THE COGNITIVE SCIENCE OF RELIGION:
A CASE FOR ATHEISM

P. Wesley Edwards

2017

1 This is a substantially revised version of a 2016 thesis (Edwards). In addition to reorganization, rewording, and several corrections, this version also includes a number of expanded discussions, as well as fuller responses to some important opposing points of view.
# TABLE OF CONTENTS

1. INTRODUCTION .......................................................................................................................... 2

2. COGNITIVE SCIENCE OF RELIGION: BRIEF OVERVIEW AND STATE OF THE FIELD ......................................................................................................................... 5

   What is the Cognitive Science of Religion? ............................................................................. 5
   A CSR Explanation for Religion .............................................................................................. 13
   The CSR Standard Model ........................................................................................................ 22
   Epistemic Consequences .......................................................................................................... 23

3. OBJECTIONS TO AN ATHEISTIC INTERPRETATION OF CSR ............................................ 25

   Introduction ................................................................................................................................ 25
   Alvin Plantinga's Foundational Work ....................................................................................... 26
   Critique of Classical Foundationalism ...................................................................................... 26
   Basis for Theism as Justified Properly Basic Belief .................................................................. 29
   Warrant and Burden of Proof ...................................................................................................... 32
   Evolutionary Argument against Naturalism (EAAN) .............................................................. 36
   Summary .................................................................................................................................... 41

   EAAN-Based CSR Critique ........................................................................................................ 41
   The Bayesian-Based Irrelevance Critique .................................................................................. 46
   Expanding the Irrelevance Argument ......................................................................................... 52
   Summary .................................................................................................................................... 60

4. CRITICAL ASSESSMENTS OF THE OBJECTIONS .................................................................. 62

   Critique of “Reformed Foundationalism” ................................................................................. 62
   Defeating Warrant ...................................................................................................................... 64
   A Reidian-Based Defense of Naturalism .................................................................................... 68
   Evolution and the Cost of Truth ................................................................................................. 73
   Humean Bootstrapping .............................................................................................................. 85
   Summary .................................................................................................................................... 90

5. BUILDING UPON THE CRITICAL ASSESSMENTS ................................................................ 91

   NOMA ........................................................................................................................................ 92
   A Brief Overview of Bayes’ Theorem ......................................................................................... 93
   On Recognizing a Good Theory ................................................................................................. 98
   Duhem-Quine and Hypothesis Bundling .................................................................................. 100
   Good Theory Virtues ................................................................................................................. 104
   Is Science Relevant to the Supernatural? ................................................................................ 110
   Do Natural Explanations Compete with Supernatural Explanations? .................................. 113
CSR and the God Hypothesis ................................................................. 115
Burden of Proof .................................................................................. 115
Parsimony and Predictive Success ..................................................... 119
CSR and the Rational Justification of Theism .................................... 130
Summary ............................................................................................. 135

6. LOOKING FORWARD ......................................................................... 138

7. WORKS CITED .................................................................................. 143
ABSTRACT

The cognitive science of religion (CSR) has risen to prominence in the 21st Century as the latest in the long history of natural explanations for religious belief. It is, however, unique in at least two ways: it is well supported empirically, and it explains and predicts the broad contours of the universal phenomenon of religious belief, including its most counterintuitive aspects. These unique features create fresh insight into certain contentious questions within the epistemology of religion. This paper focuses specifically on CSR’s epistemic significance for the rational justification of theism, and defends the view that CSR's findings significantly diminish this justification. After reviewing the historical context of CSR, I examine some of the most powerful arguments against the idea that CSR supports atheism. I then attempt to build upon important work from a number of areas—argumentation theory, the philosophies of science and religion, and crucially, evolutionary theory itself—in order to identify a set of epistemological tools and methodologies, which, when used in combination, appears to show that CSR does indeed significantly reduce the rational justification for theism.
CHAPTER 1

INTRODUCTION

Belief in supernatural entities—whether tree spirits, ancestor spirits, witches, or the Abrahamic God—has been as universal a phenomenon as art, body ornamentation, and music, and for at about as long (Mithen 174–78). While there has been a long, rich and important history of theologically oriented philosophical thought, from Aristotle and Aquinas to Swinburne and Plantinga, it is significant that the vast majority of humans across the globe have held their supernatural beliefs on the same basis that they hold their many other intuitive beliefs—that is, as givens, as beliefs held unreflectively. As we will see, a number of theistic scientists and philosophers see in this observation the very hand of God. They argue that this universal pattern points to something like Thomas Reid's *sensus divinitatus*, an inborn (though perhaps sin-corrupted) God-sense. So understood, they argue, theistic belief turns out to be completely rational. Yet, as we will also see, others argue that this very same pattern is not just explicable entirely in naturalistic terms, but in fact provides powerful evidence against God's (or gods') existence, or at least against the rational justification for belief in such entities.

Only recently have naturalistic explanations for religious belief been rooted in evolutionary-based accounts of human cognition—accounts that purport to show innate cognitive biases that predispose humans to believe not just in supernatural agents, but supernatural agents whose characteristics are constrained in telltale ways. Interestingly, within the interdisciplinary field in which this debate is developing, a field known as the
cognitive science of religion (CSR), both the theistic and atheistic camps agree not only on the underlying data, but also on much of the evolutionary framework used to explain that data. Where they differ is in their interpretation of the epistemic significance of these findings.

This paper focuses specifically on the epistemic significance of CSR for the rational justification of theism and defends the view that CSR's findings significantly undermine this justification. Chapter 2 will provide a brief historical survey of CSR as well as the current state of the field. Chapter 3 will outline and distill the main philosophical objections to an atheism-supporting interpretation of CSR’s findings and in the process survey the state of the field in the relevant areas of the philosophy of religion. As that chapter unfolds, we will discover Alvin Plantinga's profound influence on the shape and force of these objections. Chapter 4 will critically assess these same objections and argue that this debate can be advanced through the combined application of the following: argumentation theory relating to burden of proof and pragmatics, a few key insights from the philosophy of science, and crucially, an often-overlooked aspect of evolutionary theory itself. Chapter 5 will develop this combined approach in order to extend Chapter 4’s critical assessments. This extended assessment, it will be argued, forces us to conclude that CSR does indeed seriously undermine the rational justification of theism.

---

2 “Atheism” has some formal distinctions, one of which is between positive and negative atheism. The latter is commonly referred to as “agnosticism,” or sometimes, “non-theism.” What all types of atheism share, however, is an absence of belief in the Abrahamic God or any other supernatural god. See (Martin) for a thorough treatment of these varieties of atheism. For convenience, and given the focus of the present work, I will use the terms “atheism” and “atheist” in the broader, absence-of-belief sense.
for theism. Chapter 6 will identify and respond to a number of objections that can be raised to this approach.
CHAPTER 2

COGNITIVE SCIENCE OF RELIGION: BRIEF OVERVIEW AND STATE OF THE FIELD

The focus of this paper is philosophical, not scientific. The purpose of this chapter is to introduce the key concepts of CSR along with some of the reasons it is taken so seriously in the scientific community. Our focus, however, is not with mounting a scientific defense of the field’s findings, but rather with assessing the epistemological consequences of CSR if its findings were true.

What is the Cognitive Science of Religion?

Jonathan Jong gives perhaps the most general definition of CSR when he describes it as “a naturalistic research programme that aims to provide general explanations for the cross-culturally recurring collection of psychological phenomena (e.g., beliefs, behaviours) associated with supernatural agents” (Jong 521). While this is arguably too broad, he does follow it with a clarification that helpfully constrains our present discussion: “CSR is the study of human persons, not divine ones; it is the study of people’s concepts of gods, not the gods themselves” (Jong 522). A more focused definition of the field comes from one of its founding contributors and the person who coined the term “cognitive science of religion,” Justin Barrett (Barrett, “Exploring”). CSR, he tells us, “brings theories from the cognitive sciences to bear on why religious thought and action is so common in humans and why religious phenomena take on the features that they do” (Barrett, “Cognitive Science” 768). This helpful definition makes
clear that a proper understanding of CSR requires an understanding of its more general progenitor: cognitive science.

Cognitive science emerged as a reaction to behaviorism, the view that the mind is an indivisible, general-purpose processor that contains no “ready-made reaction[s]” or “innate ideas” (qtd. in Pinker 20). For the behaviorist, heredity simply has nothing to do with behavior, which is entirely formed by stimulus-response conditioning—by the environment. As such, the mind is thought of as a kind of “blank slate,” a canvas on which is painted whatever social, political, and religious teachings a child happens to be exposed to. Under this view, even a child’s talents are environmentally determined. She becomes that painting, as it were. Of course, this would imply that culture is completely autonomous from any inborn properties of the brain. This “blank slate” perspective, still alive in the social sciences, can trace its roots directly to John Locke and the fight against racism and sexism during a time when such views were commonly justified by appeal to supposed innate limitations (Pinker 18–23).

This mind-as-blank-slate view had been on a collision course with a separate historical track. The conceptual walls that have divided the human realm from the rest of the universe have been falling since Newton’s time. Newton showed that the physical laws on earth were the same as those in the heavens. Lyell erased the distinction between the dynamic, creative geological past and what had been assumed to be a completely static present. Others, in different ways, began breaching the walls separating life forms

3 I draw these examples, including the helpful “walls” metaphor, from (Pinker).
from each other and even from non-living matter. This finally culminated in the greatest hammer blow yet seen against these old conceptual barriers: Darwin’s bottom-up, algorithmic explanation of life’s complexity and apparent design. A remaining wall, still in the process of coming down, “divides matter from mind, the material from the spiritual, the physical from the mental, biology from culture, nature from society, and the sciences from the social sciences, humanities and arts” (Pinker 31). It is at the base of this final wall that both cognitive science and evolutionary psychology appear on the scene.

Cognitive science enters the picture during the middle of the twentieth century as a reaction to the then-dominant blank slate view of mind (Pinker 31). What followed is known as the “cognitive revolution” (Barrett, “Cognitive Science” 769). The ideas characterizing this revolution are (1) that minds, rather than being passive blank slates, are instead innately full of content and pro-action, and (2) that we can begin to explain the mental in terms of the physical by drawing on certain computational concepts, especially the notion of interacting, specialized, “modules,” described variously as “inference systems” (Boyer 17), “cognitive domains” and “intelligences” (Mithen 37), and “mental tools” (Trigg and Barrett 5).

As innate characteristics of human cognitive architecture, these modules are human universals. As such, they create the cognitive framework within which cultural variability forms. To be sure, there is wide and effectively open-ended variability in human cultural expression. Yet our shared cognitive architecture constrains aspects of that variability in telltale ways. In language, for example, our grammatical “programs” limit the ways in which humans build language. So, while there is no practical limit to the
number of human languages that can exist, their grammatical structures—their building blocks—reflect the telltale constraints imposed by our cognitive architecture. For example, unrelated languages are observed to “have verbs, objects, and pre– or postpositions to start with, as opposed to having the countless other conceivable kinds of apparatus that could power a communication system” (Pinker 37).

Crucially, our human cognitive modules are considered innately content rich, which is to say that they are not only “pre-programmed” with certain strategies or rules of thumb, but also “pre-loaded” with a considerable amount data—information—relevant to those algorithms (Mithen 43). This view represents a fundamental shift from the passive blank-slate view to an active, pre-configured, multiple-specializations view. It is a view that begins to shed light on why, for example, children can learn languages so quickly, without conscious effort at all, and with very limited input (Mithen 44). But where do these modules come from, and what accounts for why they do what they do?

Enter evolutionary psychology, “the study of the phylogenetic history and adaptive functions of the mind” (Pinker 51). If Darwin had shown how the apparent design of a bat’s wing could emerge from the mindless algorithm of natural selection and descent with modification, then why not the mind itself? In the late 1980s and early 1990s, this new school of thought began arguing that evolutionary theory, if applied to the emergence of the human mind, would predict that the mind should be more like a “Swiss army knife” than the undifferentiated, passive “sponge” of the blank-slate model (Mithen 43). Furthermore, this specialized, modularized mind should be expected not
only to be pre-programmed with rules and information. It should also be proactive, animated by various pre-programmed goals (e.g., *eager* pattern seeking).

Why would evolution lead to such expectations? Because such characteristics are far more adaptive than the passive, general-purpose behaviorist alternative. For example, consider the split-second, life-and-death decisions that creatures have to face on very limited information. A general-purpose, unbiased information processor would have to analyze far too many variables and possibilities in order to assess the best course of action *in time*; on the other hand, a collection of specialized modules, each with pre-loaded information and pre-programmed rules and goals (operating largely unconsciously) could produce a survivable response quickly enough to make a difference (Mithen).

This Swiss army knife analogy can be seen as representing one end of a spectrum of views, with the other end occupied by the mind-as-undifferentiated-sponge view. Of course, extremes rarely capture reality, and this case is no exception. Modularity, for example, may not take firm shape until after early childhood and the influence of culture. More importantly, were our mental modules really as non-interacting as the blades of a Swiss army knife, then human creativity, behavioral flexibility, and unpredictability would be extremely difficult to explain (Mithen 58–59). If the modularity view really does capture something about the nature of the human mind, then there must be some interaction between these modules.

One model of such cross-module interaction is that it is mediated by yet another module, but one which can be thought of as sitting on top of the rest—a meta-
representational module (Mithen 59). Regardless of the specific metaphor, the key point is that knowledge is seen as flowing between these specialized modules, allowing different cognitive tools to work on information outside the domains for which they had originally evolved. One effect of this cross-domain data flow would be the “mixing up of knowledge about different types of entities in the real world—knowledge which would have been ‘trapped’ in separate cognitive domains” (Mithen 177). In fact, this “mixing” ability may be behind not only our human obsession with analogical reasoning and metaphor, but even art and music.

Steven Mithen, for example, proposes that art is the result of cognitive fluidity between once strictly separated modules. His idea is that art results from the fluidity that emerged between the natural history module (interpreting indicators or “symbols” like hoof prints), the social module (including communication), and the technical module (producing artifacts from mental imagery). As the walls between these modules became increasingly porous, one side-effect might well be the emergence of art as “artifacts/images with symbolic meanings as a means of communication” (Mithen 163). But if art could arise as a side-effect of the evolution of cognitive fluidity in humans, what of other uniquely human phenomena, such as religion? Mithen, an archaeologist, notes that the evidence suggests that the emergence of complex art (around the Upper Paleolithic) was concurrent with the emergence of belief in supernatural beings and an afterlife (Mithen 174–78).

It is worth noting at this point that the use of cognitive science and evolution to “explain” religion is by no means the first attempt to understand it in natural terms. The
suggestion that such natural explanations exist goes back at least as far as the sixth century BC (van Inwagen 128) when Xenophanes famously said, “if oxen, horses, and lions . . . could fashion works as men do, horses would paint horse-like images of gods and oxen ox-like ones” (Xenophanes 5). Later, David Hume would introduce one of his books by saying that, with regard to religion, “there are two questions in particular which challenge our attention, to wit, that concerning its foundation in reason, and that concerning its origin in human nature” [emphasis added] (Hume 14). Closer to our own era, actual natural theories of religion began to emerge, such as those by Freud, Marx, and Feuerbach; but these early efforts have all been justly criticized as being either too vague or too untestable (van Inwagen). Elsewhere, however, something more promising had been developing. The historical track beginning with Darwin’s ground-breaking work, and eventually leading to modern evolutionary psychology, began to point the way to a fundamentally new and different kind of natural explanation for religion, one that would, for the first time, be empirically well grounded.

The sustained application of cognitive science and evolutionary psychology specifically to religious thought and behavior did not begin until the early 1990s with such works as Guthrie’s *Faces in the Clouds*, followed somewhat later by works that emphasized the role of evolution, such as Boyer’s *Religion Explained*. It was during this period (2000) that Justin Barrett would introduce the term “cognitive science of religion” (Barrett, “Cognitive Science” 780), and succinctly describe the field as drawing “upon the cognitive sciences to explain how pan-cultural features of human minds, interacting
with their natural and social environments, inform and constrain religious thought and action” (Trigg and Barrett 4).

CSR has inherited many of its key assumptions from this rich historical context, such as the rejection of the “blank slate” view in favor of one that sees the mind as composed of a collection of collaborative, content-rich, innate cognitive modules that are independent of culture—at least in terms of their basic settings—and which form the bases of our many well-documented universal cognitive biases. In addition, these specialized modules are assumed to have evolved as part of a quick response system that operates largely outside of our conscious awareness, delivering its outputs to our much slower reflective system for higher-level synthesis and evaluation (Trigg and Barrett 5). These assumptions form part of the CSR Standard Model. But we still need to understand how this model of cognitive architecture specifically explains the ubiquity of belief in supernatural agency—that is, what puts the “R” in CSR?

From an evolutionary perspective, religion can be seen as (1) a directly beneficial adaptation, (2) a result of meme selection, or (3) a by-product—a side effect—of the normal operation of cognitive faculties evolved for other purposes (Wood 734).

---

4 Relevantly, evidence suggests such higher-level, conscious deliberation may be less like a balanced reasoner, and more like a “spin doctor” or lawyer, weaving stories to integrate and defend these deliverances (Pinker 43).

5 “Meme” is adapted from the word “gene” and is the idea (introduced by Richard Dawkins in The Selfish Gene) that concepts can “evolve” in the somewhat Darwinian-like environment created by human brains interacting through language. Like genes, these memes evolve not for the specific advantage of their hosts (though mutually beneficial situations can and do occur), but for their own survival—memes, much like genes in this sense, are “selfish.” What helps some memes proliferate while others die off seems to depend on various unique features of human cognitive architecture. It’s worth cautioning here that modern colloquial usage of “meme” has drifted considerably from its original definition.
worth noting that these evolutionary interpretations are not necessarily mutually exclusive. For example, while our receptivity to certain ideas may be seen as a side-effect of our cognitive architecture, once such an idea is “captured” by our minds through some happenstance, its subsequent effects may be directly adaptive or maladaptive for the host (and/or for the meme) and can therefore be directly selected for or against, further shaping its development. With this qualification in mind, we will see that the by-product interpretation does much of the heavy lifting in CSR as it appears to best explain the data. It is, as a result, another important part of the CSR Standard Model (Jong 522). With this conceptual background in hand, let’s take a look at a typical way in which these ideas have been deployed in CSR to “explain” the phenomenon of religion.

A CSR Explanation for Religion

We humans, much more so than any other species, are able to decouple cognition from the here and now—that is, from the stimuli of our immediate environments—and entertain what-if scenarios. This is central to our ability to plan and to process information provided by others. In particular, this ability to decouple allows us to fire our “inference engines” (yet another way of describing our content rich, specialized cognitive modules) without the actual presence of physical stimuli. This has many consequences, some unintended. Pascal Boyer, for example, argues that “Supernatural concepts are just one consequence of the human capacity for decoupling representations” (Boyer 131).

Boyer goes on to argue that it is the combination of specialized inference systems and decoupled cognition that may best explain why so much of human behavior has no clear adaptive value. As an example, he points to the universal phenomenon of music.
Our uniquely evolved auditory cortex has areas that specialize both in pure tones and complex signal analysis—for language processing. Music amounts to a super-stimulus of this peculiarly human language processing system, providing “purified and therefore intense doses of what usually activates it” (Boyer 132). It is critical to note here that, under this view, we did not evolve a propensity for music; rather, music is a “cultural product that is particularly successful because it activates some of our capacities in a particularly intense way” (Boyer 132). This means that our enjoyment of music is a kind of side effect of capacities evolved for other purposes—capacities that are susceptible to artificial or inadvertent stimulation.

Other examples include certain features of human art and body decorations, which, for example, are almost universally characterized by bilateral symmetry. As it turns out, bilateral symmetry signals good genes in a mate, making it one of the universal markers of physical attraction. It also characterizes the effect of seeing a face that is looking directly at you, which is highly relevant to our predator/prey and social interaction modules (Boyer 133). Other relevant analogies should spring to mind at this point, such as naturally occurring opioids, which just happen to fit receptors designed for other internally produced molecules; and artificial enhancements to sexual characteristics, such as artificial lip coloring, which provide much heavier signal doses than would be found in nature.

A particularly relevant study helps make all this tangible. Herring gull chicks instinctively peck at a red spot on their parents’ beaks. Experiments show that if a stick with not one, but three red stripes, is presented to the gull chicks, they will begin pecking
in manner far more frenzied than seen under natural conditions. This artificial situation stimulates their mental wiring much more intensively than would a natural beak with a typical red spot. As Ramachandran puts it, "If herring gulls had an art gallery, they would hang a long stick with three red stripes on the wall; they would worship it, pay millions of dollars for it, call it a Picasso, but not understand why they are mesmerized by it" (Ramachandran 780).

Regardless of whether such artificial super-stimuli are deliberately concocted or stumbled upon in nature (such as with opioids) the key point is that the underlying architecture, which is “designed” for one purpose, is by virtue of that design vulnerable to being commandeered, or even parasitized. A simple example of such commandeering of a process well-adapted to another purpose is the way an electric light can lead a moth to spiral to its death (Dawkins, *The God Delusion* 201). For much of the moth’s evolutionary history the sun and moon would have been the most frequently encountered bright light sources. Since these are at optical infinity, a simple hard-wired rule like “keep the object at a certain angle to the left,” could have had selective advantage for navigation. This is true even if this rule results in a death spiral for locally encountered intense natural light sources (e.g., lava, flames) but only if these are sufficiently rare on average so as not to offset the net advantage.⁶

What does any of this have to do with the emergence of religious beliefs? Humans—thanks to our capacity for decoupled representations and cross-module

---

⁶ It is useful to keep in mind here that *individuals* do not evolve; rather, populations evolve over time as gene frequencies change.
thinking—can concoct or stumble upon supernatural concepts that super-stimulate our inference systems. Once such a concept captures our rapt attention, it is more likely to be mentally rehearsed, retained, and transmitted.\(^7\) In particular, certain types of supernatural concepts specifically activate those inference systems with the most intense emotions, especially those related to social interaction and morality (Boyer 135). More than simply being oddly fascinating, they matter.

Crucially, this evolutionary psychological perspective explains why the supernatural concepts we encounter in so many cultures turn out not to be the products of reflection about what would best explain the existence of the gods or the universe in some abstract sense. Instead, what we find are supernatural concepts that are immediately practical and are comprised of entities with whom one can deal. Moreover, these supernatural “agents” are invariably represented as having human-like minds, complete with goals, intentions, and memories. In other words, they can be partners in social interaction and exchange.

Indeed, for many of the world’s belief systems theology does not exist at all. Examples can be found from the ancient Greeks to modern hunter-gatherers like the Kwaio. Mario Vegetti, for example, tells us that in “the whole body of [Ancient Greek] beliefs and narratives that deal with the gods, those referring to the creation of the world

\(^7\) Just imagine, for example, many generations of a human population sitting around a fire, sharing stories and memories. This seems an almost perfect environment for the emergence of concepts that are particularly effective at super-stimulating our cognitive modules. As the storytellers, perhaps unconsciously, note the reactions of the listeners they would adjust accordingly at their next re-telling. At the same time the listeners themselves would likely retell the most engrossing aspects to others, doubtless with their own embellishments. Over many generations (of the story, not necessarily the humans), this could be expected to create stunningly engrossing (if not entirely coherent) oral traditions.
and of human beings play anything but a central role.” In fact, he notes that with few exceptions, “they do not even exist” (Vegetti 255). Similarly, the modern Kwaio seem remarkably uninterested in questions about the nature of the spirits, such as where they come from, and how they do what they do—this despite the elaborate rituals they have for dealing with these spirits. As Roger Keesing reports, the “few who bother to think about such matters only do so as a result of being prompted by an anthropologist” (qtd. in Boyer 140). And when they do respond, their answers are wildly inconsistent, which one would expect if such answers are the products of post hoc rationalizations.

The fact that supernatural agents are invariably intentional agents goes to something else peculiarly human, which was memorably captured by David Hume: “There is a universal tendency amongst mankind to . . . find human faces in the moon, armies in the clouds; and by a natural propensity, if not corrected by experience and reflection, ascribe malice and good will to everything that hurts or pleases us” (Hume). Evolution has equipped not just us but many species with cognitive systems that discriminate between animate and inanimate motion. However, humans, along with some other species (to varying degrees), do something much more elaborate: they adopt what Daniel Dennett calls the intentional stance (Dennett 109–10). This is an inference system(s) that eagerly detects intentional agency. “Agents” here refers to entities that hold a limited set of beliefs, have specific desires, and are assumed to act in more or less predictable ways based on those beliefs and desires (Dennett 110). This intentionality

8 A term he uses in preference to Theory of Mind (ToM).
makes them socially relevant, particularly in “moral-laden” (Clark and Barrett, “Reformed” 181) areas of human social interaction—areas that engage such inborn emotions as obligation, trust, and loyalty. These, in turn, underpin human exchange and cooperation. (Note how far we have come from a passive, behaviorist model.)

