one's evidence entails that  $p$

EXACTING EVIDENTIALISM does not by itself deliver the result that biased beliefs are irrational. This is because it is about the conditions under which one has justification to believe that  $p$ , not about the conditions under which a belief that  $p$  is justified. In other words, it is about *propositional* rather than *doxastic* justification. For all EXACTING EVIDENTIALISM says, one's belief that  $p$  may be rational even though it is the result of bias, provided that one's knowledge entails that  $p$ , even if one's belief isn't based on that entailing knowledge. However, proponents of EXACTING EVIDENTIALISM will not accept this verdict. They will supplement the view with a theory of doxastic justification. An obvious supplement is the theory that one's belief is justified just in case it is knowledge.<sup>28</sup> And this view – the 'knowledge norm of belief' – can easily explain why bias-infected beliefs are not rational: a biased etiology usually precludes a belief from being knowledge.

At least one version of externalist evidentialism appears to be able to account for the irrationality of biased beliefs, then. Externalists should not

---

<sup>28</sup> Sutton (2005, 2007), Littlejohn (2013, 2019), Williamson (2013, forthcoming)

be too sanguine, however. As we have already noted, most epistemologists will be unhappy with EXACTING EVIDENTIALISM, the knowledge norm of belief, and similar truth-entailing views. If anything is consensus in epistemology, it is that false beliefs can be rational. If you mistakenly believe that the white table in front of you is red because, unbeknownst to you, it is illuminated by a red light source, then you are no less rational than your counterpart in the next room who is looking at a red table. If EXACTING EVIDENTIALISM and the knowledge norm are taken to be claims about *rationality*, then, they will be roundly rejected. If they are divorced from rationality and merely taken to be claims about what one ought to believe, then most epistemologists will reject them on the grounds that one ought to be rational.<sup>29</sup> This raises a question: is there a plausible externalist evidentialism that explains why biased beliefs are irrational whilst at the same time delivering the result that false beliefs can be rational? We should not assume a positive answer. It is incumbent on externalist evidentialists to show us that there is. Call this the ‘problem of false beliefs’.

Finally, consider again one of the problems that internalist evidentialism faces: the problem of selective exposure. An evidentialism according to which one should only believe what one knows doesn’t face this problem, since a belief that is the product of selective exposure to the evidence will not usually be a knowledgeable belief. However, a less demanding externalist evidentialism will face the problem. Take, for example, a view according to which one’s evidence consists of one’s knowledge, and one may believe that  $p$  if and only if  $p$  is sufficiently probable on one’s evidence above some threshold  $n = <1$ . This view faces the problem of

---

<sup>29</sup> One consequence of demanding views like EXACTING EVIDENTIALISM and the knowledge norm is that one ought sometimes to have irrational beliefs (Hughes 2017, 2019b). Elsewhere I have argued that the best solution to this problem is to accept that there can be epistemic dilemmas: situations in which one is subject to conflicting epistemic requirements. See Hughes (2019a).

selective exposure as much as internalist evidentialism does. And like the internalist, the externalist will have to complicate their view in an *ad hoc* fashion to deal with it.

## §IX. Alternatives

Whether it is given an internalist or an externalist spin, evidentialism struggles to handle bias. Internalist evidentialists face the problem of inaccessibility, the problem of laundered biases, and the problem of selective exposure. Externalist evidentialists fare somewhat better, but still face the problem of selective exposure and also face the problem of false beliefs. In light of all this, it is tempting to think that we should give up evidentialism altogether. But are there non-evidentialist theories that do better?

It seems that there are. Consider *dispositionalism*. This is the view that whether a belief is rational or not depends on the dispositions that it manifests. Maria Lasonen-Aarnio argues that a belief is rational just in case it manifests a disposition that is conducive to knowledge: the disposition results in knowledge across a range of normal cases.<sup>30</sup> Lasonen-Aarnio argues that your belief that the table in the room is red is rational since it is the result of a knowledge-conducive disposition: in the relevant counterfactual worlds, you would have only believed that the table is red if you had known that it is red. We can plausibly add that those beliefs of Blixa's that are the result of the influence of hindsight, confirmation, or implicit bias are *not* the manifestation of knowledge-conducive dispositions. Blixa forms (and would form) these beliefs even when they

---

<sup>30</sup> Lasonen-Aarnio (2010, forthcoming). See also Williamson (2017, forthcoming)

do not amount to knowledge. If so, then dispositionalism is able to account for the irrationality of biased beliefs and to solve the problem of false beliefs. It also solves the problem of selective exposure without having to resort to *ad hoc* manoeuvres. A belief that is caused or sustained by selective exposure to the evidence, it is natural to think, is likewise not a belief that manifests a knowledge-conducive disposition. Other theories, such as those that focus on epistemic virtues may also be capable of solving these problems whilst marking biased beliefs as irrational.<sup>31</sup> We should investigate and find out.

