

When psychology undermines beliefs

Derek Leben

This paper attempts to specify the conditions under which a psychological explanation can undermine or debunk a set of beliefs. The focus will be on moral and religious beliefs, where a growing debate has emerged about the epistemic implications of cognitive science. Recent proposals by Joshua Greene and Paul Bloom will be taken as paradigmatic attempts to undermine beliefs with psychology. I will argue that a belief p may be undermined whenever: (i) p is evidentially based on an intuition which (ii) can be explained by a psychological mechanism that is (iii) unreliable for the task of believing p ; and (iv) any other evidence for belief p is based on rationalization. I will also consider and defend two equally valid arguments for establishing unreliability: the redundancy argument and the argument from irrelevant factors. With this more specific understanding of debunking arguments, it is possible to develop new replies to some objections to psychological debunking arguments from both ethics and philosophy of religion.

Keywords: Belief; Cognitive Science of Religion; Debunking; Moral Psychology

1. Introduction

With the expanding cognitive sciences of morality and religion has come a growing debate about the relevance of these fields for moral and religious beliefs. Psychologists Joshua Greene and Paul Bloom have recently claimed that their theories carry important consequences for the evaluation of moral and religious beliefs, respectively. Greene writes that recent work casts “doubt on deontology as a school of moral thought” (2007, p. 36), and Bloom claims that “while it is true that nothing from empirical study of human psychology can refute religious belief, certain theories can challenge the rationality of those who hold such beliefs” (2009, p. 125). In response, debunking arguments of this kind have been challenged by both philosophers and psychologists (Barrett, 2007; Berker, 2009; Dean, 2010; Murray, 2007). The aim of this paper is to try and resolve some of the confusions that have

Derek Leben is an Assistant Professor at the University of Pittsburgh at Johnstown.

Correspondence to: Derek Leben, University of Pittsburgh at Johnstown, Philosophy, 223B Biddle Hall, 450 Schoolhouse Rd., Johnstown, PA 15904, USA. Email: leben@pitt.edu

arisen in this important debate by revealing the structure underlying various psychological debunking strategies. The result will be a set of conditions for when psychological discoveries can have direct evaluative implications for beliefs.

Although the focus will be on moral and religious beliefs, the scope of this debate concerns beliefs in general. There is room here for a great deal of neutrality about the ontological status of beliefs. For our purposes, it should suffice to say that beliefs are attitudes about how the world is that can be true or false. I take (1)–(4) to be characteristic of normal beliefs:

- (1) Saturn is gaseous.
- (2) My car is black.
- (3) Napoleon was defeated at Waterloo.
- (4) There is a robin outside my window right now.

I will consider characteristically religious beliefs to be those like (5)–(6), postulating entities like ghosts, spirits, and gods:

- (5) When people die, their spirits go to an afterlife.
- (6) A perfect being with human-like mental states created the universe.

Rather than moral beliefs in general, I will be concerned with specific categories of moral beliefs, including consequentialist and deontological beliefs. Consequentialist beliefs are exemplified by their concern with the consequences of actions for the well-being of conscious animals, as in (7)–(8):

- (7) Texting and driving is wrong because it tends to cause injuries and deaths.
- (8) Giving to charity is right because it tends to increase happiness.

Deontological beliefs, on the other hand, focus instead on rights and duties, including claims about just punishments (9), fair rewards (10), and victimless crimes (11):

- (9) We have a duty to punish murderers, even if the world ends tomorrow.
- (10) People are born with the right to be given an education.
- (11) Gay sex is wrong, even if done in private.

This is not suggesting that deontologists necessarily endorse these particular beliefs, only that (9)–(11) are the *type* of beliefs characteristic of deontology. A controversial assumption made here is that all of the examples above are genuine beliefs. This assumption will be further discussed in section 2.3.

Before turning to the structure of debunking arguments, section 2 will provide a brief summary of Bloom and Greene's proposals for psychological mechanisms underlying certain religious and moral beliefs. Section 2 also lays out some important initial assumptions that I will be making throughout the paper, and why evolutionary considerations are being set aside. Section 3 will then provide the reasoning behind the first two conditions for a debunking argument: eligible beliefs must be (i) evidentially based on an intuition that can be (ii) explained by a psychological mechanism. However, not just any explanation is sufficient, and a third condition for a debunking argument is needed: (iii) the mechanism must be unreliable. Various notions of reliability are present throughout the literature, but section 4 will

categorize them and show two good arguments for establishing unreliability. With a proper metric of reliability in place, section 5 shows how it is possible for authors like Bloom and Greene to respond to an objection of circularity which has been raised by both Barrett (2007) and Berker (2009). Finally, the last section describes cases where beliefs are *not* evidentially based on intuition, and how to properly tell the difference. This provides the final condition for a debunking argument: (iv) any other evidence for the debunked belief(s) must be a rationalization.

While a good deal of the material here is already present in the literature, the novel task is to systematize it in a way that unifies arguments from different areas of philosophy and psychology. Revealing the underlying form of debunking arguments also presents new and strong responses to the motley crew of objections that have sprung up regarding debunking arguments in psychology, especially the circularity objection.

2. Examples and Assumptions

2.1. *Bloom and Greene's Proposals*

Cognitive psychology and neuroscience use information-processing mechanisms, including lower-level mechanisms at the level of neuroscience, to explain the intuitions underlying moral and religious beliefs. The researchers in cognitive science are eager to admit that this work is still in its infancy. However, this should not prevent a discussion about the relevance and implications of developments in these new fields. If anything, it should promote an excitement about what potential benefits lie in the future, and what the stakes of this research are. This section will be a brief advertisement (far from a review) for the kinds of cognitive mechanisms at the heart of Bloom and Greene's proposals.¹

Recent work in the cognitive science of religion has focused on mechanisms for attributing agency and intentions. Humans, like most animals, have a cognitive mechanism for identifying agents and distinguishing animals from other sorts of object. This mechanism might contribute to supernatural beliefs in several ways. It may be "hypersensitive," in that it produces a large number of false positives, and thus over-attributes agency to an environment (Barrett, 2004; Guthrie, 1993). In addition to agency, children and adults also have the ability to attribute intentions, an ability provided by theory of mind (TOM) mechanisms. Bloom (2004, 2007) has suggested that a result of the modularity between TOM and object detection mechanisms is that children develop into "natural born dualists," forming beliefs such as that one possesses a soul distinct from the body which can survive harm to the body itself. Furthermore, children and adults are likely to extend these conceptual notions of agency and intention to other domains (Bering, 2006; Bering & Bjorklund, 2004; Kelemen, 2004). As Bloom summarizes:

The proposal here is that there are certain early-emerging cognitive biases that give rise to religious belief. These include body-soul dualism and a hyper-sensitivity to

signs of agency and design. These biases make it natural to believe in Gods and spirits, in an afterlife, and in the divine creation of the universe. (2007, p. 150)

Turning to moral psychology, one early set of findings in this area suggested that emotion plays a large role in what adults and children find morally charged as opposed to conventional (Damasio, Tranel, & Damasio, 1990; Haidt, Koller, & Dias, 1993; Rozin, Lowery, Imada, & Haidt, 1999). Yet while Haidt and others group *all* moral intuitions as the product of emotion, Greene and colleagues used tasks designed to carefully distinguish deontological from consequentialist beliefs, such as evaluating trolley problem scenarios. Following the results of both fMRI and response-time measures to these scenarios (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), Greene's "dual-track" theory of moral judgment proposes that deontological intuitions are best explained as the product of mechanisms for emotional responses, while consequentialist intuitions are the product of mechanisms for higher executive functions such as "reasoning, planning, manipulating information in working memory, [and] controlling impulses" (Greene, 2007).