So eagerly does this detection system operate in humans that Justin Barrett described it as a HADD, or Hypersensitive Agent Detection Device. For humans—a species whose reproductive success had depended not only on detecting predators and prey, but also on managing the complex machinations of social hierarchy—false negatives would have been far more damaging than false positives. In other words, a HADD (rather than one tuned to a middling setting) would have served our ancestors well: Believing that some agent’s intentions are behind almost any situation would allow one to trigger one or more mental modules that rapidly deliver a rich suite of complex responses, such as deception, appeasement, creating obligation by giving something of value, etc. If one wrongly “detects” an intentional agent where none exists, then some energy is expended, perhaps some embarrassment is incurred, but no serious harm is done. However, if one errs the other way, and assumes there is no such intention when in fact there is, then the price can be much higher: death by predation (from inside or outside the species), or being outmaneuvered in the social hierarchy (through deception, cunning, conspiracies, etc.). This asymmetry in consequences is what led to the evolution
of the *hypersensitivity* in HADD, leading us to overdo things, to project intention where none exists, even on such random events as the weather.\(^9\)

Not only do humans impute intention on a hair-trigger, we seem to be unique relative to even very recent hominid ancestors in our penchant for producing cross-category conceptual combinations. While humans have produced some very fanciful imaginings—totemism (animal-human hybrids), conscious mountains, trees with memory, etc. (Boyer)—our nearest relatives, despite evidence for some decoupled cognition, show no signs of such cross-category fantasizing or mythical/religious thinking (Mithen). Yet, while we humans have spun quite a menagerie of mythical creatures and supernatural concepts the world over, we do so only in a carefully circumscribed way. Not every fantasized concept has “staying power”—that is, captures our attention, gets mentally rehearsed, and is ultimately transmitted. Boyer argues that our inference systems are tied to certain, at least partially innate ontological categories. These are *abstract* categories, like Animal, Tool, and Person, each of which entails a host of implicit (subconscious) inferences that are automatically generated and applied to any newly hypothesized specific instance of that category (Boyer 60–61). Importantly, the most attention-grabbing concepts are those that *violate* some aspects of an ontological categories, but not so many that the category no longer applies. That is, they are

\(^9\) HADD, in combination with our other inference engines, might give some insight into the often rococo nature of religious ritual found throughout the world. Dennett reminds us of a famous Skinnerian experiment in which pigeons were put on a random reinforcement schedule: a click followed by a pellet release would occur on a completely random basis. The result: pigeons developed highly complex “dances,” which built up over time as their eager cause-and-effect-seeking inference engines were mistakenly linking their at-the-moment behaviors to a *random* event. When what seemed to work before failed to work, new behaviors were added until the new sequence “worked.” (Dennett 118).
minimally counterintuitive (MCI). By preserving the category, all of its rapidly produced implicit expectations remain; yet by violating it in particular, limited ways, our interest is piqued. This category-preserving aspect is what helps separate the fascinating from the uninterestingly bizarre: Successful mythical and supernatural concepts are those that (1) can be identified with pre-existing ontological categories, which allow them to inherit that category’s rich, automatically generated inferences; but (2) add some special category-violating features, typically with important social implications. This combination evidently super-stimulates our inference engines. For humans, it is a bit like the three red stripes on a stick that so enthralled the herring gull chicks. As an example, Boyer cites the belief, among the Uduk-speakers of Sudan, that some ebony trees can recall all conversations held under the shade of their leaves. Critically for this concept’s staying power is the highly socially relevant notion that such “hot mic” discussions can be recovered through divination (Boyer 69). If this sounds like some weird, random combination of categories, Boyer points out that the concept is “not the product of an unbridled imagination; [it supports] precise inferences within narrow constraints” (Boyer 69). In particular, the Uduk listening tree is an example of something in the Plant ontological category with a very special (and socially relevant) characteristic (Boyer 62). Other examples include ghosts, Person category plus no material body; zombies, Person category plus no cognitive function; and omniscient God, Person category plus special, socially relevant cognitive powers like “knows all” (Boyer 63–64).

A final key CSR concept is something Justin Barrett describes as Theological Correctness or TC (Barrett, “Cognitive Science”). In an experiment he conducted,
Christians were asked to consider some hypothetical situations in which a disaster was about to occur, such as a ship approaching an iceberg. They were then asked how, in their prayers, they would ask God to intervene in this course of events. Keep in mind that these Christians fully understood the theological doctrine that God is omnipotent, which means that He can do anything, unconstrained even by natural law. They could, for example, pray that God keeps the ship afloat with a torn hull, or gives passengers the ability to survive freezing water temperatures until they are rescued. The results of this experiment are telling: despite the subjects’ theological conception of omnipotence, they usually said that they would pray that God change the captain’s mind about the course he is taking (Boyer 140–41). Boyer observes that if God is an instance of the Person category, then our spontaneous intuition (i.e., unconscious, rapid inference from this specialized module) is that a person can change someone’s mind but cannot intervene in the laws of physics and biology. Boyer notes that this “expectation would be irrelevant if God’s great powers were the most salient aspect of the God-concept. The expectation [of limited powers] is activated only because people represent God as a Person category agent who interacts with them” [emphasis in original] (Boyer 142). Note how this relates to our earlier observation that theology is not even a part of many of the world’s religious beliefs and practices, as such concerns are not part of the inferences generated by the Person category.

10 We will let pass here the important question of whether it is even coherent to posit the existence of an omnipotent being that, despite its limitless power, is nonetheless required to take action in order to implement its will. We will touch briefly again on this point in the last chapter.
The CSR Standard Model

With this overview of how CSR concepts are typically deployed, we are now in a better position to summarize the CSR Standard Model as it applies to the expected characteristics of “successful” supernatural agent concepts:11

- They are minimally counter-intuitive (MCI), which is to say they are “optimized for recall and transmission” (Murray and Goldberg 184). As we saw from Boyer, while supernatural beliefs are of counter-intuitive agents, they are not so different from an existing (innate) ontological category as to lose that category’s rich, automatically generated inferences.

- They are the outputs of cognitive systems evolved for detecting agency, which for evolutionary reasons, are biased toward hypersensitivity (HADD), that is, biased toward false positives over false negatives.

- In addition to their being intentional agents with their own desires and beliefs, successful supernatural agents are socially relevant: We can enter into exchange with them; and they are able to help us in our dealings with other members of our group.

- If there is an espoused theology at all—a sense of theological correctness—it appears to be secondary to, and sometimes in conflict with, the intuitive beliefs actually held. This can be revealed in practical situations (such as the iceberg experiment) where our category-based, anthropomorphic, automatic inferences assert themselves over our theoretical notions.

Under this model we would expect our cognitive systems to produce not specific supernatural beliefs or theologies, but rather a general susceptibility to an enormously wide range of supernatural beliefs bounded by the above conditions.

On a conceptual level, CSR’s empirical approach has the important consequence of allowing research to proceed without a fully fleshed-out definition of “religion.” This

11 This is adapted from (Leech and Visala) and (Murray and Goldberg).
is because it investigates ostensible religious phenomena on a case-by-case basis (Trigg and Barrett 6). As we will see later, this “bottom-up” approach resonates with one of this paper’s key themes. However, to place some bounds on our discussion, we will understand a “religious” belief to be “belief in the existence of supernatural agents.” This is a slight modification of Graham Wood’s CSR-oriented definition that refers not to agents, but entities (Wood 735). This modification is also consistent with Clark and Barrett’s definition of “gods” as “any supernatural intentional agents whose existence would impinge upon human activity” (Clark and Barrett, “Reidian” 652).

With our brief introduction to CSR complete, including some reasons why it is taken so seriously in mainstream science, we can now begin to examine its significance to the philosophy of religion—in particular, its epistemological implications for theism (again, assuming CSR is true).

Epistemic Consequences

It is important to bear in mind that even if CSR were to accurately and fully explain the universality and contours of religious thought and practice, it is by no means clear that this would undermine theism or the rational justification for religious belief. Whether or not explaining religious phenomena in CSR terms (or any other natural terms) “explains it away” is a philosophical question whose answer must be defended and not assumed—and its answer is by no means obvious. For example, Justin Barrett is not

12 Barrett and Trigg claim that this limits the impact of epistemic problems with religious belief since such problems may be limited to that particular religious belief rather than to all of them (Trigg and Barrett). We will later identify reasons to believe that the negative implications of such epistemic problems are indeed more general than they suppose.
only a Christian, but a founding scientific contributor to this new field and someone who has powerfully argued for a theism-compatible (indeed, theism-

supporting) interpretation of CSR’s findings, as we will see below. The philosophical question of CSR’s epistemic significance for theism is, therefore, timely and important.
CHAPTER 3

OBJECTIONS TO AN ATHEISTIC INTERPRETATION OF CSR

Introduction

With our brief overview of CSR in hand we are now in a better position to see how these concepts were originally deployed to argue/imply that theism is not rationally justified. Surprisingly, fully developed arguments by atheists based on CSR have been rare; instead, the field’s purported atheistic implications have often been implied or assumed. Justin Barrett, for example, observes that writers “in this area typically do not directly attack religious belief but leave plenty of reasons for believers to be suspicious” (Barrett, “Is the Spell Really Broken?” 60). Similarly, Michael Murray and Andrew Goldberg note that “many evolutionary scientists present their views as if indeed they do explain religion away by showing it either to be false outright, or at least superfluous” (Murray and Goldberg 194). It is as if such atheists believe that any sufficiently successful natural explanation for religion is necessarily incompatible with rational belief in God: once the natural case has been made, further argument is unnecessary. The result has been that philosophically disposed theistic writers, in order to develop their rebuttals, have often had to first formally develop the atheists’ implicit arguments.14

13 Boyer, for example, memorably likened supernatural belief to poetry, in that they both give to “airy nothing a local habitation and a name” (Boyer 4). See also (Atran). 14 See for example (Murray).
As we will see below, the philosophical challenges to the idea that CSR supports atheism are serious challenges indeed, and deserve careful consideration. Arguably, Alvin Plantinga’s contributions to this debate are why these arguments are as strong as they are. His pioneering epistemological criticisms of classical foundationalism and naturalism have deeply influenced the philosophical discussion surrounding CSR.

Alvin Plantinga's Foundational Work

Critique of Classical Foundationalism

Hume’s famous dictum, “A wise man . . . proportions his belief to the evidence” (Hume), has been cited and paraphrased time and again in critical thinking discussions. Many atheists are quick to point to this principle when explaining their insistence that a lack of sufficient evidence is precisely why belief in supernatural beings is unjustified. In a similar vein, but from the perspective of ethics, W.K. Clifford intoned, “it is wrong always, everywhere, and for anyone, to believe anything on insufficient evidence” (Clifford 152).

To be sure, many theists, at least implicitly, accept this evidentialist premise, where we can understand “evidentialism” to be “the thesis that, if one justifiably believes P then one must have evidence that adequately supports P” (Smith 1). As Plantinga points out, theists arguing in the natural theology tradition are, in fact, making the claim that belief in God is justified precisely because evidentialist standards are met (Plantinga, “Is Belief in God Properly Basic?” 41). As we will see, Plantinga does not reject the need for rational justification, only the idea that evidence alone is necessary to achieve it, and it is here that Plantinga parts company with natural theologians, such as Richard Swinburne.
Does evidentialism really insist that no belief is rational unless it is justified by reference to other justified beliefs? If so, it would be open to a charge of *reductio ad infinitum*. However, evidentialism comes in various forms, including coherentism-based approaches, which are not subject to this particular charge (Berker). For many others, the terminus of the justificatory chain involves some form of *foundationalism*, which holds that certain beliefs are *basic*—that is, can be rationally held without reference to other evidentially justified beliefs.

But what makes a proposition *properly* basic? That is, what criteria should be used to admit certain beliefs into this privileged group while prohibiting others? In characterizing classical foundationalism (CF), Plantinga superimposes the early and modern foundationalists’ views into the following principle: Proposition P is properly basic for person S if and only if, for person S, P is self-evident, incorrigible, or evident to the senses (Plantinga, “Is Belief in God Properly Basic?” 44). More formally,

\[(1) \; P_{pb} \leftrightarrow P_s \lor P_i \lor P_e \]

where by “self-evident” he means that as soon as one grasps a proposition, one *immediately*—as in unmediated by other propositions—sees that it is true. In clarifying his use of “evident to the senses,” he distinguishes what he understands to be Aquinas’ less tentative notion from Descartes’ more qualified one. It is this distinction that separates the early from the modern foundationalists. Whereas Aquinas might recognize “I see a tree” as being basic on the grounds of its being directly apprehended—that is, evident to the senses—Descartes would insist that only the subjective *seeming* to see a
tree would qualify as truly basic, and it is in this latter sense that Plantinga means a proposition is *incorrigible* (Plantinga and Wolterstorff 58).\(^\text{15}\)

With this conceptual background in hand, we can return to our biconditional (1). Here, Plantinga points to what he sees as a crucial hidden assumption, which we can uncover by decomposing (1) into the following conditionals:

\[
\begin{align*}
(2) & \quad P_s \lor P_i \lor P_e \rightarrow P_{pb} \\
(3) & \quad P_{pb} \rightarrow P_s \lor P_i \lor P_e
\end{align*}
\]

Statement (2) asserts something, which, for argument’s sake, Plantinga is willing to concede: For person S, *if* proposition P is self-evident or incorrigible or evident to the senses, *then* it is properly basic. However, (3) says something quite different: For S, proposition P is properly basic *only if* proposition P is self-evident or incorrigible or evident to the senses. It is with this that Plantinga takes issue.\(^\text{16}\)

Plantinga asks how, under this model, our everyday beliefs about the past—including our own past—are rational? For example, my belief that I was sitting at this desk an hour ago, is (according to Plantinga) neither self-evident, nor incorrigible for me,\(^\text{17}\) nor evident to the senses. Neither is this belief derivable from other basic beliefs. Nonetheless, it *is* basic to the person with the memory—that is, it is believed

\[\text{---} \]

\(^{15}\) For an excellent broader treatment of philosophical incorrigibility, see (Raff).\(^{16}\) Note that in addition to (1) CF includes the corollary thesis that rational beliefs must either be logically derivable from properly basic beliefs as described in (1), or at least be probable with respect to them.\(^{17}\) In the sense that “p is incorrigible for S if and only if (a) it is not possible that S believe p and p be false, and (b) it is not possible that S believe \(\neg p\) and p be true” (Plantinga and Wolterstorff 58).
spontaneously and without reference to anything else. This hardly seems irrational even if it is, under certain circumstances, unreliable (Plantinga and Wolterstorff 80).

Plantinga’s most powerful charge, however, is that classical foundationalism’s central principle is self-refuting. In particular, he points out that it is not basic by its own criteria, nor is it derivable from other properly basic propositions. On what basis, then, is it believed? In other words, to accept (1) is to violate it (Plantinga and Wolterstorff 60). Plantinga’s critique focuses on conditional (3): For a person S, only those beliefs that are either incorrigible, self-evident, or evident to the senses, can be accepted as properly basic. What he calls the Evidentialist Objection to belief in God is the charge that belief in God fails to meet this conditional. Plantinga believes that he has defeated this objection by defeating the rationality of CF on its own terms. By arguing that this evidentialist criterion of rational justification is itself irrational, he opens the door to what he hopes is a much better alternative approach—one that puts theism on a defensible rational footing. Specifically, he argues that belief in God is in fact properly basic.

**Basis for Theism as Justified Properly Basic Belief**

In clarifying and developing his notion of properly basic belief, Plantinga points out that while such beliefs may not meet the CF conditional that he rejects, they are not (necessarily) groundless or unjustified. Rather, their justification is rooted in an experience “characteristic of a certain sort,” typically in combination with other circumstances such as the absence of disease or other factors known to affect perceptual reliability for that particular type of experience (Plantinga, “Is Belief in God Properly Basic?” 44–45). These circumstances can be thought of as requisite, justificatory
preconditions. If those preconditions are met, then the basic belief in question becomes a justified, non-inferential belief—that is, it is properly basic, and, consequently, rational. To be sure, as he recognizes, such properly basic beliefs are defeasible, but they are *prima facie* justified.

Plantinga claims that there are just such justifying preconditions for rational, prima facie belief in God. Here, he draws on Aquinas and especially Calvin. Aquinas, Plantinga reminds us, claimed that knowing “in a general and confused way that God exists is implanted in us by nature” (qtd. in Plantinga, *Warranted Christian Belief* 170). Calvin expanded this into the notion that God created us with a predisposition—a *sensus divinitatus*—to believe that various actions and objects are the result of God’s handiwork (Plantinga, *Warranted Christian Belief* 172). For example, when we marvel at the intricacy and complexity of the universe, the biosphere, or even the beauty of a flower, the proposition that we are seeing the result of God’s handiwork comes to us easily, or is at least easily believed once the proposition is encountered. In other words, such experiences constitute justifying preconditions for the proper functioning of the *sensus divinitatus*. He notes too that we seem similarly predisposed to relational beliefs involving God, which produce feelings such as gratitude and guilt (Plantinga, “Is Belief in God Properly Basic?” 46).

The idea that belief in God is properly basic is a core claim of *reformed epistemology*, a term that acknowledges its connection to Calvin’s reformed theology (Smith 3). Plantinga calls his particular Calvin-based model the “Extended Aquinas / Calvin (A/C) Model” (Plantinga, *Warranted Christian Belief*). Reliabilism is the view
that belief can be justified on the basis of a cognitive process’s \textit{reliability}, even if such reliability depends on facts completely \textit{inaccessible} to the believer’s conscious awareness. Reliabilism is an instance of externalism. In contrast, internalism generally refers to the view that a believer does or at least can have access to the justificatory basis of her beliefs. This distinction bears on Plantinga’s particular brand of reliabilism, known as \textit{proper functionalism}, since it is, according to Plantinga, mainly a theory of warrant rather than justification. One way to understand his distinction between justification and warrant is by thinking of justification as that subtype of warrant that derives from reasons the believer can access. In other words, justification is internalist warrant (Altschul).

Plantinga’s attempt to undermine classical foundationalism and establish a rational basis for justified belief in God has been criticized as creating a Pandora’s Box of “justified” beliefs, a criticism that Plantinga himself dubs The Great Pumpkin Objection (Plantinga and Wolterstorff 74). Why, in other words, can one not declare belief in the Great Pumpkin to be a properly basic belief? Plantinga’s rebuttal to this is illuminating. He maintains that establishing \textit{hard} criteria for proper basicality is impossible since it can be approached only bottom-up, rather than top-down. Such a bottom-up approach, however, would necessarily be an inductive one, for which he suggests the following approach. First, assemble a relevant set of basic-belief examples, that is, clear, not-so-clear, and counterintuitive examples of belief-condition pairs for beliefs considered basic under those conditions. And second, develop and test hypotheses regarding necessary and sufficient conditions for establishing the basicality of those beliefs under their associated conditions (Plantinga and Wolterstorff 76). Note that for such a scientific-like approach
we would have to be prepared to revise our belief-condition example sets in light of those test results. He concludes, therefore, that the most we can hope to achieve is the development of criteria for prima facie—and therefore defeasible—justified basic belief (Plantinga and Wolterstorff 77).

What are the implications of such an inductive, scientific-like approach? Given that we are choosing a relevant set of examples prior to such argumentation and testing, we should not expect universal agreement on that starting example set. Any criteria I build up based on my starting set of examples may not work on your starting set of examples, and vice versa. But then what would the criteria be for choosing the starter set, since it precedes the inductive process? Why, Plantinga asks, should a Christian accept an atheist’s sample set or vice versa (Plantinga and Wolterstorff 77)? He fully recognizes that this means believers and non-believers would both be equally within their epistemic rights to reject each other’s claims as to whether or not certain beliefs are properly basic, including such claims as “belief in God is properly basic” (Plantinga and Wolterstorff 78). But this is also why the Christian is epistemically within her rights to reject The Great Pumpkin as properly basic. Note that Plantinga is not arguing for the superiority of his approach over the atheist’s, only for logical parity. But this is all he needs to do—at least at this first step. The next step is to place the burden of proof on the atheist.

Warrant and Burden of Proof

Plantinga’s notion of warrant is strikingly simple. He takes a reliabilist view regarding the warrant of some belief B, which is generated by some belief-forming process P: If P is designed to be aimed at truth, is functioning normally, and is producing
B in an appropriate setting for P (not one that, for example, is rigged to create illusions), then the belief is warranted. Following Calvin, Plantinga suggests that in the case of Christian beliefs, this P is the *sensus divinitatus*, a basic-belief producing process designed by God to be truth-tracking with regard to belief in God. If Plantinga, and like-minded theists are right about all of this—that is, right about their theological beliefs and their beliefs in the *sensus divinitatus* and its relationship to warrant—then B would be warranted independent of any evidential considerations.

Following Gregory Dawes and Jonathan Jong, we can capture this more formally:

\[ a \land w \rightarrow w \]

where “a” represents the much longer conjunction of Christian beliefs exclusive of their beliefs about warrant, and “w” represents their set of beliefs about warrant. While this seems to be an utterly trivial truth—"if a is true and w is true, then, well, w is true”—Plantinga takes it further by combining it with the assertion that Christian beliefs are *basic* with respect to *warrant* (Dawes and Jong 2). As discussed earlier, if a belief is basic, it is not inferred from other beliefs, but is “immediate” in the sense of being the unreflective, spontaneous effect of certain experiences in the context of relevant preconditions. If belief in God is basic in this sense, then, according to Plantinga, it enjoys some degree of immunity to even very persuasive evidence. Why? Because as a properly basic belief, it is neither based upon evidence, nor warranted *because* it is based upon evidence. Rather, its warrant derives from its being the output of a properly functioning belief-producing mechanism, which is designed to track truth and which is operating under normal conditions. As such, it is not subject to disconfirmatory evidence
in the same way that non-basic beliefs are, such as beliefs in scientific theories. To illustrate this, Plantinga uses the example of an innocent person convicted by genuinely overwhelming evidence—publicly available and intersubjectively verifiable—that he slashed someone’s tires (Plantinga, Where the Conflict Really Lies 359). Despite this evidence, however, the accused clearly remembers spending the afternoon elsewhere busily not slashing tires. That memory-based belief is basic (i.e., spontaneous and unreflective) for this person irrespective of the evidence, and if it is properly basic, then he is prima facie rationally justified in believing in his own innocence despite even strong evidence to the contrary.

While the overwhelming evidence would understandably be persuasive to the jury, it would also be understandably unpersuasive to the accused. Unlike the jurors, who have access only to external evidence, the accused’s basic-belief-forming faculty—in this case, his memory—has produced a basic belief, which single-handedly overwhelms the effects of all other evidential considerations combined. This analogy helps illuminate how certain pieces of evidence against Christian theism can be, from an atheist’s perspective, maddeningly unpersuasive to thoughtful believers.\(^\text{18}\)

In developing his argument, Plantinga clarifies some aspects of his notion of rationality. In addition to its dependence on proper function, which is to say that one’s belief cannot be rational if there is dysfunction in any of the relevant faculties. there is

\(^{18}\) Dawes and Jong point to the example of the Argument from Evil (Dawes and Jong 3). Their point is well-taken. Indeed, when we also consider various Old Testament passages, the Argument from Evil would seem to be an utterly decisive case against one of the Christian God’s core attributes: His perfect goodness. Yet it regularly fails to impress believers who just “know” God is “good.”
also a sense in which these faculties operate in a kind of process chain, from “sensuous imagery” and other experiences, like memories, to belief formation. He divides this chain into internal and external rationality, where external rationality describes the experience-creating faculties, and internal rationality describes the belief-forming faculties that respond to these experiences. Internal rationality, when functioning properly, produces beliefs that cohere logically with that person’s other beliefs. For example, someone who truly believes that she is made of glass may well be internally rational (Plantinga, *Warranted Christian Belief* 112). While her external rationality may be failing her in this case, her internal belief-forming faculty may nonetheless be responding normally, producing a belief that coheres with her other beliefs. This internal/external distinction helps us understand how, under Plantinga’s model, we can uncouple belief formation from the present moment, thereby allowing us to have rationally justified doxastic experiences associated with memories and a priori beliefs (Plantinga, *Warranted Christian Belief* 110–11).

