

## **HOW IRRELEVANT INFLUENCES BIAS BELIEF\***

Yuval Avnur  
Scripps College

Dion Scott-Kakures  
Scripps College

“All doctrines are to be suspected which are favoured by our passions.”

—David Hume, *Of the Immortality of the Soul*

### **1. Introduction**

Epistemic standards dictate that justification of our beliefs can be subject to defeat: some sorts of evidence, once acquired, provide reason to alter our attitudes, reduce our confidence, and perhaps assume a more agnostic stance. For example, I may come to believe, during a night of heavy drinking with colleagues, that my boss despises me. On my walk home the next morning, having sobered up, I realize that the conversation provided me with no good reason after all: it was all just drunken babble, exaggerated and sloppily processed. I have reason now to readjust my attitude towards the proposition that my boss despises me. In general, the claim that your source is unreliable (or anti-reliable), or that you were cognitively impaired, can defeat whatever justification you had (if you ever had any), or thought you had, for a belief, and this constitutes reason for a change in attitude.<sup>1</sup>

Irrelevant influences on belief have been much more controversial in the epistemology literature. Such influences seem to pose a distinctively genealogical challenge in which a belief token is criticized on the basis of its broader causal origins.<sup>2</sup> So, a subject may sincerely assert something and the retort is: “You (just) believe that because you’re her father” or “You believe that because you had a strict Catholic upbringing” or “You believe that because you went to Oxford.” There has been much discussion of the alleged legitimacy of such challenges and their connection to other, less controversial sorts of defeater.

Roger White puts the issue so: “The question is whether having formed a view as to whether  $p$  on the basis of reasons of the usual sort, information concerning the cause of your belief might constitute a further reason to change your opinion on the matter” (2010, 574). White is largely unimpressed with the case for answering “yes.” He rejects the alleged distinctive epistemological significance of such genealogical worries because either, upon examination, such worries amount to variants of familiar concerns (e.g., radical skepticism or disagreement), or because they present no real epistemological worry.<sup>3</sup> Similarly, Adam Elga (ms.) asks “when one learns that a belief reflects the influence of irrelevant factors, how much should that reduce the strength of the belief?” (p. 2). His answer is that you should decrease confidence in the belief only if the irrelevant factor indicates that you have failed to meet your own epistemic standards in forming (and perhaps sustaining) the belief.<sup>4</sup> Otherwise, Elga suggests, genealogical concerns about a belief are, if not entirely misplaced, just instances of other, familiar worries (again, such as radical skepticism or disagreement). Though these conclusions might seem surprising to those unfamiliar with current epistemology (and perhaps even to those philosophers writing on the topic even a few years ago<sup>5</sup>), some of White’s and Elga’s considerations are quite powerful. Since irrelevant influences do often *seem* to threaten our epistemic status, it isn’t obvious what message to take home about genealogical challenges to belief.

We will argue that information about irrelevant influences can, in typical cases, undermine a belief’s epistemic standing, or call for attitude revision. So there is good reason for taking such charges as “You just believe that because you were raised Catholic” seriously; the answer to White’s question is “Yes,” at least when it comes many such irrelevant causal influences, including the standard cases in the literature. In particular, we will argue that desire, or interest, is often, charitably understood, at the root of such charges as “you just believe that because you were raised Catholic,” and when one acquires good reason for believing that desire or interest has figured—in a particular fashion—in coming or continuing to believe that  $p$ , one ought to “change [one’s] opinion on the matter.” That is, such reasons typically undermine, or defeat whatever justification one had (if any).

The thought that desire or interest can warp inquiry and render us unreliable, and sometimes unjustified, has a long and distinguished philosophical pedigree.<sup>6</sup> But it is also familiar in ordinary life. One might reply to an intransigent and impermeable interlocutor, “You’re not following me because you don’t want to hear what I’m saying.” And we are apt to think, of the parent who has just asserted of his daughter “She’s the finest dancer in the company,” that “You just believe that because you’re her father.” Such a thought is clearly a challenge and a criticism. We think that desires that are near the surface (in this case, simply the desire that she be the best dancer) render the father untrustworthy in this arena. Other cases are less obviously linked to desire; and these are the cases that have figured prominently in recent discussions. They are often termed cases of “nurtured belief,”<sup>7</sup> and the irrelevant influence in such cases is presumably exposure

to a particular environment (rather than some other). For example, we may hear “You only believe that the Pope is infallible when opining *ex cathedra* because you were raised in a strict Catholic home” (hereafter *Catholic*), or “You only believe that Quine failed to undermine the analytic/synthetic distinction because you went to Oxford in the 50’s” (hereafter *Oxford*). But while such challenges do not wear the implicated desire on their sleeves, we will argue that they are charitably and plausibly interpreted as implicating desire in the etiology of such beliefs. Thus, the relatively straightforward cases, such as that of the proud father (which everyone can agree involve defeat or epistemic failure) are importantly related to the difficult cases, such as Oxford, which recent philosophers have puzzled over.

If genealogical worries, or irrelevant influence, ever raise serious epistemic problems, there is presumably some explanation for this. Otherwise it is just a “brute fact,” or basic principle, that genealogical influences defeat justification. But the story cannot be *merely* that, had one’s history gone otherwise, one would have believed otherwise, as Sher (2001, 66) suggests. As we will argue, on grounds different from White (2010), the mere counterfactual fact need not defeat (it may indicate a mere “positional” influence on belief (section 2)). Similarly, genealogical worries, at least in difficult cases like Oxford, cannot be explained merely by appealing to the unreliability of genealogy. That is, what explains why Oxford and Catholic defeat, when they do, is not *just* that a having a Catholic upbringing or an Oxford education is not a reliable indicator of what is true, as Cohen (2000) sometimes suggests (e.g. p. 18). The missing, additional story, which makes sense of both the easy and the more difficult cases, is the role of desire. Genealogical worries, and irrelevant influences on belief, are uniquely epistemologically significant when, and because, they entail a special, specific *process* of belief formation under the influence of desire.

No doubt, there are many versions of the appeal to nurtured beliefs that do not plausibly implicate desires. For example, the challenge might be that, since you were raised Catholic, you’ve never been exposed to the arguments against the pope’s authority, or to other perspectives on the history of the papacy, or maybe even to the full variety of other religions. Thus, your upbringing merely put you in a position to receive or avoid evidence or arguments, and that explains why you have arrived at your conclusion; had you had a different upbringing, you would have believed differently. This is the challenge that you are epistemically sheltered, that you believe as you do because you aren’t sufficiently well-placed to appreciate the relevant evidence.<sup>8</sup> In the next, second, section we aim to distinguish this sort of influence from the influence we take to be epistemically worrisome. We are neutral on whether this sort of “positional” influence calls for doxastic readjustment. Attention to the nature of positional influence will be useful in diagnosing and explaining a number of current disagreements about irrelevant influences. It also helps to clarify the importance of desires in characterizing the distinct epistemological worry raised by genealogy. Being raised in a Catholic household is an unreliable indicator of whether Catholic doctrine is true. And

yet, that mere fact does not itself establish that Catholic belief formed in that way is unjustified, given the possibility that a Catholic upbringing had merely a positional influence.

It is only when desires associated with a Catholic upbringing play a more specific, pernicious role in belief formation that genealogy becomes worrisome. In the third section, we characterize in some detail this sort of influence, which, we will argue, demands doxastic readjustment; we term this a “directional” influence. This section appeals to some psychological work in detailing the dynamics of such an influence. With the positional/directional distinction made, our argument proceeds as follows: That one’s belief has been directionally influenced by desire is both a plausible and charitable understanding of a charge of genealogical influence (section 4) and a defeater (section 5). Therefore, charges of genealogical influence often successfully constitute defeaters. This is so even in difficult cases such as Oxford (discussed in section 4), in which desire may not seem obviously relevant. In the final section, we reply to objections.

## 2. Positional Influence

Ultimately, we will suggest that criticism on the basis of irrelevant influences can sometimes be plausibly interpreted as implicating desire. Assuming that the presence of desire can be a defeater, we conclude that irrelevant influence can defeat. But this assumption, while it has some intuitive force, might well be doubted. One might wonder: Why, exactly, does learning of the presence of a desire ever defeat one’s justification? Grant for the moment that the irrelevant influences in Oxford and Catholic implicate desires. There are at least two ways to characterize the influence of desire in such examples. One of those ways might be epistemically unproblematic. In this section we explain this first, perhaps innocuous sort of influence. This will serve to clarify the more worrisome, second sort of influence, as well as to explain why one might have written off the epistemic significance of irrelevant influences altogether.

On the first construal of irrelevant influences, we might say that examples such as Oxford and Catholic involve “positional” influences. Such influences put us in a particular position, rather than some other, to possess certain reasons. These reasons may or may not be good, but what we learn when we learn of the irrelevant influence is that it is in virtue of these reasons that we come to form our beliefs. Desires are one potential source of positional influence. For example, consider

**President:** I loathe my college president and would love to learn that she is involved in something that threatens her position. While walking back from the office, a breathless colleague runs to up to me and says, pointing to the president’s office window, “I can’t believe it! Look! The president is doing something shameful.” Delighted, I scramble to the window, look in and come to believe that she’s involved in something that could get her fired.

President is a case in which the positional influence of the relevant desire is mediated by an intention to come to possess certain information or epistemic reasons. This is, of course, not an essential element of positional influences (whether they are desires or something else). For example:

**Pizza:** I want a pizza and, so, drive to *Pizza N' Such* to pick one up. While waiting in line, I encounter a thoughtful and trustworthy colleague who, after surveying the crowd, surreptitiously whispers to me that the college president has been seen doing shameful things and that the trustees will take up the matter at their next meeting.

In this case, were it not for the desire for a pizza, I wouldn't now believe that the president is in deep trouble. That desire has, again, put me into a position in which I come to have various reasons whereby I come to a view. What I come to believe in these cases is contingent upon my possession of a desire, and in that sense my coming to believe as I do seems a matter of accident or luck. But that fact alone need not call into question the justification of my belief; in this positional role, at least in these cases, desire itself appears to raise no worries for a belief's epistemological credentials.

