When and why people think beliefs are "debunked" by scientific explanations for their origins

Dillon Plunkett[1,2,†], Lara Buchak[3], and Tania Lombrozo[1,4]

[1]Department of Psychology, University of California, Berkeley
[2]Now at Department of Psychology and Center for Brain Science, Harvard University
[3]Department of Philosophy, University of California, Berkeley
[4]Department of Psychology, Princeton University
[†]Corresponding author: plunkett@g.harvard.edu

Abstract

How do scientific explanations for beliefs affect people's confidence in those beliefs? For example, do people think neuroscientific explanations for religious belief support or challenge belief in God? In five experiments, we find that the effects of scientific explanations for belief depend on whether the explanations imply normal or abnormal functioning (e.g., if a neural mechanism is doing what it evolved to do). Experiments 1 and 2 find that people think brain-based explanations for religious, moral, and scientific beliefs corroborate those beliefs when the explanations invoke a normally functioning mechanism, but not an abnormally functioning mechanism. Experiment 3 demonstrates comparable effects for other kinds of scientific explanations (e.g., genetic explanations). Experiment 4 confirms that these effects derive from (im)proper functioning, not statistical (in)frequency. Experiment 5 suggests that these effects interact with people's prior beliefs to produce motivated judgments: People are more skeptical of scientific explanations for their own beliefs if the explanations appeal to abnormal functioning, but they are *less* skeptical of scientific explanations of opposing beliefs if the explanations appeal to abnormal functioning. These findings suggest that people treat "normality" as a proxy for epistemic reliability and reveal that folk epistemic commitments shape attitudes towards scientific explanations.

Keywords: Belief Debunking; Epistemology; Folk Epistemology; Explanation; Experimental Philosophy; Scientific Communication

Introduction

Nietzsche (1908) claimed that "comparative ethnological science" definitively explained the origin of belief in God and that "with [this] insight into the origin of this belief all faith collapses" (p. 164). Freud (1927/1961) suggested that religious beliefs derive from wishful thinking, and that recognizing this fact must "strongly" influence our attitudes toward the belief that God exists. More recently, some have argued that belief in God ought to be abandoned in light of theories that suggest religious belief is an evolutionary adaptation (or the byproduct of adaptations; e.g., Bering, 2011). The underlying assumption in each case is roughly this: If some belief (for example, that God exists) can be traced to a process that does not necessarily track the truth—such as wishful thinking or historical accident—then we have reason to doubt that the belief is true.

Philosophers debate whether and when explanations like these—which account for some belief by appeal to psychological, neurological, evolutionary, or cultural processes—in fact challenge the truth of the beliefs that they (purport to) explain (e.g., Joyce, 2006; Nichols, 2014; Singer, 2005; Street, 2006; Wielenberg, 2010; Wilkins & Griffiths, 2013). For example, Nichols (2014) defends what he calls *process debunking* arguments, which reject a belief as unjustified after explaining that it was produced by an "epistemically defective" belief-formation process, such as wishful thinking. But, the idea that an explanation for holding some belief potentially "debunks" that belief is not restricted to academic philosophy; examples from the popular press abound. For example, neuroscientific explanations for religious belief often make headlines, sometimes with an implication that such explanations challenge the beliefs themselves. Consider one newspaper's headline: "She thinks she believes in God. In fact, it's just a chemical reaction

taking place in the neurons of her temporal lobes" (Hellmore, 1998). The implication seems to be that a belief explained by appeal to mere chemistry is somehow defective.

In the current paper, we investigate whether and why scientific explanations for why people hold beliefs can influence confidence in those beliefs. Specifically, are scientific explanations for beliefs "debunking"? We begin with a brief review of philosophical literature on whether and when scientific explanations *ought* to be debunking. We then describe prior empirical work investigating how people assimilate scientific information, as well as research on how new information leads to belief revision. This work provides a broader context for generating hypotheses concerning the case we investigate: the consequences of receiving scientific explanations for belief.

Debunking explanations within philosophy

In philosophy, a "debunking argument" against some claim *X* is an argument that takes the following form (see Kahane, 2011):

> *Premise 1:* Our belief that *X* is true is explained by some process which is not truth-tracking with respect to *X*. (The process would result in our believing *X* regardless of whether *X* is true.)
>
> *Premise 2:* If we learn that we would currently believe *X* is true whether or not *X* is actually true, we should abandon our belief that *X* is true.
>
> *Conclusion:* We should abandon our belief that *X* is true.

For example, if you believe that exposure to sunlight is extremely dangerous, and then learn that you are infected with a virus that causes its hosts to believe that sunlight is extremely dangerous, you no longer have reason to believe that sunlight is extremely dangerous and should abandon

that belief. In brief, debunking refers to challenging a belief by appeal to the process by which a belief is formed, rather than directly presenting counterevidence to the belief.

Philosophers are particularly interested in debunking arguments in the context of evolutionary explanations for moral and religious beliefs. If we can explain our belief that stealing is wrong in terms of the evolutionary fitness of holding that belief, rather than the truth of that belief, then that belief appears to no longer be supported (Joyce, 2006; Street, 2006). And, some argue, all or most moral beliefs can be given such an explanation. The same is sometimes held of religious belief: If a propensity to believe in God is explained by the evolutionary fitness of that propensity (even if God does not exist), we may have greater reason to doubt our own belief in God.

Debate about the success of evolutionary debunking arguments has centered on whether the discovery of explanations for these beliefs really *should* undermine our confidence in them (see, e.g., Copp, 2008; Wielenberg, 2010; Enoch, 2011; Wilkins & Griffiths, 2013; Jong & Visala, 2014; FitzPatrick, 2015). To our knowledge, the corresponding descriptive questions have not been addressed. Do people tend to *think* beliefs are undermined by scientific explanations for their origins? If so, when and why is this the case?

The psychology of "debunking"

Within psychology, research has investigated how to "debunk" scientifically unfounded beliefs, such as the belief that the MMR vaccine causes autism (e.g., Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Importantly, this psychological usage of the term "debunking" is much broader than the target of the current paper. Psychological research focuses on how to bring about belief revision generally, whereas debunking arguments (in the philosophical sense) involve a challenge based on the process by which some belief is formed. This more specific

form of debunking has not been investigated empirically, but the broader body of work on belief revision provides compelling hints about why people might treat scientific explanations for belief as debunking.

First, even young children can track the reliability of an information source in deciding what to believe (Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007). Similarly, adults track the credibility of human sources and are most likely to revise their own beliefs when those beliefs are contradicted by trustworthy sources (Guillory & Geraci, 2013). Moreover, it has been shown (e.g., with mock jurors; Fein, McCloskey, & Tomlinson, 1997) that a particularly effective way to get people to discount information is to make them suspicious that the source provided the information for an ulterior motive. Generalizing from information sources "outside the head" to psychological or neuroscientific belief-formation processes themselves, these findings suggest that a person's confidence in some belief could shift upon learning the belief is tied to a credible belief-formation process or a "suspicious" one.

Second, research on how people update beliefs in light of new evidence has shown that retracting the basis for belief in some proposition *X* does not always weaken people's belief that *X*, and that receiving evidence for some proposition *X* does not always strengthen people's belief that *X*. For example, providing evidence for some position can generate a backfire effect (Cook & Lewandowsky, 2011) or generate belief polarization (Nyhan & Reifler, 2010): Evidence for *X* can lead people to endorse *not-X* more strongly than before (e.g., Batson, 1975; Schwarz, Sanna, Skurnik, & Yoon, 2007). This is especially likely when people have positions that are initially strong and that they are motivated to maintain, such as those that relate to their cultural identity (Kahan, 2010). Given that beliefs about religion, morality, and science—the domains that we

explore here—have the potential to fall into this category, we might expect a debunking argument to *increase*, rather than *decrease*, confidence in the belief that it explains.

In sum, much is known about belief revision in general, but the psychology of debunking arguments is almost entirely unexplored. On the one hand, the literature on source credibility suggests that scientific explanations for belief may be debunking if (and only if) they raise suspicions about the source of the belief (in this case, the belief-formation process involved). On the other hand, research on the backfire effect and belief polarization suggests that debunking explanations could have the opposite effect; this is especially plausible if people take an explanation for belief to be threatening. On either view, it becomes important to identify what it is that makes a belief-formation process suspicious or threatening.

At one extreme, people might take all psychological or neuroscientific explanations for a belief as casting suspicion on the truth of the belief—perhaps because they focus on the proximal basis for the belief, and not on the features of the world in virtue of which the belief is true. At another extreme, people might only treat a belief-formation process as suspicious if it is *explicitly* identified as epistemically defective. This is plausible if the threshold for "suspicion" is high; perhaps the belief-formation process needs to be unequivocally untethered from the true state of the world. As we show below, our participants fall between these two extremes: Scientific explanations are debunking when they explicitly tie some belief to an epistemically defective process, but people are also sensitive to whether the explanations merely imply some epistemic defect by suggesting that the process is not functioning properly (i.e., as it evolved to function). We also test whether these effects depend upon participants' antecedent endorsement of a debunked belief, when it might be most threatening. Our experiments thus shed light on

what it is about scientific explanations for belief that makes them debunking in some cases, but not in others.

Overview of current studies

In Experiment 1, we test whether participants are responsive to explicit information about the epistemic status of a belief-formation process. Specifically, we ask participants how the protagonist of a vignette should respond to a (neuro)scientific explanation for one of his beliefs, where the explanation appeals to a process that is described as reliably truth-tracking or as reliably inaccurate. We find that responses depend on the epistemic status of the mechanisms invoked, with truth-tracking mechanisms reinforcing belief and those that are epistemically defective undermining belief. However, we also find that participants treat epistemically *neutral* explanations for belief as reinforcing. In Experiments 2-4, we therefore narrow our focus to explanations that are epistemically neutral (in the sense that brain regions are not described as truth-tracking). We test and find support for the hypothesis that *normality* in the belief-formation process is treated as a proxy for truth-tracking, where the relevant sense of normality (as shown in Experiment 4) involves proper functioning. Finally, in Experiment 5, we investigate implications for judgments with greater social and practical relevance, including attitudes toward hypothetical scientific discoveries, and we focus on how these interact with participants' antecedent beliefs. (Data and analysis scripts for all experiments are available at http://github.com/dillonplunkett/debunking.)