## §X. Conclusion

The possibility of unconscious cognitive biases has significant implications for epistemology. The irrationality of beliefs influenced by such biases cannot be explained by internalist evidentialism and raises significant challenges for externalist evidentialism. By contrast, at least some non-evidentialist theories handle them smoothly.

## §XI. References

Arkes, H. Faust, D. Guilmette, T. & Hart, K. 1988. "Eliminating the hindsight bias" *Journal of Applied Psychology* Vol 73(2): 305-307

Bacon, F. 2008. *The New Organon*. Cambridge Texts in History of Philosophy, Cambridge University Press.

---

<sup>31</sup> E.g. Sosa (2007).

Baehr, J. 2011. "Evidentialism, Vice, and Virtue" in Dougherty, T (ed.) *Evidentialism and its Discontents*. Oxford University Press

Bernstein, D. 2011. "Hindsight bias from 3 to 95 years of age" *J Exp Psychol Learn Mem Cogn*. Vol 37(2): 378-391

Bishop, M. 2000. "In Praise of Epistemic Irresponsibility" *Synthese* Vol. 122(1/2): 179-208

Bishop, M. & Trout, JD. 2004. *Epistemology and the Psychology of Human Judgement*, Oxford University Press

Brownstein, M. 2019. Implicit Bias. Zalta E. (ed.) *The Stanford Encyclopedia of Philosophy*

Camerer, C. *et al.* 2018. "Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015" *Nature Human Behaviour* 2: 637-644

Cohen, S. & Comesana, J. 2013b, "Williamson on Gettier Cases in Epistemic Logic and the Knowledge Norm for Rational Belief: A Reply to a Reply to a Reply" in *Inquiry: An Interdisciplinary Journal of Philosophy* Vol. 56 (4). pp. 400-415

Cohen, S. & Comesana, J. forthcoming. "Rationality and Truth" in Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon*. Oxford University Press

Conee, E. & Feldman, R. 2004. *Evidentialism: Essays in Epistemology*. Oxford University Press

Conee, E. & Feldman, R. (2011). "Response to Baehr" Dougherty, T (ed.) *Evidentialism and its Discontents*. Oxford University Press

Dorst, K. manuscript. "Rational Polarization"

Dougherty, T. 2011. *Evidentialism and its Discontents*. Oxford University Press

Fischhoff, B. Gonzalez, R. Lerner, J. & Small, D. 2005. "Evolving Judgements of Terror Risks: Foresight, Hindsight, and Emotion" *J Exp Psychol Appl*. Vol 11(2): 124-139

Flores, C. & Woodard, E. manuscript. "Epistemic Vigilance: In Defense of a Norm on Evidence-Gathering"

Gigerenzer, G, et al. 2001. *Simple Heuristics That Make Us Smart*. Oxford University Press

Gilovich, T. Griffin, D. & Kahneman, D. (eds.) 2002 *Heuristics and Biases: The Psychology of Intuitive Judgement*. CUP

Greco, J. 2005. "Justification is not Internal" in *Contemporary Debates in Epistemology*. Blackwell.

Hedden, B. 2019. "Hindsight Bias is not a Bias" *Analysis* Vol. 79(1): 52-61

Hughes, N. 2017. "No Excuses: Against the Knowledge Norm of Belief" *Thought* Vol. 6(3): 157-166

Hughes, N. 2019a. "Dilemmic Epistemology" *Synthese*. Vol. 196(10): 4059-4090

Hughes, N. 2019b. "Uniqueness, Rationality, and the Norm of Belief" *Erkenntnis*. Vol 84(1): 57-75

Kelly, T. 2008. "Disagreement, Dogmatism, and Belief Polarization" *Journal of Philosophy*. Vol 105(10): 611-633

Kornblith, H. 2012. *On Reflection*. Oxford University Press

Lasonen-Aarnio, M. 2010. "Unreasonable Knowledge" *Philosophical Perspectives*. Vol. 24: 1-21