Since our focus is on directional influence, we can remain neutral on whether a positional influence *per se* can pose a unique epistemic worry. The forgoing will help us to characterize that directional influence by contrast. It also helps to explain why some epistemologists have rejected the idea that irrelevant influences are epistemologically significant. For example, White (2010) seems to have positional influences in mind in his influential discussion of irrelevant influences. He writes, "There is nothing problematic about being lucky in obtaining *evidence* for one's belief (21)," and

Even if the coin toss determines which moral upbringing I receive which in turn determines the conclusions I draw, the reasoning I engage in as result of the toss and subsequent nurturing may be perfectly cogent . . . Other things being equal, sound reasoning results in justified belief. It is entirely possible for chancy events to result in a process of sound reasoning. (22)

The coin toss is an irrelevant influence that puts him in a position to reason cogently and determines the inputs to his reasoning. White argues convincingly that such positional influences (as we are calling them) need not be defeaters. He also considers a variation of the case, in which the flip of the coin doesn't determine whether he is raised in a way that allows him to reason properly, but rather is part of his moral reasoning itself. In this case, the coin flip determines the course of things downstream from the inputs to his reasoning. He writes:

A large part of what is disturbing about finding the coin in my head is that it seems I can't really be engaging in moral *reasoning* at all in this case. While it is hard to say what moral reasoning consists in, random coin tossing is not it. I'm not responding to relevant factors. I'm not really appreciating the force of various considerations. All of that is an illusion . . . The discovery of the coin shows that I never have been justified in any of my moral convictions. (p. 22)

In this second case, the influence of the coin is not merely positional. Rather, it is internal to one's reasoning process, and it effectively randomizes it. And this, White allows, can constitute a defeater. In later sections we discuss what it means to be internal to one's reasoning, how desires can play such a role, and how this defeats justificatory status. But it is important to note that White recognizes such defeaters in his essay, the principle thesis of which is that irrelevant influences are not epistemologically significant. It seems, then, that White is identifying irrelevant influences of the sort invoked in the standard cases, such as Oxford and Catholic, with positional influences. And, clearly, this would explain his deflationary regard for the epistemic significance of such influences. Below, we suggest that the identification of irrelevant influences with positional influence is a mistake, since there are more charitable interpretations of the standard charges of irrelevant influences that imply something beyond a merely positional influence. But we have no objection here to the claim that when an influence is merely positional, it need not constitute a defeater. Let us now consider the other sort of influence that an irrelevant factor, and in particular a desire, can have on belief. This other factor is not altogether unlike White's second, "internal" coin flip.

### 3. Directional Influence

In this section we characterize another sort of influence exerted by desire on belief-forming processes (we construe desire broadly, to include all interests and aversions). A *directional influence*, as we will call it, causes our *handling of evidence* to favor a particular, predetermined outcome, where the desires that determine the favored outcome go beyond merely interest in believing truth. This contrasts with a positional influence, which, as we have seen, causes us to come to possess, or puts us in a position to appreciate, a certain set of evidence in the first place. Directional influence has been discussed by psychologists as "motivated reasoning" and by philosophers as "self-deception."<sup>9</sup> But while the phenomenon is familiar, the precise ways in which such processes undermine epistemic status has received too little attention.<sup>10</sup>

Though the effects of desire on belief are familiar, the facts are sobering: the correlation between subjects' regard for their intelligence and I.Q. tests is .3 (Borkenau and Liebler, 1993); the correlation between a person's judgment of his own physical attractiveness and others' judgment is .24 (Feingold, 1992); the correlation between athletes' regard for their talents and their coaches' regard is .16 (Felson, 1981). Tellingly, Brown and Dutton (1996), in a survey of the then existing literature, conclude that "self-perceptions become more biased when the trait is highly desirable or undesirable" (1290) and conclude that "on virtually every conceivable positively valued trait the majority of people think they are better than others" (*ibid.*). Thus, the presence of desire, at least in the self-regarding sphere, is an indicator of bias, and the degree of bias covaries with strength of motivation.

Such biasing effects are widespread. For example, consider Ziva Kunda's pioneering early work on motivated reasoning 1987, 1990. In her widely cited study, non-coffee-drinkers and coffee-drinkers considered (fabricated) research that alleged serious health dangers for heavy coffee-drinking females. Women coffee drinkers judged the evidence to be significantly weaker than did men or non-coffee-drinking females. When the effects of heavy coffee consumption were presented as much less serious, women coffee drinkers rated the study no less persuasive than did the others. In short, the more serious the threat to interests, the more pronounced is the effort to defend those interests.

What are the psychological processes that mediate these effects? There is more than one way for desire to bias belief, and we don't want to count them all as directional. For example, one's interests may merely make one hypothesis more salient, where it may not have occurred to one to test that hypothesis otherwise. Given the well-known hypothesis testing biases,<sup>11</sup> such an influence will amount to a bias in favor of an outcome determined by one's interests.<sup>12</sup> For example, in Kunda's case, the female coffee drinkers may simply—due to its hedonic attractiveness—generate the hypothesis “Perhaps this study is flawed” which is then confirmed via familiar processes (Kunda 1990, 494–95). But it isn't clear that one's desires, in such a case, affects the way in which evidence is brought to bear on a hypothesis in a way that is distinct from positional influence. That is, it isn't clear that merely by bringing a hypothesis to one's attention, a desire (or any other influence) does much more than bring a particular set of considerations to light. A similar effect could result from, say, an interlocutor's asking us “Is p true?” or from attending a seminar concerning the CIA's assassination of JFK. In such cases, the influence is on which hypothesis is tested, rather than how that hypothesis is tested. Accordingly, it will not always be clear that a bias due to the subject's motivations should be counted directional, rather than positional.<sup>13</sup>

In contrast, the psychological processes that exert a directional influence are more light-fingered and pernicious, and this is plausibly the case with Kunda's female coffee drinkers. As Kunda herself put it: “[w]hen confronted with evidence that has implications for optimistic beliefs, people evaluate it in a self-serving manner, applying more stringent criteria to evidence with less than favorable implications to the self” (1987, 636; cf. Mele 1997, 1983). Desire or interest directs inquiry, and specifically the way evidence is brought to bear on a hypothesis and toward the acceptance of a doxastic target. Consider a case described by Alfred Mele:

Our desiring that p may lead us to interpret data as supporting p that we would easily recognize to count against p in desire's absence. For example, Sid is very fond of Roz, a college classmate with whom he often studies. Wanting it to be true that Roz loves him, he may interpret her refusing to date him and her reminding him that she has a steady boyfriend as an effort on her part to play 'hard to get' in order to encourage Sid to continue to pursue her and to prove to her that his love for her approximates hers for him. As Sid interprets Roz's

behavior not only does it not count against the hypothesis that she loves him, it is evidence for the truth of that hypothesis. (1997, 94)

We can ornament the case. Imagine that Sid, considering the possibility that Roz loves him, comes by what he takes to be evidence in favor of that hypothesis. (In addition to the “evidence” mentioned by Mele, Roz laughs at his jokes, smiles at him in class, etc.) Still, and understandably, such evidence strikes Sid as non-dispositive. So, reflecting, he generates some more general theory concerning the ways in which study relationships frequently blossom into romantic ones in hot-house dormitory environments. This, he then brings to bear in generating additional supporting evidence. Of course, there is the contrary datum mentioned by Mele, as well as the fact that Sid has recently encountered Roz and her erstwhile (so Sid believes) boyfriend, Rocco, enjoying an apparently splendid evening at a local pub. But, Sid thinks, Roz is such a sweetheart; she would feel guilty about her waning affections for Rocco and try to make a show of her fondness for him. When Beppe, a friend, tells Sid that Rocco and Roz *really* are a couple, Sid smiles at the thinness of Beppe’s alleged evidence and notes the fact that Beppe has recently been dumped by his own romantic partner. It’s unsurprising, thinks Sid, that Beppe would be envious of Sid’s new relationship with Roz.

We take doxastic behavior like Sid’s to be wholly typical of cases of directionally biased reasoning. Think of the alcoholic, convinced she’s a convivial social drinker, the doting parent, convinced his daughter is not using drugs, and all in the “teeth of the evidence,” as we are apt to say. Here, desire’s role goes beyond generating a hypothesis to test; rather, there seems to be a “motivational engine” within the subject’s reasoning or his evaluation and interpretation of the evidence. Sid credulously accepts confirming data, while he subjects data that strike us clearly as contra-indications of his preferred hypothesis to creative but nonetheless withering critical scrutiny. Sid’s desires play a role in his evaluation of matters that bear on his question (Does she love me?) in ways powerfully suggestive of there being an aim that is not truth: the installation of some favored belief, a belief that satisfies his desires or that defends his interests. Desire directs Sid’s inquiry toward a doxastic target or range of targets,<sup>14</sup> and not merely by triggering an otherwise uninteresting hypothesis (cf. Talbott, 1995, 60).<sup>15</sup>

Hypothesis testers have many “tactics”: biased memory searches, the use of various inferential rules (Lewicka, 1992), the generation of more general theories from which a target conclusion can be inferred (Klein and Kunda, 1991). The thwarting of one produces another in pursuit of a doxastic target. The tactics a motivated reasoner like Sid uses are no different from those made use of in standard, accuracy-guided reasoning (Trope and Liberman 1996; cf. Kruglanski, 1996). This makes the directional influence of desire all the more difficult for the reasoner to detect.

We need a more detailed account of this doxastic behavior. Talk of guidance and direction by desire is powerfully suggestive of the intentional. But certainly

such cases are not generally to be explicated by appeal to some intentional project.<sup>16</sup> Sid is not intentionally aiming to come to believe that Roz loves him. For a deeper, general explanation, we appeal to a recent and influential account of lay hypothesis testing offered by a number of psychologists (Lewicka, 1989, 1992; Friedrich, 1993; Trope and Liberman, 1996; Trope, Gervy, and Liberman, 1997) and developed by Mele in his account of self-deception (2001).<sup>17</sup> These views have it that “[h]ypothesis testing is motivated by pragmatic concerns of the subject” (Trope, Gervy and Liberman, 1997); i.e., how intensively and in what manner an individual tests a hypothesis will reflect her desires, interests and values.

Hypothesis testers aim to settle questions of the form “ $p$  or  $\sim p$ ?” Such testing is effortful and obviously involves costs to the subject in the form of time and effort devoted to hypothesis testing.<sup>18</sup> According to these “pragmatic hypothesis testing”<sup>19</sup> accounts, there is another sort of cost to the settling of questions: the cost of anticipated errors. Motivation for hypothesis testing is provided by the subject’s regard for the costs of being mistaken. A cognizer who faces a question aims to reach a “confidence threshold” at which time testing is terminated. There are costs associated with settling a question in favor of  $p$  when  $p$  is false (false positives), and there are costs associated with resolving a question in favor of  $\sim p$  when  $p$  is true (false negatives). With respect to many questions, the cost of false positives, on the one hand, and the cost of false negatives on the other, will not be identical but will be asymmetric. The costliest error, the “primary error,” James Friedrich terms it, is the error the individual is chiefly motivated to avoid. This primary error is fixed by the interests, values, and desires, of the cognizer. Importantly, these asymmetric error costs generate not only asymmetric confidence thresholds, but affect the way evidence is brought to bear in reaching one’s level of confidence. The upshot of this is directionally biased hypothesis testing.