The dual-track theory has been further supported by subsequent studies. For instance, manipulating emotional levels by introducing disgusting smells and surroundings in a room (Schnall, Haidt, Clore, & Jordan, 2008) or showing a comedy sketch (Valdesolo & DeSteno, 2006) can significantly influence deontological moral intuitions. Judgments to punish, even at high cost to oneself, can be traced to emotional influence (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Additionally, emotional impairments found in patients with psychopathy result in more consequentialist intuitions (Koenigs et al., 2007).

As empirical claims, these have each received a fair amount of empirical challenges which will not be addressed here. This is because my case does not itself stand or fall with Greene and Bloom's proposals; they simply serve as excellent illustrations for present discussion. The question for this paper is whether these empirical proposals, if true, have any epistemic consequences for beliefs.

2.2. Evolutionary Considerations

There is no shortage of evolutionary considerations in the cognitive sciences of morality and religion. The best evolutionary accounts combine a plausible description of the environment of evolutionary adaptedness with quantitative models of how organisms pursuing a moral or religious strategy will outperform peers without these beliefs. Some recent proposals for the evolutionary benefit of religious beliefs include managing the "terror" which will occur in animals with self-reflective capacities (Greenberg et al., 1990), assisting social communication (Sosis, 2006), or ensuring a level of security and law, like an invisible police force (Wilson, 2002). Most recent evolutionary accounts of religious belief build on the established biological phenomena of kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971), to show how organisms who expanded on these strategies into a wider form of "strong reciprocity" could be more

successful than uncooperative or defecting organisms (Gintis, Bowles, Boyd, & Fehr, 2003).

Evolutionary explanations of morality have been employed by Joyce (2001), Ruse (1986), and Street (2006) to make debunking arguments. Greene (2007, p. 60) has also incorporated these evolutionary accounts into his own debunking argument, using their “ultimate” explanations to supplement his “proximal” theory. The psychological debunking arguments discussed here take much from the evolutionary debunking arguments employed by Ruse, Joyce, and Street. These “arguments from evolutionary history” generally claim that a belief *p* is debunked when it is discovered to be the product of fitness pressures unrelated to its content. In other words, if moral beliefs can be explained as the result of kin selection or reciprocal altruism, this undercuts our justification for believing them. Similarly, belief in God is undermined if it can be explained as the product of pressures for increasing security among human groups.

Theists like Barrett (2004) and Murray (2007) have been quick to point out that arguments from evolutionary history do not conclusively undermine religious beliefs. Barrett and Murray grant that religious beliefs may be caused by evolutionary pressures quite unrelated to the content of these beliefs. Yet this is not sufficient to debunk such beliefs for two reasons. First, the fitness pressures themselves may have been ultimately caused by God, where God designed the environment to favor organisms disposed to believe in him. This response is summarized by Murray:

Perhaps God set up our environment and the course of evolutionary history in such a way that we come to have cognitive tools that lead us to form beliefs in a supernatural reality If that is the way things work, then my beliefs would have a connection, albeit an indirect one, to the target of the belief, and a connection of that sort would not undermine the justification of the belief. (2007, p. 396)

Second, a belief being an adaptation to some environmental pressure does not necessarily make it *insensitive* to other features. It may be the case that mechanisms selected for the task of ensuring security turn out to *also* be reliable for the task of believing in God. Barrett and Murray both point out that there are externalist epistemological theories willing to grant justification to beliefs of this sort. Therefore, something more besides just an evolutionary history unrelated to the content of belief *p* is necessary to undermine belief *p*.

Ruse, Joyce, and Street each have more sophisticated versions of the evolutionary history argument that will influence the debunking arguments discussed later. However, I wish to leave evolutionary considerations at this point and focus on psychological ones. One reason for this separation is that evolutionary considerations are not *necessary* for a psychological explanation, nor are they necessary for psychological debunking arguments of the kind presented by Greene and Bloom. I intend to show that psychological debunking arguments can stand on their own, independently of evolutionary considerations. Therefore, I omit further discussion of

evolutionary histories from this paper, without denying that they play an important role in the ultimate explanation of moral and religious beliefs.

2.3. The Homogeneity of Beliefs

Bloom and Greene's treatment of moral and religious beliefs assumes that they are genuine beliefs, and have no special evidential status by virtue of their content. In other words, beliefs about the existence of gods and duties are not significantly different from beliefs about the planet Saturn and my car. Let us call this assumption the "homogeneity of beliefs."

Mason (2010) has noted that the homogeneity of beliefs is a substantial and controversial position which involves at least three meta-ethical and meta-religious commitments. First, it must be cognitivist, where beliefs (5)–(11) are genuine attitudes about the world that can be true or false. This is rejected by non-cognitivists about religious and moral beliefs, who claim that (1)–(4) are different in kind from (5)–(11), either because they are really imperatives like "don't be homosexual!" or exclamations like "boo homosexuality!" or fictional discourse like "let's pretend that homosexuality is forbidden." Additionally, the homogeneity of beliefs must admit the possibility of real moral facts, and ones that can be objective in some mind-independent way. As Mason correctly describes, these are some "heavy-duty meta-ethical assumptions," without which "moral intuitions would be immune to debunking" (2010, pp. 451–452).

The homogeneity of beliefs position is less controversial in the domain of religion, where cognitivism and realism about the existence of gods and souls is normal. It is more challenging in the domain of morality, where the homogeneity position is rejected by many contemporary ethicists, and possibly even Kant, who is Greene's primary target. This has led philosophers who make evolutionary debunking arguments, such as Joyce and Street, to specify that their target is moral realism, as opposed to various non-cognitivist, anti-realist, and constructivist positions. I see nothing preventing psychologists from making a similar qualification, with the important difference that moral realism itself is not the target, but particular moral realist beliefs.² It is important to note that moral realism is far from a straw man. Not only do many ethicists explicitly defend this position, but most laypeople may also be realists about their moral beliefs. In a series of studies by Goodwin and Darley, they found that:

[Most] people take beliefs about the wrongness of such transgressions to be quite objective—almost as objective as beliefs about matters of everyday or scientific fact, and more objective than beliefs about social conventions (i.e., appropriate dress wear and manners) and beliefs about matters of taste. (2010, p. 169)

Therefore, even if Bloom and Greene's debunking arguments are limited to those who accept the homogeneity of beliefs assumption, this is far from a straw man. Beyond this, it is an open question exactly how many people accept the homogeneity of beliefs and which beliefs are actually undermined.