Experiment 1

In Experiment 1, participants read about a person, Michael, who learns that one of his beliefs elicits a particular pattern of brain activity. They were then asked to indicate how his confidence in that belief should change in response to learning this information.

In the *reliable* condition, Michael also learns that the pattern of brain activity is associated with true beliefs, supporting the inference that his belief was produced by a truth-tracking process. In the *unreliable* condition, Michael learns that the pattern of brain activity is associated with false beliefs, supporting the inference that his belief was produced by an epistemically defective process. Finally, in a *neutral* condition, participants learned only that the observed pattern of brain activity was associated with beliefs in that domain (e.g., religion, for belief in God).

This design had multiple aims. First, the experiment tested the assumption that people are sensitive to explicit information about the epistemic status of a belief-formation process, such that learning that a belief was formed by an epistemically defective process should decrease confidence, while learning that the belief was formed by a genuinely truth-tracking process should increase confidence. While this finding would be consistent with research on source credibility, to our knowledge it has not been shown before. Second, the experiment tested two competing hypotheses about the impact of epistemically neutral explanations: that people view such explanations for belief as irrelevant to the confidence one should have in that belief, or that people find such explanations "debunking."

The beliefs that we employed varied in domain (scientific, religious, moral) and in prevalence (common, controversial). We varied domain to ensure diverse stimulus items and thereby investigate the generality of any effects. Within each domain, we identified one belief that was common (i.e., perceived to be held by most people) and another that was controversial (i.e., with lower perceived prevalence, closer to 50% of the population). This manipulation was motivated by prior work on meta-ethical commitments, which has found that moral beliefs that are widely endorsed are more likely to be treated as objectively true than are controversial moral

beliefs (Goodwin & Darley, 2012; Heiphetz & Young, 2016). In light of this work, we speculated that meta-epistemological commitments might also vary with the (perceived) prevalence of a belief. In particular, it could be that controversial beliefs are more vulnerable to debunking.

We focused on neuroscientific explanations not only because of the attention they garner in the popular press, but also because previous research has found that the inclusion of neuroscientific information can affect how non-experts assess the quality of explanations (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008; Weisberg, Taylor, & Hopkins, 2015; Hopkins, Weisberg, & Taylor, 2016), and can also influence judgments of scientific rigor and moral responsibility (e.g., see Schweitzer, Baker, & Risko, 2013).

Method

Participants.

One-hundred-seventy-three adults (72 female, 101 male, mean age 32) were recruited through the Amazon Mechanical Turk marketplace (MTurk) and participated for pay. In all studies, participation was restricted to users with an IP address within the United States and an MTurk approval rating of at least 95% based on at least 50 previous tasks. An additional 49 participants were excluded prior to analysis for failing to complete the experiment (n = 10), reporting they might have previously participated in a similar experiment (n = 17), or failing a catch question designed to ensure close reading of the materials (n = 22).

Materials and methods.

Each version of the task involved a single target claim from one of three domains: science, religion, and morality (see Table 1). For each domain there were two possible claims, one *common* and one *controversial*. For example, the common scientific claim was "Some

Table 1: Claims used in Experiments 1-3 and the Supplementary Experiment

	Common	**Controversial**
Scientific	Some diseases are caused by microorganisms called 'germs' that can infect a host organism	Humans evolved via natural selection and share common ancestry with many other species
Religious	There is a God	Every person has a soulmate or life partner who has been preselected for him or her by God or some other spiritual force in the universe
Moral	Killing an innocent person is morally wrong	Killing animals for human consumption is morally wrong

diseases are caused by microorganisms called 'germs' that can infect a host organism." The controversial scientific claim was "Humans evolved via natural selection and share common ancestry with many other species." We confirmed in a post-test that participants did think the common claims were more widely accepted than controversial claims (see Results).

Participants first reported the extent to which they agreed with all six investigated claims, as well as six other claims matched for domain and approximate prevalence. For each participant, one of the six claims was then selected at random to be the target claim.

Participants were randomly assigned to one of the three epistemic conditions (*reliable*, *unreliable*, or *neutral*) and read a corresponding vignette. In each vignette, Michael, a participant in a psychology experiment, learns that a region in his brain—the "posterior striatum cortex"—was active when he considered his belief about the target claim. Michael subsequently learns additional information about that region. In the *reliable* condition, Michael learns that the posterior striatum cortex is associated with accurate beliefs. In the *unreliable* condition, Michael learns that it is associated with inaccurate beliefs. In the *neutral* condition, Michael learns only that it is associated with beliefs of a certain kind (moral, religious, or scientific). The vignette in

the *reliable* and *unreliable* condition was as follows (where text specific to this example, a common religious belief, is in bold for the reader):

> Michael decides to participate in a psychology experiment that involves having his brain scanned by a functional magnetic resonance imaging (fMRI) machine. During the scan, the researcher asks him a series of questions, including one about whether **there is a God.**
>
> Michael believes the following claim, and tells the researcher this when he is asked.
>
> CLAIM: **There is a God.**
>
> After the experiment, the researcher tells Michael that there was activity in his posterior striatum cortex when he expressed his belief that **there is a God.**
>
> Michael later reads in a reliable textbook that activity in the posterior striatum cortex is associated with [true/false] beliefs. When a person expresses a belief, and doing so is accompanied by activity in this brain region, the belief is usually [correct/incorrect] (even if the person expressing it has [low/high] confidence that it is true).

In the *neutral* condition, the last paragraph instead read:

> Michael later reads in a reliable textbook that activity in the posterior striatum cortex is associated with beliefs related to **religion**. For example, when a person expresses a belief that **there is a God,** there is usually activity in the posterior striatum cortex.

Next, participants were asked the following question:

> What effect do you think learning these facts should have on Michael's belief about whether **there is a God**? Specifically, should it make him more confident that it is false that **there is a God** or more confident that it is true that **there is a God**?

Answers to this question were given on a seven-point scale ranging from *Much more confident that it is false* (recorded as -3) to *Much more confident that it is true* (recorded as 3).[1]

Participants next reported what they thought their *own* reaction would be if they imagined themselves in Michael's position, and were asked to estimate the prevalence of the six investigated claims among Americans. Issues related to the former questions are revisited more cleanly in Experiments 3 and 5, and are therefore not reported here.[2] The latter questions were included to verify that common claims were thought to be more prevalent than the controversial claims, and this was indeed found to be the case.[3] Finally, at the end of this and all subsequent experiments, participants were presented with an instructional manipulation check

---

[1] We also asked participants how they predicted Michael's belief *would* change (and did the same in Experiments 2-4). "Would" responses were very similar to "should responses" and are reported for all experiments in the Supplementary Materials.

[2] In Experiments 1 and 2, we initially hoped to investigate whether participants would say that their own beliefs should (and would) change in the same way that they thought Michael's should (and would). However, any participants who had the opposite initial belief as Michael (e.g., did not themselves believe in God, but read that Michael did) were then considering two pieces of contradictory evidence (e.g., brain activity associated with false beliefs in one person who believes in God and in one person who disbelieves). Experiments 3 and 5 avoid this issue because participants were asked to consider only a general finding about people who either share or deny their belief (as opposed to specific findings about Michael and themselves).

[3] We assessed perceived prevalence by asking participants to report how many of 100 representatively sampled Americans would endorse each belief. Common beliefs were judged to be significantly more prevalent than controversial beliefs, $t(172) = 29.85$, $p < .001$. (For the common scientific, religious and moral beliefs, the individual means were 86, 67, and 92, respectively, and the means for the corresponding controversial beliefs were 56, 51, and 24.)

(Oppenheimer, Meyvis, & Davidenko, 2009) and asked to provide demographic information and feedback on the experiment.

Results

Effects of experimental conditions.

Responses were analyzed with an ANOVA using epistemic condition (3: reliable, neutral, unreliable) and perceived claim prevalence (2: common, controversial) as between-subjects factors (see Fig. 1). To maximize the number of observations per cell, we collapsed across the three different domains of explained belief (scientific, religious, moral).

These analyses revealed a significant main effect of epistemic condition, $F(1, 167) = 26.29$, $p < .001$, $\eta_p^2 = .24$. Participants in the *reliable* condition judged that Michael's confidence

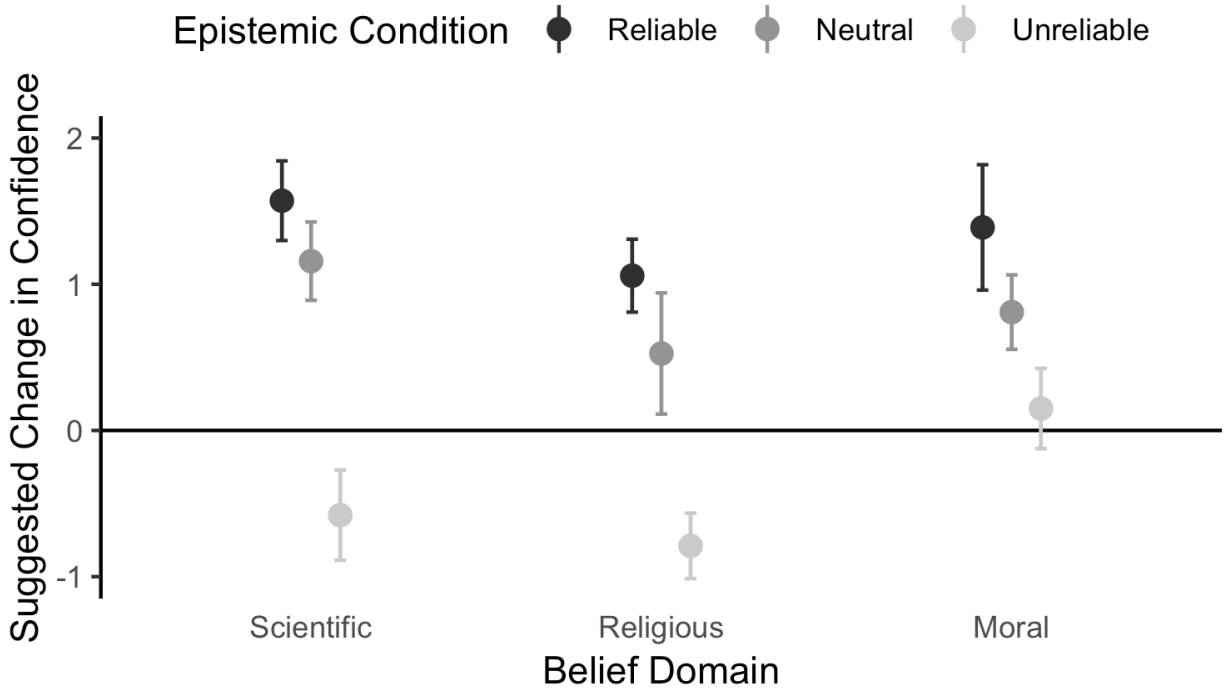

Figure 1: Experiment 1 results (error bars: 1 SEM). Participants indicated that a person should become more confident in a belief associated with a truth-tracking (epistemically reliable) brain region and less confident in a belief associated with a (epistemically unreliable) brain region linked to false belief. However, an explanation that was intended to be epistemically neutral (merely being associated with a region known to be associated with beliefs in that domain) was also judged belief-reinforcing. Results were consistent across belief domains with one exception. Participants did not report that explanations that appealed to an epistemically unreliable process should undermine moral beliefs (e.g., that murder is wrong).