First, consider confidence thresholds. When false acceptance is more costly, acceptance thresholds will be higher than rejection thresholds. As a result, the hypothesis will have to be highly probable to be accepted, but only moderately improbable to be rejected. In contrast, when false rejection is more costly, rejection thresholds will be higher than acceptance thresholds, so that the hypothesis must be highly improbable to be rejected, but only moderately probable to be accepted (Trope, Gervy, and Liberman, 1997, pp. 118–20).

Second, consider the way in which evidence is handled in generating one’s confidence levels. When Sid intensively tests friendly hypotheses in what amounts to a search for support, this reflects the high relative cost of a false negative; when he intensively tests threatening hypotheses in what amounts to a search for disconfirming data, this reflects the high cost of false positives. Given this, it’s likely that Sid will confirm happy possibilities while he subjects unhappy ones to withering scrutiny (Trope and Liberman, 1996). As Trope, Gervy and Liberman put it, the account predicts that:

[P]eople should engage in more extensive and analytic processing before rejecting a desired hypothesis than before accepting it. Thus evidence will often be evaluated more carefully and critically when it is inconsistent, rather than consistent with, a desired hypothesis. Evidence supporting the desired hypothesis is likely to bolster one's confidence to threshold, and may also lead to earlier termination of the hypothesis-testing sequence than evidence refuting it. (1997, p. 120)

Sid's directional bias is the result of the profound asymmetry in his error costs. What this model captures, and what is characteristic of directional influence, is the alternate credulousness and hyper-criticality that is sensitive to one's interests or error costs.<sup>20</sup>

All of this gives some substance to the notion that in such cases inquiry is *directionally* biased.<sup>21</sup> In such cases, we may speak of a subject's having a "target" belief (or range of beliefs). In an unproblematic case, when we engage in inquiry to settle a question of the form "p or not-p?" we presume that when and if inquiry is successfully completed we will have:

- a. Come to believe that p, if by our then current lights we have come to possess good and sufficient reason to believe that p; or
- b. Come to believe that not-p, if by our then current lights we have come to possess good or sufficient reason to believe that not-p.

In this way, our aim in settling a question of this form is disjunctive. But in cases of directionally biased reasoning this isn't quite so. Even if the subject believes that she is trying to settle a question and in that sense presumes that her aim is the disjunctive one above, and even if, at the personal level, that *is* her aim, the way she conducts her inquiry is directed by an additional aim: the satisfaction of her desires, or defense of her interests, which determine a target belief (or beliefs). This offers a way of distinguishing between the aim of our belief-forming processes when motivated by accuracy concerns from the aim in directionally biased reasoning. All inquiry may well be motivated but inquiry may be motivated by accuracy, or pursuit of directional aims or, no doubt, quite frequently, a combination of both.<sup>22</sup> This is not to say that in directionally biased inquiry an inquirer *intentionally* biases her inquiry in the direction of p. Rather, one's desires fix one's error costs, and one tests in such a way as to avoid one's "primary error."<sup>23</sup> Mele has suggested an "impartial observer test" for self-deception that is useful in detailing the notion of a target in directional bias in terms of the output of such a process.<sup>24</sup> Roughly, the idea is to hold the set of evidence available to the subject fixed, and consider a group of peers who lack the relevant desires (and so don't have the relevant error-costs determined by those desires), but who are otherwise similar to the subject and who are settling the same question. If the subject concludes that p, while those in the impartial group who conclude that not-p significantly outnumber those who conclude that p, then that is evidence that the subject's desires are directionally

influential and the target of the influence is  $p$ . This provides an “output test” for desire’s directional influence on the handling of presumptively relevant considerations.

A similar approach can test for directional influence in the *process* leading up to the conclusion: Where datum,  $d_1$ , is or would be initially regarded by both  $S$  and his impartial cognitive peers as strongly suggesting that  $\sim p$ ,  $S$  subjects  $d_1$  to intensive critical scrutiny, repeatedly generating possible accounts to explain away its epistemic significance, while a significant majority of  $S$ ’s impartial cognitive peers would not (or would generate far fewer such alternative explanations); or, while  $d_2$  would be regarded by a majority of  $S$ ’s impartial cognitive peers as irrelevant to the question “ $p$  or not- $p$ ?”  $S$  repeatedly generates hypotheses according to which  $d_2$  is evidence in favor of  $p$ ; and so forth.

Finally, directional reasoning is typically not transparent to the subject. Consider Sid again. He is seeking reasons that will allow him to bring his investigations to an end. Sid presumes, as a reflective reasoner, that his embrace of a conclusion is the result of mere evaluation of his epistemic reasons. In fact, his doxastic activity is shaped by desire and interest; he is in error about what animates his investigations. If he ever learns about the directional influence of his desire on his belief, he may well be surprised, or at least regard it as a new insight into his doxastic situation.

The questions we ask, the data that are salient for us, the evidence we regard as probative or misleading, all of these can be shaped by desire in such a way that there is an apparent doxastic target to inquiry. When desires have such directional influence, arriving at a particular conclusion is made more probable, and the aim of inquiry is not merely or solely accuracy or truth. It is, instead, also aimed at the avoidance of the costliest error, as they are fixed by her desires. Two things remain to be shown about directional influence: its connection to standard charges of irrelevant influences, and its epistemic significance. Those are the topics of the next sections, respectively.

#### **4. Directionally influential desire in standard cases of irrelevant influences**

The charge “You believe she’s the best dancer in the company because you’re her father” is regarded by those who level it as undermining the target’s justification in virtue of the role played by desire in the genesis or maintenance of belief. This is unsurprising; desire is an irrelevant influence on a belief since what we desire is, alas, often irrelevant to the truth. So showing (as we do in the next section) that the presence of directionally influential desire constitutes a defeater, effectively shows that irrelevant influences can defeat. But, before showing that, we turn now to charges like “You just believe that because  $X$ ” where the “ $X$ ” is canonically filled in by some nurturing environment. These are the cases that figure centrally in the literature on irrelevant, and specifically genealogical influences.

On their face, charges like these allege that such beliefs are suspect because of the presence of irrelevant environmental influences. The role of desire or interest is not apparent. Indeed, we agree that people who level such charges are typically not pointing self-consciously to the presence of desire as the culprit. Rather, they fix on the environment and take this to be the source of the objection. But, we will argue, such genealogical challenges are in fact plausibly understood as desire or interest-based challenges. That is, nurturing environments are connected with directionally influential desire. In such charges as “You just believe that because you were raised Catholic” the environment figures as a proxy for interest and desire.

Our desires are an obvious upshot of our environments broadly construed. Given what we all know about human beings (e.g., that in certain environments they tend to have or acquire certain desires) there’s good reason for believing that, e.g., *Catholic* undermines confidence when it does because it implicates the presence of relevant desires, broadly construed to include aversions, fears, hopes, and so on. The distinction between positional and directional influences is useful here since it points to two ways in which we might understand the role of such desires in cases like *Catholic* and *Oxford*.

On a merely positional reading of these worries, the natural thought is: “If I hadn’t been in position x, I would very likely believe something quite different now.” This worry is *contrastive* rather than intrinsic. And this is one reason why on a merely positional reading, we might conclude that there is no distinctive epistemic worry in the neighborhood. Any epistemic worry appears to be a matter of familiar worries arising from the realization that other informed people disagree with you. This seems to be so even if we identify a *desire* as the vehicle for the positional influence. You *want* Catholic doctrine to be right, to think that you’ve been raised by enlightened people rather than dupes, etc. And this desire put you in a position to become exposed to some arguments rather than others.

But *Catholic* can naturally be interpreted differently, in ways familiar from discussions of “nurtured beliefs.” On this interpretation of the criticism, you handled whatever evidence or reason you had in a way sensitive to the implicated desires, and this suggests you are less reliable than you otherwise presumably would be (see section 5). You believe what you do because the desires implicated in a Catholic upbringing had a directional influence on your belief formation.

Consider first a more straightforward case. If we had evidence that the environment in which we were reared systematically manipulated the evidence accessible to us, brought us to regard as powerfully probative data that were, in fact, irrelevant or misleading, brought us to regard as irrelevant or misleading data that were, in fact, powerfully probative, etc. and did so in ways that displayed sensitivity to a particular doxastic outcome, this would give us reason to reduce confidence in our conclusions.<sup>25</sup> Were we to learn that we were raised in such a “Brave New World” environment, we would likely regard ourselves as having reason to think about some relevant belief: “I seem to have good evidence for p.

But now that I've learned that my reasoning has been systematically manipulated by others with the aim of getting me to believe just what I do believe, I don't know what to believe." Such a response seems not at all unreasonable. The influence here is not merely positional, but also directional.

There are other, more ordinary ways to be manipulated: our own desires can sometimes lead to directionally biased reasoning that makes arriving at the target belief more likely. For example, it seems natural to understand our father above as subject to such manipulation in virtue of the role played by directionally influential desire in the genesis of his belief about his daughter's talent. Here, it is not an external source, but his own desires that are doing the manipulating.

Similar interpretations are available in cases of "nurtured beliefs," given the clear connection between upbringing or environment and one's desires. Where charges like *Catholic* and *Oxford* are plausibly leveled, the cognizers powerfully care about the various answers to questions they face because their desires are at stake, and this can rightly be interpreted to implicate directionally biased reasoning in the formation and/or maintenance of belief. In other words, being raised in a Catholic household or being educated at Oxford can produce asymmetric error costs and so directionally bias our reasoning. A lay person may not have the details of asymmetric error costs in mind, but we are all familiar enough with the phenomenon of directional bias to gesture at it when criticizing belief.

The positional/directional distinction helps us to anticipate another role for desire. G. A. Cohen's (2000) Oxford/Harvard-Analytic/Synthetic case is a central example in much of the literature on irrelevant influences. As we understand it, he's unnerved by the fact that those who went to Oxford *currently* take the analytic/synthetic distinction to be rightly drawn while those who went to Harvard *currently* take that distinction to be groundless. This is consistent with holding that, first, students in the relevant group originally formed the beliefs under a merely *positional* influence and that, second, the *maintenance* of these beliefs through time—from the 1950's on—is the result of directional influence that is itself the result of desire. Given the prominence of this example in the literature, it is worth considering both a directional influence in the original formation of the belief as well as a directional influence in the later maintenance of the belief.