3. Explaining Intuitions

An initial position one might take towards the relation between psychological explanations and beliefs is that there is simply no relation at all. Beliefs, belonging to commonsense psychology, are normative and intentional. Cognitive mechanisms are not. Thus, it is a category mistake to suppose that one has anything to do with the other. This section will lay out the first two conditions of psychological debunking arguments by way of responding to this initially plausible autonomy position.

Let us consider my belief that Saturn is gaseous. We may assume that the belief has two components: the proposition “Saturn is gaseous” and an attitude endorsing this proposition. The autonomy theorist will concede that perhaps my *attitude* may have some psychological explanation, but the proposition itself is about the gaseous nature of Saturn, not anything to do with me. The only things that are relevant to the proposition are facts about Saturn, because propositions have *evidence* rather than explanations. Merely providing an explanation of my attitude (but not any facts about Saturn) can have no consequences for the belief. Barrett seems to have this argument in mind when he criticizes Bloom for “a misunderstanding about explaining epistemic states in terms of biology [and psychology]” (2007, p. 61). In ethics, this may also be considered a version of the oft-repeated distinction between “is” and “ought” explanations, where the former can have no direct consequences for the latter. But this is perhaps misleading; a more appropriate description is an “is/is” distinction between two types of facts: facts about a person and facts about the content of a belief.³

The autonomy position is correct that carrying over findings from scientific psychology to commonsense psychology is not a simple matter, but there are three considerations that can respond to the autonomy position. First, psychological debunking arguments only attempt to *undermine* beliefs rather than falsify them. The autonomy theorist is correct that cognitive mechanisms are not themselves evidence about the truth or falsity of a proposition, and therefore cannot confirm or falsify a belief. To do so, one would need evidence related to whatever that proposition is about. However, there are a wide range of evaluative consequences for a belief other than simply being confirmed or falsified. Beliefs may also be undermined, when they are indirectly weakened by some information that weighs against the evidence one currently has for the belief.⁴ For example, if I discover that my doctor has been secretly accepting large amounts of money from pharmaceutical companies, this might undermine my beliefs about the necessity of prescriptions she suggests. This information has not directly falsified any of the beliefs I have formed based on my doctor’s advice, and a second opinion may still confirm the doctor’s claims. Yet, until this additional evidence comes in, the current evidence is outweighed by my discovery.

A second way to overcome the autonomy of commonsense psychology is that both moral and religious psychology attempt to explain the evidential source of beliefs, not beliefs themselves. Barrett objects to using unconscious cognitive mechanisms to explain beliefs, which comprise an attitude and a proposition. However, most moral

psychologists describe their target phenomena as moral *intuitions*. This is extremely important, since intuitions have a kind of dual citizenship in commonsense and scientific psychology. Like perceptions, they can be taken as evidence supporting the truth of a proposition. Simultaneously, like perceptions, they also have mechanistic explanations at the subpersonal levels.

There has been a large debate in philosophy about the evidential status of intuitions, but this largely focuses on whether intuitions *should* be evidence for beliefs. It would be difficult to find anyone who denies that intuitions *can be* and *sometimes are* evidence for moral and religious beliefs. Thus, I will specify cases where beliefs are evidentially supported by intuitions as the target of psychological debunking arguments. Clark and Barrett (2011, p. 662) have insisted that religious beliefs do not stand in evidential relations to anything, because they are not hypotheses. It is true that religious beliefs may not be hypotheses, but this does not imply that the beliefs cannot stand in evidential relations to anything. Plenty of beliefs, including (perhaps) that my car is black, are not hypotheses, yet they can still be supported by evidence.

One might worry that I am assuming a radical evidentialism, where every belief must be supported by proportional evidence. In response, I will point out that my proposals are entirely consistent with epistemological theories from theists like Bergmann (2006) and Wolterstorff (1983, pp. 163–164). Bergmann and Wolterstorff argue that beliefs are rational until challenged by stronger beliefs. I take it that the primary way for one belief to outweigh the other is to have stronger supporting evidence. And when challenged, I claim that many moral and religious beliefs appeal to intuitions in order to justify themselves. Cases that do not explicitly appeal to intuitions either are immune to psychological debunking, or are rationalizations (see section 6).

Finally, it is not just *any* mechanism that can undermine beliefs, but a specific kind. The autonomy theorist may point this out by noting that *every* belief presumably has some sort of psychological mechanism behind it. But obviously, not every belief that an individual has is false. Therefore, unless we are to abandon all of our beliefs, psychological mechanisms alone cannot undermine beliefs. Barrett provides a nice illustration of this argument: “suppose I believe I see a robin outside my window. You tell me you can exactly specify the neural pathways responsible for generating my belief. Does that mean the robin is not really there? Hardly” (2007, pp. 61–62). Barrett is correct that establishing the existence of just any mechanism responsible for a belief is insufficient to undermine that belief. After all, Barrett’s endorsement that a robin is outside is based on a psychological mechanism (as well as other initial conditions), yet explanations from vision science do not undermine the belief.

Some other feature is indeed needed to undermine a belief *p* beyond a mere mechanism that can be used to explain the endorsement of *p*. What the psychologist would need to show is that a mechanism producing that belief is unreliable. Including reliability into a psychological explanation extends the scope of that explanation, but not unrealistically. In fact, notions of reliability are a part of

mechanistic explanation in general. For instance, vision science does not have anything to say about whether or not my car is black. However, the scope of vision science does include general facts about the efficiency of the human visual system in normal conditions. If the belief about the color of my car is formed in conditions where the visual system is known to produce unreliable beliefs (say, seeing my car in the distance or in the dark), then the belief is indeed undermined.

4. Unreliable Mechanisms

4.1. *How to Evaluate Reliability*

How is the reliability of a psychological mechanism evaluated? One possibility is to simply assume that all unconscious cognitive mechanisms are reliable for their proper (evolutionary) function, given that they are the result of natural selection. Barrett has actually proposed this at one point: “nonreflective beliefs often correspond nicely to reality. This reliability comes from the observation that the mental tools responsible for these beliefs exist in large part because of their contribution to human survival throughout time” (2004, p. 9).