in his belief should increase, while those in the *unreliable* condition judged that Michael's confidence should decrease. Responses in the *neutral* condition fell between these values. All pairwise differences between epistemic conditions were significant ($p \leq .04$). This same qualitative pattern was observed for each belief domain.

There was also a main effect of claim prevalence, $F(1, 167) = 6.87$, $p = .010$, $\eta_p^2 = .040$, with participants advocating less belief reinforcement (or more undermining, in the *unreliable* condition) for controversial beliefs than for common ones. This effect was not replicated in any of our subsequent experiments. There was no significant interaction between epistemic condition and claim prevalence.

Belief reinforcement or undermining.

In addition to comparing responses across conditions, we compared mean responses against the scale midpoint to assess whether different epistemic conditions had reliably reinforcing or undermining effects on belief. Participants in the *reliable* and *neutral* conditions provided ratings significantly above the midpoint (*reliable*: $M = 1.36$, $t(55) = 7.29$, $p < .001$; *neutral*: $M = 0.83$, $t(58) = 4.54$, $p < .001$)—that is, they found the information "belief reinforcing" and thought Michael should become more confident in his belief. In contrast, participants in the *unreliable* condition provided ratings significantly below the midpoint, $M = -0.40$, $t(57) = -2.43$, $p = .018$—that is, they found the information "belief undermining" and thought Michael should become less confident in his belief. These patterns of effects were observed within each of the three domains, except that participants did not think that Michael should lose confidence in moral beliefs, even in the *unreliable* condition.

Discussion

Experiment 1 revealed that participants' judgments about whether a person should adjust his confidence in a belief were appropriately responsive to information about the reliability of the mechanism generating the belief. When a belief was associated with a truth-tracking brain region, participants endorsed increased confidence in the belief; when it was associated with a brain region linked to false belief, participants endorsed decreased confidence in the belief. This is consistent with the literature on source credibility insofar as it suggests that people track the reliability of an information source—even when that source is inside the head.

Curiously, and contrary to both of the hypotheses with which we began, responses in the *neutral* condition followed the same qualitative pattern as those in the *reliable* condition: Information that was intended to be epistemically neutral was taken to be belief reinforcing, a finding that we take up in Experiment 2. The same pattern of responses across epistemic conditions was found for all three domains and for both common and controversial claims (although values for controversial claims were shifted towards lower confidence).

Experiment 2

Experiment 2 examined why neuroscientific information presented in seemingly epistemically neutral terms prompted participants to advise belief reinforcement. Participants in the neutral condition from Experiment 1 were told that Michael had activity in his posterior striatum cortex when evaluating a particular claim, and that the posterior striatum cortex is associated with beliefs in that domain. We hypothesized that participants took this information to imply that Michael's posterior striatum cortex was functioning "normally," or as it should, and that this assumption of proper functioning was treated as a proxy for epistemic reliability, leading to belief reinforcement.

To test this hypothesis, we presented participants with scenarios similar to the neutral condition of Experiment 1, but specified that the relevant brain region was functioning either "normally" or "abnormally." Our prediction was that participants would treat the former as belief reinforcing and the latter as belief undermining.

Method

Participants.

One-hundred-seven adults (40 female, 67 male, mean age 32) were recruited through MTurk. An additional 18 participants were excluded using the criteria used in Experiment 1.

Materials and methods.

The six claims from Experiment 1 were employed in Experiment 2. Participants were randomly assigned to either the *normal* or the *abnormal* condition and to one of the six target claims. Participants were presented with the vignette below. As before, the target belief in this example is the common religious belief, and the text specific to it is in bold:

> A new biotech company is studying a part of the brain called the posterior striatum cortex. The posterior striatum cortex is broadly associated with **religious** beliefs. When an individual expresses a **religious** belief, the posterior striatum cortex is active. However, there is no connection between the exact pattern of activity in the posterior striatum cortex and how confident an individual is in that belief. There is also no connection between activity in the posterior striatum cortex and whether the belief is actually true.
>
> Michael knows all of this information, and decides to volunteer for an experiment being performed by the biotech company. Michael has his brain scanned by a functional

magnetic resonance imaging (fMRI) machine. During the scan, the researcher asks him a series of questions, including one about whether **there is a God**.

Michael believes the following claim, and tells the researcher this when he is asked.

CLAIM: **There is a God.**

After the experiment, the researcher tells Michael that there was activity in his posterior striatum cortex when he expressed his belief that **there is a God**. The researcher also tells Michael that the specific pattern of brain activity observed in his brain suggests that his posterior striatum cortex is working [normally/abnormally].

As in Experiment 1, participants answered the following question about how Michael's confidence in his belief should change.

What effect do you think learning these pieces of information should have on Michael's belief that **there is a God**?

This question was answered on a seven-point scale ranging from *Much less confident that it is true* to *Much more confident that it is true* (again, recorded as -3 and 3, respectively).

Results

Effects of experimental conditions.

Responses were analyzed with an ANOVA using mechanism type (2: normal, abnormal) and claim prevalence (2: common, controversial) as between-subjects factors (see Fig. 2). As in Experiment 1, we collapsed across the three domains of explained belief (scientific, religious,

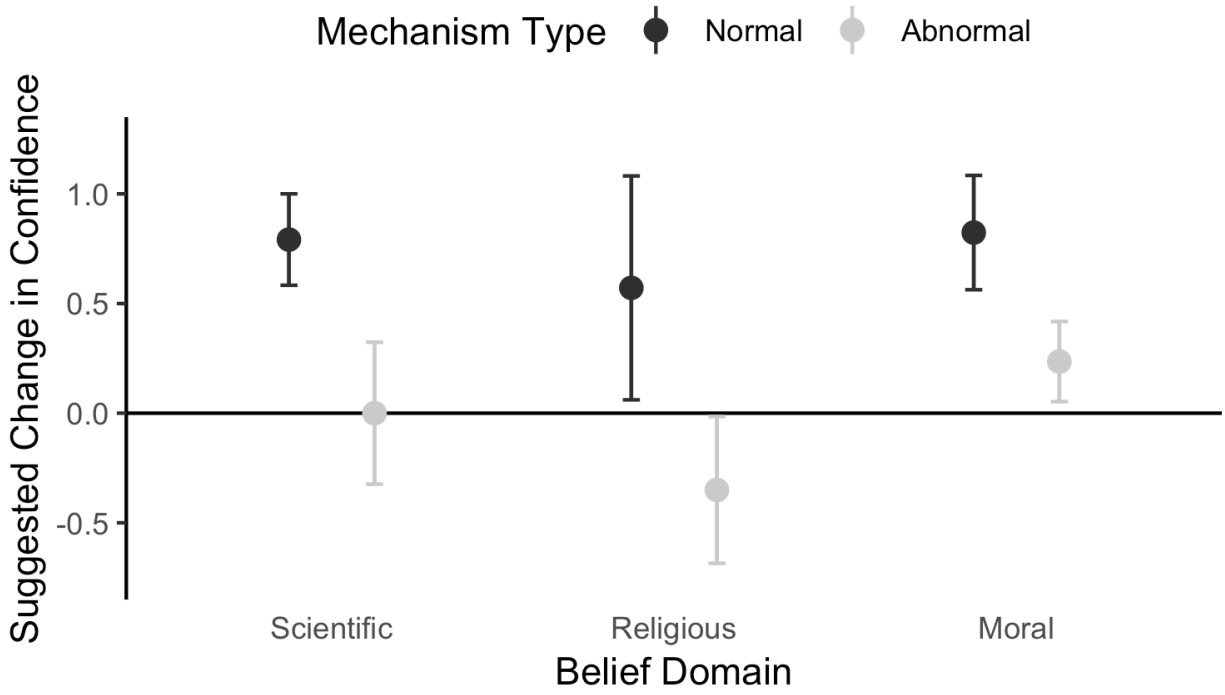

Figure 2: Experiment 2 results (error bars: 1 SEM). Participants reported that a person should become more confident in a belief upon learning that the belief is associated with activity in a brain region that is explicitly described as functioning "normally," but should not if the region is functioning abnormally.

moral). This analysis found a main effect of mechanism type. Participants in the *abnormal* condition were significantly less likely than those in the *normal* condition to judge that Michael's confidence in the target claim should increase, $F(1, 103) = 10.62$, $p = .002$, $\eta_p^2 = .095$. This pattern of responses was consistent across all belief domains. There were no other significant effects.

Belief reinforcement or undermining.

On average, participants in the *normal* condition provided ratings significantly higher than the scale midpoint, $M = 0.75$, $t(54) = 4.27$, $p < .001$. By contrast, participants in the *abnormal* condition did not differ from the midpoint, $M = -0.06$, $t(51) = -0.34$, $p = .736$. Again, these patterns were consistent across domains.