Cohen's example is effective because the analytic/synthetic distinction was, at the time, a highly charged matter. In the 1950's and '60's, and especially for students at Harvard and Oxford, acceptance or rejection of Quine's arguments was probably regarded as emblematic of a certain philosophical style and method, and so as signaling membership in or allegiance to that brand of philosophy—"on the cutting edge" or a "defender of the flame." This implicates desire because holding such doctrines is powerfully related to one's identity as a member of a group, as agreeing with one's chosen mentors, and to one's sense of one's self as competent or good.<sup>26</sup>

Formation of the belief in the first place is under at least some positional influence: A student at Harvard/Oxford is in a position to acquire evidence

relevant to the matter and, even more, there's lots of available information supportive of the favored doctrine. Moreover, the student may well want the "home" doctrine to be true (for the reasons discussed above) and, so, seek out various sources of information accordingly. But a student may also evaluate and respond to evidence in doctrine-supportive ways in a manner that reflects directional influence. For, holding the doctrine, being able to defend it, etc., marks one as a member of a group. Uncertainty or confusion about the doctrine threatens that membership. A lot can hinge on being regarded as a right-thinking member of a group. Since the group regards that doctrine as true, failure to hold that doctrine threatens one's own sense of competence and being regarded as competent. The list of interests that might be cited here is long indeed—for example, expecting to have close working relationships with mentors, wanting to be regarded as a promising student with a future in the field, etc. So, when a Harvard student in 1956 considered the question "Does Grice and Strawson's 'In Defense of a Dogma' show Quine to be mistaken?" the student would have been keener to avoid believing falsely that Grice and Strawson were right than he would have been to avoid believing falsely that Grice and Strawson were wrong. The sort of desires and interests we've pointed to provide a rationale for such asymmetric error costs and confidence thresholds.

Imagine that a student at Harvard/Oxford was uncertain as to whether the analytic/synthetic distinction is well-drawn. In a context in which holding the doctrine is emblematic of group membership and philosophical acumen, the student's very *uncertainty* on such a fundamental matter will quite directly threaten desires and interests implicating the student's competence. After all, the smart people here have figured this out; they have a grasp of why, e.g., Grice and Strawson have got it wrong/right. A competent, hardworking philosopher would have a settled view of the matter, etc. Assuming that interests having to do with self-affirmation and self-consistency are threatened by such uncertainty, the directionally biased testing of his question is hardly surprising. We have a case like Sid's.

Though the above considerations suggest otherwise, we *can* easily imagine that the original role played by studying at Harvard/Oxford and its accompanying desires was merely positional. It would depend on the details of the particular case. However, notwithstanding this, the *maintenance* of the rejection/acceptance of the analytic/synthetic distinction can itself be regarded as the result of directional influence and, so, a clear way in which the charge "You just believe that because you went to Harvard/Oxford," nonetheless raises apparently and, as we will argue in the next section, genuinely problematic epistemological worries. That charge remains plausible because it's the fact that one attended Oxford/Harvard at the time one acquired the relevant (perhaps justified) belief that explains one's more recent directionally biased maintenance of that belief. Defense of that belief is a matter of importance. As Hart, et al. (2009) put it in a large meta-analysis on the ways in which commitment to an attitude shapes which information we expose ourselves to and seek out:

Defense motivation is presumably also strengthened by the individual's commitment to the pre-existing attitude, belief, or behavior and by high relevance of the issue to enduring values. Personal commitment. . . is presumed to increase defense motivation because of the greater discomfort produced by holding an incorrect view on an important issue (Brehm and Cohen, 1962; Kiesler, 1971). Personal commitment is often conceptualized as feeling highly attached to a view (Kiesler, 1971) or contributing to feeling of ownership for a view (i.e., belief possession; see Ableson 1988). Several factors have been identified that might lead to commitment, such as sacrificing for the view (e.g., dedicating much time or effort to making a decision), freely choosing the view

(e.g., forming an attitude without coercion), and explaining the view publically or privately (e.g., defending a belief in a written essay; for review see Harmon-Jones, 2008; Olson & Stone, 2005). (558)

Hart, et al. have in mind seeking variables that co-vary with the biased selection of information, which will then permit maintenance of the belief. But, more generally, the sort of directionally biased reasoning in Sid's case might well be displayed in our Harvard/Oxford graduate's maintenance of the relevant belief. Interests having to do with competence, self-consistency, and self-affirmation are squarely at play: "I worked very hard to get a grip on the issues. Did I get it wrong? Should I now re-raise the question? I've said many nasty things in the past about folks who denied. . . I haven't bothered reading more than a few essays that dispute. . . I thought that X's defense of synonymy was antiquated. Was I wrong about that?"

In this way, criticisms like *Oxford* and *Catholic* can be charitably interpreted as implicating directionally influential desire, even if one isn't aware of some detailed, specific account of that role, and even if one's originally acquiring the belief can be defended by appeal to the relevant desire's having a merely positional role. That is, it is natural to regard these directionally influential desires as playing a role not only in the formation but also to the current *maintenance* of the belief. The implication is that these desires have shaped one's management of the evidence in ways that favor a target belief.

The idea that the charge of irrelevant influences implicates desires in the way we've described is further corroborated by the sorts of things we often say when defending ourselves against such charges: I've checked my reading of the evidence against sources that lack my desires (neutral parties); or I don't actually have the desires that you are accusing me of. We take up the epistemic significance of such charges of irrelevant influences, as well as the nature of such defenses against the charges, in the next section.

## 5. Directionally influential desires as defeaters

In the previous section, we argued that irrelevant influences often, in standard sorts of cases, implicate directionally influential desire. In this section, we

argue that the information that one's belief was (likely) formed in such a way is a defeater, and so is good reason to reduce confidence. What one often aims to point out when criticizing a believer for being subject to irrelevant influences is often a genuine defeater.

Intuitively, forming a belief by wishful thinking is irrational, and unjustified.<sup>27</sup> Forming a belief under directionally influential desire is a way of thinking wishfully. Accordingly, this sort of belief formation is also intuitively irrational, and unjustified. The task, then, is to vindicate and detail this intuition.

Consider *Catholic* again, where one is accused of having formed a belief under directionally influential desire. One might think that, if the accusation is right, the belief was never justified in the first place; or one might think the belief is justified until the believer learns that the reasoning was directionally biased. Either way, when the believer learns of the bias, she should reduce confidence in her belief, and she should no longer regard it as justified *to the same degree* as she thought it was before learning about her bias. That is, evidence that her reasoning was directionally influenced determines that she has less (or no) justification for currently believing.<sup>28</sup> It is this latter claim that we will defend in this section, not either of the antecedent, competing claims about whether she was ever justified in the first place (we are neutral on that).

For any belief that *p* which one takes to be justified, an “outweighing” defeater is some information that, evidence for, or justification to believe that not-*p*. An “undermining” defeater is some information that, evidence for, or justification to believe that one (now) lacks justification to believe that *p*.<sup>29</sup> Clearly, undermining does not entail outweighing.<sup>30</sup> This won't matter, since we claim only that evidence for directional influence is an undermining defeater, and not that it is an outweighing defeater. Let us call any undermining defeater due to evidence of directionally influential desire a ‘debunker’.<sup>31</sup> Here are two arguments that there are debunkers.

The first argument involves reliability. When you get evidence indicating that your belief-forming process is unreliable, or less reliable than you assumed, this constitutes a defeater. The degree of defeat is presumably proportional to the strength of the evidence for unreliability and the degree of unreliability thereby indicated. (This should be relatively uncontroversial to internalists and externalists alike.) Second, at least from the subject's perspective, directionally influenced reasoning is, all else equal, less reliable than non-directionally influenced reasoning, so that evidence that one's belief-forming process was so influenced is evidence that one is less reliable than one otherwise would be.<sup>32</sup> This premise requires some discussion, since it is often observed that desires themselves are not *anti-* or *contra-reliable*: wanting that *p* is no indication that not-*p*.<sup>33</sup> Why, then, does the directional influence of the desire that *p* render the process less reliable? All else equal (that is, absent any additional information about the correlation between one's wanting that *p* and *p*'s being true), believing according to one's desires is about as reliable as believing randomly, or by chance. If, as is typically the case, there is no evidence for any correlation between what one wants and

what is true, believing according to one's wishes is believing by chance, at least from the subject's own perspective. So, when desire is directionally influential, this (all else equal) reduces the reliability of the process towards chance, since the outcome of the process is driven, to the extent that the directional influence effectively guides the conclusion of the reasoning, by something merely as reliable as chance.<sup>34</sup> Therefore, evidence that a desire has been directionally influential is a debunker. Before offering the second argument, and then explaining and qualifying our conclusion, we will illustrate it with an example.

Suppose that in a logic exam you arrive at answer "P" for the first question. You then learn that you were given a reason-distorting drug that impairs logical reasoning in an introspectively undetectable way.<sup>35</sup> This information is not evidence that not-P; you have no evidence that the impairment directs reasoning towards any particular answer or the wrong answer. All you've learned is that the drug renders your logical reasoning less reliable than you assume normal reasoning to be. So, this is a defeater for your belief that p, and you should accordingly reduce confidence in your answer. This resembles directional influence, except that the drug's influence is not *directional*, since it does not bias your reasoning towards any particular answer. In the case of directional influence, your "impairment" makes more probable that you will arrive at some a conclusion that, according to our second premise above, has only a chance relation to the truth—i.e. your desires. So, it may as well be random, like the reason-distorting drug, as far as your pursuit of truth is concerned. The upshot is that whenever an influence that is not more reliably correlated with the truth than chance is introduced into your reasoning process, your reliability is brought closer to chance (unless it was as reliable as chance in the first place).<sup>36</sup>

The second argument for debunkers concerns general features of epistemic norms. In general, reason to believe that one's handling of evidence was manipulated is reason to think that one's handling of the evidence has not been *solely* responsive to truth-indicating concerns. It is reason for regarding one's inquiry as having been influenced by the manipulator's interests, regardless of whether the manipulator is someone else or oneself via one's desires. Either way, it constitutes reason for reconsidering the typical default transition from a judgment that one has good and sufficient reason for believing that p to the belief that p. For, in such a case, one's inquiry and handling of the evidence have been sensitive to desire, or one's effort to protect one's interests, rather than strictly the pursuit of accuracy or truth. This is contrary to the *point*, from an epistemic perspective, of believing what you take to be best supported by evidence or reasons.<sup>37</sup>

This worry concerning directional influence resembles some considerations raised by Hilary Kornblith (1999). As he puts it, "Ordinarily, when we reflect on our reasons for one of our beliefs, we are motivated by a desire to have our beliefs conform to the truth. (182)" But in the case of what Kornblith terms "rationalization," which we can assimilate to our directionally influenced reasoning, this is not so: "[I]n the case of rationalization, our motivation for reflecting on reasons is different... such motivations are not transparent to us... it is

precisely because of this that the process of scrutinizing our reasons may be terribly counterproductive from an epistemological point of view. (183)”<sup>38</sup> Where we have evidence that desire has been at work in shaping our reasons in directional fashion, we have reason for believing that the reasons in virtue of which we currently take ourselves to have sufficient reason to believe have been shaped by non-truth tracking considerations. Accordingly, we should reduce our confidence in the relevant beliefs: my taking my belief that *p* to be justified on the basis of my reasons, *R*, hinges on my *not* having positive reason for regarding *R* as non-truth-indicating. Debunkers defeat by providing this positive reason, just like learning that my tutors since childhood have been handsomely paid to manipulate me in order that I come to believe that *p*, or coming to learn that I’ve been hypnotized in such a way that I’m highly likely to discount or fail to notice evidence in favor of not-*p*, etc., can provide such positive reason and thereby defeat.