However, this assumption is faulty. Mechanisms with one evolutionary function are often co-opted for other functions, and therefore simply being the product of natural selection is not enough to ensure reliability for a given task. Further, even if it can be established that a task is the proper evolutionary function of that mechanism, natural selection is only a “satisficer,” and may produce mechanisms which successfully respond to task x without truly representing task x. This is not because the mechanism has broken or is being used in the wrong context, but rather, because it is designed to be systematically incorrect (McKay & Dennett, 2009). Such mechanisms confer a selective advantage because, as Barrett himself points out elsewhere (in something of a contradiction with the quote above):

[Natural selection] only favors minds that generate survival behaviors and not necessarily true beliefs. As already discussed, natural selection does not care about beliefs In fact, the illusions discussed above are part of the vast psychological literature demonstrating that human minds seem systematically to get things wrong for the sake of survival. When it comes to natural selection, Truth is expendable. (2007, p. 65)

Therefore, we cannot infer that a cognitive mechanism is reliable from the fact that it is the product of natural selection. Rather, it is an empirical question that must be independently established.

I propose that the reliability of a mechanism can be evaluated by two metrics: comparing the function of that mechanism to some objective measure of accuracy; and showing how that mechanism is improperly sensitive to the target source. For example, if we want to gauge the efficiency of a thermometer, we can use some sort of alternative measure of temperature (say, the boiling of water) and see if the thermometer marks the correct boiling point. Alternately, if the thermometer responds to information irrelevant to the target source such as the color of the room

or the gender of the person taking the measurement, it can safely be assumed that the mechanism is not a reliable one for measuring temperature, even if no alternative measure of temperature is available. Within psychology, these are known as framing effects (Kahneman & Tversky, 1981).

It is very important to add that reliability of a mechanism is always relative to a specific task. If a thermometer responds to the gender of the person taking the measurement, then this is obviously unreliable for the task of temperature measurement, but reliable for the task of gender-judgment. This also applies to psychological mechanisms. Very often, parts of the human visual system are incredibly effective for some tasks, such as detecting light, faces, and shapes. However, these same mechanisms are faulty for other tasks such as judging gradual or slight changes in stimulus (Weber's Law), especially when distracting information is present. Therefore, specifying the task of a given mechanism, as well as what information is irrelevant to that task, are important a priori assumptions to make.

4.2. The Redundancy Argument

As far as I can tell, there have been two good arguments made about the unreliability of cognitive mechanisms. The basic idea behind both is that a belief is justified only if one's evidence for it is in some way sensitive to the content of that belief (the second metric of reliability). For instance, my beliefs that Saturn is gaseous or that my car is black must be based on evidence that is in some way sensitive to facts about the chemical composition of Saturn or the reflectance properties of my car. If my evidence for the belief about my car is based on the testimony of a non-native English speaker who only knows one word for color ('black'), then my belief would still be the same despite the actual color of my car. Therefore, the belief is not appropriately sensitive to its content.

The first debunking argument from unreliable mechanisms is called the "redundancy argument" by Lillehammer (2003), and is attributed to Michael Ruse. Ruse (1986) argues that evolutionary explanations of moral beliefs successfully undermine or debunk those beliefs. His argument is more sophisticated than the basic argument from evolutionary history discussed earlier, and can successfully be carried over from evolutionary explanations to psychological ones. The redundancy argument assumes the second metric of reliability, where evidence is not properly sensitive, and uses a counterfactual test to evaluate this sensitivity. Ruse describes the basic idea: "given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act exactly the same ways. Hence the objective foundation for morality is redundant" (1986, p. 254). Lillehammer (2003) correctly notes, as I have, that the redundancy argument requires a realist conception of moral facts, and does not attempt to *falsify* moral beliefs, only to undermine them.

With all of this being said, the (unabridged) redundancy argument runs as follows:

- (1) A belief must be based on evidence that is sensitive to the content of that belief in order to be justified.

- (2) Thus, beliefs that would be held regardless of the truth-value of their content are undermined.
 - (3) Psychology shows that the evidence for belief *p* would be the same regardless of whether *p* or not *p* is the case.
- ∴ The evidence for belief *p* is not sensitive to its content.
 ∴ Belief *p* is undermined.

The third premise requires some basic counterfactual reasoning, but nothing too objectionable. Taking emotional mechanisms as responding to information about proximity, one poses the counterfactual question: if there were no real entities like rights and duties in existence, would emotional mechanisms still produce the same kind of moral intuitions? The same kind of test can be fruitfully applied to religious beliefs: if there were no God, afterlife, or soul, would the mechanisms for agency detection and TOM still produce the same kind of intuitions about these entities? The answer to both questions is arguably “yes.”

The redundancy argument is importantly different from an argument presented by Harman (1977) which claims that moral intuitions are *always* insensitive to facts. Sturgeon (1988) correctly notes that many moral beliefs do presume some sort of moral fact and there is no a priori reason to suppose that reasons for moral beliefs could never be sensitive to such facts. The argument here is a contingent claim that, in the case of deontological beliefs, the moral intuitions acting as evidence for the theory are, in fact, *not* sensitive to moral facts. Unlike Harman’s argument, the redundancy argument admits that this might have been different.

The redundancy argument is also different from the argument from evolutionary history described in section 2.2, which claims that moral/religious intuitions are unreliable because they are causally connected to environments unrelated to their content. Rather than the *actual* causal story of moral and religious intuitions, the redundancy argument compares their *possible* causal stories. Thus, as Murray and others point out, it is still possible that belief *p* indeed is the case. Yet, because the intuitive source of belief in *p* would still be present regardless of whether *p* really is the case, the belief is undermined.

4.3. *Argument from Irrelevant Factors*

Berker (2009) calls the second type of debunking argument the “argument from irrelevant factors.” This argument shares some of the same structure as the redundancy argument, but with an important difference to the third premise that requires no counterfactual reasoning. Instead, this argument claims that an unreliable cognitive mechanism is sensitive to irrelevant information, much like the thermometer that responds to color or gender. Consider the following formulation of the argument from irrelevant factors:

- (1) A belief must be based on evidence that is sensitive to the content of that belief in order to be justified.
- (2) Thus, beliefs sensitive to irrelevant features are undermined.
- (3) Feature *x* is irrelevant to the content of belief *p*.
- (4) Psychology shows that the evidence for belief *p* is sensitive to feature *x*.

- ∴ The evidence for belief *p* is not sensitive to its content.
- ∴ Belief *p* is undermined.

One clear presentation of the argument from irrelevant factors is in Sinnott-Armstrong (2007). Sinnott-Armstrong appeals to the fact that the mechanisms responsible for deontological beliefs are highly variable, depending on information like the messiness of the room or how close one is to an event. This information is presumably (though see Berker's (2009) objections below) not relevant to anyone's definition of the domain of moral properties. Just like the thermometer that provides different responses based on the color of the room, Sinnott-Armstrong concludes that the mechanisms for deontological beliefs are therefore not reliable. One might make a similar case for religious beliefs that result from a hypersensitive agency detection device: such a device responds not only to agency but also surprising noises (etc.), and thus is an unreliable guide to forming beliefs about actual agents.