Discussion

Experiment 2 found that explicitly indicating that a brain region was functioning normally led to belief reinforcement, mirroring the *neutral* condition from Experiment 1, in which normal functioning was not explicitly stated, but potentially implied. But when implied normality was contradicted by the explicit statement that an area was functioning abnormally, belief reinforcement was eliminated. These results suggest that participants judge "normal" neurological processes to be more epistemically reliable, a judgment only warranted under substantive assumptions about human belief-formation mechanisms. We return to this point in Experiment 4 and the General Discussion.

Surprisingly, we found that explicit abnormality did not reliably lead to belief undermining. On average, it neither reinforced nor undermined belief. At least two elements of the experiment could explain why this is so. First, although Michael was told explicitly in the *abnormal* condition that the activity in his brain was abnormal, he still received an implicit signal that his brain activity was—in one sense—normal: He was told that a domain-appropriate brain region was activated. Thus, in the *abnormal* condition, Michael received conflicting information about the normality of the brain activity associated with his belief, rather than exclusive indications of abnormality. Second, unlike our previous and subsequent experiments, the vignettes in this experiment indicated explicitly that there is "no connection between activity in the posterior striatum cortex and whether the belief is actually true." Partial deference to this assertion might explain why participants withdrew, but did not reverse, their attitudes towards the epistemic relevance of Michael's abnormal brain function.

Experiment 3

Our first two experiments found that neuroscience explanations that invoke or imply normal functioning are taken to support the beliefs that they explain, and that neuroscience

explanations that invoke or imply abnormal functioning are treated as irrelevant to, or undermining of, the beliefs that they explain. Experiment 3 aimed to replicate and extend these results in two ways: investigating whether the same phenomenon occurs with other kinds of scientific explanations for beliefs (genetic, cognitive, or developmental), and investigating whether it occurs for first-person judgments, in which participants reported how *their* belief would change in response to a scientific explanation for their belief (or the opposing belief).

Method

Participants.

Two-hundred-fifty-eight adults (119 female, 139 male, mean age 32) were recruited through MTurk. An additional 250 participants were excluded following the criteria employed in Experiments 1-2, or for failing an additional reading comprehension check.[4]

Materials and methods.

Experiment 3 employed the same target claims as Experiments 1-2, but considered a broader range of beliefs by including the negation of each claim in addition to its affirmation. For instance, for the common religious claim we varied whether Michael believed in the existence of God or denied the existence of God. For each participant, one of the six claims was selected at random, as was the valence of Michael's belief (whether he believed or denied it).

---

[4] This increase in exclusion rate reflects the addition of the more stringent reading comprehension check that ensured participants had paid attention to the details of the vignette. Notably, the overall exclusion rates we observed in this and other experiments are consistent with what has been seen previously on MTurk; difficult comprehension check questions have been observed to exclude nearly 40% of participants (Downs, Holbrook, & Sheng, 2010). Of 420 participants who completed the experiment and did not report they might have previously participated in a similar experiment, 33 were excluded for answering the simple catch question incorrectly (8%) and an additional 129 excluded for failing the stringent reading comprehension check (31%). Given the high exclusions, we also analyzed data from all participants who completed the experiment. We found the same significant effects as those we report below. The only differences were additional main effects of explanation discipline, neither of which interacted with our primary manipulation of normality.

Each participant read a vignette in which Michael considers the target claim and then reads a scientific explanation for belief in (or rejection of) the target claim. The manipulation of primary interest was whether the scientific explanation invoked a *normal* process or an *abnormal* process. Explanations also varied in whether they appealed to *neuroscience, genetics*, *cognitive psychology*, or *developmental psychology* (see Table 2). For example, in the *neuroscience* condition, participants read a vignette like the one below. (The target belief in this example is the common religious belief and details specific to it appear in bold. The words that varied between participants depending on the valence of Michael's belief appear in brackets.)

> Michael comes across the following claim on a website:
>
> CLAIM: **There is a God.**
>
> Michael has not given a lot of thought to whether **there is a God**. But, if he were asked what he thinks about the claim he just read, he would say that he believes that it is [true/false].
>
> Michael next reads the following fact in a book:

In the *normal* condition (in the *neuroscience* condition), the provided explanation read:

> FACT: People are more likely to [believe/reject] this claim if they have "Type M neural activity" in the ventral striatum cortex in their brain, which is the type of activity normally observed there.
>
> Michael trusts the book and believes the fact that he just read.

In the *abnormal* condition, the explanation read:

> People are more likely to [believe/reject] this claim if they frequently have "mini-seizures" in the ventral striatum cortex in the brain: this involves an abnormal pattern of activity.
>
> Michael trusts the book and believes the fact that he just read.

Table 2: Explanations used in Experiment 3

	Normal	**Abnormal**
Neuroscience	People are more likely to [believe/reject] this claim if they have "Type M neural activity" in the ventral striatum cortex in their brain, which is the type of activity normally observed there.	People are more likely to [believe/reject] this claim if they frequently have "mini-seizures" in the ventral striatum cortex in the brain: this involves an abnormal pattern of activity.
Genetics	People are more likely to [believe/reject] this claim if they do not have a mutated acfga2 gene – in other words, if their acfga2 gene is normal.	People are more likely to [believe/reject] this claim if they have a mutated acfga2 gene – in other words, if their acfga2 gene is abnormal.
Cognitive Psychology	People are more likely to [believe/reject] this claim if they engage in "Type M lexical processing" when reasoning, which is the type of processing normally engaged in these cases.	People are more likely to [believe/reject] this claim if they exhibit "cognitive biases in lexical processing" when reasoning, which is a type of processing that's abnormal in these cases.
Developmental Psychology	People are more likely to [believe/reject] this claim if they did not suffer from an attachment disorder as a child – that is, if their parental attachment was normal.	People are more likely to [believe/reject] this claim if they suffered from an attachment disorder as a child – that is, if their parental attachment was abnormal.

As in Experiments 1 and 2, participants reported how Michael's confidence in his belief should change. Responses were given on the same seven-point scale, but the data were coded relative to Michael's initial belief: Responses of *Much more confident that it is false* were recorded as -3 for participants in the *accept* condition and as 3 for participants in the *reject* condition, and vice versa for *Much more confident that it is true*. Participants were further asked to assume that the explanation provided was true, and to report how *they* would revise their beliefs upon learning the explanation. These responses were also solicited on the same scale and coded relative to each participant's initially reported belief about the claim.

Results

Effects of experimental conditions.

Responses about Michael's behavior were analyzed with an ANOVA with mechanism type (2: normal, abnormal), claim prevalence (2: common, controversial), and explanation discipline (4: neuroscience, genetics, cognitive, developmental) as between-subjects factors (see Fig. 3). As in Experiments 1-2, we collapsed across the three domains of explained belief, and we additionally collapsed across the valence of Michael's belief (whether he accepted or rejected the target belief).

We found a main effect of mechanism type, $F(1, 242) = 10.30$, $p = .002$, $\eta_p^2 = .039$. Participants judged that Michael should increase his confidence in his belief if the explanation he received for it appealed to a normal process, but not if the explanation appealed to an abnormal process. These effects were consistent across different belief domains (as in Experiments 1-2) and the valence of Michael's belief. There were no other significant main effects nor interactions.

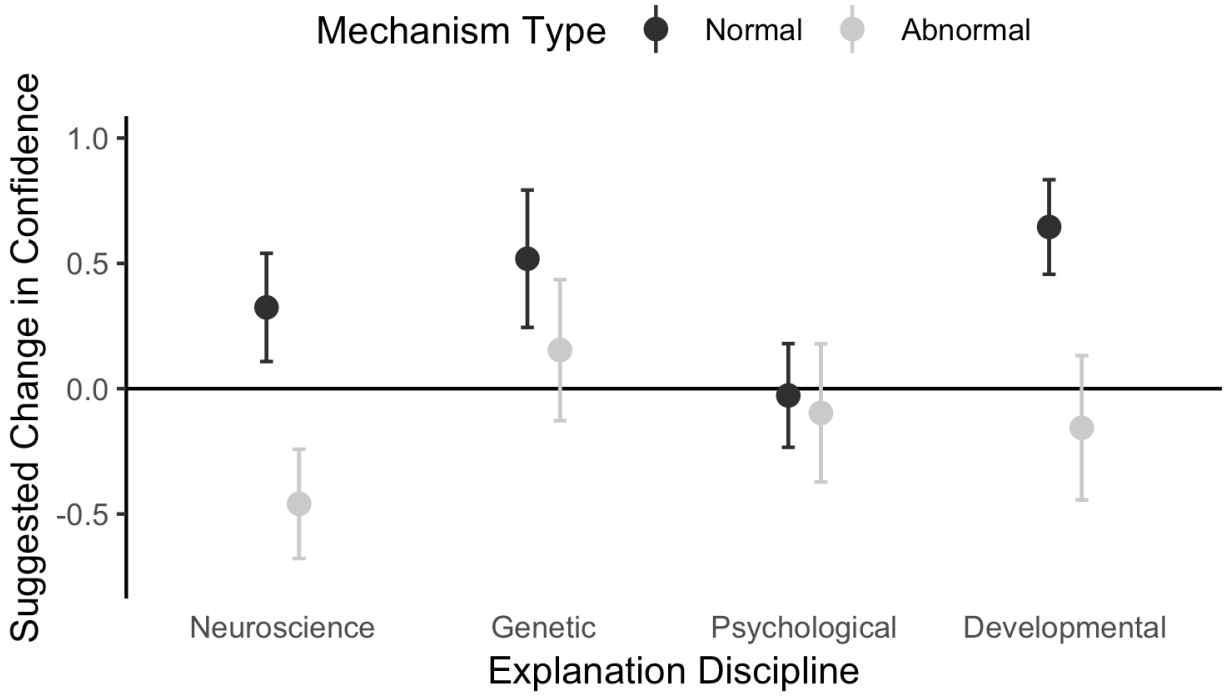

Figure 3: In Experiment 3, as in Experiment 2, participants reported that a person should become more confident in a belief upon learning that the belief is associated with a "normal" process, but not an "abnormal" one (error bars: 1 SEM). This pattern was generally consistent across different types of scientific explanation.

Belief reinforcement or undermining.