We have now offered two arguments for the view that evidence for directionally influential desire is a defeater, or a debunker. In the remainder of this section, we give details and qualifications.

Defeaters relate to justification, but there are different sorts of justification. The standard distinction is between *doxastic* and *propositional* justification, where having the former implies having the belief, while having the latter does not. Having propositional justification amounts to having what it takes, epistemically or evidentially, to form a justified belief. One can have propositional justification to believe *p* while lacking any belief that *p*, or while believing *p* in such a way (e.g. for the wrong reasons or no reasons at all) that one lacks doxastic justification. If one has, all things considered, sufficient evidence, or good reason, to believe *p*, then one has propositional justification to believe *p*. What having evidence is, exactly, and how to properly “base” belief on one’s propositional justification so that one has a justified belief is a complicated issue (see Turri 2010) that we will not take up.

One puzzling feature of debunkers is their relation to propositional justification. Similar puzzles apply to learning that you’ve been under the influence of a reason-distorting drug. Even when you learn that you were under this influence, your original conclusion may still *seem* right, and you may still seem (to yourself) to possess *the same* arguments and evidence. This might tempt one to suppose that propositional justification is left undefeated. Rather, and oddly, one becomes unable to have a doxastically justified belief on the basis of this propositional justification once one learns about the drug or the directional influence.

Here, a second distinction, due to Goldman, might be helpful: “The *ex post* use [of justification] occurs when there exists a belief, and we say *of that belief* that it is (or isn’t) justified.” (1979, 21) “The *ex ante* use [of justification] occurs when we . . . ignore . . . whether such a belief exists. Here we say of the *person* . . . that *p* is (or isn’t) suitable for him to believe.” (21) *Ex post* justification is, for our purposes, equivalent to doxastic justification. But we remain neutral on the relation between propositional and *ex ante* justification, because we are

noncommittal about whether, or how, propositional justification is defeated in such cases. A debunker, as we have defined it, certainly concerns *ex ante* justification: whether *p* is a suitable thing for you to believe is in question, regardless of whether you actually believe it any longer. We remain neutral on propositional justification because it isn't clear whether one's *evidence* for the target belief is reduced. What is clear is only that one's ability to properly *handle* the evidence is now under suspicion; one loses trust in one's reasoning. Those who think that this can only happen when one's propositional justification has also been defeated may interpret this accordingly, and forgive our overabundance of caution.<sup>39</sup>

The forgoing detail explains an element of debunkers that has been noticed in the literature. Elga (ms), for example, notes that defeaters of this sort involve alienation from one's own reasoning: *from the subject's perspective*, there may still *seem* to be the same degree of evidence for the original belief, though the subject feels that she cannot trust the evidence, or not in the same way or to the same degree as before. This sort of alienation from one's (seeming) evidence may strike one as impossible, and something along these lines has inspired an objection to debunkers.

White (2010) considers my belief that my fuel tank is full, where my only evidence is that my fuel gauge indicates "Full." If I am now told that the fuel gauge is stuck at "Full," the justification of my belief is undermined because my justification was solely grounded on a datum that is not an indicator of the truth. White plausibly suggests that there is an alternative explanation of why the gauge reads "Full," one that, as he writes, "screens off" (585) the justificatory import of my gauge reading "Full" to my belief. But, White argues, this model doesn't apply to the case of irrelevant influences. Unlike in the fuel gauge case, I don't regard my believing that *p* as an indicator of *p*'s truth: to believe *p* is to take *p* to be true. I don't adopt such a weirdly alienating attitude to my own belief states. As White puts it: "I don't think over whether *p* and then upon coming to the conclusion that *p* think, '*p*, and now I believe that *p*. Smart fellow that I am it is unlikely that I would believe that *p* if it wasn't true. So this is further evidence that *p*,' and thereby increase my confidence that *p*" (585). Moreover, even if this were an appropriate way to reason, it would show that an alternative explanation of my believing that *p* would at best undermine this additional "extra bump" I get from the consideration that I believe that *p* (*ibid*).

Let's grant that I do not regard my believing that *p* as further evidence for *p*. Still, this objection, if applied to our account of irrelevant influences as debunkers, seems to presume an inapt analogy in the fuel gauge scenario. What we've argued above is that my *reasons* or evidence for believing that *p* are screened off from providing a basis for believing that *p*. We have not argued that my believing that *p* no longer indicates to me that *p* is true. The apt analogy is that the needle on the gauge is my evidence, not my belief. There's nothing *absurdly* alienating about taking (or not taking) my reasons to be good bases for believing that *p*.

Before considering other objections, we must qualify our conclusion in two ways. First, we have argued that *directionally influential* desires are debunkers,

but the presence of desires *in general* is another matter. In particular, we have no objection to the idea that desires can sometimes have a salutary effect on our epistemic status.<sup>40</sup>

It is widely accepted that motivational states figure broadly in both decision-making and reasoning.<sup>41</sup> The conceit that we are better deliberators without the influence of desire or emotion has been called “Descartes’ error” (Damasio 1994). Indeed, desire may well be the key to, for example, the persistence required for discovery. So, the role of desire in inquiry and belief formation is not invariably pernicious, and can be a positive practical and epistemically unproblematic influence. Such positive influences are *positional*. For example, one’s interests can result in persistence and this may result in possession of evidence one would otherwise lack. One sometimes presses on in hopes of discovery due to one’s desires. Such positional influence may well have a salutary effect on inquiry (e.g. in President). So, the influence of desires is not always unhappy. We have argued only that when one gets evidence that a desire *has played a directional role*—which is often the intended upshot of charges of irrelevant influences like *Oxford* and *Catholic*—one ought to reduce confidence in one’s belief.

We will discuss the difficulty of distinguishing the salutary from the pernicious role of desire, from the subject’s point of view, in the next section. For now, we note that, since the presence of a strong desire (obviously) raises the likelihood of that desire having a directional, pernicious influence, evidence of such a desire is, in the absence of further information, evidence for directional influence. Thus, when one is accused of having strong relevant desires, one is in effect being presented with purported evidence that one’s reasoning is directionally biased.

The second qualification is that, like any defeater, debunkers are themselves defeasible. That is, in answer to a charge like Catholic, one could appeal to further facts that undermine the force of this attempted underminer. One may have done one’s *epistemic due diligence*: consulted uninterested parties, sufficiently re-examined the evidence, even undergone psychotherapy to rid oneself of one’s bias, etc. This could, in some cases, do much to weaken the force of the debunker. This point serves to highlight the fact that defeaters in general, and debunkers in particular, are matters of degree. If the implicated desire or its effect on my overall belief-forming process, all things considered, is rather weak, then the expected reduction in my reliability will be proportionately modest. It may, in some such cases, be reasonable to retain *most* of one’s confidence. This will of course depend on the details of the case.

Consider the case in which you believe your daughter is the best dancer in the troupe. Someone challenges you: you just believe that because she’s your daughter. This presents a debunker, or a *prima facie* reason to reduce confidence. But, you might reply, you’ve anonymously solicited unbiased (unrelated) judges and they’ve all agreed your daughter is the best; all the other dancers’ parents also think that your daughter is the best; you’ve pored over data about audience reactions to the various dancers, and you’ve had yourself hypnotized to forget that she is your daughter, and then were asked who you thought the best dancer

was, and still, she seems best. In short, you've done your epistemic due diligence. If this is all the case, it seems you've done much to defeat the debunker. Similarly, you might reply to the challenge that you've been given a reason-distorting drug by pointing out that you've taken an antidote.<sup>42</sup>

However, the above scenario is rather artificial, and people can seldom (honestly) appeal to such due diligence. Wilson and Brekke (1994) and others (e.g., Wilson, Centerbar, and Brekke 2002, Ahlstrom-Vij 2013) are generally pessimistic about defeating debunkers.<sup>43</sup> They think the cognitive resources required for this are immense, and that the desire that potentially distorted your reasoning in the first place also potentially distorts your attempt to compensate for the bias. If this is right, then the charge of irrelevant influence, when it implicates a debunker, will typically undermine your justification to some extent even when you've made efforts to undo the bias. In the next section, we address this pessimism about defeating debunkers as an objection to our view.

## **VI. Objections considered**

We have seen that one way to counter the charge of irrelevant influence, understood as a debunker, is to claim to have done one's epistemic due diligence. But one could doubt whether such attempted rebuttals ever succeed. Whether such pessimism is ultimately warranted is a good question, which we cannot address here. For the sake of argument, we can grant the pessimism to see where it leads. What follows is that one cannot avoid some defeat, or reduction in confidence, in light of evidence for directionally influential desire. Though this is confined to cases in which one learns about, or is accused of, some potential directional influence, it may cause worry about our view.

The worries, or objections, depend on a couple of assumptions: First, there is always a desire within the vicinity of an investigation. Why would I ever bother trying to settle any question that bore no relation to my desires? Second, if desire is always present, so is the potential problem of directional influence. To use Wilson and Brekke's nice turn of phrase, how can we tell if "mental contamination" (1994) is present? And as Paul Thagard remarks, in contrasting a positive from a pernicious role for emotion, "I know of no psychological way of distinguishing between the positive emotion of emotional coherence and the similar emotion of self-promotion" (2006, 257). From the first-person perspective, desire's directional influence is typically invisible; directional influence can often be subjectively indistinguishable from positional influence.

If desire is ubiquitous, and if every desire is potentially directionally influential, then one should *almost always* reduce confidence. So, the worry goes, irrelevant influences point to general, radical skeptical considerations. In that case there is not much new of interest here, just another radical skeptical puzzle. Furthermore, it seems to follow that if I took challenges such as *Oxford* and *Catholic* seriously, I'd have to reduce confidence in everything I believe, and of

course no one should be expected to do that. Any specific charge of irrelevant influences is just one instance of a general phenomenon, rather than being evidence for some specific problem with a particular belief. So irrelevant influences don't pose a unique, significant problem after all.

The objection's appeal rests on a conflation of the *intensity* of our desires in some cases (and the consequent immensity of the influence they can have on our beliefs) with the *ubiquity* of such desires that can significantly affect our beliefs in this way. The objection correctly observes that, in any given inquiry that could be directionally influenced, there *might* be a hidden, directionally influential desire at play, even if it seems to the inquirer that there is none. But this is not equivalent, in its epistemic import, to a situation in which an inquirer has *sufficient evidence for* believing that there is such an influence. There are two points here: the insufficiency of mere possibility and the lack of sufficient evidence. We take these up in turn.