The unreliability of agency detection devices is challenged by Clark and Barrett, who point out that there are many times where "human experience testifies" that agency detection and theory of mind mechanisms are indeed reliable:

You walk through the mall and person-beliefs pop up instantly and regularly. You lecture to a group of students and find yourself believing that you are lecturing to minded persons (and then attributing intentions to them in various circumstances). When you check out at the grocery store, you attribute both agency and intention to the clerk. (Clark & Barrett, 2011, p. 666)

Clark and Barrett's point is that mere *sensitivity* to irrelevant factors is not enough to make a mechanism unreliable, so long as it is not *actually* responding to these factors. Returning to the thermometer example, let's say that the gender-biased thermometer actually gives the correct temperature for men, but not for women.⁵ Clark and Barrett would insist that the thermometer is still reliable, but only in certain ideal conditions (only when men are taking the measurement).

The problem with this response is that I have already specified that debunking arguments are targeting beliefs whose evidence is exclusively based on intuitions. In order to determine whether the thermometer does indeed give the correct reading for men, we would need some independent measure of temperature to evaluate the thermometer (the first metric of reliability). However, this requires investigating the content of beliefs themselves, which we cannot assume (especially in the cases of moral and religious beliefs) psychologists have access to. Given this epistemic limitation, there is simply no way to sort out the ideal conditions in which the mechanism is indeed reliable, and I maintain that beliefs caused by this mechanism are still undermined.

5. The Circularity Objection

Now that the structure underlying psychological debunking arguments is exposed, we are in a place to develop a new and satisfying response to a strong objection of

circularity that has been raised against psychological debunking arguments by both Berker (2009) and Barrett (2007). The idea of the circularity objection is that establishing the reliability of a mechanism requires some objective knowledge about how it succeeds and fails. However, having objective knowledge about the failure of a mechanism assumes that the content of a belief it results in is false. Therefore, there is no non-question-begging way for psychology to undermine a belief before outside evidence about that belief is provided.

Barrett alleges that the psychologist who tries to undermine religious beliefs is guilty of exactly this kind of circularity: “do we count detections of ghosts, spirits, and gods as mistakes or as accurate detection? If as mistakes, we beg the question at hand. The same problem confronts other hypothesized cognitive systems responsible for promoting belief in gods” (2007, p. 69). Berker provides the same objection regarding moral beliefs, expressing skepticism about the possibility of any objective standard of accuracy in the moral domain:

It is question begging to assume that just because emotional processes in other domains consist in heuristics, therefore emotional processes in the moral domain consist in heuristics. How can we proclaim these emotional processes to be quick but sloppy shortcuts for getting at the moral truth unless we already have a handle on what the moral truth is? (2009, p. 317)

The positive alternative that Berker and Barrett advocate might be called a “retroactive error theory,” because it concludes that psychological explanations can only retroactively show why beliefs are undermined, but can never have direct epistemic consequences for a belief before its content is falsified.

In response to the circularity objection, it should first be recognized that this is a serious and valid objection to any attempt to undermine beliefs based on the *first* metric of reliability: comparing the output of a mechanism to some objective standard of accuracy. If this is the argument for why moral and religious psychology undermines beliefs, it surely must assume that these beliefs are already false in order to establish an objective standard of accuracy. However, if I have described the redundancy and irrelevant factors arguments correctly, neither of these employs the first metric of reliability. Instead, they both depend on the *second*: showing that a mechanism is improperly sensitive to its target source. Importantly, this does *not* involve any claims about the existence of the target source, only consensus about the features of what the target source *would* have if it were to exist.

Consider the redundancy argument: in order to show that agent detection mechanisms would still cause beliefs in souls, spirits, and gods, it is not necessary to assume that these beliefs are false. Following Ruse, all we need to do is to engage in counterfactual reasoning about the mechanism in question. Specifically, take the agent detection mechanism and imagine what beliefs it would cause in a world where souls, spirits, and gods do not exist. Now, imagine what beliefs they would cause in a world where such entities *do* exist. If the beliefs are identical, then according to the redundancy argument, the intuitive evidence for religious beliefs is

improperly sensitive. But such reasoning does not take on any assumptions about which of these worlds we *actually* live in. Now, consider the irrelevant factors argument: following Sinnott-Armstrong, it is possible to look at which factors emotional mechanisms are sensitive to and ask whether these factors are relevant to the task of making moral decisions. If not, then the mechanism is improperly sensitive to the content of moral beliefs. Yet, nowhere does this presuppose that rights and duties *do not exist*. All it involves is asking which features are relevant to these beliefs.

Although it may appear obvious that the messiness of a room is irrelevant to beliefs about rights and duties, Berker insists that there is no way to establish whether this information is truly irrelevant outside of philosophical debate. Going back to the formulation of the argument from irrelevant factors, Berker alleges that establishing premise 3 is, at the very least, philosophically contentious. If a deontologist contends that proximity is actually a relevant factor to moral properties, then it is unclear how one can respond without returning to philosophical stipulations: “so basically we have just recapitulated the same old battle of intuitions over the plausibility of consequentialism versus deontology in our evaluation of which sorts of factors are and are not morally relevant” (Berker, 2009, p. 325).

Berker is correct that philosophy is playing an important role here in helping to demarcate exactly what the domain of religious and moral beliefs is, revealing exactly which features are and are not relevant. Yet, this does not remove any of the *work* that psychology or neuroscience is doing in the argument, as Berker suggests. As I have formulated it, psychology and neuroscience must establish premise 4, while premise 3 can be established by philosophy (and anthropology and linguistics). Using framing effects to make reliability claims *does* require some prior agreement on the domain of religious and moral beliefs and their similarity to other contexts of belief. However, this is not as demanding a requirement as it may initially appear to be. One fascinating thing about Greene’s argument is that deontologists like Kant famously reject that any of their beliefs are based on emotional influence. Prior to investigation, there may even be a consensus that the messiness of a room is irrelevant to beliefs about duties and rights. I suspect that most deontologists explicitly reject this. Why is this not sufficient to establish the third premise? This consensus might be the result of armchair investigation of concepts, or it might also be the result of empirical survey work into the use of such concepts in the community.

6. Rationalization

6.1. Evidence Not Based on Intuition

Let us concede at this point that psychology might undermine beliefs in cases where the evidence an individual has for a belief is explicitly based on some intuition, and an unreliable cognitive mechanism can explain that intuition. One might object at this point that the beliefs Bloom and Greene are targeting are still immune from this

argument, because theists and deontologists often provide evidence for their beliefs aside from mere intuitiveness. Barrett describes this objection:

Theists generally do not believe in gods, nor do they justify their beliefs, on the grounds that such beliefs conferred a selective advantage in our evolutionary history. As no weight rests on this foundation, to remove it does no harm to these beliefs. (2007, p. 62)

It is true that few theists give hyperactive agency detection as a reason for their beliefs.⁶ Many deny that their beliefs are based on intuition at all. Instead, theists often claim that their religious beliefs are based on long chains of testimony that originate in divine revelation, no different in kind from non-religious historical beliefs. A select few even give empirical or logical arguments for their beliefs such as the cosmological, design, and ontological arguments. Since these beliefs are not explicitly based on intuition, Barrett would insist that they are shielded from being undermined by the psychology of religion.