With these ideas in place, Plantinga reaches a key conclusion regarding burden of proof. Since theistic belief—including belief in a God-created, religious belief-forming mechanism—is warranted provided that it is true, and is also basic with respect to warrant, then the burden is on the atheist to show that it is untrue. In meeting this burden, the critic can object to Christian belief in one of the following ways: by asserting that its content is false, or by asserting that it is unjustified regardless of its content’s truth,
perhaps by arguing against the proper functioning of the belief-forming mechanism. This is his *de facto* versus *de jure* distinction.¹⁹

So far in our discussion, Plantinga first attempts to create a space for rational, justified theistic belief by critiquing classical foundationalism, and then attempts to shift the burden of proof for theism onto the atheist by way of his conditional definition of warrant and his proper functionalist account of rationality. This is all part of Plantinga’s reformed epistemology. As we will see, he is not content to merely retreat behind these formidable new theistic fortifications. He goes on the offense by arguing that without theism, an atheistic (i.e., exclusively naturalistic) interpretation of one of science’s most successful theories, biological evolution, necessarily leads to an extreme form of global skepticism.

**Evolutionary Argument against Naturalism (EAAN)**

Plantinga’s argument against naturalism by way of evolutionary theory is as provocative and stimulating as it is famous. This is his second major contribution to the debate explored in the present work. To fully appreciate his argument, it is important to understand its scope. At the outset, Plantinga makes clear that his EAAN is neither about proving naturalism wrong, nor about proving that unguided evolution could not produce beings that really do have reliable belief-producing mechanisms. Rather, it is that while naturalism and evolution could be true, they cannot both be “sensibly believed.” To use

¹⁹ We will, along with Plantinga, though strictly in the interest of space, focus on the *de jure* rather than the *de facto* sense of rational-justification.
his example, all of one's beliefs could in fact be false, but one could not sensibly believe this to be the case (Plantinga, Where the Conflict Really Lies 310).

In building his EAAN, Plantinga challenges the naturalists’ presumption that our belief-producing faculties are reliable, given that they are the product solely of naturalistic evolution; in other words, evolution is foresightless and completely unguided (Plantinga, Where the Conflict Really Lies 311). Consistent with Plantinga’s usage of the term, naturalism is understood to entail materialism, which is to say that it entails, among other things, the belief that humans are nothing more than the sum of their physical constituents (Plantinga, Where the Conflict Really Lies 318). More generally, naturalism assumes that there are no non-natural (e.g., supernatural) entities, such as gods or souls. For Plantinga, this conjunction of metaphysical naturalism with the view that we and our cognitive faculties have arisen by way of the mechanisms and processes proposed by contemporary evolutionary theory—gives us reason to doubt two things: (a) that a purpose of our cognitive systems is that of serving us with true beliefs, and (b) that they do, in fact, furnish us with mostly true beliefs (Plantinga, Where the Conflict Really Lies 316).

---

Or, in Gilbert Ryle's famous phrase, that there is no "ghost in the machine."

In fact, there is an important distinction between metaphysical and methodological naturalism, in that the latter does not make assertions regarding the existence or non-existence of supernatural entities. Rather, as a methodological principle, it operates as if they do not exist in order to have the best chance at uncovering new knowledge. For example, if supernatural agency were always invoked first, we would likely never have moved from the “curse” theory of disease to the current germ theory of disease. However, despite his use of the term “metaphysical naturalism,” Plantinga’s argument does not seem to hinge on this distinction as it assumes only that evolution in no way purposeful. This can be true regardless of one’s metaphysical assumptions about supernatural entities.
He develops this argument by identifying and defending four key premises, which, if all are true, lead to his conclusion.

Premise 1: "Darwin's Doubt," which Plantinga expresses in conditional probability terms as

\[ P(R/N&E) \text{ is low} \]

where R is the reliability of our belief-forming faculties (BFFs), N is naturalism, and E is evolution. So, the probability P of R, given both N and E, is low (Plantinga, *Where the Conflict Really Lies* 317).

Why should we believe Premise 1? Plantinga recognizes that evolution can indeed be expected to produce adaptive behavior geared toward enhancing survival; however, behavior and belief are two different things. To be sure, survival requires some accurate tracking of the actual physical environment in order to provide survival-enhancing responses, whether those responses are finding food or avoiding becoming food. But, he claims, beliefs need not enter into this at all. (Indeed, many creatures, like the jellyfish, almost certainly have nothing like “beliefs,” at least in any ordinary sense of the word.)

More to the point, if beliefs do exist in an appropriately advanced creature, then they do not need to reflect true propositions even as they help produce effective survival-enhancing behavioral responses to environmental cues (Plantinga, *Where the Conflict Really Lies* 326–28). This disconnect between behavior and belief content means, according to Plantinga, that belief content is not subject to (i.e., is not “visible” to) evolutionary pressures one way or the other; consequently, beliefs are as likely to be objectively true as false (Plantinga, *Where the Conflict Really Lies* 330–31). Plantinga
seems to be arguing that the naturalist has no reason to suppose that the contents of belief are anything more than free-floating epiphenomena: “Content simply arises upon the appearance of neural structures of sufficient complexity; there is no reason why that content need be related to what the structures indicate [about the environment], if anything” (Plantinga, Where the Conflict Really Lies 331).

This disconnect between belief content and behavior leads directly to the issue of cognitive reliability. Regardless of how we formally define a "reliable" belief-producing mechanism, it should at least produce significantly more true beliefs than false ones. For the purposes of illustration, Plantinga conservatively offers a reliability criterion of three-times as many true beliefs as false, which, given his particular use of the Principle of Indifference, makes the probability that one thousand independent beliefs would reach this threshold an astronomically small 1 in $10^{58}$ (Plantinga, Where the Conflict Really Lies 332–34).

Premise 2: "Anyone who accepts (believes) N&E and sees that P(R/N&E) is low has a defeater for R" (Plantinga, Where the Conflict Really Lies 341). Here, "defeater" is understood to be a rationality defeater, that is, a belief that cannot be rationally held at the same time as the belief for which it is a defeater (Plantinga, Where the Conflict Really Lies 359). In other words, if you come to believe a defeater, say $D_B$, for belief $B$, then rationality constrains you to either believe its denial ($\neg B$) or to be agnostic toward it (Plantinga, Where the Conflict Really Lies 341).

Premise 3: "Anyone who has a defeater for R has a defeater for any other belief she thinks she has, including N&E itself" (Plantinga, Where the Conflict Really Lies 345).
This requires little elaboration, since N&E are beliefs like any other, produced by the same putative belief-producing mechanism.

Premise 4 and conclusion: "If one who accepts N&E thereby acquires a defeater for N&E, N&E is self-defeating . . .” Therefore, “N&E can’t be rationally accepted” (Plantinga, Where the Conflict Really Lies 345).

The conclusion means that the atheistic, naturalistic perspective, coupled with belief in biological evolution, leads to what Plantinga calls a "crushing skepticism" (Plantinga, Where the Conflict Really Lies 345). Indeed, if all of our beliefs, including those that underpin basic logical principles, are products of unreliable cognitive faculties, then argument of any kind cannot even get off the ground. After all, the very notion of “argument” presupposes the reliability of our cognitive faculties.

It is important to note that Plantinga fully recognizes that P(R/N&E) may not fully describe the situation. There may be other facts that one knows, any of which can amount to a "defeater deflector (D)," which, when incorporated, can snap the conditional probability of R to a very high level. Indeed, P(R/N&E&D) could jump all the way up to 1. Nonetheless, he correctly argues that in choosing such a deflector, certain principles surely apply. For example, one cannot reasonably be allowed to use the belief itself as the deflector. After all, if we allowed R = D or R ⊨ (entails) D, then defeat of any belief would become impossible, since P(R/N&E&R) would always be 1 (Plantinga, Where the Conflict Really Lies 349).
Summary

Plantinga has attempted to create a rational foundation for theistic belief. His approach involves, first, creating a space for theism by undermining the rational ground of classical foundationalism and with it the basis for the Evidentialist Objection to theist belief; second, shifting the burden of proof away from the theist through his truth-conditional notion of warrant and proper-functionalist account of rationality; and finally, creating a rationality-saving need for theism as an antidote to the radical skepticism that results from the conjunction of an atheistic world-view with the theory of evolution. It is difficult to overstate the influence of these ideas on the shape of the philosophical debate surrounding the justification of theism in the context of CSR. Justin Barrett, one of CSR’s founding contributors, and others, build directly upon this work.

EAAN-Based CSR Critique

Justin Barrett and Ian Church leverage Plantinga’s EAAN in an effort not only to undermine any atheism-supporting interpretation of CSR’s findings, but also to bolster the EAAN itself by providing it with empirical support. To appreciate their analogical argument, it is worth reviewing some relevant demographics. Belief in supernatural entities is not merely common throughout the world; it characterizes the overwhelming majority of the world’s population. While it is true that the number of atheists is
significant in absolute terms, in relative terms the numerical disparity between atheists and theists is very large indeed.\textsuperscript{22}

With this in mind, they build their argument on an analogy, which is intended to mirror the poor analogical reasoning of atheists. Given atheists’ small-minority status, they argue, atheists see their situation like that of a partygoer in the company of her many friends at a nightclub. Once there, this partygoer discovers that her assessment of the other guests’ attractiveness is radically different from those of her companions. However, much alcohol has been flowing, and this sober partygoer attributes the disparity to the alcohol’s distorting effects on the reliability of her friends’ cognitive functions. Barrett and Church believe this analogy captures many atheists’ thinking in the light of CSR’s findings.

As we have seen, the current consensus in CSR holds (1) that our predisposition to embrace supernatural beliefs is a side-effect of the normal operation of our cognitive faculties and (2) that these same beliefs are theologically incomplete (Barrett and Church 315). This is certainly consistent with the plethora of religious and supernatural concepts we observe in the world. As we have also seen, the normal operation of our evolved cognitive tool-suite evidently involves an agency detector, or HADD, the sensitivity of which reflects the relatively low cost of false positives as compared with the high (even

\textsuperscript{22} WIN-Gallop used the question, “Irrespective of whether you attend a place of worship or not, would you say you are a religious person, not a religious person or a convinced atheist”\textsuperscript{(International)? The responses indicate that 13% of the world were convinced atheists in 2012. However, Pew Research points out some of the challenges in identifying atheists in statistical studies (Lipka). Nonetheless, this percentage is low enough that even a substantive adjustment upward would still preserve atheists’ small-minority status.
fatal) cost of false negatives (e.g., it is generally better—cheaper, if you will—to mistake a branch for snake than a snake for a branch). One consequence of this cost asymmetry is the evolved human tendency to over-attribute *agency*, which, according to CSR, can lead to the attribution of supernatural agency when natural agency is not apparent. Are these supposed false-positive outputs of HADD analogous to our partygoer’s friends’ false-positive outputs of their alcohol-biased attractiveness detecting faculties?

Barrett and Church’s answer is instructive. While we have *independent* reasons for thinking alcohol negatively affects the reliability of our senses, what independent reasons do we have for supposing that our belief in the supernatural is similarly unreliable? They argue that without the atheist’s prior metaphysical commitments, there are in fact none whatsoever. After all, how can we assess the reliability of our god(s)-belief generating faculties unless we have independent knowledge of the presence or absence of such entities when these belief-forming faculties (BFFs) are in operation? (Barrett, “Is the Spell Really Broken?” 63) What would serve as the benchmark? For all we know it might be the atheist who is *under*-detecting God rather than the theist who is over-detecting Him (Barrett and Church 315). The bottom line, they claim, is that the original analogy fails because the partygoers can sober up, which provides the reference frame by which they can see that they had been misled by the effects of alcohol. However, with regard to belief in God, we have no similar reference frame—we never sober up, as it were (Barrett and Church 322).

Yet, as constructed, this analogy does fall short in at least one important way: Our people-related BFFs produce fairly detailed perceptions of the other guests in all but the
most extreme cases of intoxication. In contrast, our religious BFFs produce only the vaguest perceptions of the divine—so vague, in fact, that its products bear nothing in common across the world’s cultures. Why would this be the case, since *ex hypothesi* it exists specifically to *be* a God sense? Barrett and Church do appreciate that this is cause for theists to wonder why their presumably God-designed BFFs produce such indistinct and theologically nonspecific notions of the supernatural. Nonetheless, Barrett and Church claim that this does not affect their argument. Regardless of how vague such deliverances may be, for the theist they at least point in the *general* right direction—even if only “through a glass darkly.” On the other hand, the atheist is committing to the idea that these same deliverances are not simply vague but patently false, and false on such a global and historical scale that they our faculties have been producing systematically false beliefs in the overwhelming majority of humanity for most of its history. So, while the theist may have to admit that our BFFs are *imprecise* regarding religious belief (for whatever reason), the atheist must commit to the idea that this mechanism is so error prone that it produces systematically misleading results. And this, they argue, has catastrophic epistemic consequences for the atheist.

It is here that they adapt Plantinga’s general EAAN argument to the particular case of atheist-CSR. They begin by analyzing what they see as the atheists’ relevant set of commitments:

---

23 Plantinga’s answer is that it “has been compromised, weakened, reduced, smothered, overlaid, or impeded by sin and its consequences” (Plantinga, *Warranted Christian Belief* 184). This is a particularly stark example of an ad hoc hypothesis that cannot be tested independently of the theory it is intended to defend, a subject we will explore in some detail in a later section.
There are no gods, souls, and afterlife.

CSR-BFFs typically produce beliefs in gods, souls, and the afterlife.

Hence, CSR-BFFs typically produce false beliefs.

BFFs that typically produce false beliefs are unreliable.

Hence, CSR-BFFs are unreliable.

Beliefs formed by unreliable faculties lack warrant.

CSR-BFFs produce beliefs about human minds (including conscious beliefs, desires, emotions, and their relationship to actions), the causal properties of the natural world, and so on.

These beliefs (from (7 and 5)) lack warrant (Barrett and Church 316).

If this argument holds, then the atheist is forced into skepticism toward all her beliefs, not just religious beliefs, which is the same self-refuting result Plantinga had reached.

Barrett and Church do anticipate the rebuttal that BFFs may be unreliable only in the contexts in which religious beliefs are formed, but they quickly dismiss this, noting that unless one is able to specify the differentiating conditions between the two cases, such a move would amount to special pleading (Barrett and Church 316). On this point, they cite Michael Murray’s more developed argument in which he observes that HADD is generally reliable in some contexts but not others, which we know because HADD operates in conjunction with other cognitive tools (Murray 171). For example, we first hear something like a creak in the floor and imagine an agent at work, but then engage other cognitive tools to investigate and adjust that belief as necessary. This whole multi-
tool process is generally quite reliable with regard to agency. How, he asks, can the atheist critic argue that the context of religious belief formation is different from such ordinary cases without begging the question about the existence of religious entities?

For Barrett and Church, therefore, the theist’s interpretation of CSR’s empirical data avoids both the corrosive skepticism and the apparent question-begging of the atheist’s interpretation of that same data. Significantly, by drawing upon the empirical findings of CSR, Barrett and Church are not just applying Plantinga’s philosophical insights to guide data interpretation. They see it as a two-way street: “work done in the armchair is often all too susceptible to work done by other people in other armchairs. What the recent deliverances of CSR do, however, is give the theist concrete empirical data in support of Darwin’s doubt” (Barrett and Church 321).

The Bayesian-Based Irrelevance Critique

As we have seen, Plantinga considers the notion of a reliable BFF as one that produces significantly more true results than false. Jonathon Jong and Aku Visala examine this notion more formally; in particular, they examine how the concept of reliability is deployed in arguments that use evolutionary explanations generally (not just in the CSR case) to undermine theism, so-called Evolutionary Debunking Arguments or EDAs (Kahane).

They lay the groundwork for their case with some important clarifications. First, they note that attempts to use evolutionary explanations like CSR to debunk theism, can, at most, bear on the justification of theistic belief, not its veracity. The distinction is important, since otherwise one risks committing the so-called genetic fallacy: asserting
that the object of an unjustified belief—the proposition believed—is false solely because the belief is unjustified (Jong and Visala 245). Second, they analyze the notion of a BFF’s being “reliable” in terms of whether or not it is “off-track,” a concept built upon Robert Nozick’s concept of a truth-tracking process, where p is the proposition about which S’s belief-forming process has produced a belief. This process would be off-track if the following set did not hold:

1. p is true
2. S believes p
3. if p were not true then S would not believe that p, and
4. if p were true then S would believe that p (and would not believe that not-p) (qtd. in Jong and Visala 244)

In other words, if S’s belief-forming process produced belief in p regardless of whether or not p is true, then by the preceding it would be off-track.

We can recognize Boyer’s argument as an instance of an EDA: one that seeks to debunk theism by showing that theistic belief formation can be explained entirely in naturalistic terms. In terms of truth-tracking, Boyer’s form of argument implies that finding a wholly natural explanation for a BFF means that its reliability runs afoul of (3) above. But is this sufficient to debunk the object of belief? As we have seen, Barrett and Church claim that such a conclusion is possible only if one imports certain naturalistic metaphysical assumptions. Jong and Visala elaborate on this by pointing out that finding a wholly natural explanation would amount to finding an explanation for one sufficient cause, but not necessarily the only one. One or more other causes could be operative at
the same time, causes that meet the reliability conditions outlined above. Even more importantly, these other causes might exist not only at the same level as additional, sufficient proximate causes, but also at higher levels of explanation.

Jong has argued elsewhere that religious explanations, like scientific explanations, occur at different levels, up to and including the ultimate level (Jong 528). For example, in biology the proximate cause of sex is pleasure, though its ultimate cause is reproduction (Pinker 54). In theology, a first cause argument would be an example of an ultimate explanation, which can sit above and subsume all other explanations, including naturalistic ones. In such a case, there is no necessary conflict between religious belief and a purely scientific explanation of that belief, since the latter simply operates at a lower explanatory level. In the case of CSR, a theist might assert that God is the ultimate cause of religious belief, and this ultimate cause is effected through a long chain of progressively lower-level causes, which, in the present context, would include all the naturalistic mechanisms posited by CSR. Under such a view, scientific evidence simply does not bear on the question of God’s existence, nor necessarily on the rationality of believing in His existence.

Assuming for the moment that the burden of proof is on the skeptic (an assumption we will later challenge), it now seems that we need something more than just one sufficient natural explanation for theistic belief in order to debunk its justification. Perhaps in order to debunk S’s belief, we need to show that p is more likely to be false
than true given that S believes that p.\(^\text{24}\) This can be expressed in conditional probability terms as

\[
P(\sim p \mid S \text{ believes } p) > P(p \mid S \text{ believes } p), \text{ or} \]

\[
\frac{P(\sim p \mid S \text{ believes } p)}{P(p \mid S \text{ believes } p)} > 1
\]

We can then use Bayes’ Theorem to re-express the numerator and denominator as follows:\(^\text{25}\)

\[
P(\sim p \mid S \text{ believes } p) = \frac{P(S \text{ believes } p \mid \sim p) \times P(\sim p)}{P(S \text{ believes } p)}
\]

\[
P(p \mid S \text{ believes } p) = \frac{P(S \text{ believes } p \mid p) \times P(p)}{P(S \text{ believes } p)}
\]

Substituting these expanded expressions into the earlier ratio, which cancels the \(P(S \text{ believes } p)\) term, leaves us with

\[
\frac{P(S \text{ believes } p \mid \sim p) \times P(\sim p)}{P(S \text{ believes } p \mid p) \times P(p)} > 1
\]

There are three ways to assess the left-most term with regard to a belief-forming process, depending on which of the following ways the generated beliefs are thought to correspond to reality: (1) completely uncorrelated, (2) negatively correlated, or (3) positively correlated. What Jong and Visala point out is that if it is uncorrelated—that is, if the BFF is non-truth-tracking in the sense of its being indifferent to truth—then the left-

\(^{24}\) What follows is an inequality-based variation of Jong and Visala’s use of Bayes’ Theorem.

\(^{25}\) For an overview of Bayes’ Theorem, see the Bayes’ Theorem Explained section, below.
most term becomes “1,” since the numerator and denominator would be the same. But in that case, our original expression would end up reducing down to just

\[
\frac{P(\sim p)}{P(p)} > 1
\]

This means that in the uncorrelated case, debunking \( p \) amounts \textit{only} to showing that the prior probability that \( p \) is false is greater than the prior probability that \( p \) is true. But that means that considerations associated with the belief-forming process have become completely irrelevant—at least in this uncorrelated case. The whole question simply falls entirely back onto the philosophical arguments we have historically used (Jong and Visala 247). So, to the extent that CSR has shown that a BFF is uncorrelated to truth, it shows that if \( S \) justifies her belief \textit{solely} on the BFF, then her belief is unjustified. Of course, if \( S \) cites other sources of justification, which seems likely, then CSR would not be making much of a contribution one way or the other. The debate would simply revolve around her reasons for belief, just as it had before the advent of CSR.

Recall that there are two other possibilities besides the uncorrelated case: positive correlation (i.e., truth-tracking), and negative correlation (“falsehood-tracking”). In the latter case, instead of the left-most ratio above becoming “1,” it would become something greater than 1. Indeed, it could become arbitrarily large depending on how “perniciously deceptive” (Jong and Visala 248) we think the belief-forming process is. In terms of

\[\text{26 This is because the numerator expresses the probability that S believes p given that p is false, while the denominator expresses the likelihood of this same belief given that p is true. But if the truth of p has no bearing one way or the other on S’s believing that p, then it can have no differential impact on the numerator vs the denominator; therefore, as expressed, they would simply cancel out.}\]
conditional probability, this is the situation where if p is false, then it is more likely that S will believe it to be true; and if p is true, then S will be less likely to believe it. Depending on how we assess these likelihoods, the ratio could become even astronomically large.

Using Jong and Visala’s example of 0.9 and 0.0001 we would get

\[ 9000 \times \frac{P(\sim p)}{P(p)} > 1 \]

This would mean that the prior probability for p’s truth could be much higher than the prior probability of its falsehood, and yet the inequality would still hold; that is, the belief produced by this process would still more likely be false than true—a deceptive process, indeed.

But does this describe the findings of CSR? Jong and Visala think not: “Perhaps in the future, scientists will posit cognitive mechanisms that generate religious beliefs especially if they are false (and rarely if they are true), but until then, we need not worry about EDAs that posit falsehood-tracking or perniciously deceptive cognitive processes” [emphasis in original] (Jong and Visala 249).

We will examine in a later section some reasons to suspect that they have dismissed this scenario too quickly.\(^{27}\) For the moment, however, let us accept their assessment of the non-applicability of the negative correlation case and return to the uncorrelated case in which a believer’s sole justification for belief is the reliability of the

\[^{27}\text{For now, it is worth noting that examples of negative correlations between belief and the truth of that belief have been well documented in non-religious cases, such as our human tendency to deny palpably true statements when they threaten us in some way, e.g., diagnoses of terminal illness, or ego-damaging evidence of short-comings (Dawes and Jong 8–9).}\]
BFF itself. Does such a situation describe actual theists? Interestingly, it does: it seems to correspond to epistemological reliabilists, such as reformed epistemologists. This group justifies its belief entirely on the basis of these processes’ presumed truth-tracking abilities, not on evidence. Recall that reformed epistemologists explicitly reject evidentialism. To be sure, evidentialists may, along with reliabilists, agree that a belief may have been *originally* produced by processes unrelated to evidence, such as religious experience, but only evidentialists justify their belief on later-acquired evidence. For this reason, Jong and Visala point out that reformed epistemologists may be vulnerable to EDAs to the extent that such EDAs focus on justification (Jong and Visala 256). Later, we will argue that this vulnerability can be expanded.

Expanding the Irrelevance Argument

Jong and Visala’s relevancy critique is powerful if somewhat limited in scope. Josh Thurow takes a different approach. In developing his case, he first works to clean up what he sees as important vulnerabilities in the current critiques of the atheist-CSR position, and then iteratively develops and critiques his own understanding of that position. He begins by proposing a general schema intended to capture the general form of the (implied) atheist-CSR argument:

P1: If theory T is true, then religious beliefs are produced and sustained by process P.

P2: Process P is unreliable and does not make use of good evidence when it is used to form and sustain religious beliefs.
P3: If the process by which a belief is formed and sustained is unreliable and does not make use of good evidence, then that belief is unjustified.

C: If theory T is true, then religious beliefs are unjustified (Thurow 83).