First, the *mere possibility* that, on any given topic, I have a directionally influential desire is no obstacle to our view. The similarities between irrelevant influence and general, radical skeptical considerations that appeal to merely possible scenarios are few indeed, despite some suggestions in White (2010) and Elga (ms). For, the scope, source, and remedy for debunkers are significantly different from general skeptical worries. Consider scope: Global skepticism purports to defeat justification for all ordinary, perceptual beliefs. But directional influence is not a real possibility in many cases, including ordinary, perceptual beliefs. For example, though I may desire that the cup on the table be filled with coffee, my usual method of ascertaining whether that is the case is not, as far as any of the empirical data suggests, susceptible to directional bias. Perceptual beliefs and their immediate entailments seem largely immune.<sup>44</sup> So, we can set aside a vast swath of beliefs from the "skeptical" worry induced by irrelevant influences.

The source of debunkers is also unique. They do not defeat due to the *mere possibility* that we are in a situation in which our beliefs are false or unjustified. Rather, debunkers arise from *positive evidence* that something is epistemically amiss.<sup>45</sup> Debunking happens when, on the basis of learning about our upbringing, environment, or other irrelevant factors, we gain some evidence that our belief forming process is unreliable. This is analogous to *getting evidence* that you have been given a reason-distorting drug, not to merely noting that it is possible that you have been given a reason-distorting drug. That I just believe this because she's my daughter is analogous to a situation in which I learn of some evidence that I am a BIV, not a situation in which I appreciate that I merely could be a BIV. So, the source of skeptical worries, insofar as they are mere *possibility* claims, is entirely different from the source of debunkers.

Relatedly, the remedies one might suggest for skeptical worries are entirely different from those that may assuage worries due to debunkers. Skeptical worries would go away if we accepted the idea that we have a default entitlement to believe that our faculties are reliable, so long as we lack evidence to the contrary.<sup>46</sup> But this does nothing to address the charge concerning a belief about a daughter's

dancing ability, which *does* appeal to evidence that one is unreliable. Rather, “you just believe that because you’re her father” might be answered by an appeal to epistemic due diligence—perhaps the father has checked with experts who have no personal connection to his daughter and they agree. Radical skeptical arguments leave no room for appeal to such contingent facts in defense of one’s beliefs, e.g. “but all the other people around me say they agree that here is a hand.”<sup>47</sup> So the potential remedies for debunking and global skeptical worries are entirely different. Debunkers are not radical skeptical worries, and their epistemic significance does not stem from the mere *possibility* that we are less reliable given the presence of a desire.

Setting the mere possibility of directional influence aside, the second point in our reply concerns sufficient evidence. Recall that the degree to which one should reduce confidence in the cases we have been discussing is proportional to the strength of the evidence that one’s desire is likely to have played a directional influence and the magnitude of the influence thereby indicated (which is often proportional to the strength of the desire). Let us grant that many educated people know that we have desires of which we are unaware, and that those desires can bias our reasoning. So we have *some* background evidence that, for any given belief, such desires exist. Still, on any arbitrarily chosen question, the chances that I have a *strong* desire about the issue, and that this desire has had a *strong* directional influence is, in the absence of any other evidence, relatively small. There is no general evidence that we typically have strong, directionally influential and inaccessible desires on most topics of investigations. So, the level of defeat for any random belief and absent specific information about the case is rather low. Our view does not call for any *substantial, general* reduction of confidence across the board, and whatever slight reduction of confidence is called for is hardly an extreme revision of current practice on most topics. Debunkers, as we have described them, constitute significant defeaters only when there is good, specific evidence for a strong, directionally influential desire in a particular case. That we sometimes have hidden and directionally influential desires merely calls for some caution along the lines of: “Who knows? I may not be as reliable about this as I seem to be right now.”

Furthermore, though desires can be hidden and potentially directionally influential, the desires threatened by, e.g., fear of spousal infidelity, fatal illness, professional incompetence, and the like are far more likely to be directionally influential in reasoning than, say, fears about the weather that threaten my going on a picnic tomorrow. Granted, given our imperfect reliability about our own desires even in the latter sorts of cases, concerning the weather, there is *some* low, background chance that a desire has been perniciously involved. Still, the corresponding reduction in confidence will likewise be low, since all indications show merely that there is a slightly increased chance that you are unreliable. (For this reason, such typically unproblematic desires do not often figure in charges of irrelevant influences.) The recommendation for some mild humility in light of

the fact that our processes are not always transparent is hardly shocking, and not at all implausible.<sup>48</sup>

Another, altogether different objection encountered in the recent literature, for example in White (2010), Elga (ms), and discussed in Schoenfeld (2014), is that the challenge of irrelevant influences cannot be distinguished from the challenge of disagreement. All one learns in such cases is that other people, who aren't similarly influenced, do (or perhaps could) disagree. Therefore, irrelevant influences do not present a *unique* epistemological category of defeat. After replying to the content of this objection, we will attempt to diagnose it.

Disagreement and debunking have a different extension: there are cases of debunking without disagreement, and disagreement without debunking. As Vavova (ms) points out, you can learn that your upbringing, environment, or other causal factors make it more probable that your reasoning was biased without thereby learning that anyone disagrees with you on the issue. For example, the relevant belief could be on a subject matter that no one else alive has ever considered. (And, presumably, you knew before and independently of learning of your bias that one *could* disagree with you.) Conversely, there can be cases of disagreement (and certainly of *possible* disagreement) that arguably involve defeat of justification without any appeal, implicit or otherwise, to directional influence. Christensen's (2007) example of a disagreement about what a 15% tip amounts to on a bill, for instance, is a case in which neither of the disputants have an antecedent desire about the tip amount. So, since there can be debunking without disagreement, and vice versa, the two are not the same phenomenon.

Furthermore, the nature of defeat by disagreement, if there is such a thing (as some philosophers doubt) seems different from the nature of debunking. There are at least some cases of disagreement in which one believes *p*, learns that another person believes not-*p*, and this constitutes evidence for not-*p*. But this seems never to be the case with debunking. That one's desires have biased one's enquiry about *p* does not constitute evidence for not-*p*; it undermines rather than outweighs.<sup>49</sup>

However, there is some connection between disagreement and debunking, as some philosophers (Vavova (ms) and Mele, implicitly, in his (2001)) have pointed out. Learning that someone else disagrees can sometimes be a good indication that one's belief formation was the result of irrelevant influences, and this may, depending on the details, defeat one's justification. So, disagreement can have a *heuristic* relation to debunking. Similarly, when irrelevant influences have led one to form a belief under directional influence, it should be expected that one will disagree with others who are not under the same influences. This relation between disagreement and bias, along with the conflation of positional and directional influences, may go some way to explaining why some philosophers have conflated the two.

The final objection we will consider, expressed by White (2010) in writing and by many others in conversation, is that when irrelevant influences are brought to bear on the justificatory status of a belief, this is *merely* a matter of one's

reassessing the evidential support for the belief. If so, there is nothing unique about debunkers. As White puts it, "One can be inspired to reassess one's beliefs in all sorts of ways. The interesting question is whether the causal background of these beliefs can have epistemological relevance itself." (p. 3)

Facts about irrelevant influences on your belief can lead you to reassess your evidence, without introducing new first-order evidence or suggesting that you lack evidence. Rather, the reassessment concerns the way that you have managed your evidence, and seems to be a unique sort of reassessment, one with its own set of challenges and problems (see Wilson and Brekke (1994) and Kornblith (1999)). Once I'm considering genealogical worries about a belief of mine ("Perhaps I really shouldn't have given so much weight to Hilda's remark," "How could I have simply ignored Greta's remark?" "Why didn't I consider the possibility that . . .?"), I am in a process of re-evaluating the epistemic credentials of my belief. But I am still doing this *on the basis of* a concern about the etiology of my belief, even if the reassessment deals, at this point in the process, with a synchronic worry about my current evidence. The fact that my desire has the content that it does will, presumably, play a role in my reassessment, since the desire determines a potential target in directionally influenced reasoning. And the assessment of how likely it is that the irrelevant factor (e.g. my upbringing) has given rise to a powerful desire is also an essential part of the reassessment, so my considerations involve the details of the etiological worry. That is, the reassessment must involve consideration of whether and how the irrelevant influence (e.g. one's upbringing) led to one's handling the evidence in the way that one did. Since this essentially involves genealogical factors, it is not just any sort of reassessment of one's evidence.

So, reassessment of evidence due to genealogical worries of the sort implicated in charges of irrelevant influences is essentially different from just any reassessment of evidence; we still have a unique sort of epistemic worry. Furthermore, it should never have been expected, and is not a condition on our interest in the topic, that irrelevant influences defeat justification in a *brute* or *immediate* way, unmediated by any further epistemological story whatsoever. Likewise, if you learn that you formed your belief at 21,000 feet, you may come to suspect that you were a victim of hypoxia (Elga, ms), and this should lead to some reassessment of the evidence for your belief. The source, in this case, of your epistemic worries is the evidence for hypoxia, so the answer to the question "can information about the causal origins of your beliefs having to do with the oxygen levels defeat your justification?" is clearly "yes." Our view is that, similarly, information about irrelevant influences on your belief can defeat your justification, at least when those influences indicate that your belief formation was conditioned by a directionally influential desire.

Whether irrelevant influences are *sui generis*, or best construed as one of several different ways to undermine justification, has not been our concern. Rather, we have been arguing that desires as irrelevant influences can be defeaters, and they do not defeat in the same way as skeptical defeaters or defeaters

due to disagreement. What we hope to have established is the epistemological significance of debunkers.<sup>50</sup>

## Notes

\*Earlier versions of this paper were presented to the Epistemology Workshop at the University of California, Irvine and to the Claremont Philosophy Work in Progress Group; we are grateful to all of the participants for their comments. We owe special thanks to Sinan Dogramaci, Paul Hurley, Dustin Locke, Al Mele, Karl Schafer, Katia Vavova, and Rivka Weinberg for their comments on various drafts of the essay.