Dean presents the same objection to Greene's argument against deontological beliefs on the grounds that moral philosophers like Kant do not always rely on intuitions about cases to support their beliefs, but also "theoretical elegance and consistency with the nonmoral aspects of a philosophical system" (2010, p. 52). In fact, Dean insists that "it is not until years into the development of his overall philosophical system that [Kant] begins deriving specific moral duties" (p. 52–53). Therefore, because the categorical imperative and other deontological rules are not explicitly based on an intuition that can be subject to psychological explanation, all of the beliefs generated by such rules are similarly immune.

Greene is aware of this response and emphasizes that not *all* deontological beliefs are immediately or indirectly based on intuitions. "What I am proposing," he states, "is that [theoretical elegance and empirical argument] is not how *characteristically* [emphasis added] deontological conclusions *tend* [emphasis added] to be reached and that, instead, they *tend* [emphasis added] to be reached on the basis of emotional responses" (Greene, 2007, p. 65). Greene is therefore making a claim about a particular class of deontological beliefs. So while the categorical imperative might be a belief that deontologists hold based on consistency with general facts about human reasoning, the vast majority of deontological beliefs are those like: "a murderer must always be killed, even if he or she has genuinely reformed" or "masturbation is not a permissible way to use one's body." Greene (2007) and Rey (2007) both note that a similar situation exists with religious beliefs, where a select few (e.g., the universe has a beginning) might be capable of being supported independently of intuitions, but the vast majority involve more specific beliefs about myths, supernatural entities, and rituals with metaphysical consequences. It is these more specific beliefs which are the target of potential debunking arguments from psychology.

6.2. *Rationalization in Controlled Settings*

Granting this point, a large practical difficulty now arises in how to draw the difference between moral and religious beliefs that *are* based on unreliable cognitive

mechanisms versus those based on independent evidence. The difficulty is that when evidence for a belief is either difficult to produce or nonexistent, individuals are prone to rationalize. Rationalizing can be described as a process of ad hoc reasoning about a belief that is already held, usually appealing to information currently available to the rationalizer. This means that in cases where a belief is in fact based on an unconscious cognitive mechanism, people will provide reasons for their belief that are often indistinguishable from reasons based on genuinely content-sensitive evidence.

Rationalization is most striking in controlled experimental and clinical settings, where it is either called “confabulation” or “misattribution.” The term ‘confabulation’ is typically reserved for rationalizations provided by patients with brain damage. Patients with memory deficits from Korsakoff’s syndrome may confabulate events, patients with blindness from Anton’s syndrome may give confabulations for their perceptual errors, and patients with neglect of one side of the body (hemineglect) may give confabulations for their unusual appearance and actions. In more pathological contexts like dementia and schizophrenia, systematically false beliefs, reasons, and stories are termed “delusions,” and it is often difficult to demarcate confabulation and delusion. Some researchers (e.g., Langdon & Turner, 2010) have suggested that confabulations and delusions occur along a continuum.

The difficulty in distinguishing confabulations from delusions reflects a general vagueness in the definition of confabulation, including a difficulty in distinguishing confabulations from normal reasons and beliefs. I do not have a developed account of confabulation to offer here, but an important necessary feature is that reasons provided for a confabulation are not independent of, and likely would not exist without, an unreliable psychological mechanism.⁷ Thus, confabulations may be true, but only by luck (like the broken clock which is right twice a day). Beyond this, confabulations also appear immune to revision in a way that makes them stand out from other beliefs. Turnbull, Jenkins, and Rowley propose that a confabulation is held “in spite of clearly presented evidence to the contrary” (2004, p. 6). Whether this insulation from revision is unique to confabulation or carries over into more normal cases of rationalization is an open question.

The same kind of clear rationalizations that patients with brain damage provide are also present in many social psychology experiments with perfectly healthy subjects. In these cases, rationalizations are more often called misattributions. Clear cases of misattribution include attributing one type of emotional arousal to a more cognitively salient source. For example, subjects will attribute the fear of a high bridge to the sexual arousal caused by an attractive interviewer, or the pleasure of smiling to a funny video. Perhaps the most famous and striking case of misattribution is a now classic study by Nisbett and Wilson in which subjects were presented with four pairs of stockings and asked which one was the best quality. They were not told that the four pairs were all identical. As Nisbett and Wilson describe:

Subjects were asked to say which [of four identical stockings] was the best quality and, when they announced a choice, were asked why they had chosen the article

they had. There was a pronounced left-to-right position effect, such that the right-most object in the array was heavily over chosen. For the stockings, the effect was quite large, with the right-most stockings being preferred over the left-most by a factor of almost four to one. When asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article. (Nisbett & Wilson, 1977, pp. 243–244)

In this study, the subjects who picked the stocking on the far right *did* give what appeared to be independent reasons for their belief that the rightmost stocking was superior. For example, one subject might have said: “this stocking has the most fine-grained nylon.” In the controlled setting of the experiment, it is clear that the evidence people are providing is not only false, but also not the real cause of their beliefs.

Confabulation and misattribution are relevant to psychological debunking arguments because it has been proposed that even in cases where people do not explicitly give intuitions as evidence of their moral and religious beliefs, the evidence they do provide are rationalizations of intuitions. Rey writes that he “find[s] it hard to believe that anyone really buys” the standard arguments for God’s existence (e.g., the ontological argument), and proposes instead that what appear to be reasons theists give to support their beliefs are actually a form of “sincere self-deception” (2007, p. 244).⁸

Haidt and colleagues presented subjects with vignettes like the following:

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love? (Haidt, 2001, p. 814)

As Haidt (2001) describes, most people answer that it was morally wrong for the brother and sister to make love, but then have difficulty providing reasons to support this. Since the usual reasons to disapprove of incest do not apply in this case (inbreeding, emotional harm, effects on others), it becomes difficult. Yet subjects still insist that the behavior is wrong even when their reasons come to an end. Haidt concludes that this suggests that the judgment is really driven by intuitions, with reasons being rationalizations. Greene draws similar conclusions, but specifically about deontological beliefs and reasons (rather than *all* moral reasoning): “what should we expect from creatures who exhibit social and moral behavior that is driven largely by intuitive emotional responses and who are prone to rationalization of their behaviors? The answer, I believe, is deontological moral philosophy” (2007, p. 63).

How can this claim possibly be substantiated? In controlled settings, it is a much more tractable problem to isolate and identify the real causes of beliefs and behaviors. Haidt’s vignettes are carefully designed to remove all the usual reasons people might appeal to, but this is not possible in real-life settings. Further, it is important that

with almost all cases of confabulation and misattribution, the beliefs defended by subjects are objectively known to be false. Experimenters know that the attractiveness of an interviewer does not change, and that the comedy clips and stockings are all the same. But if debunking arguments are to offer more than a retroactive error theory, they must provide some method for distinguishing rationalizations from genuine reasons.