Lasonen-Aarnio, M. 2019. "Guidance, Epistemic Filters, and Non-Accidental Ought-Doing" *Philosophical Issues*, Vol. 29 (1): 172-183

Lasonen-Aarnio, M. forthcoming. "Virtuous Failure and Victims of Deceit" in Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon*. Oxford University Press

Littlejohn, C. 2013. "The Russellian Retreat" *Proceedings of the Aristotelian Society*. Vol. 113(3): 293-320

Littlejohn, C. forthcoming "A plea for epistemic excuses" in Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon*. Oxford University Press

Miracchi, L. 2019. "When Evidence Isn't Enough: Suspension, Evidentialism, and Knowledge-First Virtue Epistemology" *Episteme* Vol. 16: 413-437

Nabokov, V. 1948. "Signs and Symbols" *The New Yorker* May 15<sup>th</sup> 1948

Nickerson, R. 1998. "Confirmation bias: A ubiquitous phenomenon in many guises". *Review of General Psychology*. Vol 2(2): 175-220.

Payne, Keith B. 2001. "Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon", *Journal of Personality and Social Psychology*. Vol: 81: 181–192.

Peecher, M. & Piercey, D. 2010. "Judging Audit Quality in Light of Adverse Outcomes: Evidence of Outcome Bias and Reverse Outcome Bias" *Contemporary Accounting Research*. Vol. 25(1): 243-274

Piller, C. 2016. "Evidentialism, Transparency and Commitments" *Philosophical Issues*. Vol. 26(1): 332-350

Plous, S, 1991. "Biases in the assimilation of technological breakdowns: Do accidents make us safer?" *Journal of Applied Social Psychology*. Vol. 21(13): 1058–1082

Pohl, RF. Bender, M. & Lachmann, G. 2002. "Hindsight bias around the world" *Exp Psychol*. Vol 49(4): 270-282

Pronin, E. 2009. "The Introspection Illusion" in Zanna MP. (ed.) *Advances in experimental social psychology*. Vol. 41: 1-68

Puddifoot, K. 2016. "Accessibilism and Implicit Bias" *Pacific Philosophical Quarterly*. Vol. 97(3): 421-434

Quattrociocchi, W, Scala, A, & Sunstein, S. 2016 "Echo Chambers on Facebook" Available at <http://dx.doi.org/10.2139/ssrn.2795110>

Roese, N. & Vohs, K. 2012. "Hindsight Bias" *Perspectives on Psychological Science*. Vol 7(5): 411-426

- Salow, B. 2017. "The Externalist's Guide to Fishing For Compliments" *Mind*. Vol. 127(507): 691-728
- Saul, J. 2013. "Scepticism and Implicit Bias" *Disputatio*. Vol 37: 243-263
- Siegel, S. 2017. *The Rationality of Perception*. Oxford University Press
- Smithies, D. 2019. *The Epistemic Role of Consciousness*. Oxford University Press
- Sosa, E. 2007. *A Virtue Epistemology*. Oxford University Press
- Srinivasan, A. 2015. "Normativity Without Cartesian Privilege" *Philosophical Issues* Vol. 25 (1): 273-299
- Srinivasan, A. forthcoming. "Radical Externalism" *The Philosophical Review*
- Sutton, J. 2005. "Stick To What You Know" *Nous*. Vol 39(3): 359-396
- Sutton, J. 2007. *Without Justification*. MIT Press.
- Sweeney, P. & Gruber, K. 1984. "Selective Exposure: Voter Information Preferences and the Watergate Affair" *Journal of Personality and Social Psychology*. Vol. 46(6): 1208-1221
- Williamson, T. 2013, "Response to Cohen, Comesana, Goodman, Nagel, and Weatherson on Gettier cases in Epistemic Logic" in *Inquiry: An Interdisciplinary Journal of Philosophy* Vol. 56 (1). pp. 77-96
- Williamson, T. forthcoming. "Justifications, excuses, and sceptical scenarios" in Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon*. Oxford University Press

Williamson, T. 2017. "Ambiguous Rationality". *Episteme*. Vol. 14(3): 263-274

Wilson, T. 2002. *Strangers to Ourselves*. Harvard University Press

Wilson, T. D., and Brekke, N. (1994) 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations', *Psychological Bulletin*, 116 (1), 117–142.