Participants in the *normal* condition provided responses significantly above the scale midpoint, $M = 0.34$, $t(131) = 3.07$, $p = .003$. Conversely, participants in the *abnormal* condition gave responses that were not significantly different from the scale midpoint, $M = -0.17$, $t(125) = -1.26$, $p = .208$. These patterns were consistent across domains.

First-person judgments.

To examine participants' first-person responses (what participants reported *they* would do upon learning the explanation), we performed an additional ANOVA with the first-person responses as the dependent variable. We also incorporated participants' prior beliefs about the target claim. Participants were classified into three groups based on their reported attitude towards the target proposition: those who *agreed* with Michael (i.e., endorsed the target), those

who *disagreed* with Michael, and those who were *ambivalent* (i.e., responded at the scale midpoint). Because few participants were ambivalent, and to facilitate interpretation, this analysis included only participants who agreed or disagreed with Michael. To increase the number of participants in each cell, we also pooled data across common and controversial claims.

We performed a 2 (mechanism type) x 4 (explanation discipline) ANOVA with participants' belief (2: agreed, disagreed) as an additional between-subjects factor. We found a significant interaction between mechanism type and participant belief, $F(1, 226) = 7.66$, $p = .006$, $\eta_p^2 = .035$; see Fig. 4. Participants reported that explanations that appealed to abnormal functioning would be *more* reinforcing for their own beliefs, as long as those explanations were

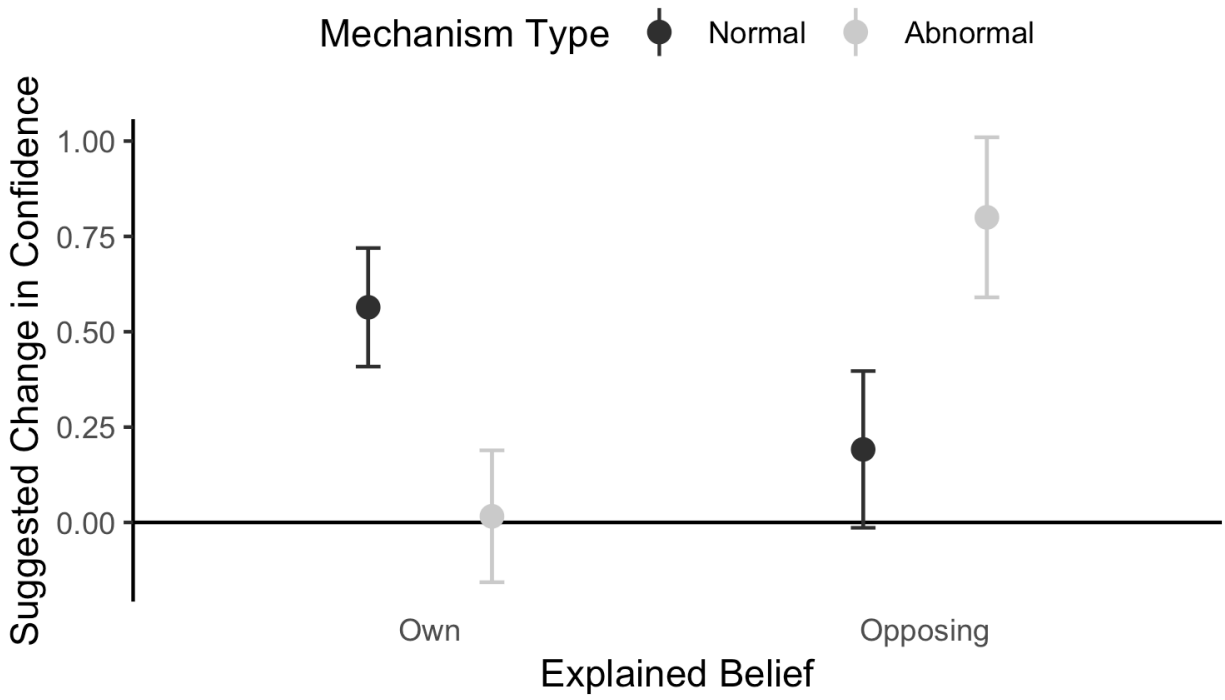

Figure 4: In Experiment 3, how participants indicated that how they would respond to a scientific explanation depended on whether the mechanism in the explanation was described as functioning normally or abnormally and whether they shared the explained belief or rejected it (error bars: 1 SEM). Participants thought that both explanations of their belief in terms of "normal" processes and explanations of the opposing belief in terms of "abnormal" processes would reinforce their belief, but not that "abnormal" explanations of their belief or "normal" explanations of the opposing belief would.

for the *opposing* belief, Welch's $t(99.65) = 2.07$, $p = .041$, but the pattern was reversed if they read explanations for *their* beliefs (Welch's $t(131.47) = -2.36$, $p = .020$). There were no significant main effects nor other significant interactions. As with third-person judgments, patterns were consistent across domains and belief valence.

Discussion

Experiment 3 replicated the key finding of our first two experiments: Neuroscientific explanations that invoked normal functioning led to belief reinforcement, while those that invoked abnormal functioning did not. Moreover, this same effect was observed with other types of scientific explanation for belief, specifically genetic, cognitive, and developmental explanations. Experiment 3 also found that effects of normal and abnormal scientific explanations for belief are not limited to third-person judgments: Participants reported that their own beliefs would respond differently to scientific explanations depending on whether the mechanism in the explanation was described as functioning normally or abnormally.

Experiment 4

Experiments 1-3 support the idea that a scientific explanation for belief can be either reinforcing, neutral, or undermining, depending on whether the belief-formation mechanism that the explanation invokes is normal or abnormal, where (ab)normality is either explicitly stated or implied. It is not entirely clear, however, what facet of "(ab)normality" drives these effects. Experiment 4 aims to tease apart two senses of normality.

Conceptual analyses and empirical work suggest that people's concept of normality includes a statistical component, but also a prescriptive component (Bear & Knobe, 2016; Hitchcock & Knobe, 2009; Wachbroit, 1994). To illustrate, we might say "perfect pitch is abnormal" because perfect pitch is uncommon (statistical abnormality). By contrast, we might

say "myopia is abnormal" because that is not how eyes "should" function (prescriptive abnormality). Furthermore, prescriptive normality can be interpreted relative to different purposes: We might say that morning sickness involves the body behaving abnormally in the sense that for purposes of personal comfort it should not behave that way; but we might say that morning sickness involves the body behaving normally in the sense that, for purposes of protecting a developing pregnancy, it should.

Thus, a mechanism involved in belief-formation may be normal in that it is statistically normal. Or it may be normal in that it is prescriptively normal relative to some purpose, such as the function it evolved to perform, or truth-tracking (a purpose that people often desire for their belief-formation processes).

Experiments 1-3 suggest that people take "normal functioning" (e.g., of a brain region) to imply truth-tracking. Experiment 4 tested whether this implication arises from evidence of statistical normality or of prescriptive normality by using explanations for belief that invoked mechanisms that could be common and/or prescriptively normal (in this case, relative to evolved function). These features were varied independently in a 2 x 2 design: Certain brain activity associated with religious belief was described as either common or uncommon and as indicating that part of the brain was either "doing what it evolved to do" or operating defectively. Our prediction was that the epistemic consequences of offering a scientific explanation track prescriptive rather than statistical normality (e.g., that people are more likely to endorse beliefs if they are associated with a brain region "doing what it evolved to do," but no more likely to endorse beliefs that are merely associated with common brain activity).

Experiment 4 had an additional aim: to use more naturalistic stimulus materials to confirm that the effects from Experiments 1-3 are not restricted to highly artificial cases. We thus

used explanations that were designed to be realistic and plausible, focusing specifically on the increasingly common case of neuroscientific explanations for religious belief, and using language inspired by media coverage of such research. We also directly measured whether participants found the explanations to be realistic and plausible.

Method

Participants.

One-hundred-ninety-six adults (98 female, 98 male, mean age 38) completed the experiment through MTurk. An additional 207 participants were excluded for failing to complete the experiment (n = 6), reporting they might have previously participated in a similar experiment (n = 12), or answering any of three reading comprehension questions incorrectly (n = 189).

Materials and methods.

Participants began the experiment by indicating the extent to which they believed in God on a 1-7 scale. Then, they read the following background information:

> In the last few decades, many scientific studies have found a connection between religious belief and activity in the temporal lobes in the brain. For example, certain kinds of activity in the temporal lobes are more common in people who believe in God. Further, those kinds of activity are more likely to be seen while people are having mystical or religious experiences, and some studies have found that stimulating the temporal lobes can cause religious experiences.

Next, in a 2x2 design, participants were randomly assigned to receive an explanation that invoked a mechanism that was either common or uncommon (i.e., statistically normal or

statistically abnormal) and either properly functioning or improperly functioning (i.e., prescriptively normal or abnormal). For example, in the *common*/*proper* condition, participants read the following (with words in bold varying between conditions):

> Suppose a major new study comes out that confirms all of the research mentioned on the previous page. Using data from thousands of people from around the world, it clearly shows that people are more likely to believe in God if a particular pattern of activity is seen in their temporal lobes. The study also reveals that this type of activity is very **common:** It is seen in a **large majority** of people around the world. The activity seems to reflect **the operation of a brain system doing what it evolved to do** – that is, it reflects the **proper** functioning of an evolved biological system. Measuring for this pattern of activity makes scientists about 40% better at predicting whether a person believes in God.
>
> Now imagine a person named Michael, who is one of the last participants in the new study. Michael believes in God, and during the study the researchers tell him that he exhibits this pattern of brain activity in his temporal lobes. They also tell him what the study has found (that this pattern of activity is more often seen in people who believe in God, that it is very **common**, and that it reflects the **proper** functioning of a particular brain system).

In the *uncommon* conditions, participants read that the pattern of activity was "uncommon" and "seen in a minority of people." In the *improper* conditions, they read that it reflected "improper" functioning and "a defect in the operation of a brain system."

As in previous experiments, participants reported whether they thought Michael should become more or less confident in his belief in God.

Participants then answered the following two questions on seven-point scales ranging from "very implausible" and "very unrealistic" to "very plausible" and "very realistic":

> How plausible do you think results of the new study are?
>
> Regardless of how plausible you think the results of the new study are, how realistic do you think the study is? In other words, whether or not you would believe the results if you heard about them, how likely do you think it is that you might see a news report about a study with these results?