While his argument could have broader applicability, he focuses his analysis on the CSR Standard Model, which emphasizes the HADD and by-product views that we examined earlier. In this context, he distinguishes two belief types—B1: god did this, and B2: god exists (Thurow 84). This distinction is another way of expressing Barrett’s earlier claim that CSR cannot be used to make a case for atheism without importing its particular metaphysical assumptions. In other words, HADD may very well show that our belief-forming mechanisms are unreliable regarding the causes of poorly understood events, predisposing us to genuinely misattribute their causes to supernatural agency; however, it does not follow from this that all such attributions are false in the case of religious beliefs. B1 may well be false in some or even most circumstances, but it does not follow that B1 is always false, or that B2 is false just because B1 is. This way of looking at the issue also reemphasizes Jong and Visala’s claim that atheist-CSR arguments can, at most, challenge justification—not veracity.

However, Thurow notes that as it stands, such attempts to undermine the unreliability premise, P2, are inadequate. Consider the following truth-tracking formulation: “(SG) If there were no gods and we were to form beliefs about gods using process P, we would still believe, via P, that some kind of god exists . . .” (Thurow 86–87). He notes that this formulation is indeed problematic as a reliability test, but only in the case of inductive belief forming processes, not of basic belief forming process. So, to
the extent that CSR explains basic belief forming processes, then P is indeed unreliable
with regard to belief in God, a point which echoes Jong and Visala’s regarding the way in
which reformed epistemologists seem particularly vulnerable to CSR’s findings. One
important aspect of Thurow’s reliability test is that it does not require that we know prior
to using our faculties that they are reliable or that their output is false, only that we have
evidence that the faculty under consideration would produce a false basic belief
regardless of that belief’s veracity (Thurow 89).

This would seem to address Plantinga’s and Barrett and Church’s worry about
naturalist question-begging, whether in terms of presuming atheism, or in presuming the
unreliability of our cognitive faculties in specifically religious contexts. However,
consider Barrett’s objection to P2 from a different standpoint. A process can be
“reliable,” but still produce incorrect results. Thurow uses the thought experiment of
one’s happening upon a mechanism that makes astronomical predictions based on
principles that applied in the past, but no longer, or perhaps apply now, but depend on
correct inputs. Thurow intends this to be analogous to our basic-belief producing, CSR-
based faculty in that both are examples of mechanisms producing “outputs.” The CSR-
based faculty produces basic beliefs, while the found device produces astronomical
predictions. Such an astronomical device is still “reliable” even if it now produces
inaccurate results, according to Thurow, because the device’s designer based it on
principles that she knew to be true at the time she designed it. Similarly, if God did in fact
design our CSR-based BFF, then it would be reliable in precisely this sense, despite any
currently inaccurate outputs. His point is that just because a process produces an incorrect
result, such as false basic beliefs, it does not necessarily mean it is “unreliable.” This would mean we have an example of something analogous to a “reliable” BFF failing his reliability test (SG), which undermines P2.

Yet, a further thought experiment suggests that this potential problem for the atheist-CSR position can be eliminated. Thurow asks us to imagine a world in which all humans are strongly disposed to believe in life on Mars. In this world, we discover that this predisposition is a by-product, a side-effect, of the normal operation of our cognitive tools—tools that had evolved to increase survivability in our ancestral, earth-bound environment. Despite the universality of this Martian-belief disposition, it does seem that this “cognitive science of Martians” discovery should lead us to suspend belief in life on Mars pending independent evidence of its existence. Notice that even if our belief in life on Mars had been prima facie justified based on the universality and even basicity of this belief, the new discovery creates a defeater for that initial justification. We should, therefore, suspend belief in life on Mars pending independent evidence. If we find any, then it may lend strong support to the reliability of our Martian-sensitive cognitive faculties, but not before. Thurow, then, has not eliminated the need for independent evidence; rather, he has shown in this thought experiment that the burden for producing such evidence seems to fall squarely on the believer, given the discovery of a CSR-like predisposition to belief.

Incorporating the preceding considerations, Thurow shapes his strongest argument for the atheist-CSR position, his CSR Process Defeater Argument:
PD1: If theory T is true, then religious beliefs are produced and sustained by process P, which is a basic belief-forming process.

PD2: Process P has the following feature: if religious beliefs were not true (i.e. no god existed), then P would still produce religious beliefs.

PD3: If the process by which a belief p is formed and sustained is structured in such a way that if p were false, the process would still generate belief that p (and the process is not an inductive argument), then we should suspend judgment about the reliability of that process with respect to p in the absence of independent evidence for the reliability of the process.

PD4: If we should suspend judgment about whether the belief-forming process we use is reliable with respect to p, we should suspend judgment about p.

PD5: If we should suspend judgment about p, then we are not justified in believing p.

PD6: There is no independent process to validate the reliability of P (from P1[PD1]).

C: If theory T is true, then our religious beliefs are not justified (Thurow 90).

Thurow also reminds us here that CSR theories address not only the origin of religious beliefs, but also why they are sustained. As such, the simple charge that the atheist-CSR position commits the genetic fallacy misses an important part of the CSR
challenge, namely, the unjustified sustaining of religious belief. To be sure, if one initially holds a belief based, say, on having ingested some hallucinogen, but then later acquires good evidence, then at that later time one becomes justified in holding that belief. But what if, despite the fact that such evidence for God exists, this particular believer remains unaware of it? If she continues to base her belief on her initial, unjustified belief-forming process—drugs in this case—then her belief remains unjustified, even though justifying evidence does exist. Thurow uses this example to draw our attention to two types of justification: doxastic and propositional. In our current example, the evidence-unaware believer is propositionally justified, since the evidence exists, but not doxastically justified, since she does not base her belief on that evidence. This reinforces Jong and Visala’s two earlier points, which were (1) CSR can address only the justification of belief, not its veracity; and (2) even if CSR thoroughly explains religious belief in naturalistic terms, it would not eliminate other explanations, such as God as ultimate explanation. To this Thurow has added the following clarification: (3) a sufficient natural explanation, such as CSR, also fails to establish that a belief is not propositionally justified—that is, even if no one is currently aware of the justifying evidence.

But the crucial problem Thurow sees in his carefully reconstructed CSR Process Defeater argument is this: It does not address the reliability of the processes people actually use to justify their religious beliefs. “People believe because: they think the Bible is reliable, they think they have witnessed, or know others who claim to have witnessed certain miracles, certain prayers get answered, their life has been changed for
the better since believing, the world seems so carefully designed . . .” And though the testimony of elders and social pressure may play a significant role, “people normally accept the testimony of elders because they think the elders have good reasons” (Thurow 92–93).

Let us say that CSR does show that people would still believe in some kinds of supernatural agents even in the absence of the belief processes they actually use. Thurow’s point is that CSR still entirely misses the question of whether or not those processes are reliable, and, consequently, fails to show that their beliefs are unjustified. The problem is that the processes examined by CSR may not be wholly or even principally responsible for the genesis and maintenance of such beliefs. So, while Jong and Visala point to the possibility of alternate sources of justification, Thurow has argued that these alternate sources are in fact what people actually use, rendering CSR even more irrelevant to the justificatory status of theism: If our HADD-based BFF is off-track in even a “perniciously deceptive” way, it may matter little if it is rarely relied upon in actual practice.

As I read Thurow, his point can be generalized as follows. Person S holds belief B in supernatural entity G, which is, let us say, formed and sustained by belief-forming processes \( J_p, J_l, J_i \) and \( J_{csr} \), each of which is sufficient on its own to form and/or sustain S’s belief B in G. In other words, the cause of belief B is overdetermined for person S. Consequently, the reliability of all J’s for person S—not just one or even most—would have to be undermined before concluding that B is unjustified. In the present case, if CSR has shown that person S would still hold B even if B were false, then \( J_{csr} \) would indeed be
undermined. But this would be like cutting only one of the steel wires that make up a thick steel cable suspending a piano in the air. All the other steel wires (i.e., all the other J’s), each with the ability to support the piano on its own, would also need to be cut. In fact, Thurow tells us, it is these other “wires,” the non-CSR J’s, that most believers actually use to form and/or sustain their belief.

Consider the following analogy. Let us say that we are in a room in which there is a realistic holographic projection of an apple. Let us assume first that there is in fact an apple in the room, though it is completely hidden within the holographic apple projection. When a person in the room, who is unaware of the projector, is asked why he believes there is an apple, he explains that he smells it and he sees it. If we assume that either of these would be sufficient to sustain his belief, then his justification is overdetermined. Consequently, if we remove the real apple, he would continue to believe it is there since the projector would continue to project. This means his sight-based justification is indeed unreliable under the current conditions. However, his other justifying reason for belief is smell and it is (under these conditions) reliable, as he would no longer smell the apple once it is removed. If he continues to believe based solely on sight, then indeed his belief has become unjustified. However, if we discover that he abandons his belief as a result of no longer smelling the apple, and further discover that most people base their belief on smell and would also abandon their belief in these circumstances, then of what relevance is the holographic projector discovery to the question of rational justification of belief?
Summary

Now that we have examined the genealogy and shape of the current leading criticisms of the atheist-CSR position we should take a moment to take stock of where we are, and then chart our path forward. As we have seen, Plantinga argues that belief in God is properly basic, contra classical foundationalism. Indeed, his argument provides strong (though not completely impervious) shielding for such belief against even powerful, intersubjectively verifiable evidence, as we saw in his tire slashing example. He then takes a reliabilist oriented, proper functionalist view of rationality in the way he truth-conditionally defines warrant. This, he argues, shifts the burden of proof for God’s existence onto the atheist. Finally, through his EAAN, he goes beyond simply arguing that theism is as rational as its naturalist alternative to arguing that theism is in fact the far more rational choice, since otherwise one slides into the bog of radical skepticism.

Others build upon this foundation to argue specifically against atheist-CSR. In so doing, they make the charge of question-begging against claims that HADD is “wrong” or “over-detecting” whenever it produces specifically supernatural beliefs. They also raise the related EAAN-based objection to the claim that our religious beliefs are unreliable, since we cannot argue that our CSR-based religious beliefs are unreliable without rendering all our beliefs unreliable. Importantly, their efforts bolster Plantinga’s argument by bringing CSR-based empirical evidence to bear.
The other powerful line of criticism against CSR-based atheistic arguments can be called the explaining-is-not-explaining-away objection.28 This includes roughly two basic types of criticism, which can be grouped under the following fallacy categories: the genetic fallacy, and the single-cause fallacy.29 The genetic fallacy as applied to atheist-CSR was expressed in a variety of ways, from the simple, “At most CSR addresses how we come to believe in God, not whether He exists” to the more formal Bayesian-based argument that CSR is entirely irrelevant to the question of God’s existence (at least in the uncorrelated case). The single-cause fallacy objections emphasize the point that one cannot assume that a well-established sufficient cause for religious belief, such as the natural one posited by CSR, is the only cause. There may well be others, each of which can separately originate and sustain belief. Indeed, these other causes might come from the same explanatory level, or from higher (even ultimate) levels. In fact, it was argued that the belief-forming and sustaining processes most believers actually use are not even addressed by CSR.

The criticisms in both of these broad categories can also be recast as special cases of the burden-of-proof objection: CSR, no matter how well proven, fails to meet the burden of proof in showing that God does not exist or that belief in Him is unjustified.

28 I have adapted this concise phrase from (Jong)
29 These categories are being used here as conceptual tools of convenience and should not necessarily be thought of as mutually exclusive and jointly exhaustive.
CHAPTER 4

CRITICAL ASSESSMENTS OF THE OBJECTIONS

Given the foundational role that Plantinga’s reformed epistemology and EAAN have played in the current debate, it is important to assess its key assumptions before critically examining any arguments built upon them. Specifically, we will examine both his claim that belief in God is rationally justified because it is properly basic, and his truth-conditional concept of warrant.

Critique of “Reformed Foundationalism”

Recall Plantinga’s discussion of the Great Pumpkin Objection, in which he argued that he has at least as much justification in rejecting that and accepting Christian belief as the atheist has in doing the converse. His argument was based on his first rejecting classical foundationalism as a self-refuting position and then gesturing toward an inductive approach for establishing prima facie, properly basic beliefs. At that stage he was arguing only for rational justificatory parity and that each community is responsible to its own starter set of examples. Michael Martin is keen not to let this pass without pointing out the relativistic consequences of such an approach. He observes that were we to take this seriously, we would have to allow any community to claim rational parity as well—from Heaven’s Gate practitioners to fairy believers. All that would be required is that a community’s beliefs be basic to their members (Martin 272). While Plantinga says
precisely this with regard to his particular community. Martin shows how this position leads to a rather extreme form of rational relativism.

The issue is not with basic beliefs as such, but Plantinga’s inclusion of belief in God as an example of such a belief. Recall that he had used his proper functionalist account of rationality to argue that before we can trust our perceptual senses, we must know that they are neither defective nor operating under inappropriate conditions. This sense-perception analogy is intended to capture how he thinks our supposed sensus divinitatus works. Martin, however, argues that the analogy is particularly weak: “Part of the justification for believing that our perception or memory is not faulty is that in general it agrees with the perception or memory of our epistemological peers . . . as well as with our other experiences” (Martin 274). However, the contexts of memory and sense perception are markedly dissimilar from the context of religion, where the lack of agreement between epistemological peers is extreme. Indeed, the very same triggering conditions for religious “basic” belief (e.g., reading the Bible, beholding a flower, or admiring the intricacies of the natural world) can and do trigger a wide range of mutually exclusive “basic” beliefs—including atheistic ones (e.g., beholding such other natural “wonders” as parasites and the agonizing deaths they cause, or even the apparently God-

—

30 The assertion should be defended by quoting him at length: “The Christian will of course suppose that belief in God is entirely proper and rational; if he does not accept this belief on the basis of other propositions, he will conclude that it is basic for him and quite properly so. Followers of Bertrand Russell and Madelyn Murray O’Hare may disagree; but how is that relevant? Must my criteria, or those of the Christian community, conform to their examples? Surely not. The Christian community is responsible to its set of examples, not to theirs” (Plantinga and Wolterstorff 77). While he was comparing only those two communities, his statement opens the floodgates of relativism across all communities. To see this, simply replace “Christian,” “atheist,” “Bertrand Russell” and “Madelyn Murray O’Hare” with the relevant counterparts from any real or imagined community.
sanctioned evil deeds as recounted in the OT). It is precisely this lack of agreement among epistemological peers that precludes belief in God from being considered a properly basic belief (Martin 275).

Martin also points out that Plantinga’s attack on classical foundationalism seems to discount the contemporary epistemological landscape, now dominated by more modern versions of foundationalism. For example, most now include memory as properly basic (Martin 271). Therefore, his attack on classical foundationalism, even if successful, does not necessarily undermine foundationalism generally. Moreover, as we have already seen, there are non-foundationalist epistemologies, such as coherentism (Berker), which if correct, would further weaken Plantinga’s argument.

Given these issues, it appears that Plantinga has failed to establish that belief in God is *properly* basic (though it may well be basic in the sense of being produced non-reflectively). Consequently, it cannot be admitted into that privileged, evidence-resistant group. But even setting this issue aside, how effective is his truth-conditional, warrant-based approach to shifting the burden of proof?

**Defeating Warrant**

Plantinga claims that the unbeliever bears the burden of proof regarding the truth-value of theistic belief, a claim he bases on his truth-conditional definition of warrant. Dawes and Jong reject this, arguing that warrant is precisely what the unbeliever needs to address; in fact, it is *all* she needs to address. They concede that fully natural explanations of religious belief are insufficient for undermining the justification of religious belief. (Recall, for example, our earlier discussion of over-determined
justifications and multiple levels of explanation.) Rather, their argument is based on taking a different approach toward the notion of reliability.

Rather than attempting to define reliability at the outset, they come at it from the other direction, that is, by identifying ways to recognize instances of unreliability. (Dawes and Jong 8). Freed from the requirement for a positive definition, they argue that the atheist has only to show that our cognitive faculties are unreliable specifically in those circumstances in which they generate religious belief. As they remind us, this circumstance-specific approach is consistent with the way in which we think of the reliability of our perceptual senses (Dawes and Jong 7).

If their approach works, it would address one part of the EAAN, which claims that we cannot argue on the one hand that our cognitive processes are unreliable when they produce religious beliefs, and on the other hand use those same indicted processes to accept non-religious beliefs. Recall that Barrett and Church had dismissed scope-narrowing as nothing more than special pleading, at least until some positive criterion of reliability could be found for separating religious belief-forming conditions from non-religious ones (Barrett and Church 322). However, Dawes and Jong’s approach shows that such a positive criterion is not necessary; one need only recognize relevant cases of unreliability. Just as one can recognize a wrong answer or non-answer without knowing the right answer, so too can one identify particular instances of unreliability without a
fully articulated general theory of reliability. This certainly seems correct. I may not know, for example, how an airplane flies, yet still know that whatever the answer is, it will not be that it is being carried by winged angels.

This way of looking at unreliability is relevant in another way. Even if, under Plantinga’s model, Christian belief is not warranted in virtue of its explanatory power or evidence (because it is basic with respect to warrant), it is still based on a causal mechanism—something akin to Calvin’s sensus divinitatus. As such, it is vulnerable to a better competing explanation. To illustrate this, Dawes and Jong use the case of a professor’s belief in the existence of a whiteboard, which she can see in her classroom. This belief, they note, is basic in Plantinga’s sense: it is neither held as an explanatory hypothesis, nor warranted on that basis. Rather, its warrant derives from the fact that it was spontaneously formed by sensory and cognitive faculties that are typically reliable in situations such as this. But now this professor receives the following memo: The “Physics Department are conducting experiments in selected lecture theatres using remarkably realistic holographic images of both whiteboards and whiteboard markers, without informing lecturers. . .” Let us say that based on this and other factors, it is possible that this particular professor is using just such a classroom (Dawes and Jong 10). The suggestion that such new information should have no effect on the professor’s belief

31 Interestingly, Plantinga himself makes essentially this same point following his critique of classical foundationalism. He argues that while the theist may not have a ready, clearly formulated alternative to classical foundationalism to justify her basic beliefs, this hardly means none exists or that it is irrational to proceed without such an alternative (Plantinga and Wolterstorff 75).
seems clearly incorrect. As they point out, she should now at least want some independent evidence to offset the warrant-undermining effects of this new information.

Compare this example with Plantinga’s earlier tire slashing example where we saw that even very strong external evidence may have understandably little force in undermining a first-hand, memory-based *basic* belief. What is the difference between that example and the current whiteboard example? Dawes and Jong’s answer, I suggest, would be that the former uses the accused’s memory-based basic belief in order to attack the *veracity* of the prosecutor’s belief, despite the (intersubjectively verifiable) evidence to the contrary. It is, in Plantinga’s terminology, a *de facto* argument. The latter example, however, attacks the *warrant* of the professor’s belief. It is a *de jure* argument to the effect that in this particular case, the relevant BFF is not operating under conditions appropriate to it—*regardless* of whether or not the belief is true (the whiteboard may in fact be real). Moreover, by specifically attacking warrant, we logically preclude the believer from appealing to his beliefs *regarding warrant* to rebut an argument that is, after all, challenging precisely that (Dawes and Jong 12)

They note too that their approach is not vulnerable to the question-begging objection, since no presumption of truth or falsity is being made. One is not *a priori* excluding explanations, supernatural or otherwise; one is simply comparing two competing explanations for belief formation and then making an inference to the best explanation. This avoids the need to import any atheistic metaphysical assumptions. Indeed, because theirs is specifically a warrant-defeating argument, it is the Christian and
not the atheist that faces the charge of question-begging if any appeal is made to beliefs about warrant (Dawes and Jong 15).

Dawes and Jong’s analysis effectively undermines Plantinga’s intriguing attempt to use his definition of warrant to shift the burden of proof onto the skeptic. At the same time, their use of inference to the best explanation and their negative approach to the concept of reliability accomplishes the following: (1) it undermines the charge of question-begging when the naturalist uses CSR’s findings to argue that our BFFs are unreliable specifically in the case of religious beliefs; (2) it answers Barrett and Church’s objection that treating religious contexts and non-religious contexts differently amounts to special pleading; and (3), it begins to call attention (through the notion of inference to the best explanation) to a hidden assumption in the single-cause fallacy charge: while there may indeed be many justificatory legs to a belief, they are not necessarily equally likely, even though they may all be logically possible. We will explore the significance of this point later. First, we will examine some surprising implications of Plantinga’s own arguments, which, in their own ways, are as self-defeating as he claims naturalism is.

A Reidian-Based Defense of Naturalism

Michael Bergmann critiques the EAAN by using Plantinga’s own Reidian-based epistemological argument. Earlier, we reviewed Plantinga's analysis of the conditional probability that our cognitive faculties (R) are reliable given naturalism (N) and evolution (E), expressed as \( P(R/N&E) \). Recall that he argued that if this conditional probability is low, then we have a defeater for R, which leads to the self-defeating result that all our beliefs, including our beliefs in N and E, are unreliable. Bergmann argues, however, that
even if we grant this low probability assessment, we still do not have a defeater for R given N and E. He does this by examining Plantinga’s own separate critical analysis of an atheistic argument from evil. Recall that additional background knowledge can change a conditional probability from low to high—Plantinga’s “defeater deflector” (D)—provided that D neither equals nor entails R. All relevant propositional (i.e., inferential, non-basic) knowledge can be incorporated in this fashion; so, we can simplify our earlier expression by collapsing N, E, D and all other relevant knowledge into k, which gives us $P(R/k)$.

This can be used to express Plantinga’s earlier tire slashing example, where R represents the accused’s innocence. Since, in that example, we are talking only about propositional knowledge, both the jury and the wrongly accused person can agree that $P(R/k)$ is low (Bergmann 623). The difference, however, is that the accused has additional non-propositional information, in this case, her memory. Because of this, and despite the low probability of her innocence based on propositional knowledge, she is nonetheless reasonable in believing in her own innocence.

The analogy with Christian belief that Plantinga intends with his tire slashing example seems clear: even if all the propositional evidence we have for God’s existence makes the probability of His existence low, the theist might nonetheless have a kind of seeming, a spontaneous, non-propositional basic belief that trumps all propositional knowledge. Bergmann points out, however, that this very same argument can be used against Plantinga’s EAAN. Bergmann’s insight is based on Reid’s famous “commonsense” epistemology, which so influenced Plantinga’s. Reid’s thinking on non-inferential basic beliefs is that they are rooted in our natural commonsense. This natural
commonsense generates basic beliefs directly from experiences, like sense perception, and not from other beliefs. Reid argues further that this same commonsense faculty not only generates basic beliefs, but also actively rebels against contradictory concepts—a kind of “immune response”—by generating emotions like ridicule (Bergmann 624).

This is consistent with Plantinga’s views regarding properly basic beliefs. The problem it presents for his EAAN, however, is that the naturalist evolutionist can use precisely this same line of reasoning to reject Plantinga’s low probability assessment of our cognitive processes’ reliability. Just as the theist can invoke non-inferential, commonsense, spontaneously generated beliefs to override the weight of propositional evidence against that belief, so the naturalist evolutionist can refer to her own non-inferential beliefs to override arguments like Plantinga’s (Bergmann 625). As Michael Martin puts it, “since what is self-evident is relative to persons, a classical foundationalist could argue that [the CF biconditional definition of proper basicality] is self-evident and that if Plantinga were sufficiently attentive, the truth of [that definition] would become clear to him” (Martin 270–71).

This point is related to an overlooked confusion in Plantinga’s argument, which is best illustrated by examining his own thought experiment regarding a “hypothetical species,” one which

is cognitively a lot like us: members of this species hold beliefs, make inferences, change beliefs, and the like. And let us suppose naturalism holds for them; they exist in a world in which there is no such person as God or anything like God. Our question, then, is this: what is the
probability that their cognitive faculties are reliable? (Plantinga, *Where the Conflict Really Lies* 329–30)

Plantinga’s thought experiment is designed to give us and the other species comparable cognitive histories. Yet, in assessing this imaginary species’ cognitive reliability, we seem compelled to conclude, via the EAAN, that \( P(R/N&E) \) for these creatures is indeed low, at least with regard to belief content. Plantinga intends this analogy to show why we similarly cannot regard our own cognitive processes as reliable under naturalism.