6.3. Rationalization in Uncontrolled Settings

If the main identifying features of confabulations are insulation from counter-evidence and reliance on an unreliable mechanism, then there should be ways to discover confabulations even in uncontrolled settings. In Haidt's study, the moral beliefs of subjects were shown to be insulated from revision by pointing out that their reasons did not apply to the vignette. In normal contexts, moral reasons could be challenged by *hypothetical* dumbfounding. If a belief about the immorality of gay marriage is supported with: "it will discourage traditional marriage," one might challenge this by asking whether evidence such as the absence of an increase in marriage counseling or divorce rates in a country with gay marriage would show the belief to be false. I suspect that many ethicists have had experiences like this, as well as the following: after reading Peter Singer, a student claims that giving to charity cannot be morally required because there is no way to know where the money is going. The instructor then asks: "so, *if you did* know where the money was going, then it *would* be morally required?" The student then pauses, says "um. . ." and looks desperately for some other reason. These hypothetical dumbfoundings are good reason to think that the original reason given was not the *real* reason for their belief.

The other side of revealing rationalizations in uncontrolled settings involves showing that an intuition is *dependent* on an unreliable mechanism. One excellent method of discovering this is prediction: if a belief is *better predicted* by an unreliable cognitive mechanism than by the evidence a person gives for that belief. Greene (2007) gives an example of a hypothetical woman named Alice who goes on many dates and gives various reasons for who she likes and dislikes. Over time, Alice's friend eventually notices that a *better predictor* of who she likes or dislikes is how tall they are, and the friend even uses some statistics to show this. It is certainly possible that this is a coincidence, but given that we should favor a theory when it explains more of the data, a more likely explanation of Alice's behavior is that she has a genuine preference for tall men and is rationalizing. Similarly, it could be a coincidence that emotional responses are as good (or better) at predicting deontological intuitions than the categorical imperative, but a more likely explanation is that deontological beliefs are caused by the emotions and the categorical imperative is a rationalization. The same method carries over to religious beliefs: if agent detection mechanisms are a better predictor of intuitions underlying these beliefs, then sources such as the cosmological argument or historical arguments from divine revelation might be dismissed as rationalizations as well.

Is it the case that psychological mechanisms are better predictors of deontological and religious intuitions? I suggest that they are. None of the standard arguments for the existence of God are convincing or influential enough to be the source of religious beliefs for more than a handful of people. The fact that theists tend to reject the same type of historical argument in other religions suggests that this, too, is a poor predictor of religious beliefs. In other words, Christians who rely on the Bible as a historical source will reject the Vedas as a historical source. On the other hand, agent detection devices can predict a large range of religious beliefs.

Similarly, while the categorical imperative predicts many deontological intuitions, there are important ones that it fails to account for. Most deontologists agree that it is permissible to lie to a rampaging murderer or kill one person to save the whole world. While deontological moral *theory* cannot predict these deviations, they can be explained by the dual-systems theory. Therefore, in these cases, psychological mechanisms are better predictors of the intuitions behind these beliefs than the reasons believers may give.

7. Conclusion

Resistance to psychological debunking arguments comes in many varieties. One might object that beliefs and cognitive mechanisms exist in different domains (intentional versus non-intentional), or that debunking arguments cannot undermine one belief without undermining all of them, or that the methods of debunking arguments are viciously circular. It has been my hope that a closer examination of the structure of debunking arguments can provide systematic responses to each of these objections, and more.

The key to linking the domains of commonsense psychology and cognitive mechanisms has been intuition, which can simultaneously serve as the explanandum of a mechanistic theory⁹ and the evidence for a belief. When it is specified that a mechanism which explains this intuition must be unreliable for forming beliefs about the target domain, one no longer needs to worry about discarding all beliefs (since presumably, many are based on mechanisms which are reliably sensitive to their target domain). The question then becomes how to establish reliability in a non-circular way, and two methods for this were proposed: showing that the truth of a belief is redundant (it does not play a role in explaining the intuition); and showing that the mechanism is sensitive to irrelevant factors. Neither of these methods is question-begging in the way that Berker and Barrett allege.

Finally, a large challenge for debunking arguments is to identify when the evidence for a belief is genuinely based on intuition, given the rampant phenomenon of rationalization. One might worry that this makes debunkers no better than Freudians who insist that a woman's ambitions are based on penis envy, despite her emphatic claims to the contrary! However, I am optimistic that rationalizations can indeed be detected outside of controlled settings. Sometimes, they are clear due to hypothetical dumbfounding. Other times, rationalizations may be apparent from the fact that an

underlying cognitive mechanism is a better predictor of beliefs and behavior than the evidence provided by an individual. In the previous section, I agreed with Rey and Greene that it is plausible that many religious and deontological beliefs fit these criteria. At the same time, a clearer set of criteria for establishing rationalization in uncontrolled settings is an area where more work can and should be done.

After all this qualification, one might worry that the list of candidates for psychological debunking arguments has been shortened to the point of being uninteresting. While it is certainly true that these conditions limit the scope of debunking arguments, the targets remain extremely popular both among academics and the general public. In some opinion polls, roughly half of Americans oppose same-sex marriage and over two-thirds believe in an afterlife. If the arguments presented here are valid, it is not the case that psychological articles with titles like “Disgust sensitivity predicts intuitive disapproval of gays” or “The natural emergence of reasoning about the afterlife as a developmental regularity” are without consequence for these beliefs.

Acknowledgements

would like to thank Jonathan Hricko, John Waterman, Jeff Maynes for their helpful comments and suggestions on earlier drafts. Thanks also to Martin Rice and the audience at a faculty seminar at the University of Pittsburgh, Johnstown in November, 2011.

Notes

- [1] I wish to emphasize that these proposals are taken as mere examples from two large fields where not all researchers are interested in either supporting or debunking moral or religious beliefs. Also, since my task is not to provide a review of work in these fields, this section will necessarily be superficial and will skim over a large body of work and many distinctions (i.e., the by-product versus adaptation debate in the cognitive science of religion) that would require much more space to properly discuss.
- [2] This allows for Greene to be skeptical about deontological beliefs and still be a realist about the contents of utilitarian beliefs, since he maintains a non-intuitive basis for these beliefs.
- [3] There are many varieties of the “is/ought” distinction depending on what you take “ought” explanations to be. Traditional versions usually take the distinction to be a semantic or epistemic one, where moral claims are imperatives or emotional expressions and thus derive no consequences from factual descriptions. Since we are specifically targeting a cognitivist and realist approach to moral claims, the “is/ought” distinction mentioned here is a metaphysical claim that the content of moral beliefs is metaphysically distinct from facts about human psychology.
- [4] A careful reading of Greene and Bloom reveals that they have this consideration in mind. In the quotes used in the introductory section, Bloom clearly states that “nothing from empirical study of human psychology can refute religious belief,” although the empirical work can “challenge [their] rationality” (2009, p. 125). Greene also talks about “casting doubt” (2007, p. 36) on deontological beliefs rather than directly refuting or falsifying them.