Results

Responses were analyzed with a 2x2 ANOVA with prescriptive normality (2: proper functioning, improper functioning) and statistical normality (2: common, uncommon) as between-subjects factors (see Fig. 5). We found a main effect of prescriptive normality, $F(1, 192) = 18.52$, $p < .001$, $\eta_p^2 = .094$, and no main effect of statistical normality, $F(1, 192) = 0.43$, $p = .51$, nor interaction, $F(1, 192) = 0.94$, $p = .33$. Participants reported that Michael should become more confident in his belief in God if he learned that belief is associated with a pattern of activity that indicates proper functioning (testing against the scale midpoint, $M = 0.39$, $t(108) = 3.22$, $p = .002$), but indicated that he should become less confident in his belief in God if he learned that it was associated with improper functioning ($M = -0.45$, $t(86) = -3.03$, $p = .003$). By

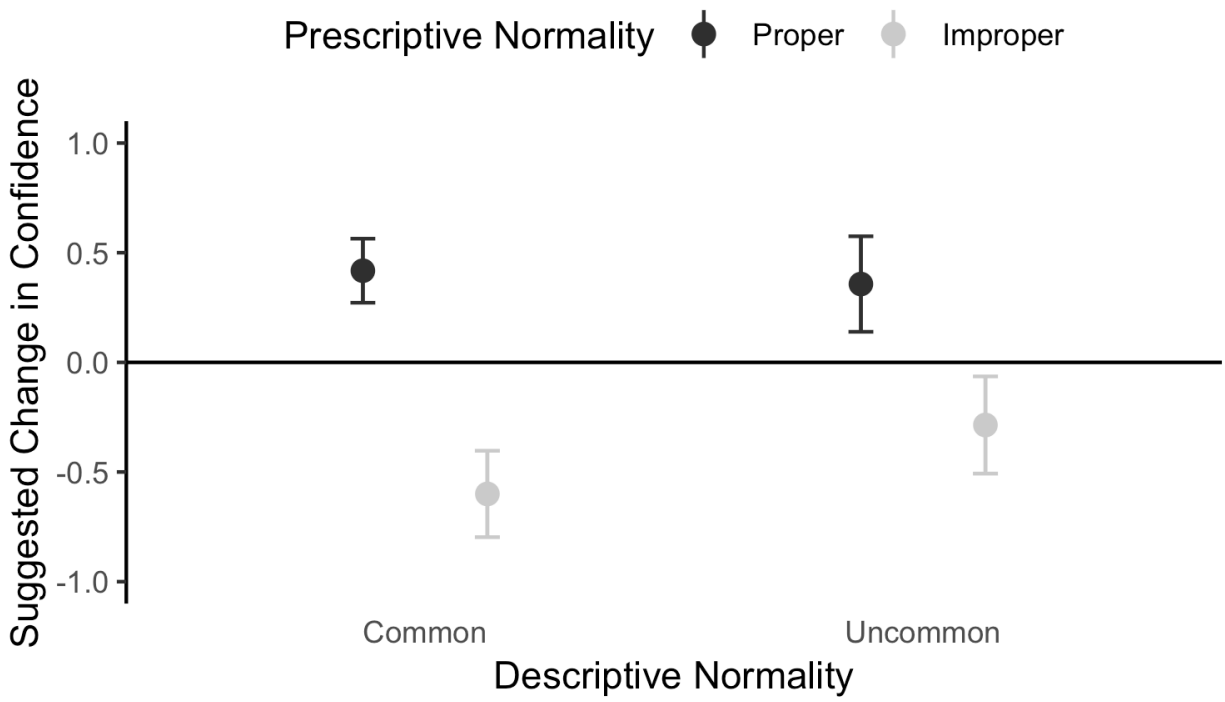

Figure 5: Experiment 4 results (error bars: 1 SEM). Participants reported that a person should become more confident in a belief associated with a pattern of brain activity that indicates "proper functioning." By contrast, it made no difference whether a belief was associated with a common or uncommon pattern of activity.

contrast, it made no difference whether Michael learned that his belief is associated with a common or an uncommon pattern of activity.

Participants found the scenarios plausible and realistic (*M*: 0.26 and 0.63, respectively, on 7-point scales from "very implausible/unrealistic," coded as -3 to "very plausible/realistic," coded as 3). Although participants found the explanations that appealed to improper functioning less plausible than those that appealed to proper functioning, $F(1, 192) = 5.93$, $p = .016$, $\eta_p^2 = .030$, they found them no less realistic than explanations that appealed to proper functioning, $F(1, 192) = 1.56$, $p = .21$. Explanations using common and uncommon mechanisms were not judged differentially plausible, $F(1, 192) = 0.012$, $p = .91$, or realistic, $F(1, 192) = 0.067$, $p = .80$.

To ensure that patterns of effects were not a consequence of differences in the perceived plausibility of the proper and improper functioning mechanisms, we modeled responses with an

ANCOVA that included plausibility as a covariate (in addition to factors for the two experimental manipulations). We observed the same results: a main effect of proper functioning, $F(1, 191) = 18.36$, $p < .001$, and no other significant effects nor interactions, including any significant relationship between plausibility and how people thought Michael should revise his belief, $F(1, 191) = 0.11$, $p = .74$.[5]

Discussion

Experiments 2-3 showed that people think explanations for the origin of a belief that invoke a "normal" mechanism should increase one's confidence in the explained belief. Experiment 4 refined this discovery, providing evidence that the effect is driven by prescriptive—not statistical—normality. Whether an explanation appealed to a properly or improperly functioning process determined whether or not it was regarded as belief reinforcing. By contrast, it made no difference whether the explanation appealed to a common or rare mechanism.

Notably, as in Experiments 2-3, participants were given no indication that the function of the operative neurological process ("what it evolved to do") was to produce true beliefs. Thus, our participants appeared to be making the substantive assumption that merely being associated with a prescriptively normal process (a brain region "doing what it evolved to do") makes a belief more likely to be true. Equally important, participants did not make the analogous assumption that being associated with a statistically normal process makes a belief more likely to be true.

---

[5] As in Experiment 3, many participants were excluded for answering at least one reading comprehension check question incorrectly. However, all results were qualitatively identical and all significant effects remained significant at the 95% confidence level even if data from all participants were included in the analyses.

Experiment 4 also confirmed that participants found the provided explanations realistic, and it verified that perceived plausibility did not drive observed effects. However, a concern about ecological validity might remain: Scientific explanations for belief, in the popular press or elsewhere, are rarely explicit in referencing proper or improper biological functioning. In an additional experiment (reported in Supplementary Materials), we replicated the basic design of Experiment 3, but with *implied* rather than explicit abnormality. Instead of indicating that activity in a brain region was "normal" or "abnormal," the text specified that it involved "Type I neural activity" versus "mini-seizures," where we took the latter to imply abnormality. We found evidence that the former case (which did not imply abnormal functioning) ranged from reinforcing to neutral, whereas the latter (which implied abnormality) ranged from neutral to undermining, depending on whether the belief involved affirmation or negation of a target claim. The case of "mini-seizures" was selected from a popular press article that discussed the connection between spirituality and temporal lobe epilepsy (Hagerty, 2009). We thus have reason to believe that the kinds of scientific explanations that laypeople encounter in the popular press and elsewhere do imply (prescriptive) normality or abnormality, and that this has epistemic consequences.

Experiment 5

Experiment 5 goes beyond our first four experiments by investigating whether scientific explanations that invoke "normal" or "abnormal" functioning influence consequential real-world judgments, such as responses to scientific discoveries. We were also interested in whether such responses would be influenced by whether or not participants held the explained belief, as suggested by Experiment 3.

In Experiment 5, participants read about the discovery of a neuroscientific explanation (which invoked either normal or abnormal functioning) for belief in God or for atheism. They were then asked to make a series of judgments about that discovery, including whether it would be appropriate to teach in a science class and whether it would be important to see it replicated before accepting it. Previous work has found that participants are more skeptical and critical of findings that challenge their prior beliefs (e.g., Lord, Ross, & Lepper, 1979; Edwards & Smith, 1996) or their identity (e.g., Greitemeyer, 2014; Munro & Munro, 2014; Nauroth, 2015), including neuroscientific findings specifically (Scurich & Shniderman, 2014). The results of Experiments 1-4 suggest that explanations that appeal to abnormal functioning are viewed as threatening to the beliefs they explain. Accordingly, we predicted that participants would report greater skepticism towards explanations that implied abnormal function if they endorsed the explained belief (e.g., atheists would be more skeptical of explanations that appeal to abnormal processes if those explanations explained atheism, rather than theism).

Religious belief was chosen for two reasons: We anticipated that participants would have strong but variable beliefs, and religious beliefs are relevant to many contemporary debates about science education and policy.

Method

Participants.

Five-hundred-thirty-nine adults (228 female, 309 male, 2 other gender or declined to specify, mean age 31) were recruited through MTurk. An additional 102 participants were excluded following the criteria employed in Experiments 1-2.

Materials and methods.

Participants began by reading a short vignette describing the discovery of a scientific explanation for a particular belief. As in Experiments 1-2, these explanations all appealed to neuroscience. Discoveries were randomly varied in whether they explained *theism* or *atheism*, and whether that explanation appealed to *normal* or *abnormal* functioning in the brain. To explore a wider range of cases, the explanations also randomly varied in whether they involved the *presence* or *absence* of some pattern of neurological activity. As an example, the *abnormal presence* explanation for *theism* is reproduced below:

> Suppose a team of scientists discovers that people are more likely to believe in God if the ventral striatum cortex in their brain suffers from mini-seizures. Mini-seizures reflect an abnormality in that brain region. The team has retested and confirmed the discovery in two additional experiments.

Participants then reported their agreement, on a seven-point scale, with seven statements chosen to detect increased skepticism or dismissal of explanations that were more threatening to participants' own convictions (see Table 3). On a final screen, participants identified their own position on a theism-atheism scale.