Bergmann points out, however, that with regard to this hypothetical species, Plantinga is arguing from the outside looking in, as it were. Bergmann concedes, for the sake of argument, that the EAAN applied to Plantinga’s hypothetical species example gives us a defeater regarding belief in this species’ cognitive reliability; however—and this is key—it is a defeater for *our* belief in *their* cognitive reliability, not our own. Why? Because in looking at that species from the outside (as we are not members of their species), we have access *only* to propositional knowledge about them (Bergmann 625). When we compare this analogy with Plantinga’s earlier tire slashing example we see that his low assessment of the alien species’ cognitive reliability is like the *jury’s* low assessment of the *accused’s* innocence; however, in the case of assessing our own cognitive reliability, we are more like the accused assessing her *own* innocence, since we have access to relevant non-propositional knowledge.

Even with this inner access to non-propositional knowledge, is this memory-based basic belief indefeasible? Clearly not, and Plantinga admits as much. In his tire slashing example, he recognizes that there could be evidence so strong as to justify doubting one’s
own memory, a development which would constitute a defeater for this basic belief. He
does not explore the implications of this for his argument, however, remarking only that
in such a situation “what we have is a strong conflict between memory and these other
sources of belief” (Plantinga, Where the Conflict Really Lies 180). Yet, as we will see,
CSR’s findings highlight the importance of exploring precisely this point. What’s more, a
defeater may also come from other non-propositional beliefs, such as, “Basic belief in
God is the product of an Evil Demon” (Bergmann 625). The point, of course, is not to
defend such alternative explanations, only to note that such a non-propositional belief
would constitute a defeater for one or more other non-propositional beliefs.

Bergmann has used Plantinga’s own Reidian-based argument in one context to
undermine his conclusions in this one. First, he used Plantinga’s argument against an
atheistic argument from evil in order to undermine the EAAN argument. Then he used
Plantinga’s argument that belief in God is properly basic in order to show that precisely
the same line of argument can be made that belief in human cognitive reliability is a
properly basic belief, contra EAAN. Finally, he reminded us that basic beliefs need not
be consistent, and can even serve as defeaters for other basic beliefs.

It is important to keep in mind that Bergmann’s critiques do not depend on
actually embracing Reid’s epistemology. Rather he is showing that this epistemology can
be used to defeat the EAAN. This was done in much the same way that the EAAN itself
assumes the truth of naturalism in order to undermine it. But there is another, even more
fundamental direct criticism of the EAAN, namely, that it is based on a fundamental
misconstrual of how evolution actually works.
Evolution and the Cost of Truth

We saw earlier that a valid critique of some atheist-CSA arguments was that they committed the single-cause fallacy, either by assuming that a natural explanation automatically debunks a supernatural one, or by ignoring the possibility of different levels of explanation. As we will see, this important way of thinking about multiple causes cuts both ways.

It appears that Plantinga’s EAAN has its own kind of levels-of-explanation confusion. According to Paul Griffiths and John Wilkins, Plantinga is inappropriately casting complementary explanations from distinct explanatory levels as competing alternatives at the same level (Griffiths and Wilkins 5). Recall that Plantinga had argued that adaptive behavior and belief content are completely independent things. Physical traits and behaviors either increase or decrease the likelihood of reproductive success, and are therefore subject to evolutionary selective pressures. Yet if belief content is independent of these, how can it be influenced by evolutionary pressures at all? Evolution, in other words, can reward fitness-tracking but not truth-tracking. But Griffiths and Wilkins argue that this is not a sensible distinction, and they use a levels-of-explanation argument to show why.

As I read them, evolutionary explanations operate at multiple explanatory levels. A more general explanation can be consistent with a variety of lower-level explanations. One such lower-level explanation would answer the question, “Which aspect of some particular trait is the target of selection?” while a higher-level explanation would answer the question, “Which trait is more fitness enhancing?” Treating explanations from each of
these levels as competing alternatives would be like asking whether archaeopteryx’s non-aerodynamic feathers were originally selected for keeping it warm (versus, say, catching insects) or for enhancing fitness. To see how this levels-of-explanation confusion relates to the distinction between belief-content and behavior we need to first understand the role of cost in evolution.

Arguably, Griffiths and Wilkins’ most important contribution to this discussion is their reminding us of a central evolutionary concept often overlooked in arguments such as Plantinga’s: cost. Costs create constraints. This familiar concept is seen in many areas of human endeavor, and is often captured by such terms as “diminishing returns”—the point at which the benefits of further improvement are outweighed by the costs of making that incremental improvement. “Cost” can be energy, material expenditure, or even deleterious side effects. Importantly, it can also be time. For example, consider an investor. Ideally, she would eliminate all uncertainty before making a particular investment, but the window of time is limited. If she waits too long, the opportunity is lost, and with it possibly her financial survival. Imagine too that she actually has access to all of the analytical tools and information she needs to perfectly assess the investment and eliminate all risk, but doing so would take too much time and/or other resources that, if employed in this perfection quest, would far outweigh the investment’s best-case return. In order to financially survive, she develops some imperfect, but much-better-than-chance rules of thumb. From a levels-of-explanation perspective, this partial retreat

32 Should the following finance-oriented example fall short in any way, I hasten to point out that it is my own and not Griffiths and Wilkins'.

---

74
from perfect accuracy would not change that fact that she is still after the best possible investment choice. She is not now after something else. We might wonder why some particular rules of thumb work for her, even though in certain situations, they can and do result in bad investment decisions. Likely, the details of her rules reflect historically contingent facts about her local resource and threat environment. Were they different, her particular rules would likely be different, but that is because they would be adjusting to track the same thing as well as possible given those different constraints.

This same notion of constrained, cost-sensitive optimization also applies to biological evolution. “Improvements,” however we define them for a particular species in a particular environment—speed, size, etc.—do not continue forever toward some idealized state, such as the ability to run at near light speeds. Why then should we suppose this would be any different for the evolution of our cognitive faculties’ truth-tracking ability? This insight not only begins to link our cognitive faculties directly to evolutionary forces, but also leads us to expect that these faculties would not be perfect trackers of truth. Indeed, our beliefs are the output of a set of cognitive adaptations. Those adaptations are not designed to produce only true beliefs, or to produce all the relevant true beliefs on every occasion. But this is not because they are tracking some property other than truth. It is because they are tracking truth in a constrained manner. (Griffiths and Wilkins 7)

Yet another constraint arises from the intrinsically probabilistic nature of many cognitive tasks. When forced to make a decision under conditions of uncertainty, the
risks of false positives and false negatives move in opposite directions. One type of error can be reduced, but only at the cost of increasing the other. Evolution, therefore, would be expected to favor those who maximize expected “payoffs” (outcomes) by optimizing the balance between the two types of errors. In other words, it would select for those creatures that act or "form an action-guiding belief" so as to maximize the expected payoff (Griffiths and Wilkins 7–8). In other words, “a tendency to commit a fallacy may evolve, not merely as a trade-off between cost and accuracy, but in order to improve accuracy given cost constraints. Cheap fallacy can, in the right circumstances, deliver more truth than costly validity” [emphasis added] (Vlerick and Broadbent 189).

Moreover, to the extent that such expected payoffs are asymmetric, evolution would be expected to drive toward biased action-guiding heuristics (or “rules of thumb”)—such as the bias that puts the “H” in HADD.

Accordingly, Griffiths and Wilkins analyze the evolution of cognitive reliability as a straightforward instance of decision making under uncertainty. A creature has to “decide” whether or not to act given some environment signal—a signal which is a less-than-perfect indicator of some actual situation (say, being in the sights of a predator). The reward/punishment effects of making this decision involve “assessing” the net payoffs of each of the four possible combinations of action and signal accuracy: whether to act or

---

33 The expected payoff is the resulting net benefit—positive or negative.
34 It cannot be stressed enough that such evolutionary analyses should never be construed as implying that creatures (even humans in most cases) are making such payoff calculations consciously. Rather, the analyses describe evolutionary pressures on populations that, over time, lead to a high frequency of individuals that behave as if such calculations were being consciously made.
not act, and whether the signal is veridical or not (Griffiths and Wilkins 8). Griffiths
and Wilkins’ analysis makes clear that truth-tracking cannot sensibly be an alternative to
fitness tracking—payoffs are higher when truth is tracked and lower when it is not. In
other words, fitness tracking is truth tracking, but subject to the realities of less-than-
infinite resources and time: "Organisms track truth optimally if they obtain as much
relevant truth as they can afford, and tolerate no more error than is needed to obtain it"
(Griffiths and Wilkins 9).

What is not clear, however, is whether any of this actually addresses Plantinga’s
specific charge. Plantinga himself recognizes that truth must be accurately tracked
somehow—the evolution of adaptive behaviors could hardly have occurred otherwise.
The issue is whether belief content reflects that tracking or is instead completely
untethered from it as Plantinga suggests. In their discussion of payoffs, Griffiths and
Wilkins have wisely assessed these payoffs in terms of the four possible combinations of
action and signal accuracy, rather than belief content and signal accuracy; however, they
have not articulated a strong evolutionary link between action and belief content. It is just
this link that goes to the heart of Plantinga’s argument. Nonetheless, their insightful
analysis lays the critical groundwork that can lead us directly to this strong connection.
How? Once again, the answer is directly related to cost.

35 Think of the net payoff, P, as being the result of the benefit, B, minus the Cost, C, of acting: P =
B – C. This is then assessed for the four combinations of Act/Veridical, Act/Not-Veridical, Not-
Act/Veridical, and Not-Act/Not-Veridical, typically represented in a 2x2 matrix.
Plantinga’s argument is based, in part, on his assertion that beliefs do not produce action on their own, but only in combination with desire (Plantinga, Warrant and Proper Function 225). He then looks at the range of possible independent desires and beliefs that could combine to produce behavior (Plantinga, Warranted Christian Belief 235). This combinatorial approach is analogous to considering the range of possible x’s and y’s that, when added together, would produce the number “4.” We could get this with 1+3, 2+2, 6+ (-2), etc. Given such a limitless range of possible belief-desire input combinations that could produce a particular behavior, it certainly seems that the (Vast) majority of beliefs would be false on statistical grounds alone. With our cost-based understanding of evolution, however, we can uncover the problem with Plantinga’s approach: It amounts to an idealized characterization of the evolutionary problem, one which abstracts away the all-important cost component. William James’ observation that we are sometimes forced to make epistemic decisions on insufficient evidence nicely (if ironically) captures the issue at hand (James).

As we have seen, Plantinga’s is a reductio challenge: he is assuming a naturalist framework in order to undermine it. However, if we assume such a framework, then belief content cannot be treated as the free-floating costless mental phenomenon that dualists seem wont to imagine. Rather, from a naturalist perspective, belief content must be seen as entirely determined by physical brain structures and activity. Accordingly, we have to ask what it would take for evolution to select for a cognitive system that consistently produces beliefs uncorrelated or negatively correlated with truth even as it consistently produces adaptive behavior.
In addressing the negatively correlated case, Griffiths and Wilkins argue that a second truth-tracking cognitive mechanism would be necessary in order to make the adjustments necessary to ensure any emerging belief remains false, even as the evolving behavior becomes increasingly adaptive (Griffiths and Wilkins 6–7). This would certainly result in a more complex neurological design. In particular, it would require more energy in much the same way that a more complex, inefficiently written computer program would require more electrical energy than would a less complex, more efficient one—for the same output. Such a cognitive design would, in short, be more costly relative to the simpler, more direct, belief-truth-tracking approach.

Their point is very well taken, though it seems to focus on the case of maintaining a deceptive cognitive faculty, rather than one that is simply orthogonal to truth. That is, their second cognitive mechanism would be needed to keep beliefs false, as if falsity were the objective. It seems that a more general case for linking belief content to behavior can be made by incorporating and combining the concepts of predictive success and evolutionary gradualism.

The more truth-tracking belief content is, the greater the number of diverse situations in which it would produce behaviors with positive payoffs, and the fewer with negative payoffs. In other words, better belief-content truth-tracking would provide a higher behavioral “return on investment,” where the “investment” is the energy-cost of creating and maintaining the necessary cognitive structural complexity associated with creating and sustaining that belief. Note that adding Plantinga’s belief-desire distinction leads to the same result since desire, like any other cognitive state, incurs costs to build
and maintain. Therefore, the same complexity-reducing and applicability-expanding pressures would line desires up with true belief in precisely the same way.

Significantly, this applicability maximizing concept mirrors one of the attributes of a good scientific theory. Among other virtues (which we will examine below) good scientific theories are more parsimonious, that is, they explain more with less. They are, in other words, *simpler* relative to the range of phenomena they explain. The best theories have the highest predictive success with the simplest theoretical structure precisely because this feature reflects their having avoided encumbering themselves with ad hoc hypotheses when faced with findings that violate their predictions. In the limit, adding unrelated ad hoc hypotheses would eliminate a theory’s predictive power while dissolving into little more than a detailed *description* of the phenomena it is intended to explain.36

While it is certainly easy to imagine a false belief that would fortuitously produce survival-enhancing behavior in a particular situation, it is far less easy to imagine the *step-wise, cost-sensitive* evolutionary path that could plausibly create and maintain such a situation. One of Plantinga’s own examples is that of a person who *wants* to be eaten by a tiger, but who also happens to believe either that running away is the best way to do this, or that by running away he is really running toward it (Plantinga, *Warrant and Proper Function* 225). These are, of course, *logical* possibilities, but from the standpoint of (cost-constrained) evolutionary theory, it would be difficult to describe a step-wise

36 We will return to this point later.
selection path whereby such a situation would not only gradually emerge (remember, *each* step has to confer advantage—to “pay its way”), but also be maintained. How, for example, should we imagine this individual responding to another large predator, one that she had never before encountered? Does this person systematically reverse “running away” with “running toward” in *all* situations, or is it a situation-specific concept, such as only when tigers are involved? In the latter case we have ever more situation-specific rules and therefore far less parsimony (more cognitive complexity to manage), and in the former we just have a problem with our use of language, since it seems belief is, in fact, tracking truth after all.

As the preceding discussion shows, the problem with a collection of situation-specific false beliefs is not only that it requires a more energy-intensive structure by virtue of its necessarily higher neurochemical complexity. It is also difficult to see how it would provide a competitive advantage over beliefs that happen to be even a little more truth-tracking—that is, beliefs that align more (even marginally) with real-world ontological categories like Predators, Persons, etc., and thereby enhance survival in a far *wider* range of *novel* situations, for the same complexity cost (i.e., “bang for the buck”).

Evan Fales reaches a similar conclusion but from the standpoints of deductive inference and enumerative induction (Fales 442–43). He points out that the conclusions of a deductive argument have random truth-values if the premises are false. From an evolutionary perspective, this means that false premises would make the development of *generalized* algorithms, indeed our very ability to reason, impossible. Why? Because, once again, there would be no plausible step-wise natural selection path to get there.
There could be no series of *cumulative* steps, each with increasing adaptive advantage, since there would be no “getting warmer” and “getting colder” feedback along the way. Only by tracking progress toward or away from some real-world truth could such evolution-driving feedback occur. Note that this argument is consistent with the distinction between belief content and behavior: In order for a conscious, *deliberative* algorithm to evolve, it is *belief content* in the form of true premises that would have to track truth, particularly in light of our earlier cost-sensitive, complexity-reducing discussion.

Fales also points out that we reach the same conclusion even if we consider only the enumerative implications. As a variation on Plantinga’s Tiger example, Fales uses the example of “Freddy,” which I have modified to more closely mirror the present argument (Fales 443). Freddy believes that a heavy rock is lighter than water (false), and that one can kill something by hitting it with something light (false). Freddy encounters a venomous snake, and, acting on his false beliefs, successfully kills it, happily surviving the encounter and keeping the odds of his genes’ getting into the next generation above zero—for now. Later, however, he falls into the sea and, acting on his beliefs, straps the stone to his body as a floatation device, which rather quickly reduces those reproductive odds to zero. Those holding *beliefs* closer to truth (cost adjusted) would be genetically rewarded by having a better chance of having their genes remain in the gene pool, since those *beliefs* would produce more adaptive *behaviors* in more diverse situations for the same cost (that is, the cost of maintaining the neurochemical structures associated with
that belief) compared with those holding beliefs that were less close to truth, that less accurately track reality’s ontological categories.

This point becomes even clearer when one remembers that it is not individuals that evolve, but populations based on very large numbers of interactions with the environment and with each other in many different ways over many generations. A fortuitous combination of false beliefs—whether generated non-inferentially from perception, or inferentially from a cognitive mechanism—might help in one or even a few interactions. However, over generations, across the population, and across widely disparate situations, the relatively high “cost” of producing and/or maintaining non-truth-tracking beliefs would quickly manifest itself as they are repeatedly tested against the complexities and regularities that characterize reality. This is not very different from what one observes in history when a flawed scientific model that seemed to work in a few situations is applied to ever more diverse situations over time. Importantly, one need not assume the kind of wildly false beliefs Plantinga uses in his example in order to feel the culling effects of evolutionary pressure. Indeed, beliefs do not need to be completely false at all, merely less true than they could be (for the cost). In such a situation, evolutionary pressure for improvement comes from competition with other beliefs that are better performing—even marginally—thereby driving beliefs closer to those that do a better job of carving reality at its joints, of getting the maximum payoff from the minimum cost.

To the logical critiques of the EAAN that we had earlier examined, Griffiths and Wilkins have added a powerful conceptual critique: the EAAN gets evolution wrong.
when it argues that belief content is detached from behavior and is therefore invisible to the forces of evolutionary selection. We expanded their argument to help show how cost considerations ensure that fitness-tracking would not only track as much “signal” accuracy as it can afford, but also belief content accuracy. Fales bolstered this further by pointing out that not only basic beliefs, but also our very ability to deduce, to reason, could never have evolved were the belief-inputs to those evolved algorithms untethered to truth-tracking. We have to conclude that belief content, no less than behavior, is very much a target of active, evolutionary selection.

So far, the preceding discussions have made both a negative and a positive case: On the negative side, we have attempted to refute the EAAN, but even if we were successful, this would not mean our beliefs become justified by default. With the help of Griffiths, Wilkins, Fales, and others, we made some progress toward a positive case for the reliability of our cognitive faculties and many of the beliefs these faculties produce. Thus far, however, the case seems strongest for BFFs that are directly subject to selective pressures, for example, those related to finding food and avoiding predators. But what of our more rarified scientific and philosophical concepts? Griffith and Wilkins make a crucial indirect argument in this direction when they point out that the history of science makes clear that our evolved commonsense faculties have reached a point where they can identify their own shortcomings, and can then bootstrap from there (Griffiths and Wilkins 12). Fales, as we have seen, also makes important progress by helping us to see a more direct linkage between our commonsense beliefs and those insights that derive
from our more general reasoning capabilities. Yet it seems we can do a bit more to strengthen this “bridging” effort.

Humean Bootstrapping

The EAAN argues that under naturalism and evolution none of our beliefs is reliable (assuming unpurposive evolution). Even if our preceding discussion has refuted this, it seems we have shown only that some of our beliefs are reliable, in particular, those that are directly related to survival in our ancestral environments. But then why should we have any trust in the enormous and growing edifice of human knowledge that did not even exist for most our evolutionary history, from quantum mechanics to aerospace engineering? In other words, even if fitness-tracking is belief-truth-tracking, it seems the cost-constrained, optimizing nature of evolution would produce at best a kind of “tunnel vision” into truth. Natural selection, after all, “sees” only a population’s local, immediate environment, implying that truths that it tracks may only be narrow, “shallow truths” (Vlerick and Broadbent 194). And even then, there is still the worry that the ancestral environments that produced the associated BFFs have long since changed.

So, even if we can use our understanding of evolution to justify some of our basic cognitive faculties—our “intuitive ontologies” (Vlerick and Broadbent 199) and cost-based rules of thumb—we still need to climb from this small universe of the commonsense to the burgeoning universe of modern human knowledge. How do we build such a justificatory bridge? Some, like Griffiths and Wilkins have argued that even our most advanced theories still have to stand up to the scrutiny of commonsense, but this overlooks Bergmann’s earlier point that these same commonsense faculties often actively
obstruct our efforts to grasp broader, deeper truths, particularly when those truths contradict our faculties’ basic-belief outputs. The result is that many of the now well-established “truths” of science were in fact mercilessly ridiculed when first introduced and even now continue to seem utterly counterintuitive. Consider, for example, the outrage to commonsense posed by the now extremely well-confirmed theory of quantum mechanics (e.g., a particle’s being in two places at once). Indeed, natural selection itself is not much more intuitive. As Paul Bloom put it, “we might intellectually grasp it, with considerable effort, but it will never feel right to us” (Bloom 122).

As we will see, Vlerick and Broadbent’s “Humean bootstrapping” approach not only directly addresses this worry, but also suggests a more general strategy that bears directly on the thesis of the present work. Their insight is to recognize that the cognitive reliability question should be approached from within, as it were, rather than admitting of a single, general, outside-looking-in answer that provides justification for all of our beliefs in one fell swoop. At the risk of reading into their argument, theirs can be thought of as an algorithmic approach.

The problem, which echoes Plantinga’s objection, is that there is an epistemic circularity in using our cognitive faculties to self-certify. But does that mean we cannot use these faculties in an introspective, procedural fashion? Even if they cannot self-certify, can they self-correct? While our evolved commonsense intuitions often create obstacles to grasping more scientifically accurate yet deeply counterintuitive truths, the fact remains that we have overcome them, as the progress of science demonstrates. How is this possible? If, for the sake of example, we agree with the CSR Standard Model’s
computational, modularized view of human cognition, then recall that one of these modules provides a means by which the other modules can share data. For example, our compulsive usage of analogue reasoning and metaphor appears to be a case of our using cognitive skills from one specialty area to reason about information ordinarily handled by another. This suggests that humans can override the deliverances of individual modules’ intuitive ontologies in a way that other species cannot. The idea is that humans have an evolved-in set of values (e.g., coherence, predictive accuracy, etc.) aimed at a similarly evolved-in, truth-seeking goal. These combine to create an epistemic drive, which appears as innate to humans as any of the other, more primitive universal drives that we do share with other animals (Vlerick and Broadbent 199).

This epistemic goal and its associated set of values, combined with our cross-module reasoning abilities, provide the means by which humans’ cognitive processes are able to “debug themselves” (qtd. in Vlerick and Broadbent 200). Importantly, Vlerick and Broadbent’s approach does not depend on the truth of the Standard Model’s modular view of human cognition, or, indeed, on any particular evolutionary story of how our cognitive abilities arose. Rather, it illustrates an approach that echoes Daniel Dennett’s description of bootstrapping in a different context:

[Y]ou can use your existing, imperfect, ill-understood methods of inquiry to refine those very methods, pitting good ideas against better ideas, and using your current sense of what counts for a good idea as your temporary, defeasible guide to improvement. . . . It isn't logically guaranteed to work, but so what? It is much more likely to work than
flipping a coin, and the odds get better over time” [emphasis in original] (Dennett 368).

Note the critically important “over time” phrase. Dennett uses a particularly apt analogy, which captures not only the algorithmic nature of bootstrapping, but also (I argue) something of particular relevance to the question of how one can establish cognitive reliability in a non-question-begging way, a point that will affect my particular variation on his analogy:

If someone moves to a new country and knows no one at all, how can they ever hope to establish a circle of trusted friends? One could simply declare some arbitrarily chosen group of people to be trustworthy by fiat, but it would be difficult to consider this assertion justified. Of course, these people might really be trustworthy, since, as we saw earlier, justification and objective truth can vary independently. There is another approach: One could think of this not as a problem that requires a single-step solution, but rather an algorithmic, multi-step solution, in which the results of earlier steps guide the later ones. The process starts without prior knowledge of where it will lead. (This is true of evolution’s own cumulative, multi-step, foresight-less selection process.) For example, one could choose the initial circle of friends by literally guessing, perhaps through random encounters in the town square. But this is just the entry point into the process—the first step of the algorithm—not the whole of the solution. With this randomly chosen group of people in hand, one can then gain some initial experience with them, dropping the least trustworthy and relying on those remaining to make introductions for the next round of meeting new people, people who are now at least
marginally more likely to be trustworthy. Importantly, this describes just one iteration. The idea is that this process continues, iteration by iteration, with later iterations building upon the results of earlier, less-informed ones until one has built a network of friends who can far more justifiably be considered trustworthy—despite the utterly random nature of the initial group.