1. See Christensen 2010 for a recent discussion of such “higher order” defeat. The idea that evidence for one’s own unreliability constitutes a defeater has been disputed elsewhere in the literature. See Horowitz 2014 for a recent, critical discussion.
2. By ‘genealogy’ we do not mean an influence that *determines* with absolutely certainty what one will believe. Rather, we mean it as a type of influence on belief which often probabilifies, in ways we discuss at length below, the formation of certain belief.
3. Schoenfeld (2014) defends a similar thesis.
4. See Schecter (ms) and Vavova (ms) for direct objections to Elga. They propose different accounts of irrelevant influences as defeaters.
5. Cohen (2000), Sher (2001).
6. Plato seems to have had this issue in mind, and in quite generalized fashion, in the *Phaedo*, when he recounts Socrates’ remark: “[E]verywhere in our investigations the body is present and makes for confusion and fear, so that it prevents us from seeing the truth” (103d). Nietzsche’s account of the genesis of slave morality is a rather striking case of the how such a desire- or interest-based genesis of beliefs might undermine the justification of certain claims. For a reading of this case that’s rather close to what we have in mind see Wallace (2007). Bacon (1620/2000, 44) warns against the “stain” of emotions on the understanding and the ways in which desires and hopes create “fanciful” beliefs. And Mill (1859/1978, 6) highlights the role of desire in generating moral and political opinion, citing peoples’ “desires and fears for themselves.” Mill suggests that the resulting opinion has a mere chance relation to truth, an idea that we detail below.
7. See Cohen (2000).
8. Other appeals to nurturing doxastic environments might amount to the charge that you believe that p because you’re gullible or because you haven’t thought of all the options yet or because you took a mind-altering drug. None of those involve desire, but some of them might well, in certain contexts, express genuine defeaters. Still, even though they sometimes undermine, since they don’t implicate desire, we won’t discuss them in any detail.
9. See Mele, (1997) and (2001).
10. The epistemological significance of motivated, directional reasoning is discussed in Ellis (2014).

11. There is, for example, evidence that retrieval of evidence from memory and the evaluation of evidence varies in ways that are sensitive to the hypothesis we are testing (Trope, 1996; Synder and Unranowitz, 1978).
12. Mele (1997, 94; 1987, 144ff.) gives such an account of the role of motivation in hypothesis testing. Trope, Gervy and Liberman note in this context that “desired possibilities are more cognitively accessible because they are associated with an agent’s goals . . . more pleasant to contemplate . . . [and seemingly] more probable because individuals try to achieve them” (1997, 113).
13. There are many cases in which desire or interest plays such a priming role *repeatedly* during a line of inquiry. As Kunda writes: “[W]hen motivation is involved, one may persist in asking one directional question after another . . . thus exploring all possible avenues that may allow one to endorse the desired conclusions” (1990, 495). In such cases, we may well be powerfully tempted to ascribe a hedonically favored doxastic target to the subject’s hypothesis testing. This is one way in which the boundary between positional and directional influence is not at all straightforward.
14. We say “range of targets” rather than “target” because, for example, Sid may, notwithstanding his powerful desire that Roz love him, be unable, in current circumstances, to confirm a belief with such a content. As Kunda often emphasizes, subjects are constrained by what they can plausibly justify to themselves (1987).
15. Though we cannot pursue it in detail here, a simple priming model of motivated reasoning also appears inadequate for cases of twisted self-deception (Mele, 1997) or unwelcome believing (Scott-Kakures, 2000). The pragmatic hypothesis testing account described below can make good sense of such cases.
16. This is not uncontroversial. On the dispute between intentionalists and deflationists about self-deception, see Mele (2001, ch. 2).
17. Mele terms this the “FTL” account. For more, see Mele’s (1998) application of Trope and Liberman’s account to the issue of agency in self-deception.
18. Where there is little at stake, there is unlikely to be much cause for intensive testing. In such cases, subjects do appear to utilize quick and dirty heuristic strategies (Harkness, DeBono, and Borgida, 1985).
19. This term is Trope and Liberman’s (1996); Friedrich (1993) terms a similar account “Primary Error Detection and Minimization” (PEDMIN).
20. In some cases, asymmetric confidence thresholds may seem to reflect a reasonable demand for more evidence in situations in which the stakes are very high, without contributing to directional bias. For example, suppose you learn that your friend of many years is accused of some shocking behavior. The stakes are high: If he’s guilty, your friendship, an important relationship, will be over. So, it seems reasonable, and not necessarily a sign of directional influence, if you require more evidence that you otherwise would to avoid a “false positive.” Our account is compatible with this, once it is clarified what it is you are demanding more evidence *for*. If your interests merely (i) raise the threshold for acting (with no further inquiry) as a result of error costs, this may be reasonable and need not constitute directional influence on our account. For, in this case one may well have adjusted one’s confidence to one’s evidence in entirely appropriate ways. But if, instead, your interests (ii) alter your view of what a given body of evidence establishes, this constitutes a directional influence. This, we think, is not reasonable. In (ii), you raise or lower acceptance and rejection thresholds

depending on your asymmetric error costs. In (i), these error costs do not shift the threshold for accepting or rejecting the hypothesis; what is altered is how much evidence you demand in order to *take action* given your current regard for that hypothesis.

21. Again not all motivationally biased reasoning is also directionally biased. One may be powerfully motivated to bring one's inquiry to a close and so conclude inquiry precipitately; i.e., one may be powerfully motivated to get an answer to a question rather than some particular answer. See Kruglanski and Webster (1996).
22. See Mele (2001, pp. 38ff.) for a discussion of competing forms of motivation in hypothesis testing.
23. It's an open question as to the psychological basis of an inquirer's motivation to avoid her primary error. One of us (Scott-Kakures, 2009) has argued that a plausible account can be mustered according to which "Primary Error Detection and Minimization" motivation is an artifact of cognitive dissonance reduction.
24. See Mele (2001, 106-107).
25. A few philosophers (e.g. Kelly (2005)) have disputed this, though we cannot address this here. See Horowitz 2014 for critical discussion.
26. See Steele (1988) and Aronson (1968).
27. See Nichols (forthcoming) and Van Fraassen (1988).
28. The general question of what constitutes evidence that one's belief is directionally biased is an important and complicated one that we cannot take up here. The specific evidence we are discussing is that one formed one's belief while having a powerful and relevant desire. This evidence can come in the form of the claim that the belief is due to an upbringing that makes certain relevant desires more probable, as we discussed above.
29. A "blocking" defeater (White (2010)) indicates that p was never justified in the first place. We won't distinguish this from undermining defeaters, since blocking defeaters are a subclass of undermining defeaters.
30. Vice versa is more vexed. Arguably, if you obtain an outweighing defeater, you learn that not-p, and this seems to constitute evidence that your source of evidence for p is misleading, and accordingly this constitutes an undermining defeater.
31. The term 'debunking' is sometimes used to signify an objection due to evolutionary features of the species of the believer. For example, debunking moral judgments involves the claim that the information that humans evolved to make (or be disposed to make) such judgments due to reproductive pressures rather than due to the truth of such judgments constitutes a defeater for such judgments. We are not using the term in this way here, and are remaining neutral on that topic.
32. The qualification 'from the subject's perspective' is necessary because it is possible that the subject's reasoning, when unbiased, is less reliable than chance. In such cases, the bias would improve reliability. However, the subject cannot *regard* her reasoning as less reliable than chance while regarding the resulting belief as justified. We want to set such bizarre cases aside because cases of irrelevant influences involve a thinker who doesn't take her own reasoning to be unreliable.
33. And vice versa, for cases of unwelcome believing.
34. There may be rare cases in which one's desires make one's belief forming process more reliable. For example, suppose that everything the Catholic Church says is true, and one's Catholic upbringing causes one to have directionally influential

- desires that favor believing what the church says. Still, that one's beliefs were formed under directional influence is evidence (albeit misleading, in this sort of case) that one's process was unreliable. Even reliabilists, such as Goldman (1986), hold that a reliable process does not lead to justification if (misleading) defeaters are ignored.
35. Vavova (ms) discusses this sort of case.
  36. It might be suggested that, having already formed the belief that *p*, and then having learned that one has a strong directionally influential desire that *p*, one should conclude that, at least in this case, one's desires have helped one to get at the truth. We assume this sort of bootstrapping is irrational, at least in cases not involving fundamental or basic beliefs (e.g. that one's faculties are reliable). Otherwise, similar maneuvers would be available to counter *any*, even outweighing, defeaters. For example, when you believe that *p*, and when a trusted friend presents you with new, independent evidence that not-*p*, it is surely not an appropriate response to say, "But *p* is true, so the new evidence is misleading." This would make it impossible to ever correct one's view in light of new evidence, which is presumably not a rational way to believe.
  37. This is not to say that in cases of directional influence I am wholly disconnected from how things stand in the world. Again, as Kunda emphasizes, motivated reasoning is constrained by what it is that we can plausibly justify to ourselves.
  38. Or, rather, we might say, there's no good reason for believing that in such circumstance scrutinizing our reasons will make it more likely that we will get at the truth.
  39. The case of gullibility is similar in some respects. Suppose you learn that you are extremely and unusually gullible with respect to *p*. Your evidence for *p* might still seem to be there, even though you no longer regard this as an indication that you should go on believing. So, you appear to have an *ex ante* defeater but not necessarily a propositional justification defeater.
  40. Some philosophers even argue that the influence of desire is the typical justifier for certain beliefs. For example, Alvin Plantinga (2000) has argued in various works that a natural desire to believe in God is, effectively, God's way of appearing to us, and justifies religious belief. Others (e.g. Schafer 2013) have argued for similar theses concerning morality and the appearance of "the good" through desire. We won't take up these special cases here, but note that we are focusing on standard cases such as *Oxford*, for which no story of "the appearance" of truth is available.
  41. See Owens (2000).
  42. Relatedly, the charge of irrelevant influences, understood as implicating directional influences, is just a *prima facie* consideration. Its power to defeat can be countered or entirely eliminated if the relevant desire, or the factors that were thought to make its presence probable, is not actually there. In response to *Catholic*, a decisive reply is "No, I was raised Jewish." This is not a matter of doing one's epistemic due diligence; it's a matter of whether the implied influence was there in the first place.
  43. For example, Ehrlinger, Gilovich and Ross (2005) argue that one is more likely to take one's own desires as helpful with respect to the aim of believing truly. Some of their considerations could suggest that one is more likely to regard one's desires as playing a positional rather than a directional role. The upshot is that we are often blind to the actual influence of our beliefs, making compensating