Results

Participants were classified into three categories based on their own beliefs: theists (n = 230), atheists (n = 231), and those at the scale midpoint (n = 78). We did not include participants at the scale midpoint in our analyses. To create a single dependent measure for analysis, we averaged each participant's responses to our seven questions (reverse coding the *Transparency* and *Replication* items; $\alpha = .69$). We analyzed this "composite trust" measure with an ANOVA

Table 3: Test statements used in Experiment 5

Importance	This would be an important finding
Science Class	This finding would be appropriate to teach in a high school science class
Theology Class	This finding would be appropriate to teach in a high school theology class
Acceptance	I would accept that this finding was probably true
Government Funding	I think it is appropriate for the government to fund research of this type
Transparency[a]	It would be important to know who funded this research
Replication[a]	Before accepting this finding, it would be important to see it replicated by an independent team of researchers

[a] Transparency and Replication items are reverse coded

with mechanism type (2: normal, abnormal), explained belief (2: theism, atheism), presence (2: presence, absence), and participant's belief (2: theist, atheist) as between-subjects factors (see Fig. 6). This analysis revealed an interaction between mechanism type, explained belief, and participant's belief, $F(1, 445) = 7.86$, $p = .005$, $\eta_p^2 = .018$, which reflects part of the effect that we anticipated: When an explanation is offered for atheism, atheists are more skeptical of explanations that appeal to abnormal processes than explanations that appeal to normal processes, Welch's $t(103.59) = -2.70$, $p = .008$. By contrast, when an explanation is offered for belief in God, atheists are more skeptical of explanations that invoke normal functioning, Welch's $t(93.04) = 2.27$, $p = .025$. Theists, in contrast, were generally ambivalent about mechanism type for all beliefs; Welch's $t(109.01) = 0.26$, $p = .80$ for atheism and Welch's $t(100.07) = -1.09$, $p = .28$ for belief in God. However, the direction of differences was consistent with our prediction.

Our analysis additionally found a main effect of explained belief, $F(1, 445) = 4.24$, $p = .040$, $\eta_p^2 = .014$ (explanations for atheism were regarded more skeptically than explanations for belief in God across most conditions) and a main effect of participant's belief, $F(1, 445) = 25.68$,

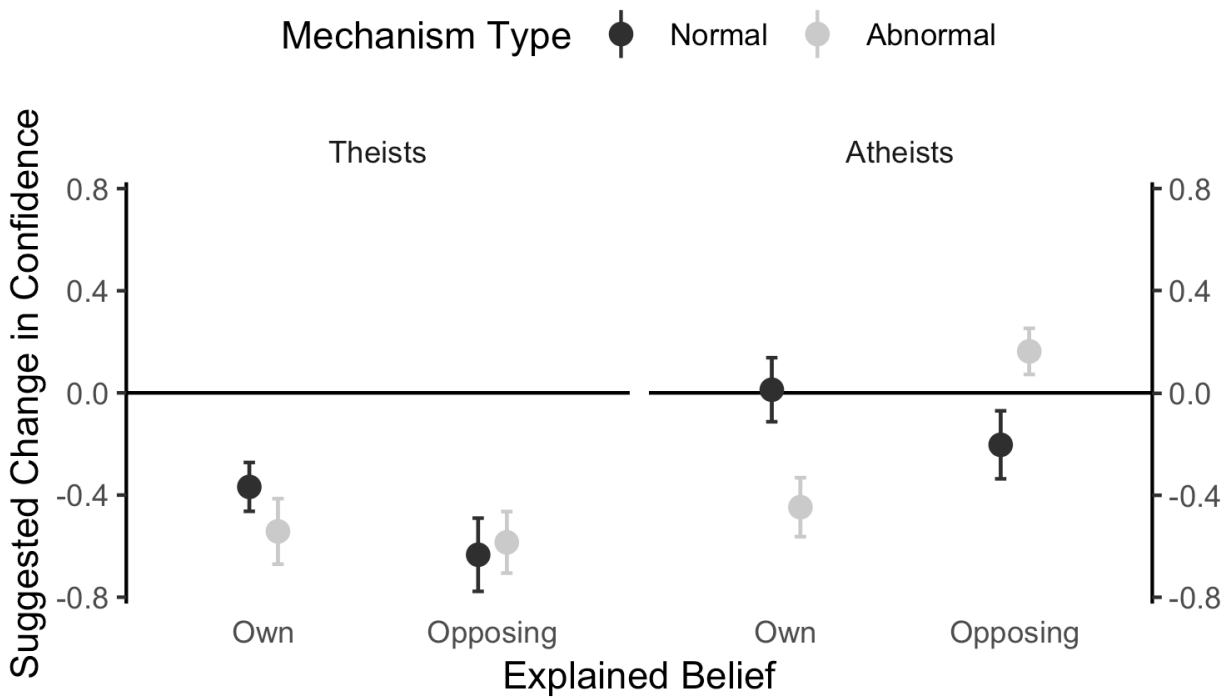

Figure 6: Experiment 5 results (error bars: 1 SEM). Participants were more skeptical of scientific discoveries that offered explanations for their own beliefs if those explanations appealed to abnormal processes. By contrast, they were *less* skeptical of explanations that appealed to abnormal processes when those explanations were of beliefs that they disagreed with. This pattern is consistent with the hypothesis that explanations that appeal to abnormal functioning are more likely to be viewed as belief undermining and, therefore, to be challenged by those who hold the explained belief.

$p < .001$, $\eta_p^2 = .049$ (theists were more skeptical of all scientific explanations, consistent with past results showing that theists in the United States have less trust in science and scientific explanations; e.g., Clobert & Saroglou, 2015; Cacciatore, Browning, Scheufele, Brossard, Xenos, & Corley, 2016). Finally, there was a significant interaction between mechanism type, participant's belief, and presence/absence, $F(1, 445) = 5.83$, $p = .016$, $\eta_p^2 = .013$, which appears to reflect that, regardless of the belief being explained, atheists were somewhat more skeptical of abnormal presence explanations (i.e., displaying an abnormal pattern of brain activity) than normal presence explanations (i.e., displaying a normal pattern of brain activity), but were somewhat *less* skeptical of abnormal absence explanations (i.e., displaying a lack of an abnormal

pattern of brain activity) than normal absence explanations (i.e., displaying a lack of a normal pattern of brain activity).

We also analyzed responses to each of the seven questions individually; results are reported in Supplementary Table 1.

Discussion

Experiment 5 found that participants were more skeptical of scientific discoveries that offered explanations for their own beliefs if those explanations appealed to abnormal processes. By contrast, they were *less* skeptical of scientific explanations that appealed to abnormal processes when those explanations were of beliefs that they rejected. This is consistent with our hypothesis that explanations appealing to abnormal functioning are more likely to be viewed as belief undermining and, therefore, to be challenged by those who hold the explained belief.

Experiment 5 goes beyond Experiments 1-4 to demonstrate that information about how beliefs are generated impacts real-world judgments, not just explicitly epistemic judgments. It also convincingly shows that responses to information about the origins of belief can be moderated by one's own beliefs. Finally, Experiment 5 addresses the possible concern that our results reflect demand characteristics: One might suspect that participants in Experiments 1-4 had determined (consciously or unconsciously) that they were expected to find explanations that appeal to abnormal functioning undermining and responded obligingly. However, participants in Experiment 5 made no judgments about what effect that explanation should have on anyone's belief, yet the results are consistent with motivated reasoning driven by a tendency to think that beliefs are undermined if explained by appeal to abnormal functioning.

General Discussion

Across five experiments, we find evidence that the effects of scientific explanations for belief are sensitive to whether invoked mechanisms are truth-tracking (Experiment 1), with the additional—and more surprising—result that "normal" functioning is treated as a proxy for epistemic reliability (Experiments 2-4). We find that this result holds for a range of beliefs (Experiments 1-4), even if abnormal functioning is only implied (Experiments 1 and 5; Supplementary Experiment), and for different types of scientific explanation (Experiment 3). We also isolate the relevant sense of (ab)normality: Beliefs associated with improperly functioning systems (i.e., systems functioning differently from how they evolved to function)—but not with statistically infrequent systems—are potentially debunked by a scientific explanation. Our final experiment (Experiment 5) additionally suggests that, because explanations for belief are taken to be threatening when they appeal to abnormality, participants whose beliefs are explained by appeal to abnormal functioning are motivated towards skepticism about the proffered explanation, and this manifests in their attitudes towards scientific importance, funding, replication, and so on.

The results of Experiment 1 support the hypothesis that people are responsive to process debunking arguments, which challenge beliefs on the grounds that they are the products of epistemically defective mechanisms (Nichols, 2014). Further, we find no evidence for the hypothesis that epistemically neutral scientific explanations of belief are regarded as threatening to belief merely because they attribute the belief to something other than its truth. We also find no evidence that explanations for belief are especially prone to generating a backfire effect or belief polarization. Instead, we find the surprising result that participants judge seemingly neutral explanations for belief (e.g., correlation with "Type I neural activity") as belief reinforcing.

These findings are broadly consistent with research on source credibility: People seem to track the reliability of a source in revising beliefs, even when that source is a belief-formation process "inside the head."

This work goes beyond prior work not only by focusing on belief-formation processes as information sources, but also in isolating what it is about those processes that generates particular epistemic consequences. Experiments 2-4 support the hypothesis that people treat prescriptive normality (relative to evolutionary function, for example) as a proxy for epistemic reliability. In other words, they appear to make the substantive assumption that because something is functioning as it "should" in some way (e.g., as it evolved to function, in Experiment 4), it is likely to be truth-tracking. Moreover, prescriptive normality appears to be inferred from sparse cues; referring to something as a "mini-seizure" has epistemic effects similar to explicitly labeling it an "abnormal" process (Supplementary Experiment).

Our results could help explain variation in the uptake of scientific explanations for belief. Previous research already suggests that scientific findings are regarded more skeptically when they conflict with prior beliefs, both because such findings are less consistent with prior commitments (Koehler, 1993) and because of motivated reasoning in response to counterattitudinal or disconfirming evidence (Lord, Ross, & Lepper, 1979; Taber & Lodge, 2006; Kahan, 2010; Nyhan & Reifler, 2010). Our results reveal an additional role for implicit or explicit assumptions about (ab)normality when such explanations involve the origins of belief, and they suggest that the effects of (ab)normality interact with an individual's own belief. As demonstrated in Experiment 5, people may be more skeptical of "abnormal" explanations for their own beliefs and more accepting of "abnormal" explanations for contrary beliefs.