While this may seem commonsensical, it is a rich analogy for our purposes. It can, for example, answer Plantinga’s global skepticism worry. Recall that Plantinga argued that it is self-refuting to believe, simultaneously, in naturalism, evolution, and the reliability of our cognitive faculties. To avoid the bog of radical skepticism he looked for his solution outside of naturalism. But note that his approach assumes the need for a single-step solution to the problem of justifying trust in our cognitive faculties (e.g., “God designed it.”). Vlerick and Broadbent, on the other hand, solve this problem not with a single-step solution, but rather a multi-step, algorithmic, cumulative justificatory process—a “bootstrapping” process. Indeed, it is precisely this foresightless, yet cumulatively improving algorithmic process that is at the very heart of biological evolution generally and its ability to create extremely complex examples of apparent design.

Understood this way, our ability to bootstrap across a range of inputs and cognitive domains would itself become a target of selection. If we add to this both our cost-sensitivity considerations and Fales’ point about the necessity of true premises in the development of deliberative algorithms, we seem to have a very clear, step-wise, selection path through which reliable, general reasoning abilities could emerge—the kind
of general reasoning that would justify our confidence in our ability to acquire scientific and philosophical knowledge that is actually counter-commonsense.

Summary

At this point we have come a considerable way in rebutting some of the key objections to the atheist-CSR position. Before exploring what force remains in those original objections, we should take stock of where the discussion now stands.

We have argued that Plantinga’s case for the proper basicality of theism fails. Independent of that outcome we argued that his notion of warrant fails to shift the burden of proof for God’s existence onto the skeptic. On a third independent track, we dealt with a number of particularly powerful, recurrent criticisms of the atheist-CSR position, including the charges of question-begging, and committing the genetic and sole-cause fallacies. We also criticized the EAAN from a number of perspectives: logically, by showing that cognitive reliability could be treated as a basic belief using Plantinga’s own criteria; and conceptually, in that it gets evolution fundamentally wrong in its overlooking the crucial role of cost. Finally, we made a positive case for the reliability not only of our commonsense beliefs, but also of our reasoning abilities generally, which gives us some reason to think that we can confer reliability onto our more derivative and often counterintuitive beliefs. This case was built on the general algorithmic idea of “bootstrapping,” which, as we will see, has even broader applicability.
While this recent literature has helped us reduce the impact of the original objections to atheist-CSR, considerable force still remains at a more general level. We have not, for example, addressed the widely held idea (even in mainstream scientific circles) that religion and science operate in essentially non-intersecting spheres. This is the view that science generally, not just CSR, simply has nothing to say on the matter of religious claims and vice versa. From an epistemological standpoint, it is the claim that religious, even supernatural beliefs generally, are simply beyond the scope of scientific methodology and its rules for determining what counts as “valid” knowledge. Rebutting this claim is, therefore, a necessary part of establishing a connection between CSR and the justification of theistic belief. We also need to address an even more general issue, namely, a pervasive skepticism regarding science’s ability to convincingly prove or disprove anything—a skepticism often used to create a logical space for theistic claims. Finally, while we may have defeated Plantinga’s truth-conditional, warrant-based attempt to shift the burden of proof, there may be other reasons to suppose that the burden still lies with the atheist. As a result, we need to make a strong positive case that the burden of proof does indeed lie with the theist.

These issues will be addressed in parallel along with the development and justification of a closely related, mutually reinforcing set of epistemological tools and methodologies. This development will draw from (1) the bootstrapping concept as
applied to certain normative epistemological principles; (2) a Bayesian and pragmatics-based analysis of burden of proof; and crucially (3), some important ideas related to the Duhem-Quine thesis, especially those related to hypothesis bundling. If this multi-pronged approach is successful, we will have justified the use of a set of methodological principles and analytical tools for assessing the impact of observational evidence on the rational justification of beliefs generally—natural and supernatural. We will then make just such an impact assessment relative to CSR’s findings. This will position us to finally address Thurow’s as-yet unanswered objection that, however well-supported CSR may be, it fails to address the processes people actually use to justify their belief.

NOMA

The topic of this paper—and any other work that attempts to use scientific findings or methodologies to examine the rational justification of religious claims—is overshadowed by the perennial question, “Is science at all relevant to religious or other supernatural questions?” A full exploration of this historically rich literature is beyond the scope of the present work, but a sufficiently strong response can be made nonetheless. It is well worth noting at the outset that the non-relevance-of-science-to-religion view is held not only by theists, or even by scientists outside of the mainstream. It is a position held by none other than the National Academy of Sciences, among other leading institutions. In one of their publications, they state that “all of science, is necessarily silent on religion and neither refutes nor supports the existence of a deity or deities” (qtd. in Fishman 815). This position is sometimes referred to as Non-Overlapping Magesteria or NOMA, a term famously introduced by Steven J. Gould (Gould). Despite this
prestigious institutional and academic support, more than a few mainstream scientists and philosophers disagree (Stirrat and Cornwell).

Yonatan Fishman provides a particularly helpful Bayesian framework within which he tries to assess the relevance of science to the investigation of supernatural claims, including religious beliefs. However, as we will see, his argument needs to be expanded in order to avoid begging some central questions, one of which is the validity of using Bayes’ Theorem on religious claims in the first place. The use of Bayes’ Theorem in other natural contexts can be defended in terms of how well it explains and integrates actual scientific practice (Fishman 816–17). While we will defend its relevance to supernatural claims in a later section, for now we will simply point to the fact that Bayes’ is used by Plantinga and other critics of the atheist-C SR position when making their own cases.\textsuperscript{37} A brief recap of Bayes’ Theorem is therefore in order.\textsuperscript{38}

A Brief Overview of Bayes’ Theorem

The following is one of the theorem’s common forms:

\[ P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\sim H)P(\sim H)} \]

Following Fishman, we will treat a supernatural claim, such as the existence of God, as the hypothesis \( H \). In other applications, \( H \) might be some scientific claim. Its

\textsuperscript{37} See, for example (Plantinga, \textit{Warranted Christian Belief} 230)

\textsuperscript{38} The following example assumes the relevance of Bayes’ to the “God hypothesis,” where, for convenience, “God” will be assumed to be the Abrahamic God. While the use of the God hypothesis in Bayes’ will be defended in a subsequent section, that defense will rely on some familiarity with Bayes’ Theorem. So, in the interest of clarity and brevity, it seems best to present the overview here, with the God hypothesis included for illustrative purposes.
**prior** probability is expressed as \( P(H) \), which is based on all the other background information we have, excluding the hypothesis \( H \) itself and the probability-impacting effects of some new piece of evidence, \( E \), relevant to \( H \). In other words, the prior probability is our “going in” assessment based on everything we knew before the new evidence is considered. Importantly, this new relevant evidence can be confirmatory or disconfirmatory of \( H \). The left-hand side of the expression, \( P(H|E) \), is called the “posterior” probability of \( H \), which is the revised probability of \( H \) given (“|”) the new evidence \( E \)—it is, in other words, posterior to (after) consideration of the new evidence.\(^{39}\)

With these building-block concepts and notations in hand, we can now explore the remaining terms on the right-hand side of the expression. \( P(E|H) \) is the probability that the new evidence would be observed if the hypothesis were true (i.e., given \( H \)). This term can be thought of as related to \( H \)’s predictive power, a means by which we can test \( H \)—for example, “Since \( E \) is predicted or entailed by \( H \), do we find \( E \)?” Of course, it is not quite so simple. Even if we do find \( E \), it may mean little until we consider the effects of the other terms in our Bayesian expression, in particular, the terms containing \( \sim H \).

If \( P(H) \) is the prior probability that the hypothesis is true (some value between zero and one), then the prior probability that \( H \) is **false**, \( \sim H \), would be expressed as \( P(\sim H) \). \( P(E|\sim H) \), then, is the probability that the new evidence would be found if \( H \) were **false**.

The importance of the \( P(E|\sim H) \) term becomes clear with a simple example: Suppose my

---

\(^{39}\) So interpreted, this might more properly be considered a statement of Bayesian Confirmation Theory, the difference being the assumed applicability of some standards of rationality. As a more thorough exploration of the literature on the derivations, distinctions, and controversies surrounding this topic is beyond the scope of the present work, we will continue to simply use the term “Bayes’ Theorem.” For a fuller treatment see (Talbott).
hypothesis is that little green fairies are the sole, proximate cause of the orbits and rotations of the planets and they intend to keep things going as they have been for at least another hundred years. So, H entails that the sun will come up tomorrow, since, ex hypothesi, fairies will be causing this for some time. Of course, observing the sunrise tomorrow is hardly cause for banner headlines about confirmation of fairies, since we fully expected this E for other reasons that have nothing to do with H. So, while P(E|H) is high, so is P(E|~H). In fact, in this particular case they are the same, which effectively cancels out the contribution of E, resulting in the posterior probability P(H|E) being the same as its prior probability P(H). More generally however, you can think of the fact that since P(E|~H) is in the denominator of our Bayes’ expression, a larger value can only put downward pressure on the whole right-hand side of the expression and therefore on the posterior probability P(H|E). In other words, a higher P(E|~H) weakens E’s contribution to the likelihood that H is true.

There is another ~H term, namely P(~H), which is multiplied by P(E|~H). The product of these helps formalize another of our intuitions, which is that the more extraordinary a claim is, the more extraordinary the evidence for it needs to be. To see this recall that P(~H) refers to the prior probability of ~H, in other words, the prior, “going in” probability that H is false based on all of our background knowledge but before considering the new evidence E. Let’s put some numbers to this to help bolster our intuition. To say a claim is extraordinary is to say that, based only on our going in background knowledge, it is very unlikely to be true. For example, let’s say P(H) is just 1% or 0.01, which is the same as saying that P(~H) is 0.99. Let’s also assume that this
hypothesis has the virtue of making not just a testable prediction, but the kind of unexpected prediction that can dramatically affect the likelihood of H. To keep it simple, let’s say that if H really is true, then the likelihood of finding E is fully 100% (1.00) but—and this is the critical measure of “unexpectedness”—if H is false, then the likelihood of finding E is only 0.01. In this situation, if we were to actually find the predicted evidence, then the posterior probability of H given E would jump all the way up to 0.50—an impressive jump given that before finding this evidence it was a mere 0.01. Still, it’s hardly overwhelming support. What if instead, H is not quite so outlandish? Instead of the long-shot 0.01 for H’s prior probability, what if it were the still small, but not quite as “crazy,” 0.15? In this situation, the posterior probability for H given the same evidence becomes 0.95. That’s quite a change from the 0.50. Keep in mind that the only difference between the two cases is H’s prior probability. We haven’t changed the nature or relevance of the evidence. In other words, the same evidence doesn’t produce the same posterior probability. This is because the “crazier” H is—that is, the lower its prior probability—the stronger the evidence needs to be to offset that craziness in order to get to the same level of posterior probability. The smaller P(E|~H), the greater the impact of that evidence. Think of it this way: if E has a near-zero chance of being found in a world where H is false, then finding E is obviously a big deal.

Let’s make the preceding discussion a bit more tangible by looking at two examples. Fishman points to the real-world case of experimental control groups, such as those given placebos as part of determining the effectiveness of a certain drug. Even if everyone who takes the real drug reports a benefit, celebrations must still wait until we
see what the placebo-taking control group reports. Think of “I feel better” as the evidence E. This is encouraging to hear from those taking the real drug; however, if we also get this same evidence from members of the placebo-taking control group, then P(E|~H) goes up. As we have seen, this puts downward pressure on the posterior probability of H given E—that is, on the likelihood that this drug is really helping. Of course, as we have just seen, the P(~H) term plays a significant role as well. If all our background knowledge going into the experiment suggested that this drug would be highly unlikely to make people feel better, then the evidence would have to be that much stronger to get to same level of confidence in the drug.

To further bolster our intuition of priors, imagine a complete stranger telling you he has three, twenty-dollar bills in his pocket. What attitude should you take toward this claim? The claim is not only logically possible; there is nothing in our background knowledge to suggest his claim would be an unsurprising scenario, no “going in” reason to suppose that it is more likely than not. Since this is a question of prior probability—that is, without new evidence not already part of our background knowledge—it seems that “agnosticism” would be the appropriate attitude. What if, instead, this stranger claimed to be an extraterrestrial with physical strength many times that of the strongest human? Like the first claim, this one is also logically possible (i.e., it is not incoherent; if it were, the prior would just be “0”). However, limited again only to our background knowledge, it seems we should assess the prior probability of this claim as being much lower than that of the first claim, despite the fact that both are logically possible. To be sure, saying the prior is low for the alien hypothesis, or even for God, is not the same
thing as saying it is false. Indeed, there may be new evidence that more than compensates for this low prior possibly, thereby raising the posterior probability—conceivably to near certainty.

Fishman suggests that such Bayesian considerations are particularly relevant to the scientific evaluation and assignment of burden of proof to supernatural claims. His argument, however, depends on the way he interprets the significance of a successful natural explanation of some purported supernatural phenomenon. In particular, he assumes that a natural explanation can be counted as part of \( \neg H \); that is, it can be counted as part of “any mutually exclusive set of alternative hypotheses” (Fishman 816). But is this a valid assumption? Based on our earlier discussion of levels-of-explanation, it would seem to be one in need of defense. Our discussion thus far has laid the groundwork for just such a defense. First, however, we need to address a critically important general point, which relates not only to Fishman’s assumption that natural explanations exclude God-based ones, but also to the central argument of this paper.

**On Recognizing a Good Theory**

One can *always* introduce some new claim to make God’s existence consistent with a natural explanation. With the contribution of such “helper” hypotheses, God can simply be cast as an ultimate explanation that complements any newly discovered natural explanation.\(^{40}\) Implicit in such an approach is the assumption that the burden rests with

\(^{40}\) Fishman shares the interesting fact that Benjamin Franklin’s lighting rod had been criticized as an effort to interfere with God’s Will (Fishman 822). This reminds us that in the (not too distant) past, God was very much considered the *proximate* cause of most unexplained phenomena.
the skeptic to show that any such “tossed in” additional hypotheses are false. Before we can move forward, we need to challenge this assumption.

The old adage that “absence of evidence is not evidence of absence” is particularly relevant here. Consider: Do we believe everything that we cannot disprove? Obviously we do not. So, despite the fact that there may be no evidence of absence for a particular claim (e.g., there is a teapot in the very center of the moon41), we nonetheless justifiably withhold belief, and in many cases even justifiably disbelieve the claim. So, while the old adage is often true, it is by no means always true. Once again, our Bayesian framework can help clarify things. Recall our fairies “theory,” which includes the prediction that the sun will come up tomorrow. The problem, as we saw, is that this evidence is no less expected if that theory is false. However, if our theory predicts (or entails) something that we would not expect to find, then finding it would provide at least some confirmatory support to the theory. This once again captures our intuitive sense that a good theory’s predictions should be unexpected: The evidence should be there if H is true, but absent if it is not. Recall our discussion of the role of P(E|~H) in this regard.

A number of traditional atheistic arguments can be cast in these terms. For example, some have argued that the presence of natural evil (e.g., crippling parasites, tsunamis, etc.) is unexpected under the omnibenevolent God hypothesis, HG, and is expected if ~HG. Finding such evil, they claim, amounts to disconfirmatory evidence that an omnibenevolent God exists. There is a rich literature on the Argument from Evil, but

41 This is a slight modification to Bertrand Russell’s Teapot in order to take into account any recent advances in technology.
suffice it to say that rebuttals to it often challenge the claim that the absence of natural evil is a prediction at all. Such rebuttals can be thought of as follows: While the Omnibenevolent God Hypothesis, \( H_G \), *on its own* predicts the absence of natural evil, \( H_G + H_A \), where \( H_A \) is some auxiliary hypothesis, does *not*. For example, if \( H_A \) is “the world-corrupting effects of Original Sin,” then we might well *expect* to find natural evil.

**Duhem-Quine and Hypothesis Bundling**

This idea of adding some auxiliary, “helper” hypothesis seems to present a general problem, one famously captured in the Duhem-Quine Thesis. Roughly, this says that no evidence can ever refute a theory, unless one can prove that no theory-saving auxiliary hypothesis exists. Some \( A \) can always be invented to save it. Philip Kitcher describes philosophers’ efforts to develop falsifiability criteria without accounting for Duhem-Quine as having resulted in what he calls “naïve falsificationism” (Kitcher 42).

As Kitcher points out, however, no theory is tested in isolation. For example, if one is testing some theory’s prediction that a particular microorganism will be present in some sample, then simply looking through a microscope to confirm its presence would indeed be the kind of test we are looking for. However, finding the microorganism using this test would “confirm” our theory only if we assume the truth of many other theories—some big, some small—such as optics theory, the proper functioning of that particular microscope, and the validity of numerous other higher-level physical, biological, and chemical theories, *themselves* confirmable only as members of their own crowded hypothesis bundles. Importantly, failing to find the microorganism would *not* refute the
new hypothesis; it would refute the *conjunction* of the target hypothesis with all the other hypotheses bundled with it.

How skeptical should this make us? Based on the historical progress of science and technology, not very. Over time we have generally been able—somehow—to “separate the wheat from the chaff” with regard to good and bad theories. Our methods evidently work well enough despite the fact that we are not always clear why they do. To help us get some insight into why they work, Kitcher builds on Duhem’s observation that “Hypotheses are tested in bundles” (qtd. in Kitcher 44). Kitcher’s critical insight is twofold: “[1] While hypotheses are always tested in bundles, they can be tested in *different* bundles. [2] An auxiliary hypothesis ought to be testable independently of the particular problem it is introduced to solve, independently of the theory it is designed to save” [emphasis in original] (Kitcher 46).

The significance of these two points cannot be overstated. Consider the long and ongoing story of human knowledge accumulation as both an integrated whole and an ongoing process. From this perspective it is clear that not only *can* we test in different bundles; we have been doing precisely this in ever new bundle assemblages, over and over again, in many different, shifting contexts within and between all investigatory fields—whether we realize it or not. In other words, tested bundles are constantly being reassembled from *different* pieces of *other* tested bundles. This is done not by explicit design but as a natural consequence of the hypothesis testing process itself. This means that humanity’s ever-growing body of factual knowledge is built up not only from
explicit testing, but also from the implicit contribution such tests make to the confirmation of the many other hypotheses bundled with it.

The problem is this: When a hypothesis bundle is confirmed, how should one allocate that confirmation across the hypotheses bundled with it? This is clearly a problem if there is just one set of hypothesis bundle. But what if I can create many bundles, each containing the target hypothesis bundled with different sets of other hypotheses? Returning to the early example for a moment: When that microscope is used in other experiments, the assumption (hypothesis) that it is not defective is now made a part of different hypothesis bundle, which is being confirmed by different evidence.

The essential insight is that an hypothesis is never tested in a vacuum, but only as one node in a large and evolving web of investigative activities spanning the full gamut of human inquiry, where each tested hypothesis bundle is inevitably composed of varying combinations of other hypotheses, themselves bundles of bundles. These ever-recombining bundles are themselves confirmed by different pieces of evidence. This picture describes a dynamic, deeply nested and thickly interconnected web of interrelated hypotheses, one that is self-correcting both in its massively parallel and multi-faceted contact with real-word data, and also in its ongoing need to maintain internal coherence. There is a powerful kind of bootstrapping effect here: As these testing activities continue, they mutually reinforce, iteratively improving the confidence of each hypothesis by increasing the number of different successfully tested bundles of which it is a member. The probability that a particular hypothesis is wrong shrinks as the number and variety of
bundle “teams” to which it belongs are successfully tested. The net effect is that over time we create an ever-wider and ever-deeper web of ever-more confirmed hypotheses.

Kitcher’s second point regarding the independent testability of auxiliary hypotheses is equally crucial. As a case in point, he examines the story of Neptune’s discovery. Newton’s celestial mechanics (part of classical physics) is comprised of a compact, tightly integrated set of hypotheses (i.e., it exhibits parsimony—more on this shortly), which together create a problem-solving strategy that can predict a wide range of often unexpected phenomena. Based on the observational data known at the time, the orbit of Uranus was predicted to follow a certain path. This prediction was, however, falsified. What to do? Throwing out Newton’s celestial mechanics at that point seemed unreasonable, particularly given its impressive record of success otherwise.

One solution to such a situation, as we have seen, is to add an auxiliary hypothesis, which would save the theory by directly accounting for this unexpected result. We could do this by adding a one-off equation that is invoked only when dealing with Uranus. Fortunately, in this case, a different approach was taken: the existence of an as-yet undiscovered planet was hypothesized, one that had the properties necessary to make celestial mechanics consistent with the observed behavior of Uranus. Why should this auxiliary hypothesis be taken seriously? First, it could be tested independently of the theory it is trying to save: just look through a telescope; no question-begging assumptions necessary.

Why was this better than adding the one-off Uranus-only equation? Because doing so would encumber the original, compact, integrated set of hypotheses with
something entirely ad hoc, that is, something unrelated to the existing problem-solving strategy. The new-planet hypothesis, on the other hand, preserves the theory’s parsimony while accounting for the new data.\textsuperscript{42} In other words, it maximized “bang for the buck.”

**Good Theory Virtues**

This Neptune example provides some key initial insights into how we might begin separating good theories from not-so-good theories. As Maarten Boudry et al. remind us, however, there are no simple rules for distinguishing science from non-science/pseudoscience that would work in all cases (Boudry et al.). Nonetheless, there are some clear cases, such as the preceding one, and from which we can make some initial, tentative inferences. We can then “bootstrap” from there as we explore and test other cases, improving our inferences as we go.

Some initial good-theory “virtues,” which we can infer from the Neptune Example,\textsuperscript{43} include its being *unified* in the sense of being composed of a tightly integrated set of hypotheses rather than some hodgepodge of unrelated hypotheses (e.g., a few equations that predict and explain almost all the data, rather than one equation specifically for Neptune, and another for another observation, etc.). This can be thought

\textsuperscript{42} In this particular example, nothing needed to be added to the theory at all. Rather, the independent confirmation of the hypothesized unexpected planet just added to the theory’s record of predictive success and, therefore, to its probability of being a correct theory. Other situations, however, might justifiably require an actual addition to the theory, though its value would, under this parsimony-optimizing guideline, have to “pay its way”—that is, pay for the *complexity* it adds to the theory by explaining much *more* than just the one aberrant data point. This typically means that it integrates with the theory’s existing hypothesis set, rather than just being some disconnected, “but if it’s Tuesday” rule. In the limit, adding ad hoc hypotheses that don’t “pay their way” eventually dissolve a theory into a mere *description* of the phenomena the theory purports to explain.

\textsuperscript{43} The following virtues are taken from (Kitcher), though the “fairy theory” example is my own effort at an illustration of his insights.
of as the degree to which a theory can “explain the most with the least.” The other virtue is independent testability. It must be not just testable, but testable independently of the problem being addressed. As an example of this last point, consider any number of conspiracy theories that claim sinister design rather than randomness lurks behind some unlikely outcome. Any evidence suggesting otherwise immediately generates ad hoc hypotheses that such evidence is part of the conspiracy. Yet these cannot be tested without assuming the truth of the conspiracy. Again, this does not mean that the theory is necessarily false, but it does undermine its justification. Other examples might include the “theory” of the Power of Prayer: many examples of confirmation are pointed to, but all examples of disconfirmation are handled by auxiliary hypotheses, none of which are independently testable.

Good theories should also be fecund: they should suggest new areas of inquiry, which are themselves confirmed. This is because the closer a theory is to accurately describing a real-world process or phenomenon, to “carving reality at its joints,” the more likely it is to have broader applicability and thereby raise fertile new questions. Finally, as we have already discussed, good theories should exhibit predictive success, in the sense of making unexpected, yet confirmed predictions, a feature which is related to its being falsifiable in our “non-naïve” sense.44

44 A particularly good example of the effect of a theory’s getting closer to real-world categories and processes is what happens when you create a taxonomy of animal life based, say, on just a creature’s size. You can do this—indeed, you can use any arbitrarily chosen feature—but the resulting taxonomy would tell you only about size; it would not give any additional insights or allow you to predict other unexpected relationships. On the other hand, if you base your taxonomy on homologies, then suddenly new insights blossom forth: the groupings turn out to make predictions that are confirmed by biogeography, genetics, fossils, etc.
Justification

Why should we accept the criteria we have identified above? It should be clear that they are not arbitrary criteria: their justification is the result of what turns out to have been a long-running, process rooted in the kind of bootstrapping we discussed. Whatever initial, even haphazard guesses we took in the early days of human inquiry, these criteria emerged and continue to be iteratively improved, in massively parallel fashion, across the gamut of human inquiry. These improvements include the shedding of methodological procedures that have not worked and the improvement of those that have. As with all inductive knowledge, such conclusions are, in principle, defeasible, but this is hardly a skeptical result since any alternative would have to be at least as effective as what it hopes to replace. And, since many independent lines of corroboration converge on the same principles—principles that have produced a vast edifice of corroborated knowledge—the odds of a fundamental reversal, while never zero, has in this case become astronomically small.