- [5] Thanks to an anonymous reviewer for this suggestion.
- [6] Although arguably this is the “sensus divinitatus” position proposed by John Calvin and defended by Plantinga (2000).
- [7] This is weaker than the claim that confabulations are *caused* by unreliable mechanisms. Also, it is not sufficient to define confabulations, since we can imagine many cases of good beliefs dependent on unreliable mechanisms. Consider a detective trying to solve a crime who has a dream which gives him an idea leading him to investigate a suspect and crack the case. Dreams are unreliable for the task of solving crimes, and the reasons the detective gives have nothing to do with the dream itself except that they were inspired by, and thus would not exist without, the dream.
- [8] Rey actually goes even further than this, and doubts whether people actually endorse the beliefs at all. He claims that “at some level they [the well-informed theists] believe this claim is false” (2007, p. 245). It is not necessary for the present purposes to follow Rey this far in order to establish the more modest (but still bold) claim that many reasons for religious beliefs are confabulations.
- [9] I have attempted to remain relatively neutral about the necessary and sufficient conditions for a psychological explanation, except for the broad claim that it somehow involves identifying a mechanism which (along with other initial conditions) produces the explanandum. However, I wish to leave the details of this vague as to leave the door open to future work in the nature of explanation. Even the brief description provided here may be incorrect. While I have been assuming a metaphysical (or ontic) version of explanation, I do not wish to deny that there is an important epistemic and pragmatic aspect to explanation of any kind in the sciences. Therefore, the first condition of this position should be understood to claim: “however psychological explanation works, a mechanism explains the endorsement of p.”

References

- Barrett, J. (2004). *Why would anyone believe in god?* Walnut Creek, CA: AltaMira.
- Barrett, J. (2007). Is the spell really broken? Bio-Psychological explanations of religion and theistic belief. *Theology and Science*, 5, 57–72.
- Bergmann, M. (2006). *Justification without awareness*. New York: Oxford University Press.
- Bering, J. (2006). The folk psychology of souls. *Behavioral and Brain Sciences*, 29, 453–498.
- Bering, J., & Bjorklund, D. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology*, 40, 217–233.
- Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs*, 37, 293–329.
- Bloom, P. (2004). *Descartes' baby*. New York: Basic Books.
- Bloom, P. (2007). Religion is natural. *Developmental Science*, 10, 147–151.
- Bloom, P. (2009). Religious belief as an evolutionary accident. In J. Schloss & M. Murray (Eds.), *The believing primate: Scientific, philosophical and theological reflection on the origin of religion* (pp. 118–127). New York: Oxford University Press.
- Clark, K.J., & Barrett, J. (2011). Reidian religious epistemology and the cognitive science of religion. *Journal of the American Academy of Religion*, 79, 639–675.
- Damasio, A., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral Brain Research*, 41, 81–94.
- Dean, R. (2010). Does neuroscience undermine deontological theory? *Neuroethics*, 3, 43–60.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution of Human Behavior*, 24, 153–172.

- Goodwin, G., & Darley, J. (2010). The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. *Review of Philosophical Psychology*, 1, 161–188.
- Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S., & Lyon, D. (1990). Evidence for terror management II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. *Journal of Personality and Social Psychology*, 58, 308–318.
- Greene, J. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 3: The neuroscience of morality: Emotion, disease, and development* (pp. 35–79). Cambridge, MA: MIT Press.
- Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400.
- Greene, J., Sommerville, L., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.
- Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. New York: Oxford University Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.
- Hamilton, W. (1964). The genetic evolution of social behavior. *Journal of Theoretical Biology*, 7, 1–52.
- Harman, G. (1977). *The nature of morality: An introduction to ethics*. Oxford: Oxford University Press.
- Joyce, R. (2001). *The myth of morality*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1981). The framing of decisions and the psychology of choice. *Science*, 21, 453–458.
- Kelemen, D. (2004). Are children “intuitive theists”? Reasoning about purpose and design in nature. *Psychological Science*, 15, 295–301.
- Koenings, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, 446, 908–911.
- Langdon, R., & Turner, M. (2010). Delusion and confabulation: Overlapping or distinct distortions of reality? *Cognitive Neuropsychiatry*, 15, 1–13.
- Lillehammer, H. (2003). Debunking morality: Evolutionary naturalism and moral error theory. *Biology and Philosophy*, 18, 567–581.
- Mason, K. (2010). Moral psychology and moral intuition: A pox on all your houses. *Australasian Journal of Philosophy*, 89, 441–458.
- McKay, R., & Dennett, D.C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32, 493–561.
- Murray, M. (2007). Four arguments that the cognitive psychology of religion undermines the justification of religious belief. In R. Genet & C. Genet (Eds.), *Evolution of religion: Studies, theories, and critiques* (pp. 393–398). Santa Margarita, CA: Collins Foundation.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Plantinga, A. (2000). *Warranted Christian belief*. New York: Oxford University Press.
- Rey, G. (2007). Does anyone believe in God? In L. Antony (Ed.), *Philosophers without gods* (pp. 243–265). Oxford: Oxford University Press.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality & Social Psychology*, 76, 574–586.
- Ruse, M. (1986). *Taking Darwin seriously*. Oxford: Blackwell.
- Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755–1758.
- Schnall, S., Haidt, J., Clore, J., & Jordan, A. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34, 1096–1109.

- Sinnott-Armstrong, W. (2007). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2: The cognitive science of morality: Intuition and diversity* (pp. 47–77). Cambridge, MA: MIT Press.
- Sosis, R. (2006). Religious behaviors, badges, and bans: Signaling theory and the evolution of religion. In P. McNamara (Ed.), *Where God and science meet: How brain and evolutionary studies alter our understanding of religion, volume 1: Evolution, genes, and the religious brain* (pp. 61–86). London: Praeger.
- Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127, 109–166.
- Sturgeon, N.L. (1988). Moral explanations. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 229–255). Ithaca, NY: Cornell University Press.
- Trivers, R.L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57.
- Turnbull, O., Jenkins, S., & Rowley, M. (2004). The pleasantness of false beliefs: An emotion-based account of confabulation. *Neuro-Psychoanalysis*, 6, 5–45.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17, 476–477.
- Wilson, D.S. (2002). *Darwin's cathedral: Evolution, religion, and the nature of society*. Chicago, IL: University of Chicago Press.
- Wolterstorff, N. (1983). Can belief in god be rational if it has no foundations? In A. Plantinga & N. Wolterstorff (Eds.), *Faith and rationality* (pp. 135–186). Notre Dame, IN: University of Notre Dame Press.

Copyright of Philosophical Psychology is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.