The finding that "folk epistemology" treats prescriptive normality as a proxy for truth-tracking is noteworthy. Although it is a near-tautology that beliefs formed via reliable processes are truth-tracking—and, indeed, reliabilist epistemologists have argued that beliefs are justified if and only if they are formed via a reliable process (Goldman, 1979)—it is puzzling how we can determine whether a belief-forming process is reliable, and it is unclear if there are any general features of a belief-forming process that can indicate reliability, aside from direct examination of that process's track-record. This is most pronounced for processes that generate beliefs in domains where it is an open question whether *any* beliefs are true, such as religion or morality. One solution that people appear to employ is to assume that a process's *prescriptive normality* reflects its *reliability*.

The assumption that prescriptively normal processes reliably produce true beliefs is only warranted under substantive assumptions. For example, in Experiment 4, participants judged beliefs that were produced by "a brain system doing what it evolved to do" to be more trustworthy. This could be justified by the assumption that natural selection selects for true beliefs. While this assumption is consistent with common misconceptions regarding natural selection (Lombrozo, Shtulman, & Weisberg, 2006; Shtulman, 2006), it is often rejected by philosophers and biologists (see Downes, 2000; McKay & Dennett, 2009). Another possibility is that natural selection, though not truth-tracking in itself, could tend to produce reliable belief-formation mechanisms (evolutionary reliabilism*;* Ramsey, 2002). (In other words, producing true beliefs is part of "doing what it evolved to do" for all belief-producing neural processes.) Evolutionary reliabilism is hotly debated by philosophers (see, e.g., Downes, 2000; Feldman, 1988; Fodor, 2002; Plantinga, 1993; Ramsey, 2002; Sage, 2004; Stephens, 2001; Stich, 1990). A

general tendency to treat prescriptive normality as a proxy for epistemic reliability thus betrays nontrivial underlying commitments.

Our findings shed light on philosophical debates in several ways. First, they provide some evidence that folk epistemology is a reliabilist epistemology of a particular kind: one that takes normal functioning as a proxy for reliability. This, in turn, suggests the importance of further psychological research on why people take normal functioning to be a proxy for truth-tracking—and the importance of philosophical debate about whether they are correct to do so. Finally, these data about folk epistemology bear on normative epistemology insofar as an epistemology ought to provide an explanation for why ordinary individuals reason as they do about belief.

Our studies have a number of important limitations. First, our participant population was restricted to individuals in the United States using a single crowdsourcing platform. While we were nonetheless able to identify interactions between participants' beliefs and their responses to the task, we did not investigate other individual differences, nor extend our findings to other populations. Second, the range of scientific explanations for belief that we considered was fairly restricted. For example, we did not consider explanations like Freud's (which appealed to wishful thinking), evolutionary explanations, or pseudo-scientific explanations. Moreover, our experiments manipulated the traits of belief-associated processes, not processes that were explicitly described as belief-producing themselves. Finally, considered together, our experiments suggest that explanations that appeal to normal mechanisms produce belief reinforcement more strongly and consistently than explanations that appeal to abnormal mechanisms produce belief undermining. Future work could investigate the basis for this asymmetry.

This work builds on two very different existing traditions: a long line of work in psychology interested in how people respond to new evidence, and more recent work in experimental philosophy investigating folk judgments concerning classic conundrums from epistemology (e.g., Beebe, 2012). However, our findings address a question that had, until now, only received theoretical discussion: why explanations for belief might be regarded as debunking. We provide the first systematic empirical investigations of whether and when people treat such explanations as debunking. We also provide an account of why this is the case grounded in the idea of normal functioning (more specifically, proper functioning), and we demonstrate that our findings extend to real-world judgments with implications for public policy, education, and the public acceptance of science.

Acknowledgements

This research was supported by a McDonnell Scholar Award and NSF grant DRL-1056712 to Tania Lombrozo.

References

Alexander, J., Mallon, R., & Weinberg, J. (2010). Accentuate the negative. *Review of Philosophy and Psychology, 1*, 297-314.

Beebe, J. R. (2012). Experimental epistemology. In A. Cullison (Ed.), *Continuum companion to epistemology* (pp. 248-269). London, United Kingdom: Bloomsbury.

Bering, J. (2012). *The belief instinct: The psychology of souls, destiny, and the meaning of life.* New York, NY: W. W. Norton.

Cacciatore, M. A., Browning, N., Scheufele, D. A., Brossard, D., Xenos, M. A., & Corley, E. A. (2016). Opposing ends of the spectrum: Exploring trust in scientific and religious authorities. *Public Understanding of Science*, 0963662516661090.

Clobert, M., & Saroglou, V. (2015). Religion, paranormal beliefs, and distrust in science: Comparing East versus West. *Archive for the Psychology of Religion, 37*, 185–199.

Cook, J., & Lewandowsky, S. (2011). The debunking handbook. Retrieved from http://www.skepticalscience.com/docs/Debunking_handbook_draft2.pdf

Copp, D. (2008). Darwinian Skepticism about Moral Realism. *Philosophical Issues, 18*, 186–206.

Downes, S. M. (2000). Truth, selection and scientific inquiry. *Biology and Philosophy 15*, 425-442.

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system?: Screening Mechanical Turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2399–2402). New York, NY, USA: ACM.

Edwards, K. & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments, *Journal of Personality and Social Psychology, 71*, 5-24.

Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Feldman, R. (1988). Rationality, reliability, and natural selection. *Philosophy of Science 55*, 218-227.

Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin, 23*, 1215–1226.

FitzPatrick, W.J. (2015). Debunking evolutionary debunking of ethical realism. *Philosophical Studies, 172(4)*, 883-904.

Fodor, J. (2002). Is science biologically possible? In J. Beilby, (Ed.), *Naturalism defeated?* (pp. 30-42). Ithaca, NY: Cornell University Press.

Freud, S. (1961). *The future of an illusion*. (J. Strachey, Ed. & Trans.) New York, NY: Norton. (Original work published 1927)

Goldman, A.I. (1979). What Is Justified Belief? In G.S. Pappas (ed.), *Justification and Knowledge*. Dordrecht: Reidel, pp. 1–25

Greitemeyer T (2014) I am right, you are wrong: How biased assimilation increases the perceived gap between believers and skeptics of violent video game effects. PLoS ONE 9: e93440. https://doi.org/10.1371/journal.pone.0093440

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition, 2*, 201–209.

Hagerty, B. B. (2009, May 19). Are spiritual encounters all in your head? *NPR*. Retrieved from http://www.npr.org

Heiphetz, L., & Young, L. L. (in press). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition*.

Hellmore, E. (1998, May 3). She thinks she believes in God. In fact, it's just a chemical reaction taking place in the neurons of her temporal lobes; Science has gone in search of the soul. *The Observer*. p. 20.

Hitchcock, C. & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 106*, 587-612.

Joyce, R. (2006). *The evolution of morality*. Cambridge, MA: MIT Press.

Jong, J. and A. Visala (2014). Evolutionary debunking arguments against theism, reconsidered. *International Journal for Philosophy of Religion, 76,* 243-258.

Kahan, D. (2010). Fixing the communications failure. *Nature*, *463*, 296–297.

Kahane, G. (2011). Evolutionary Debunking Arguments. *Noûs, 45*, 103-125.

Knobe, J. (2009). Folk judgments of causation. *Studies in History and Philosophy of Science, 40* 238-242.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56*, 28-55.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106–131.

Lombrozo, T., Shtulman, A., & Weisberg, M. (2006). The intelligent design controversy: Lessons from psychology and education. *Trends in Cognitive Sciences, 10*, 56-57.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

McKay, R. T. & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32*, 493-561.

Munro, G.D. & Munro, C.A. (2014). "Soft" versus "hard" psychological science: Biased evaluations of scientific evidence that threatens or supports a strongly held political identity. *Basic And Applied Social Psychology, 36*, 533-543.

Nauroth P, Gollwitzer M, Bender J, Rothmund T (2015) Social identity threat motivates science-discrediting online comments. *PLoS ONE 10*: e0117476.

Nietzsche, F. (1908). *Human, all too human: A book for free spirits*. (A. Harvey, Trans.) Chicago, IL: Charles H. Kerr. Retrieved from https://www.gutenberg.org/ebooks/38145 (Original work published 1878)

Nichols, S. (2014). Process debunking and ethics. *Ethics, 124*, 727-749.

Oppenheimer, D. M., Meyvis T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216–1226.

Plantinga, A. (1993). *Warrant and proper function*. New York, NY: Oxford University Press.

Ramsey, W. (2002). Naturalism defended. In J. Beilby, (Ed.), *Naturalism defeated?* (pp. 15-29). Ithaca, NY: Cornell University Press.

Sage, J. (2004). Truth-reliability and the evolution of human cognitive faculties. *Philosophical Studies, 117*, 95-106.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology, 39*, 127-161.

Schweitzer, N. J., Baker, D. A., & Risko, E. F. (2013). Fooled by the brain: Re-examining the influence of neuroimages. *Cognition, 129*, 501-511.

Scurich, N., & Shniderman, A. (2014). The selective allure of neuroscientific explanations. *PLoS ONE, 9*, e107529.

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychology, 52*, 170-194.

Singer, P. (2005). Ethics and intuitions. *Journal of Ethics, 9, 331-352.*

Stephens, C. L. (2001). When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument. *Philosophical Studies, 105*, 161-189.

Stich, S. P. (1990). *The fragmentation of reason.* Cambridge, MA: MIT Press.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*, 109-166.

Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science, 61*, 579-591.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470-477.

Weisberg, D. S., Taylor, J. C. V., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgment & Decision Making*, *10*, 429–441.

Wielenberg, E. J. (2010). On the evolutionary debunking of morality. *Ethics, 120*, 441-464.

Wilkins, J. S. and P. E. Griffiths (2013). Evolutionary debunking arguments in three domains: Fact, value, and religion. In G. Dawes & J. Maclaurin (Eds.), *A new science of religion* (pp. 133–146). New York: Routledge.