With this conceptual and justificatory base established, we can now explore some of the implications of these criteria. Among the suite of good theory virtues that Kitcher has outlined, we will focus here on two closely related ones, both of which are particularly relevant to our CSR-specific discussion: Parsimony and Burden of Proof.
Kitcher’s notion of a “unified” theory can also be recast in terms of parsimony (or Occam’s Razor). The idea is basically that a theory should have only as much complexity as it needs to correctly explain its range of phenomena. Anything more would be superfluous, anything less and the theory would simply fail. But what do we mean by “complexity”? There are at least two senses here: quantitative and qualitative (Pigliucci and Boudry). The former addresses the number of distinct entities in the theory (independent of type), and the latter addresses the number of distinct types, or ontological categories. It is in this latter sense that we will mean a theory is more or less parsimonious.

The benefit that type-parsimony provides to a theory can be captured directly in Bayesian terms. Recall that a claim’s prior probability directly contributes to its posterior probability (all else being equal). When probabilities less than one-hundred percent are multiplied, the result is a smaller value (e.g., if the prior probability of each of three independent propositions is ninety percent, then the probability that all three are true is ninety percent multiplied by itself three times, or approximately seventy-three percent). Therefore, the more independent entity types there are in a theory, the lower that theory’s prior probability, which means any new relevant evidence has that much more work to do to raise the posterior probability. Put another way, the more entity types there are, the more ways there are to be wrong.
Burden of Proof

This way of looking at the issue helps us use Bayes to better understand the nature of burden of proof. The lower the prior probability, the stronger any new evidence will have to be in order to overcome that lower starting point and give us a higher posterior probability. The implications of this become clearer when we think about how we populate our background knowledge, specifically, how hypotheses can go from being the subject of a posterior probability assessment, to being a member of our background knowledge—our priors. As Massimo Pigliucci et al. point out, when some theory becomes well established, it then becomes part of the background knowledge used in determining the priors in other contexts for other theories (Pigliucci and Boudry). For example, the Germ Theory of Disease is considered extremely well established. If a new theory were to come along that entails the falsity of that theory, the truth of which is now part of our background knowledge, then its prior probability would be correspondingly very low. Consequently, the evidence for this new, rival theory would need to be high enough to compensate. This resonates with our intuition about burden of proof: The lower the prior probability $P(H)$, the more reasonable it seems to expect the burden of proof to fall on the advocate of $H$ and not on the skeptic. Note, however, that under this pragmatic approach the burden can shift over time as evolving theories and new evidence affect our background knowledge in different ways. This raises an important general point.
“Fallacy Fork”

The Bayesian discussion thus far is not intended to suggest that there are any formal criteria on which we can neatly determine burden of proof in all cases. Indeed, the idea that one can determine formal rules for crisply identifying logical fallacies generally in real-world situations seems to run up against a “Fallacy Fork” (Boudry et al.): the more the rules reflect realistic situations, the less general their applicability, and vice versa.

This is an important point. While deciding where to assign burden of proof (BoP) can be fairly obvious in some situations, it is much less so in others. Pigliucci et al. point to two types of BoP, including prudential and evidential (Pigliucci and Boudry). Prudential BoP describes those situations where the relative costs of false-positive versus false-negative errors figure heavily in deciding BoP placement. For example, we would typically assume that some just-unearted WWII bomb is dangerous unless proven otherwise. Yet in a courtroom, we assume that a potentially dangerous criminal is not guilty unless proven otherwise. The difference is clearly prudential, or, alternatively, cost-based.45 Such considerations preclude precise, one-size-fits-all general criteria. Yet recognizing this does not mean that clear cases cannot be found, or that strong cases one way or the other cannot be made when those practical, situational considerations are carefully accounted for.

---

45 In the innocent-until-proven guilty example, the costs include those to the society as a whole, including the costs to that society’s freedoms were the burden otherwise.
Is Science Relevant to the Supernatural?

Now that we have laid the conceptual groundwork for justifying and using certain epistemological principles, such as burden of proof, parsimony, predictive success, and non-naïve falsification, we need to deliver on an earlier promissory note to defend the relevance of these to supernatural claims. First, however, we need to show that these natural-world-based methodologies do not “stack the deck” against the supernatural—that they do not presuppose naturalism.

The epistemological principles we have reviewed above have developed from our mundane, natural-world experiences. Even Plantinga draws analogies between sensing the Divine and sensing the everyday, such as our ordinary sense-perception of a tree (Plantinga, “Is Belief in God Properly Basic?” 44) Of course, none of this should be taken to mean that our epistemological principles force us to believe only in what we can see or directly observe. Science could hardly have progressed to its current state were this the case. The basic structure of the atom, for example, was demonstrated long before there was any hope of seeing one. This was possible because while we could not directly (aided or unaided) see the atom, its existence, given certain hypothesized properties, has observational consequences. Do supernatural phenomena or entities have observational consequences? Some certainly do, such as telekinesis and ESP. Fishman, for example, reminds us of recent tests of so-called “psi” powers (Fishman 824).

We can readily imagine results that would lend strong support to such claims. Similarly, we can imagine fairly unambiguous evidence for a sensus divinitatus, such as everyone’s being born with a clear disposition toward to, say, Christian beliefs—
especially among those born in cultures that had never heard of Christianity (recall our discussion of the power of predicting and finding unexpected results.). It seems reasonable to conclude, therefore, that supernatural entities and phenomena are not in principle outside the reach of confirmation using the methods we have discussed. As Fishman points out, the question of how to epistemologically demarcate between the natural and the supernatural is ill posed. The question is really just one of whether or not there are good reasons to believe whatever claim X happens to be, supernatural or otherwise (Fishman 830). To be sure, not all existence claims—natural or otherwise—can be evaluated. In order to be accessible to epistemic evaluation, such claims must have observable consequences (or predictions) that meet the following conditions: (1) they are testable independently of the theory that entails their existence; and (2) they must be publicly available—that is, they must be intersubjectively verifiable. There is nothing in any of these criteria to suggest that the deck is somehow being stacked against supernatural claims, or that their falsehood is being presupposed.

Does God’s existence have such publicly available observational consequences? Thinkers such as Plantinga believe that our brains embody something like Calvin’s sensus divinitatus, which would indeed produce publicly observable effects, such as our measurable predisposition to believe in supernatural agency. In fact, this hypothesized sensus can be thought of as a generator of a kind of religious experience: a non-inferred, basic belief on a par with the basic beliefs produced by ordinary, direct sense perception.

Michael Martin summarizes Swinburne’s useful religious experience classification scheme (Martin 155–56). One of its key distinctions is between religious
experiences of public objects and those of non-public objects, where “public” means that the presence of the object is intersubjectively verifiable under ordinary conditions, though not everyone would necessarily experience the object as a supernatural entity. The experiences of non-public objects lack this intersubjective verifiability: the object would not have been seen by anyone else, even by someone standing right next to the subject. This distinction reminds us that under any interpretation, the *sensus* is thought to be producing belief in a non-public object, regardless of whether it is producing a kind of *indirect* perception that is triggered by ordinary perception of public objects (e.g., beauty of flower), or producing direct perception.

If the *sensus* is triggered by the perception of public objects, it would amount to a kind of indirect sense of the divine, and we would have the same problem of a lack of agreement among epistemic peers that we discussed earlier. For example, we can all agree that a particular public object exists (e.g., a flower in full bloom), and we can agree on its many details, such as size, color, etc. However, as soon as we move beyond these intersubjectively verifiable observations and into the subjective experiences provoked by looking at it, confidence in any further agreement quickly evaporates. This situation is not unlike the wildly different experiences people have in response to a piece of art—despite the fact that they all agree on its directly observable properties, such as color, size, paint type, etc. However, in the case of the *sensus*, the experience is not simply one of aesthetic and emotional *states*, but of belief in the actual existence of some type of *entity*. Therefore, the question quickly turns to that of the epistemic reliability of the responsible BFF, which in this case is the supposed *sensus divinitatus*. However, as we
have seen, it is precisely this lack of agreement among epistemic peers that renders a BFF unreliable, and therefore not at all analogous to ordinary sense perception. On the other hand, if the sensus is producing a direct sense (however vaguely) of a non-public object, the epistemological methodologies we have discussed still apply provided that this non-public object produces consequences that are public, in much the same way that the unobservable atom produced observable, publicly available consequences. And, as we have seen, the theism embraced by mainstream Judeo-Christianity, and by Plantinga, clearly entails just such consequences.

It seems clear at this point that supernatural entities and phenomena are well within the reach of the methodologies we have been discussing and Plantinga’s Judeo-Christian God is no exception. Regardless of how one interprets the operation of the sensus, it seems clear that it is subject to the same epistemic principles as any other claim, supernatural or otherwise. These principles include all the tools of reason, including for example, conditional probability assessments and Bayes’ Theorem. However, even if we have shown that the same epistemological principles apply to both natural and supernatural claims, perhaps there is a more general objection, such as that embodied in the NOMA claim.

Do Natural Explanations Compete with Supernatural Explanations?

Earlier, we pointed out a key undefended assumption in Fishman’s Bayesian argument, namely, that natural explanations immediately become part of ~H, where H is the hypothesis that God exists. In other words, Fishman assumes that a natural
explanation is automatically a member of “any mutually exclusive set of alternative hypotheses” to a supernatural/God-based H (Fishman 816). We are now in a good position to address this. Let us begin by generalizing and consolidating the key points of our discussion so far:

Whenever a hypothesis, \( H_G \), is introduced in relation to another, \( H_N \), regardless of why it was introduced (e.g., as a competing hypothesis or as a complementary one), it falls into the \( \sim H_N \) set if it (1) logically contradicts \( H_N \), (2) makes predictions that contradict \( H_N \), or (3) does not “pay its way”; that is, it has elements that are explanatorily superfluous relative to \( H_N \).

This generalization captures other notions we have discussed, such as inference to the best explanation. For example, if two theories explain/predict the same data set, but one explains it with less—that is, one is more parsimonious than the other—then the other theory must have elements that are not paying their way by virtue of the existence of the other simpler but equally successful theory. This specifically addresses the levels-of-explanation objection, in which God is being posited as the ultimate explanation while natural explanations like CSR are simply the proximate means through which He implements His Plan. From a Bayesian standpoint, the prior probability of the more complex theory (which has the added type complexity, i.e., God) would therefore be lower than the simpler one. Recall that this can be understood by thinking of the extra elements (that is, the extra ontological types) as creating more ways of being wrong with no additional explanatory/predictive power to “pay its way.” Such a situation describes
condition 3 above and illustrates how, under some conditions, explaining really is explaining away. Conditions 2 and 3 capture the facts that a proposed hypothesis cannot be complementary to a theory if it makes it less successful or renders it logically incoherent. Insofar as theories involving supernatural entities (like $H_G$) meet any of these conditions, we can see that Fishman’s assumption is, in fact, well justified. With this key assumption established, we can now address the remaining open questions.

**CSR and the God Hypothesis**

By way of recap, we have addressed the NOMA concern, defused the skeptical worry that comes from the Duhem-Quine thesis, and shown how natural explanations really do compete directly with supernatural ones. Before completing our argument, we should still make a positive case that the burden of proof rests with those interpreting CSR data as supporting theism. We also still need to respond to the objection that CSR fails to address the reasons people actually use to justify their belief.

**Burden of Proof**

Recall that the assignment of burden of proof depends on background knowledge relative to a specific claim, along with a number of situation-specific, pragmatic considerations. As a result, the burden can shift over time, and even within the same conversation (Pigliucci and Boudry 4). Yet, its assignment is far from arbitrary. As we have seen, if God’s existence, or some theory entailing God’s existence like Calvin’s *sensus divinitatus*, has a very low prior probability, then the burden falls to the theist.

But why should we think this prior is very low? Recall our earlier example of two claims made under the same conditions, one of someone’s having three, twenty-dollar
bills and the other of their being a powerful extraterrestrial. That simplistic example should give some insight into the following observation: *While no theory is certain, theories are not equally uncertain.* It is in this light that Fishman seems to be saying that, given our background knowledge relative to the God hypothesis $H_G$, the prior probability of $H_G$—that is, $P(H_G)$—is *very* low relative to that of any coherent natural explanation. If this is correct, then the evidential bar for $H_G$ is much higher than that for an alternative, naturalistic, $\sim H_G$ hypothesis. The higher the evidentiary bar, the more that absence of evidence seems like evidence of absence after all.

To show how this applies to the God hypothesis, Fishman reminds us of our natural reaction to the absence of evidence for Santa Claus, whose powers are far outside anything our background knowledge would lead us to expect. Faced with the absence of evidence for *this* supernatural being, our natural response is not normally one of agnosticism (Fishman 817). More to the point, any coherent explanation for belief in Santa that has a higher prior probability would necessarily exclude the actual existence of Santa and would, therefore, be an alternative competing explanation. It seems clear that absent strong evidence for Santa, positive disbelief is justified. There is no burden to *disprove* his existence. This example not only reinforces our earlier argument that natural explanations do indeed compete with, rather than complement, supernatural ones; it also bolsters our intuition of how burden of proof is assigned and therefore how to understand the significance of absence of evidence.

Still, it is important to be cautious when attempting analogies with the God concept. We can bolster our analogy by having it more closely track our CSR-specific
discussion: Consider those who believe in Santa based on their experiencing the “Spirit of Santa” as if in a vision or as a non-reflective response to seeing a decorated tree. Imagine further that a class of drug is found, which in some doses induces various “spirit of” experiences, including those of Santa, and in larger doses, experiences of God. Given our discussion so far, this would seem to count as relevant evidence in support of any number of alternative explanations—in particular, those that specifically leave out the God / Santa hypothesis and which are, as a result, for all the reasons we have reviewed thus far, competing and not complementary explanations.

Jeff Jordan anticipates and addresses an important objection to this line of argument: If one could find a drug that induces more mundane experiences, say of a cup, this would not typically be taken as impugning cup experiences generally. Just as our faculty for perceiving cups may be led astray, so too the sensus divinitatus. In showing why this objection fails, Jordan illustrates our earlier conditional probability argument that supernatural hypotheses have extremely low prior probabilities. He points out that in the case of cups, both veridical and drug-induced experiences have physical causes; however, in the religious case, God is “wholly unlike any physical cause” and represents a “radical qualitative difference” from all other possible physical causes (Jordan 261). So, regardless of the mere logical possibility of a complementary cause like God, His radical “otherness” radically reduces the prior probability of His existence. Once again, we have to conclude not only that the burden of proof falls on those asserting His existence, but that in order to meet that burden they must produce a level of evidence that is as radically high as His prior probability is low.
It is worth noting at this point that we can understand Plantinga’s *sensus divinitatus* as a BFF created (through guided evolution) to generate solely physical responses to ordinary physical stimuli, where the physical responses are brain states that strongly predispose us to unreflectively form beliefs in the supernatural. Under this interpretation, the sensus operates entirely in the natural world, producing natural outputs from natural inputs—with no supernatural involvement whatsoever. In this sense, the only difference between the *sensus divinitatus* and a naturalistic HADD is whether or not the BFF is the product of purposive design. Yet the fact remains that the *sensus divinitatus* hypothesis thoroughly *entails* the existence of God, both in the sense of God’s having “guided” evolution (i.e., “caused” in a manner wholly unlike any physical cause) to produce the *sensus*, and His being the presumed veridical object of the *sensus’* output.\(^46\) Consequently, the arguments just raised still apply, especially that of parsimony, which we will revisit shortly.

With the burden clearly on those asserting God’s existence or the existence of something like the God-entailing *sensus*, at least some of the objections to atheist-CSR are made irrelevant, in particular, those that simply point out that CSR’s findings do not *disprove* God and consider that the end of the matter. In such cases the Santa Claus analogy really does apply: failing to disprove this being is hardly motivation for even an agnostic attitude, let alone *belief*, and for the same reasons. It is, in short, more rational

---

\(^46\) Plantinga makes clear that he does not reject biological evolution, just the idea that it is *unguided* (Plantinga, *Where the Conflict Really Lies* 129), which is why he can think of the *sensus divinitatus* as a product of physical evolution.
to believe that such a proposition is false—given its extremely low prior—than to claim a neutral uncertainty.

However, even if the placement of burden of proof were still in doubt, we can also point to the effects of assessing parsimony and predictive success. Does adding this additional ontological category (i.e., God) pay its way? If we can explain the same range of data without it, then clearly it does not. This leads us to ask how our competing theories compare.

Parsimony and Predictive Success

Even if the burden of proof rests with the theist, this burden can in principle be met. For example, if a theory entailing God’s existence makes unexpected, publicly available predictions that are confirmed, this would strengthen that hypothesis—possibly enormously—even despite its very low priors. Indeed, human knowledge has grown from just such developments in other areas. What about here? Building on our earlier discussion of comparative predictive success, we will first critically compare the sensus divinitatus with ordinary sense perception, and then apply a similar analysis to theistic and atheistic interpretations of CSR’s findings.

Sense Perception and Religious Experience

---

47 Of course, it is completely unclear how God can be regarded as an actual explanation for anything, particularly if we understand the act of explanation as one of moving us from the known to the unknown. Attempting to move from the unknown to the unknown is a characteristic not only of the God hypothesis, but of the whole panoply of human supernatural “explanations” like witches’ curses, rain gods, etc.
Earlier, with the help of Swinburne’s religious experience classification scheme, we were able to place the purported output of the *sensus divinitatus* into the category of religious experiences of a non-public object. With this in mind, let us look again at Plantinga’s likening belief in God to belief in immediate perceptions (e.g., his tree outside the window). He is far from alone in making such comparisons. Indeed, believers often compare their experience of God to ordinary sense perception. One can schematize and superimpose both types of experience as follows [adapted from] (Martin 157), where $B_S$ is a spontaneous perceptual belief (i.e., basic belief) based on sense experience, and $B_R$ is a spontaneous religious belief based on a religious experience. $C$ represents the conditions under which these occur, where $C_S$ is the appropriate set of conditions for a sensory perception, and $C_R$ for a religious one.

(1) / (1’) $B_S / B_R$ under appropriate condition set $C_S / C_R$ is likely to be true

(2) / (2’) $C_S / C_R$ obtains

(3) / (3’) My $B_S / B_R$ in (the tree outside my window) / (God) is of the kind appropriate to the underlying mechanism and conditions.

(4) / (4’) Therefore, my $B_S / B_R$ in (the tree outside my window) / (God) is likely to be true.

The differences between (1) and (1’) are important. In both cases we have to assume that there is a causal mechanism at work as well as some reality external to, and independent of, the person having the experience—call this the External Causality Hypothesis $H$ (Martin 158). We will call $H$ for Sense Experience $H_S$ and for religious experience $H_R$. In (1), the external trigger for the $H_S$ experience is intersubjectively verifiable for instances of sense perception. The same is true of the associated reliability
conditions, \( C_s \). We expect that others under the same conditions, \( C_s \), would have the same experience. What might a competing \( \neg H_s \) look like for (1), that is, a non-external causality hypothesis that is similarly intersubjectively verifiable? It seems that something other than a Cartesian brain in a vat would be difficult to construct—again, given the intersubjectively verifiable nature of evidence and conditions that support \( H_s \) and \( C_s \).

On the other hand, neither \( H_R \) nor \( C_R \) can point to anything like this intersubjectively verifiable support. Consequently, a competing \( \neg H_R \) for (1’) seems all too easy to imagine, especially since we have extensive background knowledge of \( \neg H \) causes for other objects of belief when they and their reliability conditions are similarly unverifiable in the same public, intersubjective sense—from drug effects, surgical stimulation, and various pathologies to well-understood psychological phenomena. Bear in mind that such alternative \( \neg H_R \) hypotheses, such as Freud’s wish-fulfillment hypothesis, need not be flawless or even strongly supported by evidence when the alternative is an \( H_R \) that has such low priors. It need only be palpably better than \( H_R \), again, given our current background knowledge and that knowledge’s effect on the priors.

Of course, things have changed dramatically since Freud’s time. Since the emergence of CSR’s empirically based findings, we have never had a more plausible \( \neg H_R \) hypothesis than we do now. But are we not simply begging an important question here? How do we know that such religious experiences are not actually giving us insight into some other realities? We can answer this by applying our falsifiability and predictive-success criteria: We have good reason to believe that the output of this putative BFF—
that is, this *sensus divinitatus*-like mechanism—is unreliable because the resulting beliefs are too often mutually exclusive, mirroring local cultures rather than converging on any consistent or coherent picture. This is not question begging; it is an empirical result. Things could have been different, but they are not. It is not question begging, therefore, to dismiss such alternate causal hypotheses, especially when one has independent, converging lines of evidence that *do* support both $H_S$ and $\neg H_R$ cases, such as CSR without the encumbering/complexifying God hypothesis. Again, such $\neg H_R$ explanations need not be flawless, only the best among the alternatives that can be inferred from the available data.\(^{48}\) With all these considerations in mind, how does a theist-supporting interpretation of CSR’s findings compare with an atheist-supporting one?

**Theist vs Atheist CSR**

In the specifically Judeo-Christian CSR interpretations we have examined, what does the entailed God Hypothesis $H_G$ predict? Recall that such theism, despite being of a non-public object, does have *public* observational consequences. As Fishman points out, belief in an immortal soul that survives death would seem to entail that an individual’s personality, memories, passions, interests, etc.—those properties that are *essential* to our identity as individuals—are *independent* of any physical substrate, such as our physical brains. If true, then the current findings of neuroscience, let alone CSR, would be *unexpected* (Fishman 822).

\(^{48}\) Recall that using the concept of parsimony we have already assessed and dismissed the possibility that these alternate hypotheses might in some sense complement one another rather than compete. We have also dismissed any ad hoc hypotheses that cannot be tested independently of the theory.
Under a theistic interpretation we might expect that brain damage could mute or mislead the essential “you,” since under such a dualist view, we have to imagine the soul inhabiting and using the body much as a pilot in the cockpit of a large and complex machine. However, we would not expect such alterations to *change* you in any essential way. The evidence from brain injury, surgery, and other studies have provided consistent evidence that our personalities, sense of self and identify, feelings toward others, memories, etc., can be fundamentally changed or piece-meal extinguished by physically changing the brain. Literally no aspect of what makes *you* is invulnerable to being altered by *physical* changes to the brain.

The evidence for this is extensive and often shockingly counterintuitive. Beyond the obvious cases of short and long-term memory loss (e.g., not knowing the people dearest to us) due to disease and trauma, the two most striking examples come from the effects of frontal lobe damage, and from the effects of a procedure called *corpus calloscotomy*—the so-called “split brain” surgery. Perhaps the most famous case of the former was that of Phineas Gage, who survived having a four-foot-long, thirteen-pound railroad spike propelled completely through his head, passing from below the cheekbone and through the frontal lobe before exiting the top of his head (Hauser 226). Despite retaining his use of language, body movement, perceptions, and memory, he became “a different person, [going] from courteous, responsible, and ambitious to rude, unreliable, and shiftless” (Pinker 42). One might describe him as having gone from saint to sinner.

The effects of split-brain surgery provide a very different insight. They expose not only the modular nature of our minds, but even the fact that our sense of having *one*
mind, of having a unique self (or “soul”), is illusory. The procedure involves cutting the band of fibers separating the left and right hemispheres. The result is that the “self” becomes two quite distinct selves, each with its own “free will,” desires, hopes and dreams. When experimenters give instructions only to the right (non-talking) hemisphere and it acts on those instructions, the unaware left hemisphere spontaneously confabulates a story to explain why the whole person has decided to act in a certain way (Pinker 43).49

These separate hemispheres will even answer very personality-specific questions differently, such as “What do you want to be when you grow up?” (Brogaard).

Given the epistemological tools we have discussed, such as parsimony (with its sensitivity to ad hoc helper hypotheses that do not pay their way), the best inference from such evidence is that if selective damage can piecemeal extinguish or fundamentally change who you are, and even expose the illusory façade that you have a single undifferentiated self, then the essential “you” is entirely the product of the configuration and interaction of the physical brain’s components. The destruction of the whole brain, then, would mean the destruction of the whole essence of the individual (again, from the standpoint of the most parsimonious explanation that encompasses and explains the data without conflicting with it). These neuroscientific findings are consistent with—in fact often predicted by—a naturalistic, unguided, evolutionary account of the mind’s development (Fishman 822). On the other hand, these same findings are not only entirely unexpected under a theistic (especially Christian) account, but require the creation of ad

49 To be clear, this means that the speaker is unaware that confabulation is occurring, and sincerely believes what they are saying.
hoc, independently untestable auxiliary hypotheses in order keep this ever-accumulating body of findings consistent with that theistic account.