- for the distorting effects of desires more difficult. As Van Fraassen (1988, 145) dramatizes this, “how shall we find a spot of safety?” We address this sort of concern, taken as an objection to our view, in the next section.
44. Though see Trivers (2011) for some surprising examples of bias (though not exactly motivational) about perceptual beliefs.
  45. Vavova (ms) and Schechter (ms) make a similar point.
  46. See Wright (2004) for this view on perceptual faculties, and Foley (2007) for a more general version.
  47. But see Moore (1939) for something remarkably similar to this idea. Needless to say, many epistemologists reject this as a legitimate way to address radical skepticism.
  48. Another allegedly distinct basis for holding that the debunker we describe results in general skeptical worries might appeal to epistemic desires, e.g. the desire to have an answer to a question, the desire for knowledge, the desire to have the truth, or the desire not to be wrong. It might be argued that such desires constitute something like a general motivational background in cases of reflective inquiry. If so, would not the presumptive presence of such desires deliver fully general grounds for reducing confidence in such beliefs? But, firstly, the contents of such desires appear to provide a basis for reassurance. After all, a desire for the truth is not (from the first person perspective) a desire that a particular state of affairs obtain. Indeed, desires with such contents seem to be well suited as the motivational engines of the disjunctive aim that, as we have argued, is distinctive of accuracy guided reasoning. So, the presence of such a desire does not, by itself, indicate the presence of a debunker. The influence of such desires, given their contents, is positional. This is one reason why the charge “You just believe that because you want the truth!” without some further story, is a non-starter. Of course, such desires can give rise, during the process of inquiry, to desires that would, were one to become aware of them, constitute positive evidence of a debunker. Perhaps one’s desire to have the truth is so powerful and the difficulties of securing it so patent that one comes to place one’s hopes fully in securing the truth via theory T. In that case, the presence of a desire for T’s adequacy may well constitute reason for reducing confidence.
  49. On some strictly formal accounts of evidence, this very intuitive claim requires a little nuance. Learning that you were biased may result in lowering credence in *p*, which effectively raises credence in not-*p*. This might be taken to constitute evidence for not-*p* on some definitions of ‘evidence’. But surely there is an important and intuitive difference between lowering confidence in *p* because some purported evidence for *p* has come into doubt, and lowering confidence in *p* because some first-order information to the effect that not-*p* is true has come to light. We take it that any plausible model of evidential relations must mark this distinction.
  50. In this paper we have not addressed the issue of self-consciously self-induced beliefs (such as the case of Pascal’s Wager), though White ends with such cases (609-10). Here, the relevant desire is to *believe that p*, rather than that *p*. This may result in manipulation of exposure to evidence, e.g., focusing my attention on data suggestive of *p*, spending time amongst the *p*-believers, etc. If so, the chief upshot will be positional influence, and so is not necessarily epistemologically worrisome. This is White’s conclusion. But cases of self-induced beliefs can involve directional

influences too: the desire to believe *p* could be the upshot of a desire that *p* (or some closely related content), which itself could be directionally influential. So the positional/directional distinction may also help to explain our ambivalence about self-induced beliefs.

## References

- Alstrom-Vij, Kristoffer (2013). "Why We Cannot Rely on Ourselves for Epistemic Improvement." *Philosophical Issues* 23:276–296.
- Aronson, Elliot (1968). "Dissonance Theory: Progress and Problems." In *Theories of Cognitive Consistency: A Sourcebook*, edited by R.P. Abelson, E. Aronson, W.J. McGuire, T.M. Newcomb, M.J. Rosenberg and P.H. Tannenbaum, 5–27. Chicago: Rand McNally.
- Bacon, Francis (1620/2000). *The New Organon*, Lisa Jardine and Michael Silverthorne, eds. Cambridge: Cambridge University Press.
- Borkenau, P. and A. Liebler (1993). "Convergence of Stranger Ratings of Personality and Intelligence with Self-ratings, Partner Ratings, and Measured Intelligence." *Journal of Personality and Social Psychology* 65:546–553.
- Brown, J. and K. Dutton (1995). "Truth and Consequences: The Costs and Benefits of Accurate Self-Knowledge." *Journal for Personality and Social Psychology* 21: 1288–1296.
- Christensen, David (2007). "Epistemology of disagreement: The good news." *Philosophical Review* 116 (2):187–217.
- Christensen, David (2010). "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81 (1):185–215.
- Cohen, G.A. (2000). *If You're an Egalitarian, How Come You're So Rich?* Cambridge: Harvard University Press.
- Damasio, Antonio, R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
- Elga, Adam. (ms) "Lucky to be rational."
- Ehrlinger J., T. Gilovich, and L. Ross. (2005). "Peering into the bias blind spot: people's assessments of bias in themselves and others." *Personal. Soc. Psychol. Bull.* 31:680–92
- Ellis, Jonathan (2014). "Beyond Disagreement: Directional Reasoning and Other Forms of Higher-Order Evidence." *American Philosophical Association meeting*, April 2014.
- Feingold, A. (1992). "Good-looking People are not What We Think." *Psychological Bulletin* 111: 304341.
- Felson, R.B. (1981). "Self and Reflected Appraisal among Football Players: A Test of the Meadian Hypothesis." *Social Psychology Quarterly* 44: 116–126.
- Foley, Richard (2007). *Intellectual Trust in Oneself and Others*. Cambridge: Cambridge University Press.
- Friedrich, J. (1993). "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena." *Psychological Review* 100: 298–319.
- Goldman, Alvin (1979). What is Justified Belief? In Ernest Sosa & Jaegwon Kim (eds.), *Epistemology. An Anthology*. Blackwell. 340–353.
- Goldman, Alvin I. (1986). *Epistemology and Cognition*. Harvard University Press.
- Harkness, A.R., K.G. DeBono and E. Borgida (1985). "Personal Involvement and Strategies for Making Contingency Judgments: A Stake in the dating Game." *Journal of Personality and Social Psychology* 49:22–32.
- Hart, W., D. Albarracin, A. Eagly, I. Brechan, M.J. Lindberg & L. Merrill (2009). "Feeling validated versus being correct: A meta-analysis of selective exposure to information." *Psychological Bulletin*, 135, 555–588.

- Horowitz, Sophie (2014). "Epistemic Akrasia." *Nous* 48:4: 718–744
- Kelly, Thomas (2005). "The Epistemic Significance of Disagreement". In John Hawthorne and Tamar Szabo Gendler (eds.) *Oxford Studies in Epistemology*, volume 1 (Oxford University Press), pp. 167–196.
- Klein, W. and Ziva Kunda. (1991). "Motivated Person Perception: Constructing Justification for Desired Beliefs." *Journal of Experimental Social Psychology* 28:145–168.
- Kornblith, H. (1999). "Distrusting reason." *Midwest Studies in Philosophy* 23 (1):181–196.
- Kunda, Ziva (1992). "Can Dissonance Do it All?" *Psychological Inquiry* 3:337–339.
- Kunda, Ziva (1987). "Motivated Inference: Self-serving Generation and Evaluation of Causal Theories." *Journal of Personality and Social Psychology* 53:636–647.
- Kunda, Ziva (1990). "The Case for Motivated Reasoning." *Psychological Bulletin* 108:480–498.
- Kruglanski, Arie and Donna Webster. (1996). "Motivated Closing of the Mind: 'Seizing' and 'Freezing'." *Psychological Review* 103: 263–283.
- Lewicka, M. (1989). "Pragmatic Reasoning Schemata with Differing Affective Value of a Consequent Logical Implication." *Polish Psychological Bulletin* 23:237–252.
- Lewicka, M. (1992). "Pragmatic Reasoning Schemata with Differing Affective Value of a Consequent Logical Implication." *Polish Psychological Bulletin* 23:237–252
- Mele, A. (1983) "Self-Deception." *Philosophical Quarterly* 33:365–77.
- Mele, Alfred. (1987). *Irrationality*. New York: Oxford University Press.
- Mele, Alfred (1997). "Real Self-Deception." *Behavioral and Brain Sciences* 20:90–102.
- Mele, Alfred (2001). *Self-Deception Unmasked*. Princeton, N.J.: Princeton University Press.
- Mill, J.S. (1859/1978). *On Liberty*. Indianapolis: Hackett Press.
- Moore, G.E. (1939). Proof of an external world. *Proceedings of the British Academy* 25(5):273–300.
- Nichols, Shaun. (forthcoming). "Process Debunking and Ethics." *Ethics* Vol. 124, No. 4, July 2014
- Owens, David (2000). *Reason Without Freedom: The Problem of Epistemic Normativity*. Oxford: Routledge.
- Plantinga, Alvin (2000). *Warranted Christian Belief*. Oxford University Press: New York.
- Schafer, Karl. (2013). "Perception and the Rational Force of Desire." *Journal of Philosophy* 110 (5):258281.
- Schechter, Joshua. (ms). "Luck, Rationality, and Explanation."
- Schoenfield, Miriam. (2014). "Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief." *Noûs* 48 (2):193–218.
- Scott-Kakures, D. (2000). "Motivated Believing: Wishful and Unwelcome." *Noûs* 34 (3):348–375.
- Scott-Kakures, D. (2009). "Unsettling Questions: Cognitive Dissonance in Self-Deception." *Social Theory and Practice* 35 (1): 73–106.
- Sher, George. (2001). "But I Could Be Wrong." *Social Philosophy and Policy* 18 (02):64–78.
- Steele, C. (1988). "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self." *Advances in Experimental Social Psychology* 21: 261–302.
- Snyder, M. and Uranowitz, S.W. (1978). "Reconstructing the Past: Some Cognitive Consequences of Person Perception," *Journal of Personality and Social Psychology* 36: 941–950.
- Talbott, William. (1995). "Intentional Self-Deception in a Single Coherent Self." *Philosophy and Phenomenological Research* 55:27–74.
- Thagard, Paul (2006). *Hot Thought: Mechanisms and Applications of Emotional Cognition*. Cambridge, MA: MIT Press.
- Trivers, Robert. (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York: Basic Books.
- Trope, Y., B. Gervy, and N. Liberman. (1997). "Wishful Thinking from a Pragmatic Hypothesis Testing Perspective." In M.S. Myslobodsky, (ed.) *The Mythomanias: The Nature of Deception and Self-Deception*. Mahwah: Lawrence Erlbaum.

- Trope, Y. and A. Liberman. (1996). "Social Hypothesis Testing: Cognitive and Motivational Mechanisms." In E.T. Higgins and A.W. Kruglanski, (eds.) *Social Psychology: Handbook of Basic Principles*. New York: Guilford.
- Turri, John. (2010). "On the relationship between propositional and doxastic justification." *Philosophy and Phenomenological Research* 80 (2):312–326.
- Van Fraassen, Bas (1988). "The Peculiar Effects of Love and Desire." In Brian P. McLaughlin & Amelie Oksenberg Rorty (eds.) *Perspectives on Self-Deception*. Berkeley: University of California Press.
- Vavova, Ekaterina. (ms) "Irrelevant Influences."
- Wallace, R. Jay. (2007). "Ressentiment, Value, and Self-Vindication: Making Sense of Nietzsche's Slave Revolt." In Brian Leiter & Neil Sinhababu (eds.) *Nietzsche and Morality*. Oxford University Press. 110–137.
- White, Roger. (2010). "You just believe that because . . ." *Philosophical Perspectives* 24 (1):573–615.
- Wilson, T.D. and N. Brekke. (1994). "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations." *Psychological Bulletin*. 116 (1): 117–42.
- Wilson, T.D., D.B. Centerbar, and Nancy Brekke (2002). "Mental Contamination and the Debiasing Problem." In T. Gilovich, D. Griffin, and D. Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.
- Wright, Crispin and Martin Davies (2004). "On Epistemic Entitlement." *Aristotelian Society Supplementary Volume* 78:167–245
- Wright, Crispin (2004). "Warrant for nothing (and foundations for free)?" *Aristotelian Society Supplementary Volume* 78 (1):167–212.