In the specific context of theist-CSR, there is certainly at least one auxiliary hypothesis that can be tested independently of theism: the *sensus divinitatus*. If its predictions could be confirmed, it would lend support to the underlying $H_G$ hypothesis. Indeed, this is what Clark and Barrett claim is the case: “the results from cognitive science . . . suggest empirical confirmation for Thomas Reid’s speculations” (Clark and Barrett, “Reidian” 650). But what are its predictions and how do they compare with those coming from an atheistic interpretation of the same CSR data? As we explore this, keep in mind our earlier discussion of parsimony.

Since the idea of the *sensus divinitatus* has been put forward in the context of Christian theism, we will understand God in the following discussion to be the Abrahamic omni-God. The *sensus divinitatus*, then, would be understood to have been put in place by God, perhaps through guided evolution, in order to sense Him. As such, this hypothesis would predict some tendency toward a global uniformity of belief that, at the very least, would not be *in conflict* with God’s attributes. A stronger prediction certainly seems reasonable, namely, that we would find a universal predisposition toward something more specifically akin to Abrahamic monotheism. Yet, even the weaker prediction is not met.

We do in fact find a universal phenomenon with regard to belief in the supernatural. To be sure, nobody “is born with the idea that the birthplace of humanity was the Garden of Eden, or that the soul enters the body at the moment of conception, or
that martyrs will be rewarded with sexual access to scores of virgins” (Bloom 120). But
the universal aspects of these supernatural beliefs do not even vaguely approximate the
omni-God of Judeo-Christian monotheism, or any particular theism for that matter. As
Boyer observes, the differences between the world’s religions runs much deeper than
superficial label differences, like Mormon or Taoist; it goes to the very heart of how
supernatural agents are conceived. These conceptions show little discernable overlap and
include ancestor spirits, witches, gods that die and/or are easily fooled, trees that can
remember, and living mountains that are appeased with animal hearts and fetuses.
Moreover, concepts essential to our Western notions of religion, such as Salvation, are
often completely absent (Boyer 6–10, 65). It seems the only actual universal we can
extract from this extreme variability—this cornucopia of output from our inborn
“teleological promiscuity” (Kelemen and Rosset), which is constrained only by the
contours of our inborn cognitive architectures—is the common denominator of agency.
Can auxiliary hypotheses be offered to make this extreme variability consistent
with the original (Christian) theistic prediction? Certainly. As we have seen, the Adamic
Fall has been used to explain any deviation from what would be otherwise be expected.
Alternatively, some have suggested that God’s intention was to provide only a “slight
taste of divinity,” allowing for variant cultural expressions (Clark and Barrett,
“Reformed” 175, 187). It should be clear at this point that both explanations are rather
stark examples of the kind of ad hoc, “helper” hypotheses we examined earlier. In
particular, they are obviously not testable independently of the theory they are intended to
save (i.e., Christian theism). Indeed, if such an explanatory strategy were valid, then we
would have a situation where evidence that supports theism is eagerly accepted, while evidence that contradicts it is explained away by the imagined side-effects of the Adamic Fall or the untestable intentions of an inscrutable supernatural deity. But then this would render Christian-theism (in this example) consistent with any conceivable set of findings, which of course renders it both unfalsifiable and devoid of explanatory power (and, arguably, factually meaningless).

As we saw earlier in the Celestial Mechanics example, independently testable auxiliary hypotheses can provide important insights, and can often expand the predictive power of a theory well beyond the new findings that motivated their introduction. As it turns out, some auxiliary hypotheses offered in defense of theistic models actually do suggest means for independent testing. The problem is that when those tests fail new untestable auxiliary hypotheses are introduced. For example, the Adamic Fall, in the context of a Christian CSR interpretation, does make an independently testable retrospective prediction. Helen De Cruz points out that if Original Sin corrupted our sensus divinitatus as Plantinga suggests, then we would expect to find the earliest evidence of religion to be closer to monotheism with the corrupting effects of sin driving deviations over time. Instead, the archaeological and anthropological evidence we do have reveals precisely the opposite pattern. Monotheism appears only recently, correlated with larger societies (De Cruz and De Smedt 59). Therefore, to the extent that the Adamic Fall auxiliary hypothesis is testable in the retrospective sense, it is falsified along with any theistic CSR that depends on it. Of course, without the criterion of independent testability, it is all too easy to explain this away—one is limited only by imagination, and
even failing that, one can simply say “God has His inscrutable reasons.” Indeed, as Dawes and Jong point out, when the reformed epistemologist invokes Original Sin in this way, she is undermining the reliability of the *sensus divinitatus* in the most damaging way possible: by pointing out that it is unreliable specifically in the religious-belief-forming context (Dawes and Jong 14) for which it was purportedly designed.

Is there an alternative theory that meets our epistemological conditions while explaining *at least* the same range of data that theistic CSR claims to explain? We have already addressed Plantinga’s worry that unless our cognitive faculties reflect purposive design, we would slide into radical skepticism. With that concern set aside and with the epistemological tools we now have at hand, it seems that simply *removing* the theistic element is all that is needed—nothing needs to be added back. We are simply unencumbering what we already have. On the other hand, if we leave the God auxiliary hypothesis in the theory, then we are forced to add *ever more* auxiliary hypotheses (e.g., the Fall, Satan’s direct action, God’s inscrutable intentions, etc.) on an on-going, case-by-case basis to ensure that the theory survives our ever-accumulating contrary data. The price of maintaining one unnecessary auxiliary hypothesis is having to introduce a blossoming edifice of ever more auxiliary hypotheses—some to buttress the underlying theory, others to buttress the buttresses. And, as we have seen, this not only adds complexity, it renders the whole theory unfalsifiable. However, if we simply *remove* the God hypothesis, then the whole teetering superstructure of tangled, interconnected auxiliary hypotheses needed to sustain it can simply fall away, like so much built-up ice sloughing off an airplane wing, revealing its simpler contours and allowing it to function
fully. The theory not only regains parsimony, but also recaptures other key virtues, such as falsifiability.

Indeed, unencumbered atheist-CSR not only explains the same data with much less complexity, it better predicts what we actually find. For example, recall that Plantinga had posited the sensus divinitatus in order to “explain” such things as our predisposition to relational beliefs with God, which produce feelings of gratitude and guilt (Plantinga, “Is Belief in God Properly Basic?” 46). Yet this is already specifically explained by atheist-CSR but without appeal to a supernatural entity. Recall that our social interaction inference engines produce rich, emotionally intense, morally laden, social exchange intuitions, which can be activated once our over-eager agency detector decides that some socially relevant exchange partner exists—even an invisible one that is super relevant—thanks to the combination of HADD and the super-stimulative effect of certain MCI notions as described by Boyer. Furthermore, atheist-CSR, unlike the theist version, explains not just guilt and gratitude. It also explains the universal nature of many of our moral tendencies, including those that are less than noble, and it does so without introducing untestable auxiliary hypotheses involving yet more supernatural entities and motives, like Satan and his influence. For example, a game theoretic, gene centric, evolutionary model of the origin of our core moral and social emotions not only explains the typical patterns of altruism in humans and even non-humans, but also explains why such noble impulses have such disappointing aspects, such as their very short-range: the more closely one is related to another person, the more likely one can be expected to demonstrate self-sacrificial altruism. Yet, this likelihood drops off precipitously as we
move from immediate family to cousins and finally to strangers (Pinker 241–68).\(^{50}\)

Again, atheist-CSR explains a much larger range of the observed data, without having to layer on untestable auxiliary hypotheses to deal with findings that do not fit the theory’s prior commitments.

CSR and the Rational Justification of Theism

We are now in a good position to respond to Thurow’s charge that CSR is irrelevant because it does not address the reasons people actually use to justify belief in God. Recall that Thurow argued that belief in God is supported by multiple independent sources of justification, and that therefore each of these supports would need to be knocked out in order to render a theist’s belief unjustified. We made the case that as far as the justification for belief in God is concerned, the burden of proof rests squarely with the theist. While this does bear on this multiple-justifications characterization, the thrust of Thurow’s point is less about burden than about relevance. If CSR simply misses the mark with regard to the reasons people actually use, then it is not bringing anything to the table, as it were, and the discussion falls entirely back onto the arguments that have traditionally been made.

This picture of belief as being supported by multiple sources of justification assumes that each source is independent. If the CSR model is correct, then we have

\(^{50}\) See also (Pinker 269–82) for an extended discussion of other “quirks.” For a fascinating, fuller treatment of the consequences of a gene centric, as opposed to an individual centric, view of evolution, see (Dawkins, The Selfish Gene). These consequences include specific patterns of conflict and cooperation between generations, the sexes, and even parents and their offspring, many of which could hardly be described as morally praiseworthy.
reason to believe that many, if not all of these seemingly independent justificatory legs—such as Thurow’s example of the testimony of elders—actually spring from one phenomenon: our innate CSR-based BFF. He does seem to recognize that CSR’s findings do undermine the independence of these justificatory sources, at least in the sense that “the findings of CSR might seem to show that our religious belief-forming, sustaining, and religious argument evaluating processes are all unreliable.” He refers to this as a “cunning rejoinder” [emphasis in original] (Thurow 95).

Like Barrett, Thurow’s response to this type of rejoinder is to make a version of the genetic fallacy charge. He then points out that even if we come to initially believe through a CSR-type process that some event was caused by God, we can offset that predisposition by “taking a closer look at” the event. Indeed, this is how we typically approach many of our beliefs. Thurow illustrates this with the example of the natural bias a father might be expected to feel for his son, who is accused of a crime; however, given strong enough evidence, his belief can change. Thurow concludes that while CSR may predispose some people to accept theistic arguments for no other reason than that predisposition, “it doesn’t follow that religious belief in general is unjustified because many people may still be able to evaluate the evidence in an accurate way.” (Thurow 96). What’s more, even if they do find such evidence wanting, then their belief would become unjustified not because of CSR’s findings, but because of that insufficient evidence.

It is worth recalling at this point some of our earlier comments about the nature and variety of religious/supernatural beliefs found throughout the world and throughout history. It was this same pattern that had impressed Aquinas and Calvin (both of whom
had far less data than we do now) and, through them, inspired Plantinga to explore the implications of a *sensus divinitatus*, a concept upon which he built many of his assertions, including that theism is warranted without evidence and that the burden of proof rests with the skeptic. It seems clear at this point in our discussion that CSR is highly relevant to understanding not only this same observed universal pattern of belief, but also the near-universal pattern of *an absence* of concern with justification among believers. Recall that these beliefs are generally held *unreflectively*. As we have seen, from the ancient Greeks to the Kwaio the question of justification hardly comes up at all, and when it does it is often the result of an investigator’s prompting. This suggests that for a great many if not most people, their stated justifications are not really the bases of their religious beliefs at all, but are produced post hoc to confirm the deliverances of the pre-reflective mental processes identified by CSR. In fact, this view is not only consistent with Barrett’s own experimental findings regarding Theological Correctness. (Recall the prayer experiment involving the endangered ship, which revealed that people’s *actual* conception of God differed from their stated *theoretical* conception.) It is also consistent with a significant body of research in psychology, which supports the view that humans generally tend to produce post hoc rationalizations to confirm pre-reflective beliefs and choices.\(^{51}\) CSR addresses precisely this unreflective source of belief. With this in mind, Thurow’s point, even if valid, would apply only to a small, rarified sliver of humanity’s believers—that is, people who *have* justificatory reasons for their supernatural beliefs. It seems then that it

\(^{51}\) See for example (Johansson) and (Haidt) in addition to the earlier cited (Pinker 43).
may be Thurow’s argument, and not CSR’s findings, that are largely irrelevant to explaining the world’s actual pattern of beliefs.

Once again, none of this proves that theism is unjustified, but it does significantly reduce the likelihood that it is.\(^{52}\) Consider: before the findings of CSR, Plantinga’s argument, and Thurow’s point about multiple sources of justification, would at least have had the effect of increasing the plausibility, the likelihood, that theism is a justified rational belief. After those findings, however, and coupled with the methodological tools reviewed earlier, we can see that this likelihood is now not only significantly reduced, but reduced well below the level that had existed before CSR—that is, the range and strength of available justifications is reduced, with many of them shown to be different manifestations of the same underlying process—a process with strong empirical support. Indeed, the conditions Plantinga had described as conditions under which the sensus works (e.g., the beauty of a flower) are precisely those conditions under which CSR works; yet CSR goes further, explaining not only the same data, but also contrary data, and it does this without having to introduce encumbering, untestable assumptions (e.g., the existence of supernatural agents).

We can illustrate this by revisiting Dawes and Jong’s earlier Whiteboard example, but extending it in an important way. Recall that their thought experiment involved information that put in doubt a professor’s spontaneous, unreflective belief that there is in

\(^{52}\) To be fair to Thurow, he specifically qualifies his argument as addressing only the question of whether CSR has rendered theism unjustified, deferring the question of whether CSR “supports or reduces the degree of justification of religious belief” (Thurow 79). Nonetheless, his insightful argument raises points that must be addressed as part of making the case that CSR does indeed reduced this degree of justification.
fact a whiteboard present in her classroom. This doubt led her to seek independent
evidence in order to rehabilitate her belief. What if her other justifications had included
the testimony of other professors? In that case, those testimonial justifications would fall
under precisely the same suspicion for precisely the same reason and therefore not at all
contribute to the likelihood that the belief is justified. What if it were also noted that other
professors justified their belief for a variety of other ostensibly independent reasons? In
that case, it would depend on what those other reasons are; however, if it could be shown
that many of those other reasons were also affected by this newly discovered information,
then the range and strength of those justifications would be reduced, if not eliminated
outright. What may have seemed like many independent sources of mutually reinforcing
lines of justification before the new data appeared, would have collapsed into one
suspicious source afterward.

In fact, the scope of CSR’s findings on the range of pre-existing justifications is
considerably broader than the doubt-creating discovery in the Whiteboard example.
CSR’s findings bear not only on one’s own spontaneous belief and the testimony of
others’ similarly spontaneous beliefs, but also on nearly all theism-relevant spontaneous
beliefs, ranging from the perception of the Bible’s divine nature and whether one has
witnessed a miracle, to one’s spontaneously seeing the hand of God in the intricacies of
nature, or that of the Devil in other aspects of nature (e.g., disease, earthquakes). We have
to conclude that Thurow’s important points have been addressed.
Summary

We set out in this chapter to develop and justify a set of epistemological tools and methods that can be applied to both ordinary claims, and claims about the existence of supernatural entities. We then applied these to recent findings from the cognitive science of religion and concluded that these findings significantly diminish the rational justification for theism. Our development was multi-pronged and mutually reinforcing, building upon the previous chapter by drawing together and integrating a few key insights from the philosophy of science and epistemology. These insights included theory “virtues,” which differentiate better theories from worse ones, particularly in the context of the Duhem-Quine problem. They also included the tools of conditional probability and Bayes’ Theorem, which helped clarify our thinking regarding the effects of evidence on the likelihood that some claim is true in the context of our background knowledge. We then combined these ideas with the notion of “bootstrapping”: an iterative, self-correcting approach to building justified beliefs from an entirely random starting place. The final result was our curation and justification of an epistemological tool suite applicable to both natural and supernatural theories. In the process we were able to (1) address the NOMA/relevance objection; (2) show that our toolset applies equally to natural and supernatural claims; (3) establish that the burden of proof does indeed rest with the theist-based interpretation of CSR; (4) comparatively assess theist-CSR and atheist-CSR to show that atheist-CSR is, by far, the more parsimonious explanation (leaving aside the question of whether the theist interpretation is coherent); (5) address Thurow’s important relevancy objection; and finally (6), observe that one consequence of the preceding is that
the range and strength of the justification for theism have not only been reduced by the findings of CSR, but reduced to a level well below what it was prior to the advent of this new field.

In what initially seems like a surprising move, Clark and Barrett “concede that there is no reason to appeal to a god to explain the data of cognitive and evolutionary psychology of religion” (Clark and Barrett, “Reidian” 661). Their point, however, echoes Plantinga’s, namely, that belief in God is properly basic with respect to warrant, and so is not believed as a hypothesis—it is not believed because of the evidence. Moreover, they argue, such principles as we have been discussing here lead to unacceptable consequences. For example, they assert that if we “were committed to the principle of simplicity [i.e., parsimony] with absolute devotion in all areas of human inquiry” then the “simplest hypothesis is that only I exist and that you and other ‘people’ are simply figments of my imagination” (Clark and Barrett, “Reidian” 661).

At this point in our discussion, however, it should be clear why both of these points fail. We have already shown why belief in God is not properly basic, yet to the extent that CSR is true, our predisposition for theistic belief is the product of a basic-belief-forming faculty as described by the CSR standard model, and which is indeed unreliable specifically regarding religious beliefs—a point we established without question-begging or “stacking the deck” against theism. As for their critique of a “devotion” to simplicity, we have seen that parsimony has a lower bound: a theory should shed its complexity only until doing so any further would cause it to be wrong. The goal is not runaway simplicity as if there are no other objectives, but rather to be a simple as
possible while preserving the theory’s explanatory and predictive power. We are therefore in no danger of falling into the solipsism that Clark and Barrett are warning us against, since such an extreme would fall far below this lower-bound principle, just as would such claims as “the universe was created ten minutes ago with our memories intact.” Such claims are not disprovable, but for reasons we have already discussed, there is no good reason to suppose the burden rests on skeptics of such imaginings. Even more to the point, it is difficult to see how a theistic hypothesis would shield against such merely logical possibilities. After all, perhaps an Evil Demon created only you as a brain in a vat. Perhaps, as Bergmann had earlier suggested, one takes it as a basic belief that an Evil Demon is responsible for creating false basic beliefs in God. This leaves us with Clark and Barrett’s original concession, namely, that there really is no reason to appeal to any god to explain the findings of CSR.
Throughout this paper we have taken and defended a number of definitive positions across a wide range of age-old philosophical debates, including the compatibility of science and religion, the reliability of our reasoning faculties, the relevance of evolutionary theory to theism, and even whether monism or dualism is the better characterization of our minds/souls. Despite the pedigree of this historically rich and important discussion, each of our positions was informed and connected by something quite new on the scene: the emerging data from the cognitive science of religion. In fact, one of its many fascinating insights has been that of showing us how areas once thought completely unrelated are in fact intimately connected in a complex cause and effect relationship. Examples of this include the relationship of evolution to the universal predisposition to believe in supernatural agency and (contra Plantinga) the relationship of evolution to the reliability of our self-correcting reasoning abilities.

Of course, drawing on such a wide range of philosophical areas means we have to leave many interesting related arguments unexplored, both against and for the thesis of the present work. For example, while the argument developed here has not been limited to Christian theism, there is an additional vulnerability specific to Christian theist-CSR, which stems from the traditional Christian view that “God is such a being that it is logically impossible that there be casually necessary conditions for the divine will” (Jordan 262). It seems, then, that if it is indeed God's Will that person S experience Him,
then *any* natural cause we do find cannot be traced to God, since the Divine Will requires no such necessary natural preconditions (Jordan 262). For the Christian, this further undermines any appeal to God as the ultimate cause operating through natural, proximate causes. It also creates yet another scenario under which explaining really is *explaining away.*

From the other direction, at least two objections can be developed using the methodologies discussed in this paper. The first is based on the pragmatics of assigning burden of proof. We had earlier noted that such considerations can shift the burden of proof on a case-by-case basis, preventing the development of any simple, one-size-fits-all rule. As we saw, these include evidential and prudential considerations. Pascal essentially makes just such a prudential argument in his famous Wager. The second possible objection is based on our parsimony discussion, in which we argued that the theory that explains the most with the least ontological-type complexity is the better theory. Richard Swinburne has argued that God is, in fact, the *simplest* explanation. In what follows I suggest some initial responses.

**The Return of Pascal’s Wager**

Earlier we noted that assigning burden of proof is not a simple matter, and involves both logical and pragmatic considerations. It is difficult not to be reminded of Pascal’s Wager, which could be recast in the same decision-making-under-uncertainty terms that we had earlier examined. Recall our discussion about the pragmatic *cost* considerations associated with assigning burden of proof. We used the example of what our default assumptions should be regarding a just-found unexploded WWII bomb. Since
the cost of being wrong is so high, the burden appropriately shifts to the negative position. But what of the cost of being wrong about God? By some accounts, being wrong about the bomb would be preferable to being wrong about God. Does this not undermine our earlier argument that the burden of proof lies with the theist? There are at least two reasons why it does not.

As Pascal put it, “’God is, or He is not,’ But to which side shall we incline” (Pascal 497)? Herein lies the first significant problem. As we have seen, the world is filled with a variety of gods and religions, many of which are mutually exclusive—even within Christianity. Pascal is presenting us with a false choice. The costs that need to be assessed are those that each mutually exclusive belief system imposes on those not choosing that particular system. So, the choice is not simply between unbelief and belief, but also between many beliefs. After all, to be a believer in one belief system is to be an unbeliever in all other mutually exclusive belief systems. But what are these “costs”?

Even within Christianity there is a divergence of opinion regarding whether God is really as He is portrayed in the Old Testament, namely, inclined to torment someone for all eternity merely for reaching the very defensible conclusion that He does not exist. Indeed, a literal reading of the OT would make Him responsible for what today would be considered war crimes, e.g., “neither have ye pity, slay utterly old and young, both maids and little children, and women” (The Holy Bible: King James Version, bk.Ezek. 9.6). If, instead, God’s ethics were more recognizably like our own, would He not be more offended by the “play it safe” approach Pascal seems to be suggesting? Would He not feel greater respect for an honestly arrived at mistake?
God as Simplest Explanation

An interesting objection can be made to my earlier argument against theism on the grounds that it is not parsimonious. Swinburne has argued in just these terms that God *is* in fact the simplest explanation of all (Swinburne). A full treatment of this is beyond the scope of the present work; however, we can at least draw attention to the central problem in Swinburne’s argument: The notion that God can be *both* simple *and* possess the properties of omniscience and omnipotence seems incoherent. Consider: God is understood to be conscious, to know all distinct facts in the universe, including the most complex scientific theories known and yet to be discovered, as well as all the mental states of all humans (Ostrowick). He is, in short, far, *far* more capable than our most complex computers and *complex* brains. This seems to empty the word “simple” of meaning. “God” as a label is certainly “simple,” but that label’s referent is another matter altogether. Yet the problem with this notion of “simplest” goes much deeper.

If we regard the notion of “explanation” as one of *increasing* intelligibility, of building from the known to the unknown, from things we *understand* to thing we do not, then God’s radical otherness precludes His serving as an explanation of *any kind*, let alone a “simple” one. In other words, His radical otherness means that when we use Him to “explain” something, we are attempting to build from the unknown to unknown.53 No light is actually being shed.

53 It can also be argued that His radical otherness renders Him factually meaningless (if not incoherent).
Perhaps this problem can be illustrated by way of an analogy. Suppose we are living during prehistoric times, before the natural processes behind the phenomenon of rain were known. Suppose further that I hypothesize some rather complex natural cycles to explain rain. Our tribal priestess, however, avers that she has a far “simpler” explanation: namely, the tribal rain god Dzaloc. It is not only one simple word, but all conceivable rain-related phenomena can be understood in terms of this god’s inscrutable powers and motives. She goes on to stipulate that, as matter of this god’s definition, it is *supremely* simple compared with my proposed complex processes. This approach amounts to confusing a label for our ignorance with an actual explanation. In other words, even if we cannot *understand* the nature of Dzaloc or the means by which it causes rain, we are to regard it as an *explanation* for rain—and a supremely “simple” one at that.


=294767&site=ehost-live&scope=site.


Oxford University Press, 2011.

Plantinga, Alvin, and Nicholas Wolterstorff. *Faith and Rationality: Reason and Belief in


779–80.

47.

Stirrat, Michael, and R. Elisabeth Cornwell. “Eminent Scientists Reject the Supernatural:
A Survey of the Fellows of the Royal Society.” *Evolution: Education and
Outreach*, vol. 6, no. 1, 2013, p. 33.

Swinburne, Richard. “God as the Simplest Explanation of the Universe.” *European

edited by Edward N. Zalta, Winter 2016, Metaphysics Research Lab, Stanford
University, 2016. *Stanford Encyclopedia of Philosophy*,


